



TRABALLO FIN DE GRAO
GRAO EN ENXEÑARÍA INFORMÁTICA
MENCIÓN EN COMPUTACIÓN



Aliñamento de imaxes oftalmolóxicas usando representacións neuronais implícitas

Estudante: Mateo Amado Ares

Dirección: José Rouco Maseda

Jorge Novo Buján

A Coruña, xuño de 2025.

A quienes despertaron mi curiosidad.

Agradecimientos

Agradezco a mis directores, Jorge Novo Buján y José Rouco Maseda, y a David Rivas Villar por su apoyo y ayuda durante este trabajo. También doy las gracias a mi familia y amigos, especialmente a Daniela, por estar ahí cuando hacía falta.

Resumo

O aliñamento da imaxe oftalmolóxica é útil para, entre outras cousas, revisar o avance dunha enfermidade ao longo do tempo, fusionar diferentes modalidades de imaxe ou comparar entre diferentes pacientes. O caso dos ollos é de particular interese xa que permiten a observación in-vivo de tecido neuronal e vasos sanguíneos, o que posibilita a detección temprana de certas enfermidades. Aliñar as imaxes manualmente é un traballo tedioso e complexo, polo que automatizar este proceso é de gran interese.

Neste traballo explórase o uso de redes de representación implícita aplicadas á tarefa de aliñamento de imaxes oftalmolóbicas. Neste tipo de redes, a deformación é parametrizada como unha función continua nos propios pesos da rede, coas coordenadas da imaxe móbil como entrada e a deformación correspondente como saída. Representar a deformación de esta forma ten varias vantaxes frente a representacións tradicionais discretas, como a independencia de resolución e poder prescindir de grandes bases de datos xa que se adestran mediante un proceso de optimización para cada par de imaxes. Ademais, en lugar de usar funcións de activación estándar como RELU, estudamos empregar unha función de activación sinusoidal (SIREN) que pode axudar a eliminar o sesgo cara sinais de baixa frecuencia e mapear mellor deformación pequenas e detalladas [1].

Adaptando o traballo realizado por Wolterink et al. [2], valoraráse se este método é apto para a tarefa de aliñamento de imaxes oftalmolóbicas.

Abstract

The alignment of ophthalmic images is useful for, among other things, reviewing the progression of a disease over time, fusing different image modalities, or comparing different patients. The case of the eyes is of particular interest as it allows for in-vivo observation of neuronal tissue and blood vessels, enabling the early detection of certain diseases. Manually aligning images is a tedious and complex task, so automating this process is of great interest.

This work explores the use of implicit neural representations applied to the task of ophthalmic image alignment. In this type of network, the deformation is parameterized as a continuous function in the network's weights, with the coordinates of the moving image as input and the corresponding deformation as output. Representing the deformation in this way has several advantages over traditional discrete representations, such as resolution independence and the ability to avoid large databases, as they are trained through an optimization process for each image pair. Moreover, instead of using standard activation functions like RELU, we

explore using a sinusoidal activation function (SIREN [1]), which can help eliminate the bias toward low-frequency signals and better map small and detailed deformations.

By adapting the work of Wolterink et al. [2], it will be evaluated whether this method is suitable for the task of ophthalmic image alignment.

Palabras clave:

- Imagen médica
- Imagen oftalmológica
- Aprendizaje profundo
- Registro de Imágenes
- Representaciones neuronales implícitas

Keywords:

- Medical imaging
- Ophthalmological imaging
- Deep learning
- Image Registration
- Implicit neural representations (INRs)

Índice Xeral

1	Introdución	1
1.1	Motivación	1
1.2	Obxectivos	2
1.3	Estrutura	2
2	Contexto	3
2.1	Oftalmoloxía	3
2.1.1	Anatomía do ollo humano	3
2.1.2	Imaxe oftalmoloxica	4
2.2	Rexistro de imaxes	5
2.2.1	Categorías de rexistro	5
2.2.2	Estado da arte	8
2.3	Representación Neuronais Implícitas	12
2.3.1	Aplicacións	13
2.4	Traballo proposto	14
3	Metodoloxía e planificación	18
3.1	Metodoloxía do desenrollo	18
3.2	Planificación do proxecto	18
3.3	Recursos utilizados	19
3.3.1	Software	19
3.3.2	Hardware	20
3.3.3	Conxuntos de datos	20
3.3.4	Estimación de custos	20
3.4	Seguimento da planificación	21
3.4.1	Estimación de custo real	21

4 Traballo Realizado	23
4.1 Vista Xeral	23
4.2 IDIR	23
4.2.1 Arquitectura	24
4.2.2 Método	31
4.2.3 Replicación de resultados	32
4.3 Adaptación a 2D	32
4.3.1 Proceso de Rexistro	33
4.4 Conxuntos de datos	33
4.4.1 FIRE	33
4.4.2 RFMID	34
4.4.3 Diferencias entre os datasets	35
4.5 Diseño de Experimentos	35
4.5.1 Metodoloxías Desenvolvidas	36
4.6 Métodos de Avaliación	37
4.6.1 Avaliación Cuantitativa	37
4.6.2 Avaliación Cualitativa	39
5 Experimentos e resultados	40
5.1 Vista Xeral	40
5.1.1 Descripción dos experimentos	41
5.2 Exemplos de rexistro	42
5.3 Función de loss	42
5.3.1 Planteamento	42
5.3.2 Resultados	43
5.3.3 Discusión	43
5.3.4 Conclusíons	44
5.4 Resolución da imaxe	45
5.4.1 Planteamento	45
5.4.2 Resultados	45
5.4.3 Discusión	45
5.4.4 Conclusíons	46
5.5 Regularización	46
5.5.1 Planteamento	46
5.5.2 Resultados	46
5.5.3 Discusión	47
5.5.4 Conclusíons	48
5.6 Batch size	48

5.6.1	Planteamento	48
5.6.2	Resultados	48
5.6.3	Discusión	49
5.6.4	Conclusíons	50
5.7	Estratexias de mostraxe	50
5.7.1	Plantexamento	50
5.7.2	Resultados	51
5.7.3	Discusión	51
5.7.4	Conclusíons	51
5.8	Inicialización	52
5.8.1	Planteamento	52
5.8.2	Resultados	52
5.8.3	Discusión	53
5.8.4	Conclusíons	53
5.9	Axuste dinámico do batch size	53
5.9.1	Planteamento	53
5.9.2	Resultados	53
5.9.3	Discusión	54
5.9.4	Conclusíons	54
5.10	Comparativa e resumo de resultados	54
5.10.1	Rendimento por dataset	54
5.10.2	Comparación de funcións de activación	55
5.10.3	Impacto dos parámetros principais	55
5.10.4	Limitacións identificadas	55
5.10.5	Conclusíons xerais	56
6	Conclusíons	57
7	Traballo futuro	59
7.1	Arquitecturas alternativas	59
7.2	Invertibilidade	59
7.3	Enfoque híbrido	60
A	Material adicional	62
A.1	Anexo regularization	62
A.1.1	Figuras experimentos de regularización	62
Glosario		65

Índice de Figuras

2.1	Imaxes do ollo humano, extraídas de [3]. Á esquerda, vista lateral do ollo anotada. Á dereita, retinografía do ollo anotada.	4
2.2	Exemplo de rexistro de imaxes de retina [4]	15
2.3	Visualización do campo de vectores de deformación (DFV). Á esquerda, representación mediante frechas. Á dereita, esta deformación aplicada a unha cuadrícula.	15
2.4	Proceso de rexistro de imaxes iterativo	16
2.5	Comparación de publicacións ó longo do tempo que relacionadas co rexistro de imaxes. Datos extraídos de Scopus [5], realizando as consultas: "TITLE-ABS-KEY(image AND registration) AND NOT(deep AND learning)" e "TITLE-ABS-KEY(image AND registration) AND (deep AND learning)"	17
3.1	Diagramas de Gantt da planificación do proxecto e duración real de cada fase .	22
4.1	Inicialización primera capa	25
4.2	Inicialización seguintes capas	25
4.3	Regulizador Jacobiano	28
4.4	Regulizador Hiperelástico.	29
4.5	Regulizador Bending Energy	29
4.6	Regulizador Jacobiano en 2D	29
4.7	Regulizador Hiperelástico en 2D	29
4.8	Regulizador Bending Energy en 2D	30
4.9	Clasificación dos pares de imaxes en categorías.	34
4.10	Exemplo de imaxes do conxunto de datos FIRE [6] cos puntos de control indicados. De esquerda a dereita, categorías <i>S</i> , <i>P</i> , <i>A</i>	34
4.11	Mapas de calor de mostraxe	37
4.12	Cálculo do erro de rexistro mediante a distancia euclidiana.	37
4.13	Gráfico de avaliación FIRE, [6]	38

4.14 Norma de Frobenius dunha matriz $A \in \mathbb{R}^{m \times n}$, onde a_{ij} son os elementos da matriz A	38
4.15 Exemplos de avaliación visual: (a) Composición de imaxes en checkerboard, (b) Composición de imaxes por cor, (c) Visualización dos vectores de desprazamento.	39
5.1 Métricas dataset FIRE	41
5.2 Exemplos de rexistro: combinacións de dataset (FIRE/RFMID), función de activación (relu/SIREN) e éxito.	43
5.3 Comparación de diferentes funcións de perda sobre imaxes de FIRE e RFMID . .	44
5.4 Comparación de diferentes resolucións de perda sobre imaxes de FIRE e RFMID. Menor distancia media é mellor.	45
5.5 Comparación do impacto da regularización hiperelástica sobre os datasets FIRE e RFMID para modelos ReLU e SIREN	47
5.6 Exemplos de rexistro con ausencia e exceso de regularización	47
5.7 Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID	49
5.8 Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset FIRE	49
5.9 Comparación do rendemento da rede con diferentes batch sizes sobre imaxes da categoría S do dataset FIRE	50
5.10 Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset RFMID	51
5.11 Exemplos das diferentes inicializacións ca función de activación RELU	52
5.12 Exemplos das diferentes inicializacións ca función de activación SIREN	52
5.13 Resultados da lotería de inicialización	52
5.14 Resultados de usar distinto número de fases	53
A.1 Mapa de calor cos resultados de diferentes combinacións de termos de regularización e funcións de activación sobre os datasets FIRE e RFMID	63

Índice de Táboas

3.1	Comparativa entre os servidores utilizados	20
3.2	Estimación de custos dos recursos humanos (IVE incluído)	21
4.1	Replicación dos resultados de IDIR	32

Capítulo 1

Introducción

Neste primer capítulo expónense as motivacións e obxectivos deste traballo. Ademais, detallarase a estrutura da memoria e os apartados que a conforman.

1.1 Motivación

A oftalmoloxía válese da análise de imaxes obtidas por diversos métodos para realizar diagnósticos e seguimentos precisos. Non obstante, dado que estas imaxes poden prover de distintas modalidades e ser tomadas dende distintos puntos no espazo ou en instantes separados no tempo, é preciso aliñalas para poder comparalas de xeito efectivo. O aliñamento, tamén denominado rexistro, consiste en deformar dúas ou máis imaxes co obxectivo de que as características de interese se atopen na mesma posición (superpostas). Este é un proceso tedioso e propenso a errores cando se realiza manualmente, polo que calquer mellora nel é de gran utilidade para que os profesionais da saúde, que poderán adicar máis tempo a tarefas más relevantes. Este problema é axeitadao para ser automatizado, xa que non require dun xuízo clínico, senón que se basea na comparación de características visuais das imaxes.

Existen diversas técnicas para realizar aliñamento de imaxes automático, especialmente ca chegada do aprendizaxe profundo á visión por computador para imaxes médicas, onde é habitual o uso de redes neuronais convolucionais ([CNN](#)) [7]. Estos métodos, aínda que efectivos, teñen limitacións. Unha delas é que requieren dunha gran cantidade de datos para o seu adestramento, o que pode ser un problema en campos como a medicina, onde a obtención de datos é cara e complexa. Ademais, xeralmente teñen unha precisión menor ca métodos automáticos convencionais, pese a que si que son más rápidos [8]. Adaptando o traballo realizado por IDIR [2], preténdese aplicar redes de representación implícita para o aliñamento de imaxes oftalmolóxicas, para determinar se poden superar as limitacións dos métodos anteriores.

1.2 Obxectivos

Neste traballo explorarase o uso de redes de representación implícita para o aliñamento de imaxes oftalmolóxicas, para determinar se son aptas para esta tarefa e se poden superar as limitacións dos métodos anteriores. Para iso, os obxectivos específicos son:

- Adaptar o traballo de IDIR [2] para aplicalo a imaxes oftalmolóxicas de dúas dimensións.
- Comparar o rendemento do método proposto nos conxuntos de datos de FIRE [6] e RFMID [9].
- Analizar a influencia de distintos parámetros no rendemento, en particular a influencia da función de activación SIREN.

1.3 Estrutura

Nesta sección detallarase a estrutura da memoria e os apartados que a conforman.

- **Capítulo 1: Introdución:** neste capítulo introdúcese o traballo, explicando as motivacións e obxectivos do mesmo.
- **Capítulo 2: Contexto:** neste capítulo explicarase o contexto do traballo, introducindo conceptos básicos de visión por computador e imaxes médicas, así como o estado da arte en aliñamento de imaxes.
- **Capítulo 3: Metodoloxía e planificación:** neste capítulo explicarase a metodoloxía empregada e a planificación do traballo.
- **Capítulo 4: Traballo realizado:** neste capítulo describese o traballo realizado.
- **Capítulo 5: Experimentos e resultados:** neste capítulo presentaranse os experimentos realizados e discutiránse os resultados obtidos.
- **Capítulo 6: Conclusíons:** neste capítulo resúmense as conclusíons do traballo e as súas implicacións.
- **Capítulo 7: Traballo futuro:** neste capítulo propoñeranse liñas de traballo futuro.

Capítulo 2

Contexto

NESTE apartado introdúcense o contexto relevante a este traballo que provee os conceptos básicos necesarios para a súa comprensión. Para elo describese o campo da oftalmoloxía e a imaxe médica, así como o estado da arte en aliñamento de imaxes.

2.1 Oftalmoloxía

A oftalmoloxía é a especialidade médica encargada do estudo e tratamiento das enfermedades dos ollos, incluíndo o globo ocular, a súa musculatura, o sistema lagrimal e as pálpebras. O ollo humano é un dos órganos dos que mais dependemos e maior cantidade de información sensorial aporta, así como un dos mais complexos do noso corpo [10].

A importancia da oftalmoloxía radica non só no tratamento das enfermedades oculares, senón tamén na súa capacidade para proporcionar información valiosa sobre o estado de saúde xeral do paciente. A observación directa dos vasos sanguíneos e do tecido neuronal 'in vivo' permite aos oftalmólogos detectar signos precoces de diversas enfermedades sistémicas. Por exemplo, o glaucoma, que non presenta síntomas nas súas etapas iniciais, pode ser diagnosticado mediante exames regulares da presión ocular e do nervio óptico [11]. Esta capacidade de diagnóstico precoz fai da oftalmoloxía unha especialidade fundamental na prevención e no mantemento da saúde visual e xeral do paciente.

2.1.1 Anatomía do ollo humano

O ollo encargase de captar a luz e transformala en impulsos eléctricos que se envían ao cerebro. Esta información é interpretada polo cerebro, que mediante mecanismos como a atención e a memoria, permite a percepción visual. [12] O ollo humano está composto por varias estruturas, cada unha cunha función específica que permite a percepción visual [13]. Entre elas destacan:

- Córnea e Cristalino: actúan xuntas para enfocar a luz na retina. A córnea, situada na parte exterior do ollo, proporciona maior parte da capacidade refractiva, mentres que o cristalino, unha lente flexible, axusta o enfoque para obxectos a diferentes distancias.
- Pupila e Iris: regulan a cantidade de luz que entra no ollo. O iris, a parte coloreada do ollo, expándezese ou contráese para controlar o tamaño da pupila, o orificio central.
- Retina: unha capa de células sensibles á luz (fotorreceptores) que converten os estímulos luminosos en sinais eléctricas, procesados inicialmente na retina mesma.
- Nervio óptico: transporta as sinais eléctricas xeradas na retina ata o cerebro, onde se interpretan como imaxes.
- Disco óptico: tamén coñecido como "punto cego", é a área onde o nervio óptico sae do ollo; carece de fotorreceptores.
- Vasos sanguíneos: distribúen os nutrientes e o oxíxeno necesarios á retina e eliminan os seus residuos metabólicos.

A figura 2.1 mostra estas estruturas localizadas en imaxes.

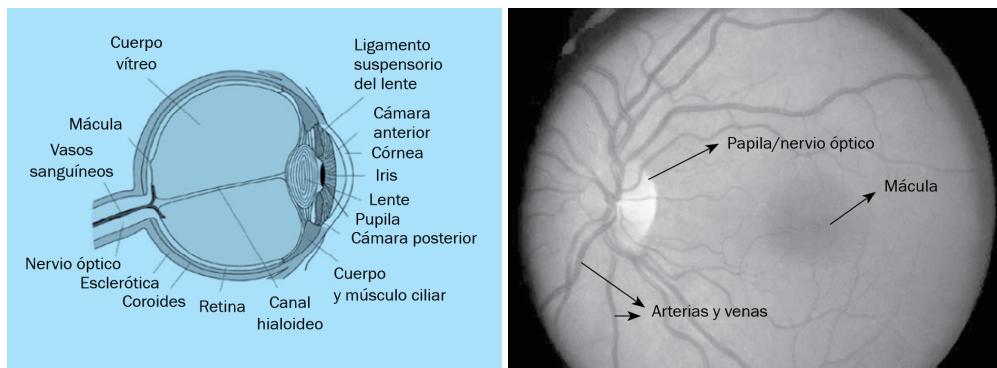


Figura 2.1: Imaxes do ollo humano, extraídas de [3]. Á esquerda, vista lateral do ollo anotada. Á dereita, retinografía do ollo anotada.

2.1.2 Imaxe oftalmolóxica

Existen diversas modalidades de imaxe médica que permiten observar o ollo, cada unha con diferentes propiedades e aplicacións. Entre elas inclúense a fotografía de fondo de ollo, a tomografía de coherencia óptica (OCT) e a angiografía con fluoresceína [14].

Este traballo céntrase na fotografía de fondo de ollo entre outras razóns polo seu uso común na práctica clínica. Isto é débese en gran parte á súa accesibilidade, requerindo equipo

mais barato e menor entrenamiento comparada cas outras modalidades. Ademais, é unha técnica non invasiva e rápida de realizar, o que a fai preferible na maioría dos casos [15].

Para realizala faise uso dunha cámara especial denominada retinógrafo, e xeralmente require da previa dilatación da pupila do paciente. Desta forma permítese maior entrada de luz nos ollos, o que provoca unha mellor visualización da retina e mellora a calidade da imaxe. Un especialista pode analizar a retinografía para detectar signos de enfermidades como a retinopatía diabética, a hipertensión ou a dexeneración macular [16].

2.2 Rexistro de imaxes

O rexistro de imaxes é un proceso que consiste en, sobre dúas ou mais imaxes, determinar a correspondencia espacial entre elas e alinealas nun sistema de coordenadas común, co obxectivo de que as características de interese se atopen na mesma posición.

O rexistro de imaxes ten utilidade en moitos campos diferentes como a imaxe satelital, xeografía, robótica... mais o campo da imaxe médica é dos mais interesantes póla súa aplicación práctica e é o que se aborda neste traballo [17]. Estas imaxes poden variar a nivel temporal, espacial, de dimensión ou de modalidade.

No ámbito da saúde un rexistro adecuado pode empregarse para comparar imaxes dun mesmo paciente tomadas en distintos momentos, en distintas modalidades ou para comparar entre diferentes pacientes. Isto permite a revisión do avance dunha enfermidade ao longo do tempo, a fusión de imaxes de distintas modalidades ou a detección de patróns comúns entre distintos individuos. A fusión de imaxes permite interpretar moito mellor a información dispoñible nelas, e é de gran axuda para guiar aos médicos na toma de decisións. Tamén é útil para correxir os movementos involuntarios do paciente durante a adquisición de imaxes, como no caso da respiración en imaxes de pulmóns, ou para a intervención guiada por imaxe (IGRT) que non podería funcionar sen a utilización axeitada de técnicas de rexistro de imaxes [18].

Ata recentemente, gran parte do traballo de rexistro facíase de forma manual por expertos con software como BigWarp [19], e dependía das habilidades do profesional para detectar as características de interese e realizar o aliñamento. Isto facía que o proceso fose lento e propenso a erros, ademais de pouco práctico para grandes volumes de imaxes.

Na figura 2.2 móstrase un exemplo de rexistro de imaxes de retina, onde se pode observar como as imaxes son aliñadas para que as estruturas anatómicas coincidan.

2.2.1 Categorías de rexistro

O rexistro de imaxes pode ser clasificado en distintas categorías segundo as súas características.

- **Segundo o número de imaxes:**

- *Par a par*: O rexistro realiza entre dúas imaxes, unha fixa e unha móbil.
- *Múltiple*: Rexístranse varias imaxes simultaneamente, buscando unha correspondencia global.

• **Segundo a modalidade:**

- *Intra-modalidade*: As imaxes pertenecen á mesma modalidade (por exemplo, dúas retinografías).
- *Inter-modalidade*: As imaxes proveñen de modalidades diferentes (por exemplo, retinografía e OCT).

• **Segundo o tipo de transformación:**

- *Ríxida*: Só permite traslación e rotación, mantendo as distancias e ángulos.
- *Afín*: Ademais de traslación e rotación, permite escalado e cizallamento.
- *Deformable (non ríxida)*: Permite deformacións locais complexas e non lineais.
- *Difeomórfica*: Transformación non ríxida que é continua, invertible e diferenciable en todo o seu dominio. Se non tén esta característica, non se pode garantir que a transformación sexa reversíbel, polo que son preferidas en moitos casos [20].

• **Segundo o grao de automatización:** [21]

- *Manual*: O usuario selecciona puntos de control ou axusta parámetros.
- *Automático*: O proceso realiza sen intervención humana, mediante algoritmos.
- *Semiautomático*: Combina intervención manual e automática.

• **Segundo a natureza da transformación:**

- *Simétrico*: A transformación é consistente en ambas direccións entre as imaxes.
- *Asimétrico*: A transformación calcúlase só nun sentido. Cando se traballa con imaxes de forma asimétrica, a imaxe de referencia denominase imaxe fixa e a imaxe que se quere rexistrar imaxe móvil.

• **Metodoloxía**

– **Técnicas clásicas** [22]

- * *Por correlación*: Optimizan funcións de correlación entre as intensidades das imaxes para atopar a mellor aliñación.
- * *Por entropía*: Baseados na maximización da información mutua, resultan especialmente útiles en rexistro multimodal.

- * *Por contornos*: Empregan características xeométricas, como bordes ou liñas estruturais, para establecer correspondencias espaciais.
- * *Por wavelet*: Utilizan descomposicións en diferentes frecuencias e resolucións para capturar información relevante.
- **Técnicas baseadas en aprendizaxe [8, 21]**
 - * *Iterativo profundo*: Estratexias que aplican redes neurais de forma recursiva para refinar progresivamente a transformación estimada.
 - * **Aprendizaxe supervisada**
 - *Totalmente supervisado*: Requiere datos de adestramento con alíñacionés precisas (ground truth), que permiten optimizar directamente a tarefa de rexistro.
 - *Debilmente supervisado*: Emprega información parcial ou indirecta, como anotacións espaciais limitadas ou métricas proxy.
 - *Supervisión dual*: Combina diferentes formas de supervisión, como etiquetas explícitas e funcións de perda auto-supervisadas, para aumentar a xeneralización.
 - * **Aprendizaxe non supervisada**
 - *Métrica de similitude*: Optimiza criterios de similitud (e.g., NCC, SSIM) sen necesidade de datos anotados, favorecendo adaptación a novos dominios.
 - *Baseado en GANs*: Emprega arquitecturas adversarias para aprender transformacións realistas entre imaxes sen supervisión directa.
 - * *Por similitude*: Modelos que aprenden funcións de custo diferenciables que maximizan a coincidencia entre as imaxes de entrada.
 - * *Aprendizaxe por reforzo*: Formula o rexistro como un problema de toma de decisións, onde un axente aprende políticas óptimas mediante retroalimentación ambiental.

As transformacións lineais globais soen representanse en matrices de transformación, onde cada elemento da matriz representa un parámetro da transformación.

No caso de transformacións mais complexas, utilizanse campos de vectores de deformación (DFVs), que permite representar deformacións locais na imaxe, facendoa moito mais flexible para representar transformacións non lineais e detalladas. Os DFVs adoitan ser representados cunha matriz de igual tamaño á imaxe, onde cada elemento representa un vector que indica a dirección e a magnitud da deformación.

Na figura 2.3 móstranse dúas formas de visualizar un DFV: mediante frechas que indican a dirección e magnitud da deformación, e aplicando a deformación a unha cuadrícula para ver como se distorsiona.

2.2.2 Estado da arte

O rexistro de imaxes médicas constitúe unha área de investigación fundamental que experimentou importantes avances nas últimas décadas. Neste ámbito, a precisión e a robustez do rexistro cobran especial relevancia, xa que son empregados para o diagnóstico e seguimento de enfermidades, así como para a planificación de tratamentos cirúrxicos. No eido da oftalmoloxía, os métodos que funcionan ben en varios dominios de imaxe médica (cerebro, pulmóns, etc) adoitan requirir de axustes para funcionar en retinas, polo que hai un estado da arte paralelo.

A evolución dos métodos de rexistro en retinografías reflicte a transición dende enfoques puramente algorítmicos cara metodoloxías híbridas, onde publicacións recentes como HybridRetina [23] mostran como para acadar os mellores resultados é beneficioso combinar ambos enfoques, aproveitando a precisión dos métodos clásicos e a adaptabilidade dos métodos de aprendizaxe automática.

2.2.2.1 Métodos clásicos

Os métodos clásicos de rexistro de imaxes médicas poden clasificarse en dúas categorías principais: Aqueles baseados en similitude de imaxe (IBR) e aqueles baseados en características (FBR). Tamén existen métodos híbridos que combinan ambos enfoques [24]. O resultado final pode ser os parámetros da transformación ou a imaxe fusionada.

Métodos baseados en similitude de imaxe O rexistro realiza comparando os valores de intensidade dos píxeles ou voxels mediante unha métrica de similitude entre a imaxe fixa e a imaxe móbil. Este enfoque tende a requerir de múltiples iteracións para converxer, nas cales calcúlase o grado de semellanza entre as imaxes e actualízanse os parámetros da transformación utilizando un mecanismo de optimización ata que se cumpran os criterios de terminación.

Os métodos de rexistro tradicionais teñen tres compoñentes principais: a métrica de similitude, o optimizador e o modelo de transformación.

A figura 2.4 mostra un diagrama do proceso de rexistro iterativo.

Métodos baseados en características O rexistro realiza identificando e emparellando características salientables entre as imaxes, como puntos, liñas ou bordes. Tipicamente, estes métodos teñen 3 pasos principais:

- **Detección de puntos de interese:** Identificación de puntos ou rexións salientables nas imaxes, como bordes, esquinas ou texturas. Para isto poden utilizarse utilizan algoritmos como SIFT [25], SURF [26], BRISK [27] ou FREAK [28].

- **Descripción de características:** os puntos detectados son descritos e comparados entre imaxes usando descriptores e algoritmos de emparellamento como FLANN [29] ou RANSAC [30].
- **Estimación da transformación:** unha vez atopadas as correspondencias, calcúlase a transformación que aliña as imaxes.

Algúns dos métodos tradicionalmente máis utilizados neste campo son **GDB-ICP** [31] e **Harris-PIIFD** [32]. Este último utiliza o algoritmo **Harris** [33] para a detección de puntos de interese, describenos con **PIIFD**, e emparéllanse usando **BBF** [34]. Finalmente, refínanse as coincidencias e escóllese a transformación (ríxida, afín ou polinomial) segundo o número de pares de puntos válidos. Sobre esta base propuxérónse varias melloras para adaptalo ao rexistro multimodal de retinas como **UR-SIFT** [35] ou **GMM** [36].

Unha vantaxe deste enfoque é a capacidade para rexistrar imaxes con grandes variacións locais ou modalidades diferentes, xa que non depende tanto da semellanza global entre as imaxes.

Outros métodos clásicos relevantes no campo da imaxe de ollo inclúen **REMPE** [37], que estima simultaneamente a pose das cámaras e a forma do ollo. Fai uso de un modelo elipsoidal para o ollo e estima a posición das cámaras con RANSAC, para logo refinala cunha variante de **PSO** [38].

Tamén existen múltiples programas que fan uso de estos métodos en ferramentas para facilitar o rexistro de imaxes, como SimpleITK [39], Elastix [40] ou ANTs [41].

2.2.2.2 Métodos de aprendizaxe profunda

Ca chegada dos métodos de aprendizaxe profunda á imaxe médica, comenzaron a empregarse redes neuronais para realizar o aliñamento de imaxes. Existe un gran interés polos métodos baseados en aprendizaxe profundo, como se reflexa no crecente número de publicacións no campo. Na figura 2.5 móstrase a evolución do número de publicacións sobre rexistro de imaxes, diferenciando entre os métodos baseados en aprendizaxe profunda e os métodos tradicionais.

Estos métodos tenden a requerir dunha gran cantidade de datos para ser adestrados, o que pode ser unha desvantaxa xa que en moitos casos non se dispoñen de bases de datos anotadas do tamaño necesario. Os métodos de aprendizaxe profunda poden ser clasificados en dous tipos según se requiran de s anotados ou non na etapa de entrenamento: supervisados (requíren anotacións) e non supervisados (non requíren anotacións) [42].

Segundo o grado de supervisión utilizado na etapa de entrenamiento, os métodos supervisados poden dividirse en supervisados ou débilmente supervisados. O rexistro totalmente supervisado

fai uso de DVF_s de referencia para supervisar o proceso de aprendizaxe, e o termo de perda adoita basearse na discrepancia entre os DVF_s de referencia e os DVF_s predichos.

O rexistro débilmente supervisado pode utilizar outras etiquetas de referencia implícitas, non baseadas en datos explícitos como os DFVs, senón que utilizan información indirecta para guiar o proceso de rexistro, como a semellanza entre as imaxes ou restriccións baseadas na forma ou límites anatómicos das estruturas. Máis de dous tipos de datos de referencia son frecuentemente utilizados para adestrar modelos de rexistro débilmente supervisados [8].

Os métodos non supervisados teñen a vantaxe de non requirir de datos anotados, o cal é unha gran vantaxe xa que un dos maiores retos para as redes de imaxes médicas é a recolección de datos de calidade para o adestramento [7]. A creación de conjuntos de DFVs anotados é un proceso laborioso e costoso, que normalmente só pode ser executado por especialistas, polo que os métodos de rexistro non supervisados son de gran interese. De forma similar aos métodos iterativos, é común empregar unha métrica de similitude entre as imaxes xunto con un termo de regularización para guiar o proceso de optimización evitando caer en transformación non realistas.

Os enfoques de aprendizaxe profundo son útiles tanto na súa capacidade de aprender a tarefa de rexistro de maneira end-to-end, como para substituír módulos concretos do proceso tradicional. Nese sentido, os métodos de aprendizaxe profunda poden ser categorizados según a tarefa do proceso de rexistro que substitúen.

Os métodos de aprendizaxe profunda poden substituír calquera destes pasos que forman os métodos de rexistro clásicos de forma independente ou en combinación.

En rexistro baseado en intensidade (IBR)

- **Métrica de similitude:** Os métodos de aprendizaxe profunda poden aprender métricas de similitude más robustas que as tradicionais. Estas métricas aprendidas poden ser más efectivas en imaxes multimodais ou con artefactos. Por exemplo, Czolbe et al. [43] propoñen dúas métricas de similitude semánticas que aprende a semellanza entre imaxes comparando as características de alto nivel extraídas. Presentan unha aproximación non supervisada que fai uso de autoencoders e outra semisupervisada que incorpora datos de segmentación.
- **Optimizador:** Os métodos de aprendizaxe profunda poden substituír o proceso de optimización iterativa tradicional por redes que aprenden a predecir directamente os parámetros de transformación óptimos. Unha aproximación común é empregar estes conjuntos de datos para optimizar unha CNN que, dadas dúas imaxes novas e non vistas, predice o DFV correspondente [44]. Durante o proceso de entrenamiento, a rede pode ter acceso aos DFVs ca deformación correcta, ou pódense obter indirectamente a través da optimización dunha métrica de similitude de imaxes.

- **Modelo de transformación:** Estes métodos aprenden representacións implícitas da transformación a través de redes neuronais, permitindo modelar deformacións mais complexas que os modelos paramétricos tradicionais. Os métodos como IDIR [2] encaixan nesta categoría, utilizando campos neuronais implícitos para representar as transformacións de rexistro.

En rexistro baseado en características (FBR)

- **Detectores de características:** As redes neuronais poden aprender a detectar puntos de interese mais robustos e repetibles que os detectores clásicos como . SuperPoint [45] introduce un detector de características baseado en redes neuronais que aprende a detectar puntos de interese e a describilos simultaneamente.
- **Descriptores de características:** Os descriptores aprendidos mediante redes neuronais poden capturar información mais discriminativa, mellorando a precisión do emparellamento posterior. Estes métodos aprenden representacións que son invariantes a transformacións específicas do dominio.
- **Emparellamento:** As redes neuronais poden aprender a realizar o emparellamento de características de forma mais robusta, especialmente en presenza de cambios de iluminación ou perspectiva. SuperGlue [46] é un exemplo de modelo que aprende a emparellar puntos de interese detectados utilizando unha arquitectura baseada en atención para capturar as relacións entre os puntos.

Métodos de regresión directa Os métodos desta rede neuronal aprende a mapear directamente desde un par de imaxes ata os parámetros da transformación, sen necesidade de optimización iterativa nin extracción explícita de características.

Tamén son denominados métodos de inferencia amortizada debido á capacidade de realizar múltiples inferencias (rexistros) tras un único proceso de entrenamento, en contraposición aos métodos tradicionais que requieren optimización individual para cada par de imaxes. Estes enfoques son útiles pola súa eficiencia computacional na fase de inferencia. Voxelmorph [47] é un dos frameworks mais utilizados no rexistro de imaxes deformable, facendo uso de modelos baseado en CNNs e que tamén permite incorporar información auxiliar (como segmentacións) se está dispoñible, mellorando así a precisión do rexistro.

Métodos como UDIR-Net [48] ou DIO [49] tamén implementan estas ideas.

GANs As GAN (Generative Adversarial Network) son un tipo de rede neuronal que consta de dous modelos que compiten entre si: un xerador e un discriminador. O xerador intenta crear datos falsos que sexan indistinguibles dos datos reais, mentres que o discriminador

intenta distinguir entre os datos reais e os datos xerados. Este proceso de competición mellora iterativamente a calidade dos datos xerados e o criterio do discriminador.

No contexto do rexistro de imaxes, as GANs poden ser utilizadas para aprender a transformación entre imaxes de forma non supervisada. O xerador pode ser adestrado para producir transformacións que aliñen a imaxe móbil coa imaxe fixa, mentres que o discriminador avalía a calidade do rexistro. Un exemplo de aplicación de GANs no rexistro de imaxes médicas é o traballo de Mahapatra et al. [50], onde se propón un modelo de rexistro de imaxes baseado en GANs que demostrou ser efectivo tanto en retinas como con resonancias magnéticas cardiovasculares.

2.3 Representación Neuronais Implícitas

A representación de coñecemento é un dos problemas más importantes na área da computación, e as redes profundas son unha das ferramentas más útiles, especialmente no campo da visión por computador. Tradicionalmente empréganse representacións discretas, onde o espazo de entrada é dividido en celdas e cada celda é asignada un valor (por exemplo nubes de puntos, matrices de píxeles ou véxeles...). Unha das principais desvantaxes destas representacións é que a súa complexidade increméntase rápidamente co número de dimensións representadas, ademais do custo de memoria asociado.

As representacións neuronais implícitas son un paradigma innovador que permite modelar sinais continuas mediante funcións parametrizadas por redes neuronais. Codifican a información como unha función continua, que mapea valores de entrada aos valores correspondientes de saída, en lugar de almacenar directamente valores de características o señales.

Representar o sinal como una función continua permite solucionar os problemas asociados á discretización e obténense outra serie de vantaxes.

As INR son moito mais eficientes debido á compresión da información que realizan de forma implícita. Ao mesmo tempo, permite un nivel de detalle non limitado pola resolución da imaxe, senón pola capacidade da rede. Ademais, as representacións continuas son diferenciables, o que permite o cálculo de gradientes e derivadas de forma analítica en lugar de ter que aproximalos por diferencias finitas. Isto tamén implica que as representacións implícitas son independentes da resolución, o que permite a reconstrucción en calquer escala espacial.

Tipicamente emprégase un MLP como arquitectura para representar a función implícita. Non obstante, o uso da función de activación ReLU tende a non obter os melhores resultados, debido a que son incapaces de representar deformacións locais sen afectar o seu comportamento global [51], polo que moita investigación diríxese a atopar alternativas que melloren a representación do sinal. [52]

Unha destas alternativas é SIREN [1], sobre a que profundizaremos mais adiante. Outras propostas inclúen [53] propón as funcións de activación gaussianas como alternativa a SIREN,

e argumenta que poden obter mellores representacións e mais robustas. [54] achega unha nova función de activación baseada en wavelets, que parece ser especialmente útil para a representación de imaxes.

As representacións implícitas poden ser clasificadas en dúas categorías: xeneralizables e sobreaxustadas [55]. As representacións sobreaxustadas céntranse en reproducir con precisión unha única sinal, mentres que as representacións xeneralizables poden modelar varias nunha mesma rede.

2.3.1 Aplicacións

As INR son utilizadas en todo tipo de campos, dende xeración de imaxes [56], pasando por reconstrucción de obxectos [57] [58] ou modelado de sinais complexas [59].

As representacións implícitas están a recibir cada vez máis atención da comunidade médica, e son especialmente útiles para as tarefas de imaxe inversa, que requiren a reconstrucción de representacións correctas a partir de datos incompletos ou ruidosos [60]. Métodos como NeRP, propuxeron o uso de representacións implícitas para a reconstrucción de imaxes de resonancia magnética a partir de datos incompletos, e obtiveron resultados comparables a métodos tradicionais [61].

NeRF fai uso de representacións implícitas para a sintetizar novos puntos de vista en escenas 3D [57], optimizando unha función volumétrica continua que modela a densidade de volume e a radiancia emitida en cada punto do espazo. Utilizan un MLP, cuxa entrada é unha única coordenada continua 5D (localización espacial (x, y, z) e dirección de visión (θ, φ)) e cuxa saída é a densidade de volume e a radiancia emitida dependente da vista nesa localización espacial. A única entrada necesaria para optimizar a súa representación é un conxunto de imaxes con poses de cámara coñecidas. Este traballo demostra que as representacións implícitas están capacitadas para modelar escenas 3D complexas con alta fidelidade visual.

As representacións implícitas tamén teñen bastante potencial no campo de planificación de traxectorias, onde se fai uso de INRs para modelar entornos e planificar traxectorias para un ou varios axentes [55]. A principal vantaxe de facelo desta forma frente á forma tradicional (algoritmos computacionalmente intensos, especialmente para multi-axentes) é a velocidade á que encontran solucións (por debaixo do milisegundo en GPUs). A maior desvantaxe é que non garanten a converxencia a unha solución óptima e sen colisións, mais os autores demostran que a calidade das traxectorias xeradas é adecuada para a maioría das aplicacións [62].

No ámbito médico, utilizanse este tipo de representacións para garantir a seguridade do paciente durante a cirurxia teleoperada e optimizar a traxectoria do robot para evitar colisións co paciente, por exemplo nas boca e gorxa [63]. Con este método, evítase a reconstrucción de mallas a partir de imaxes, que é un proceso costoso e imperfecto, e modélase mediante unha

INR a partir dos datos médicos dispoñibles. Os comandos de movemento da man do operador son tomados como entrada polo modelo, que logo de un proceso de optimización, xera unha secuencia de movementos libre de colisións que será enviada á man robótica. Tamén son utilizadas para crear reconstruccións 3D de pulmóns que mitigan as distorsións causadas polo movemento respiratorio [64].

Outro uso interesante das representacións neuronais implícitas é a compresión de imaxes. Algoritmos como COIN [65] representan os datos de entrada mediante redes neuronais implícitas (funcións que mapean coordenadas a valores RGB), logrando unha compresión eficiente e unha redución significativa do tempo de codificación en moitas modalidades.

2.4 Traballo proposto

Como mostramos neste capítulo, a pesar do potencial das representaciones neuronales implícitas en diversos dominios médicos, a súa aplicación específica ao rexistro de retinografías permanece inexplorada.

Este falta de investigación é especialmente relevante dadas as potenciais vantaxes que ofrecen as INRs, como a capacidade de modelar deformacións complexas e a súa independencia da resolución.

Baseándose no framework introducido por [2], propónse modificalo para adaptalo á tarefa de rexistro de retinografías. O obxectivo é conseguir rexistros consistentemente precisos, especialmente no dataset de FIRE que contén imaxes reais de retina.

Esta adaptación representa unha contribución novel ao estado da arte no rexistro de retinografías, explorando por primeira vez o potencial das representaciones neuronales implícitas neste dominio específico.

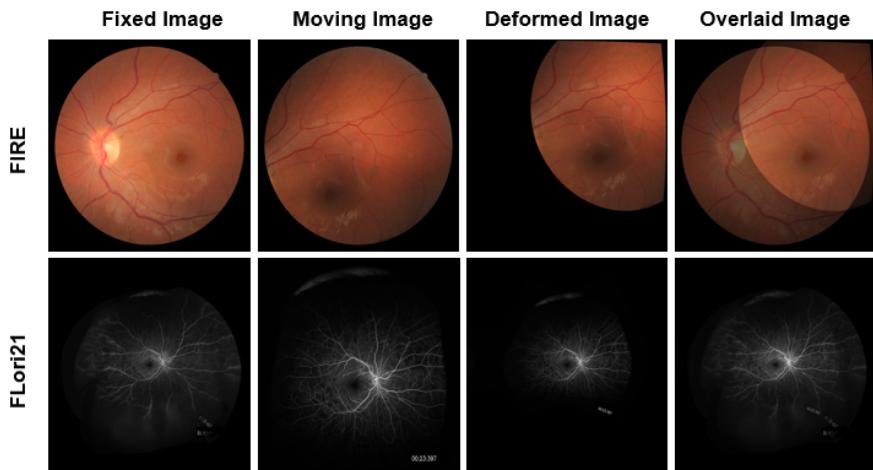


Figura 2.2: Exemplo de rexistro de imaxes de retina [4]

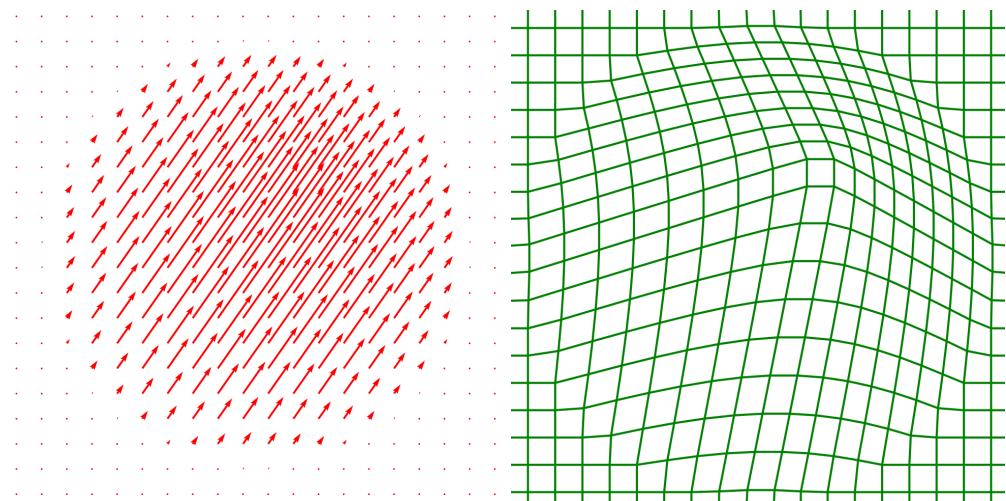


Figura 2.3: Visualización do campo de vectores de deformación (DFV). Á esquerda, representación mediante frechas. Á dereita, esta deformación aplicada a unha cuadrícula.

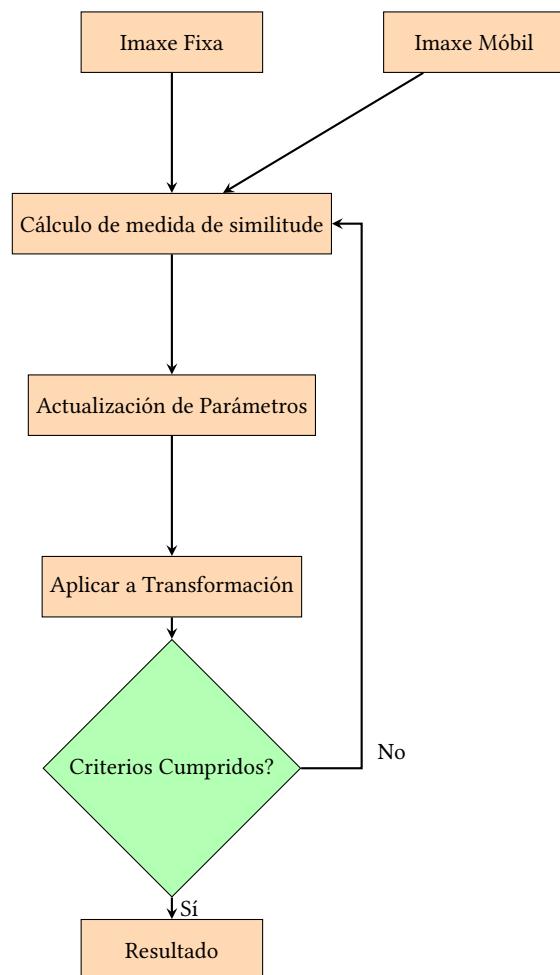


Figura 2.4: Proceso de rexistro de imaxes iterativo

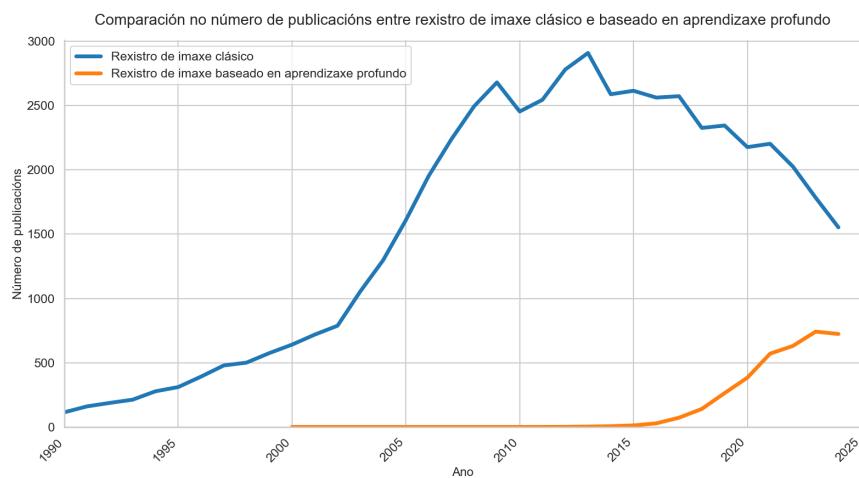


Figura 2.5: Comparación de publicacións ó longo do tempo que relacionadas co rexistro de imaxes. Datos extraídos de Scopus [5], realizando as consultas: "TITLE-ABS-KEY(image AND registration) AND NOT(deep AND learning)" e "TITLE-ABS-KEY(image AND registration) AND (deep AND learning)"

Capítulo 3

Metodoloxía e planificación

NESTA sección explícase a metodoloxía de traballo empregada para o desenvolvemento do proxecto, así como a planificación do mesmo. Ademais, describense os recursos utilizados e faise unha estimación dos custos asociados ao proxecto.

3.1 Metodoloxía do desenrolo

Ao ser un proxecto de investigación, a metodoloxía de traballo máis adecuada é unha metodoloxía iterativa e incremental, que permite poder adaptarse aos cambios que van xurdindo durante o desenrolo do proxecto. Esta metodoloxía permite obter un artefacto funcional ao final de cada iteración, o que permite obter retroalimentación constante sobre o desenrolo do proxecto. Cada iteración comeza cunha análise do que se quere conseguir, seguida das fases de deseño e codificación, e remata cunha fase de teste do produto desenvolvido.

3.2 Planificación do proxecto

O proxecto inicialmente divídese nas seguintes fases principais:

- **Revisión do estado da arte:**

- Estudo do dominio biolóxico: características das imaxes oftalmolóxicas, a súa importancia e aplicacións.
- Análise de traballos relacionados con IDIR, representacións implícitas e segmentación de imaxes oftalmolóxicas mediante redes neuronais.

Esta fase estimouse en aproximadamente 3 semanas de traballo, dada a necesidade de familiarizarse co contexto e identificar solucións previas relevantes.

- **Análise do traballo base:**

- Estudo en profundidade do código orixinal de IDIR.
- Replicación dos resultados orixinais para verificar o funcionamento correcto.

O esforzo estimado para esta tarefa foi de 2 semanas, considerando a análise do código e a posta en marcha do entorno.

- **Adaptación ao novo dominio:**

- Modificación da arquitectura para traballar con imaxes 2D en lugar de 4D.
- Implementación das adaptacións necesarias para imaxes oftalmolóxicas.

Esta etapa estímase unha duración de 5 semanas, debido á complexidade das modificacións e probas necesarias.

- **Avaliación e experimentación:**

- Deseño dunha metodoloxía de avaliação específica para o novo dominio.
- Realización de múltiples experimentos para optimizar o rendemento.
- Validación da efectividade en imaxes oftalmolóxicas.

Estímase un esforzo de 8 semanas, repartidas entre o deseño experimental, execución e análise dos resultados dos distintos experimentos.

- **Documentación:**

- Redacción da memoria final do proxecto.
- Análise e presentación dos resultados obtidos.

Para a documentación reserváronse 2 semanas, incluíndo a redacción, revisión e preparación dos anexos. A memoria será redactada ao longo de todo o proceso, pero nesta fase final revisaráse e completarase.

En total, estímase unha duración de 20 semanas para o proxecto, repartidas entre as distintas fases. Na figura 3.1 móstrase o diagrama de Gantt que resume a planificación do proxecto, indicando as fases principais e a súa duración estimada, así como o diagrama da duración real de cada fase.

3.3 Recursos utilizados

3.3.1 Software

Xa que parte do traballo consiste en adaptar un traballo previo, decidíuse empregar moito do mesmo software ca o traballo orixinal para facilitar a implementación e reproducibilidade.

O mais relevante é PyTorch, unha librería de código aberto para Python que facilita o desenrollo de redes neuronais. Utilizáronse as versións de Python 3.12.3 e CUDA 12.2. Tamén se empegan librerías de apoio como NumPy (para traballar con matrices), Matplotlib (visualización), OpenCV ou scikit-learn (manexo de imaxes).

Outro software empregado inclúe VSCode (IDE), Git (control de versións) e LaTeX (redacción de memoria).

3.3.2 Hardware

O proxecto foi desenrolado nun ordenador portátil conectado por ssh a un servidor con GPU. Utilizáronse dous servidores diferentes, un montado por min¹ e outro facilitado polo grupo de investigación VARPA (Visión Artificial y Reconocimiento de Patrones).

A gran parte dos experimentos foron realizado no primeiro, mais para poder executar o proxecto cas imaxes na súa resolución orixinal foi necesario empregar o segundo debido ás limitacións de memoria da GPU.

3.3.3 Conxuntos de datos

Para o desenrollo do proxecto empregáronse dous conxuntos de datos diferentes:

- **RFMID:** 3200 imaxes de fondo de ollo en cor con resolución 1712x1712.
- **FIRE:** 134 pares de imaxes de retinas en cor, cun tamaño de 2912×2912 pixels

Estos son descritos en maior detalle na sección 4.4.

Característica	Homelab	Servidor VARPA
Procesador	AMD Ryzen 9 5950X	AMD Ryzen Threadripper 3960X
GPU	NVIDIA RTX 3090	NVIDIA RTX A6000

Táboa 3.1: Comparativa entre os servidores utilizados

3.3.4 Estimación de custos

Os custos do hardware son ignorados xa que xa estaba dispoñible antes da realización do proxecto.

Os custos dos recursos humanos calcúlanse para un estudiante e dous titores, resultando nun custo estimado de 13.640€, IVE incluído. A táboa 3.2 mostra a estimación de custos dos recursos humanos desglosados, considerando un estudiante a 12€/hora e titores a 35€/hora.

¹ <https://blog.m19182.dev/writings/Building-my-Homelab>

Recurso	Custo por hora	Horas estimadas	Custo total
Estudiante	12€	880h	10.560€
Titor 1	35€	44h	1.540€
Titor 2	35€	44h	1.540€

Táboa 3.2: Estimación de custos dos recursos humanos (IVE incluído)

3.4 Seguimento da planificación

A planificación do proxecto foi revisada periodicamente segundo as fases do proxecto, e para identificar desviacións respecto ao plan inicial.

Pese a que nas fases iniciais do proxecto se respectou a planificación, a fase de adaptación ao novo dominio e a fase de avaliación e experimentación sufriron retrasos significativos.

A fase de adaptación ao novo dominio requiriu máis tempo do esperado debido á complexidade das modificacións necesarias para adaptar o modelo a imaxes 2D, así como á necesidade de realizar múltiples probas para garantir o correcto funcionamento do modelo adaptado. A fase de avaliación tamén se viu afectada, xa que requiriu máis tempo do esperado para deseñar unha metodoloxía de avaliación adecuada. Finalmente, a fase de experimentación requiriu máis tempo do previsto, en parte debido aos malos resultados obtidos inicialmente, que obligaron a revisar en profundidade o código e implementar novas probas para asegurar a correcta implementación. En total, isto conlevou un retraso de aproximadamente 18 semanas respecto á planificación inicial.

A fase final de análise de resultados e redacción da memoria tamén se viu afectada, aínda que en menor medida, o que conlevou un retraso adicional de 2 semanas, resultando nunha duración total do proxecto de 40 semanas, fronte ás 20 semanas inicialmente previstas.

3.4.1 Estimación de custo real

A estimación de custo do proxecto foi de 13.640€, como se indicou na sección 3.3.4. Non obstante, debido aos retrasos no desenvolvemento do proxecto, o custo real aumentou de forma proporcional ao tempo extra empregado.

Dado que o proxecto se estendeu durante 20 semanas máis do previsto, o custo real estimado ascende a 27.280€ (IVE incluído).

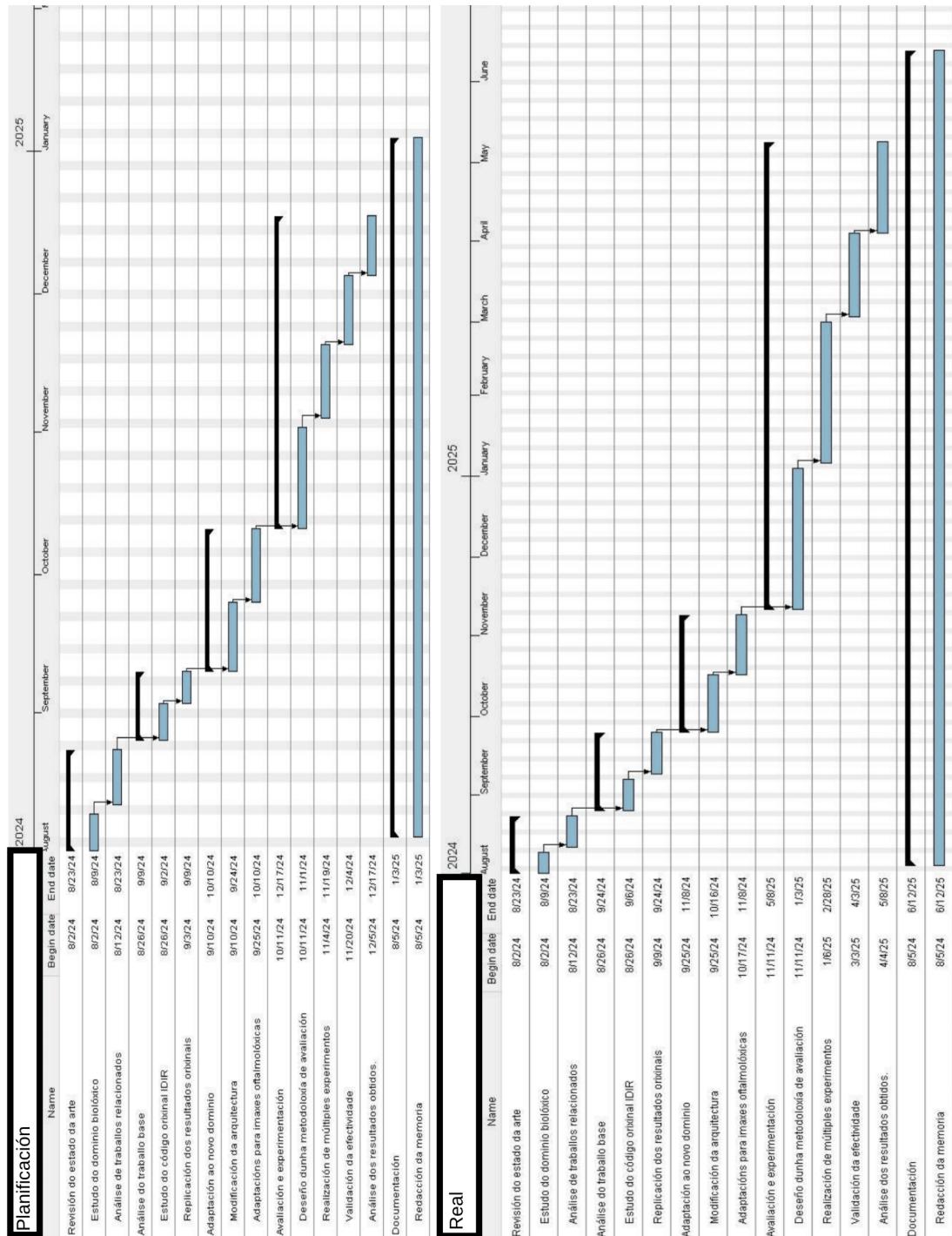


Figura 3.1: Diagramas de Gantt da planificación do proxecto e duración real de cada fase

Capítulo 4

Traballo Realizado

NESTE capítulo preséntase o traballo realizado para adaptar o framework IDIR ao rexistro de imaxes de retina en 2D. Comézase cunha vista xeral do proceso e unha descripción detallada do método IDIR orixinal, seguido da explicación das modificacións realizadas para adaptar o sistema ás características específicas das imaxes de fondo de ollo. Posteriormente, describense os conxuntos de datos empregados, o deseño experimental desenvolvido e os métodos de avaliación utilizados para validar os resultados.

4.1 Vista Xeral

O proxecto consiste en adaptar o framework de IDIR, ideado para o rexistro de 4DCT torácicas, para o rexistro de imaxes de fondo de ollo en 2D. Para isto foi necesario modificar gran parte do código, así como adaptar o proceso de rexistro e avaliación.

4.2 IDIR

IDIR (Implicit Deformable Image Registration) é un método de alíñamento de imaxes baseado en redes neuronais. A súa principal diferenza frente a unha rede convolucional tradicional é que,

O que se propón é optimizar directamente o DFV facendo uso dunha representación implícita, de forma que a deformación está representada nos propios pesos dun MLP [2].

Outros traballos como NIR [66] ou NODEO [67] propoñen métodos de rexistro similares que tamén fan uso de representacións implícitas das deformacións, aplicados a resonancias magnéticas do cerebro.

4.2.1 Arquitectura

Faise uso dun MLP de 3 capas, e determinaron experimentalmente que obtiñan mellor resultado con 256 unidades por capa que 128. Por cada epoch de entrenamento (2500 en total), 10000 puntos son muestrados aleatoriamente do espazo de coordenadas dentro da máscara. O térmico de perda é a 'normalized cross-correlation' entre os valores dos píxeles muestrados na imaxe fixa e os correspondentes da imaxe móbil. Utilizan Adam de optimizador, cun learning rate de 0.0001.

4.2.1.1 Función de activación

Unha elección estándar para a función de activación é **ReLU**:

$$f(x) = \max(0, x) = \begin{cases} x & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Non obstante, para redes de representación implícita como ca que estamos traballando, esta ten unha serie de desvantaxes.

As ReLUs teñen un sesgo cara a sinais de baixa frecuencia [51], o que significa que o modelo pode ter dificultades para representar pequenas deformacións locais no rexistro de imaxes.

[68] demostraron que a gran parte das funcións de activación utilizadas en redes neuronais (ReLU, tanh, sigmoide e todas as súas variantes) son incapaces de extrapolar función periódicas sinxelas debido á súa tendencia a converxer cara a comportamentos lineais cando se extrapolan fóra do rango de adestramento.

Existen varias formas de superar este sesgo, como preprocessar as coordenadas de entrada con funcións de activación periódicas [57] ou substituír a función de activación ReLU por unha función de activación periódica [1].

Neste traballo escollemos a segunda opción, utilizando unha función de activación periódica de tipo **SIREN**:

$$f(x) = \sin(ax + b), \quad \text{con } a, b \in \mathbb{R}$$

Unha vantaxe engadida das funcións de activación periódicas nas redes SIREN é que poden ser diferenciadas varias veces, o que expande substancialmente o conxunto de termos de regularización que se poden empregar na rede, como veremos na seguinte sección.

larger frequencies appear in the networks for weights with larger magnitudes.

Outros traballos como [57] non utiliza unha función de activación periódica, mais para a representación adecuada de zonas de alta frecuencia utilizaron codificación posicional, que xa as incorpora de forma implícita na rede con bós resultados.

Outra das vantaxes que ten SIREN é que é unha función suave ou infinitamente diferenciable, é dicir, que admite derivadas de calquer orde. Outros exemplos de funcións de activación infinitamente diferenciables son:

Sigmoide:

$$f(x) = \frac{1}{1 + e^{-x}}$$

Tangente Hiperbólica:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Softplus:

$$f(x) = \ln(1 + e^x)$$

Inicialización de pesos En [1] propuxeron unha inicialización específica para as redes SIREN, a cal consiste en inicializar a primeira capa de xeito que a función seno recorra múltiples períodos sobre o intervalo $[-1,1][-1,1]$. Isto conséguese multiplicando os pesos da primeira capa por un factor de escala ω_0 , sobre o cal recomandan $\omega_0=30$. A fórmula para a inicialización dos pesos da primeira capa é a seguinte:

$$w_i \sim U \left[-\frac{1}{n}, \frac{1}{n} \right]$$

Figura 4.1: Inicialización primera capa

onde n é o número de neuronas de entrada (o tamaño da capa anterior).

As seguintes capas inicialízanse da seguinte forma:

$$w_i \sim U \left[-\frac{\sqrt{\frac{6}{n}}}{w}, \frac{\sqrt{\frac{6}{n}}}{w} \right]$$

Figura 4.2: Inicialización seguintes capas

Desta forma asegurase que a entrada a cada activación sinusoidal está distribuída normalmente cunha desviación estándar de 1, o que debería mellorar a estabilidade e converxencia durante o adestramento da rede. Unha consecuencia desto é que, xa que os propios pesos da rede representan a deformación, inicialmente a rede comeza cunha deformacion moi similar en todos os casos, que o entrenamento deberá corrixir.

En [69] implementan unha versión simplificada de SIREN para facilitar o estudo destas, e propoñen melloras proceso de inicialización. Unha delas é utilizar a distribución Kaiming

(He) en lugar da uniforme. Tamén propoñen un método para escoller un valor de w apropriado segúin o problema a resolver.

4.2.1.2 Termos de loss

O termo de perda é a función que se optimiza durante o adestramento, e é o que guía a rede cara a unha solución óptima. Esta cuantifica a discrepancia entre a saída da rede e o resultado desexado.

Para a tarefa de rexistro de imaxes, utilízanse dúas categorías principais de métricas para avaliar o aliniamento entre imaxes: as métricas baseadas no erro e as métricas baseadas na similitude. As métricas baseadas no erro (MSE, L1...) miden as diferenzas píxel a píxel entre as imaxes, sendo más sensibles a diferenzas locais e proporcionando unha medida absoluta. As métricas baseadas na similitude (NCC, SSIM...) teñen en conta patróns estructurais e relacións estatísticas entre as imaxes, sendo más robustas fronte a variacións na iluminación e pequenos desprazamentos. [70]

Os principais termos de perda valorados para este traballo son:

- **MSE (Mean Squared Error):** Erro cadrático promedio entre á imaxe fixa e a móbil. É sensible a valores atípicos e ruido.

$$\text{MSE} = \mathbb{E}[(Y - \hat{Y})^2] = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

onde y_i é o valor do pixel da imaxe fixa, \hat{y}_i é o valor do pixel da imaxe móbil, e N é o número total de píxeles. [71]

Regulizador Hiperelástico en 2D

- **L1 (Mean Absolute Error):** Mide o error absoluto promedio. Menos sensible a valores atípicos que MSE.

$$\text{L1} = \mathbb{E}[|Y - \hat{Y}|] = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

- **Huber Loss:** Combina MSE e L1, sendo cadrática para errores pequenos e lineal para errores grandes.

$$\text{Huber}(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{se } |y - \hat{y}| \leq \delta \\ \delta(|y - \hat{y}| - \frac{1}{2}\delta) & \text{noutro caso} \end{cases}$$

onde δ é un hiperparámetro que define o punto de transición entre os comportamentos cadrático e lineal.

- **Smooth L1 Loss:** Similar a Huber Loss, pero cunha transición suave entre as rexións cadrática e lineal.

$$\text{SmoothL1}(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{se } |y - \hat{y}| \leq 1 \\ |y - \hat{y}| - \frac{1}{2} & \text{noutro caso} \end{cases}$$

- **NCC (Normalized Cross-Correlation):** Evalúa a similitude entre as dúas imaxes normalizando as súas intensidades. É invariante a cambios na iluminación.

$$\text{NCC} = \frac{\sum_{i=1}^N (y_i - \mu_y)(\hat{y}_i - \mu_{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \mu_y)^2 \sum_{i=1}^N (\hat{y}_i - \mu_{\hat{y}})^2}}$$

onde μ_y e $\mu_{\hat{y}}$ son as medias das imaxes fixa e móbil, respectivamente.

- **SSIM (Structural Similarity Index):** Evalúa a similitude estructural entre as dúas imaxes, considerando luminancia, contraste e estructura.

$$\text{SSIM}(y, \hat{y}) = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)}$$

onde μ_y , $\mu_{\hat{y}}$ son as medias, σ_y , $\sigma_{\hat{y}}$ son as desviacións estándar, $\sigma_{y\hat{y}}$ é a covarianza, e C_1 , C_2 son constantes para evitar divisiones entre cero. [71]

Debido á natureza das imaxes de retina, onde poden existir diferencias de iluminación e contraste entre as imaxes fixa e móbil, parece mais apropiado empregar métricas baseadas na similitude como NCC ou SSIM.

NCC utilizouse a implementación de <https://github.com/BDdeVos/TorchIR/blob/main/torchir/metrics.py>.

4.2.1.3 Termos de regularización

Debido a que o rexistro de imáxenes deformables é un problema mal planteadoo (ill-posed problem**), é común utilizar algún tipo de regularización sobre o DVF para evitar deformacións pouco realistas. Os métodos de rexistro baseados en redes neuronais convolucionais (CNN) representan os DVFs como mostras en una cuadrícula de véxeles, e polo tanto, solo se poden aproximar gradientes espaciais mediante esquemas de diferencias finitas (aproximar derivadas mediante cálculo numérico de diferencias entre valores adyacentes en la cuadrícula). Este é un proceso computacionalmente moi costoso e ineficiente, ademais implica errores de discretización e perdas de precisión.

Facendo uso de representacións implícitas, todas as operación son diferenciables, e os gradientes poden ser computados facilmente de forma analítica en lugar de ter que aproximalos, facendo uso da libréaría de autodiferenciación de PyTorch.

Utilizando ReLU como función de activación, a rede é diferenciable unha vez, mentres que utilizando unha función de activación periódica (como SIREN), a rede é diferenciable todas as veces que se precise. Desta forma, podemos calcular calquera número de termos de regularización e incluílos na optimización da rede.

Algúns exemplos de termos de regularización que se poden empregar son:

- Jacobian regularizer: O determinante Jacobiano da transformación ($\det \nabla \Phi$) nunha localización x é un indicador de estiramento ou compresión local. Un determinante Jacobiano negativo ou moi cercano a 0 indica que están a ocorrir dobreces e a transformación non será invertible. A matriz jacobiana é a matriz que contén todas as derivadas parciais da función de transformación (calculado mediante gradientes). O termo de regularización do Jacobiano penaliza os valores do determinante Jacobiano que se desvían de 1, tentando preservar áreas locais e evitar estiramentos ou dobreces extremas.

$$S^{jac}[\Phi] = \int_{\Omega} |1 - \det(\nabla \Phi)| dx$$

Figura 4.3: Regulizador Jacobiano

Ω representa o dominio ou rexión do espazo sobre o cal está definida a transformación Φ .

- Hyperelastic regularizer Tamén se poden engadir restriccións ao DVF con este termo proposto por [72]. Consiste en tres termos, un termo de lonxitude, un termo de área e un termo de volumen co obxectivo de controlar variacións nestes aspectos. O termo de lonxitude penaliza a variación da lonxitude dos vectores do DVF, sendo u a medida desplazamiento dun punto no espacio. A matriz de cofactores da matriz do Jacobiano da transformación controla o área. A función de máximo asegura que só as expansións que sobrepasan certo límite sexan penalizadas. O determinante da matriz do Jacobiana controla o volume, e ambas penalizan o crecemento e a contracción por igual. α_l , α_a e α_v son hiperparámetros que controlan a importancia de cada térmico.
- Bending energy penalty Pódese impoñer a suavidade da deformación empregando esta penalización proposta en [73], que require que as segundas derivadas do DVF sexan pequenas en todo o dominio, o que evita deformacións bruscas e discontinuas. Este termo non pode ser utilizado nunha rede que utilice ReLU como función de activación, xa que a segunda derivada de unha ReLU é sempre igual a 0.

Para a implementación neste traballo modificaronse todos estos termos para que funcionaran con transformacións de dúas dimensións en lugar de tres, sustituíndo o gradiente de 3 dimensión

$$S^{hyper}[\Phi] = \int_{\Omega} \left[\frac{1}{2} \alpha_1 |\nabla u|^2 + \alpha_a \varphi_c(\text{cof} \nabla \Phi) + \alpha_\nu \psi(\det \nabla \Phi) \right] dx,$$

Funcións convexas: $\varphi_c(C) = \sum_{i=1}^3 \max \left\{ \sum_{j=1}^3 C_{ji}^2 - 1, 0 \right\}^2$ and $\psi(v) = \frac{(v-1)^4}{v^2}$.

Figura 4.4: Regulizador Hiperelástico.

$$S^{bending}[\Phi] = \frac{1}{8} \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \left[\left(\frac{\partial^2 \Phi}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \Phi}{\partial y^2} \right)^2 + \left(\frac{\partial^2 \Phi}{\partial z^2} \right)^2 + 2 \left(\frac{\partial^2 \Phi}{\partial x \partial y} \right)^2 + 2 \left(\frac{\partial^2 \Phi}{\partial x \partial z} \right)^2 + 2 \left(\frac{\partial^2 \Phi}{\partial y \partial z} \right)^2 \right] dx dy dz$$

Figura 4.5: Regulizador Bending Energy

no Jacobiano por un de dúas, eliminando as derivadas parciais en z en bending energy e o termo de volume no termo hiperelástico.

- Jacobian regularizer:

$$S^{jac}[\Phi] = \int_{\Omega} |1 - \det(\nabla \Phi)| dx dy$$

Figura 4.6: Regulizador Jacobiano en 2D

- Hyperelastic regularizer:

$$S^{hyper}[\Phi] = \int_{\Omega} \left[\frac{1}{2} \alpha_1 |\nabla u|^2 + \alpha_a \varphi_c(\text{cof} \nabla \Phi) \right] dx dy,$$

$$\varphi_c(C) = \sum_{i=1}^2 \max \left\{ \sum_{j=1}^2 C_{ji}^2 - 1, 0 \right\}^2$$

Figura 4.7: Regulizador Hiperelástico en 2D

- Bending energy penalty:

$$S^{bending}[\Phi] = \frac{1}{8} \int_{-1}^1 \int_{-1}^1 \left[\left(\frac{\partial^2 \Phi}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \Phi}{\partial y^2} \right)^2 + 2 \left(\frac{\partial^2 \Phi}{\partial x \partial y} \right)^2 \right] dx dy$$

Figura 4.8: Regulizador Bending Energy en 2D

A regularización tamén ten un impacto significativo no tempo de computación, xa que require múltiples pasadas de retropropagación por época para calcular os distintos termos de penalización. Sen regularización só se fai 1 pasada para calcular o gradiente do termo de similitude da imaxe. Ca regularización do Jacobiano, ademais do termo de similitude, calcúlanse dúas derivadas (unha por dimensión) para obter o Jacobiano, resultando en 3 pasadas por época. Engadindo a regularización hiperelástica (sen termo de volume), é necesario calcular unha derivada adicional para o cofactor da matriz Jacobiana, facendo un total de 4 pasadas por época. Finalmente, ca penalización de enerxía de flexión, necesitanse derivadas segundas, o que implica 7 pasadas por época en total. No traballo orixinal de IDIR, chegaban a usar 13 pasadas debido a que traballan en 3D.

Se os termos de regularización teñen demasiada influencia sobre o termo de loss, a rede fará transformacións moi pequenas para evitar ser penalizada, o que resultará nunha transformación insuficiente. Por outro lado, se os termos son demasiado pequenos, a rede fará transformacións moi grandes, o que resulta nunha transformación irrealista e sobreaxustada. Isto é especialmente evidente no caso da función de activación SIREN, que tende a sobreaxustarse facilmente debido ao seu sesgo cara sinais de alta frecuencia. A cantidade óptima de regularización depende da parexa concreta de imaxes a alinear, polo que intentaremos determinar cal é a mellor para unha mostra de imaxes.

Ademais, pese a que os diferentes termos de regularización valoran diferentes aspectos, cabe ter en conta que tamén superposición nalgunhas das propiedades que valoran. Por exemplo, o regularizador hiperelástico pode considerarse un termo mais xeral que inclúe indirectamente penalizaciones do Jacobiano (ambos penalizan as "dobreces") e de suavidade das transformacións (como fai o bending pero en menor grado).

4.2.1.4 Learning rate e batch size

O learning rate é un parámetro do optimizador (Adam neste caso) que regula o tamaño dos axustes efectuados aos parámetros do modelo durante cada iteración de actualización. Determina a magnitud do cambio aplicado para minimizar a función de perda, afectando tanto a velocidade de converxencia como a estabilidade do proceso de aprendizaxe. Un learning rate demasiado alto pode provocar que a rede diverxa, mentres que un learning rate demasiado baixo pode resultar en converxencia lenta ou quedar atrapado en mínimos locais.

Debido á natureza da rede, o batch size utilizado ten unha relación directa co learning rate, polo que tentaremos determinar a relación óptima entre ambos.

Unha das heurísticas mais comúns para relacionar o learning rate e o batch size é a regla de escalado linear [74]. A regla indica que o learning rate óptimo debe escalarse linearmente co tamaño do batch size.

Unha forma de explicar isto é, xa que con batches más grandes temos unha mellor aproximación do gradiente real, é posible utilizar un learning rate maior sen que a rede diverxa. [75]

O batch size nesta rede representa o número de coordenadas amostradas aleatoriamente do espazo de coordenadas dentro da máscara.

4.2.2 Método

Sendo o obxectivo encontrar unha transformación espacial óptima entre a imaxe móbil e a imaxe fixa, é necesario obter a función de deformación $\Phi(x) = u(x) + x$ que mapea cada coordenada x na imaxe móbil a unha coordenada na imaxe fixa, de forma que a coordenada x na imaxe fixa corresponda anatomicamente á coordenada $\Phi(x)$ na imaxe móbil. Este problema pode ser formulado como un problema de optimización onde L_{data} é unha métrica de similitude entre as imaxes fixa (F) e móbil (M), L_{reg} é un termo de regularización na transformación Φ , e α é un termo de ponderación.

$$\hat{\Phi} = \underset{\Phi}{\operatorname{Arg\,min}} L_{data}(M \circ \Phi, F) + \alpha L_{reg}(\Phi) \quad (4.1)$$

A principal innovación que introduce IDIR[2] é que a transformación Φ está implícitamente representada na rede neuronal.

Comparado cunha CNN tradicional, esta rede non recibe valores de intensidade de píxel como entrada, senón que recibe coordenadas espaciais (continuas) e devolve unha nova coordenada. Xa que os pesos da rede definen a transformación, estos poden ser optimizados directamente facendo uso dunha métrica de similitude como función de perda.

Parametrizar a función de deformación como unha dentro dun ten varias vantaxes para o rexistro de imaxes. En primeiro lugar, a representación da transformación é continua e polo tanto independente da resolución da imaxe, grazas a iso o mesmo modelo poder ser empregado para imaxes de calquer tamaño, ao contrario dunha CNN tradicional que ten que ser adaptada para cada resolución.

Segundo, facelo desta forma permite aproveitar as capacidades de librerías como PyTorch para calcular os gradientes da transformación respecto das coordenadas. Isto permite obter gradientes más precisos que as aproximacións por diferencias finitas e permite aproveitar unha gran cantidade de literatura sobre regularización eficiente en imaxes médicas.

Terceiro, pódese modificar a función de activación empregada na rede para axustala ás

necesidades particulares da tarefa de rexistro de imaxes.

O NTK describe cómo un modelo de red neuronal responde a cambios en sus parámetros durante el entrenamiento, e dependendo da función de activación empregada, o NTK varía e a rede pode ser máis ou menos sensible a certas deformacións.

Finalmente, entrenaráse unha nova rede por cada parella de imaxes, sendo esta unha rede bastante pequena en comparación e prescindindo da necesidade de grandes conxuntos de datos para o seu adestramento.

4.2.3 Replicación de resultados

Replicáronse os resultados obtidos por Wolterink et al. [2] que se mostran na táboa 4.1.

Táboa 4.1: Replicación dos resultados de IDIR

Scan	IDIR / Replicación
1	0.76 (0.94) / 0.79 (0.92)
2	0.76 (0.94) / 0.71 (0.89)
3	0.94 (1.02) / 0.95 (1.01)
4	1.32 (1.27) / 1.32 (1.22)
5	1.23 (1.47) / 1.23 (1.46)
6	1.09 (1.03) / 1.15 (1.04)
7	1.12 (1.00) / 1.11 (0.99)
8	1.21 (1.29) / 1.20 (1.28)
9	1.22 (0.95) / 1.16 (0.99)
10	1.01 (1.05) / 1.09 (1.05)
Promedio	1.07 / 1.07 (1.08)

4.3 Adaptación a 2D

Para adaptar o modelo IDIR a 2D, é necesario modificar a arquitectura da rede para que funcione con imaxes bidimensionais. A arquitectura orixinal de IDIR ten unha entrada de 3 dimensións (x, y, z) e unha saída de 3 dimensións (dx, dy, dz), mentres que a nosa rede ten unha entrada de 2 dimensións (x, y) e unha saída de 2 dimensións (dx, dy).

Modificáronse as capas de entrada e saída da rede para que acepten coordenadas bidimensionais (orixinalmente $[3, 256, 256, 256, 3]$, agora $[2, 256, 256, 256, 2]$).

Foi necesario modificar as funcións de interpolación, que antes interpolaban valores tridimensionais e agora interpolan valores bidimensionais.

Tamén adaptáronse os termos de regularización para que funcionen con coordenadas bidimensionais, como se detalla na sección 4.2.1.3.

Finalmente, o proceso de avaliación tivo que implementarse dende 0 para adaptarse ao novo formato de imaxes e ground truth.

4.3.1 Proceso de Rexistro

Inicializase a rede e o resto de parámetros relevantes (función de activación, optimizador, métrica de loss, termos de regularización, etc).

A inicialización dos pesos da rede é especialmente relevante no caso da función de activación SIREN, que é sensible a valores iniciais. Tamén se xera un tensor de coordenadas inicial que contén todas as coordenadas dentro da máscara da imaxen fixa.

Cando comeza o adestramento por cada epoch repítese o seguinte proceso:

Mostreanse "batch size" puntos no tensor de coordenadas orixinal e estos pásanse pola rede, a cal devolve a transformación que predice para esas coordenadas. A continuación, aplícase esta transformación e calculase o loss entre a imaxe móbil transformada e a imaxe fixa. O valor de loss axústase según os termos de regularización usados, e finalmente realiza a retropropagación para actualizar os pesos da rede.

O proceso de mostraxe pode ser aleatorio ou utilizando unha estratexia específica.

Cabe destacar que a tarefa de rexistro de pulmóns é diferente da de retinografías. O tamaño das deformacións é moito maior, así como o grado de superposición entre as imaxes. Por este motivo, non podemos asumir que os parámetros óptimos para pulmóns sexan tamén os mellores para retinas.

4.4 Conxuntos de datos

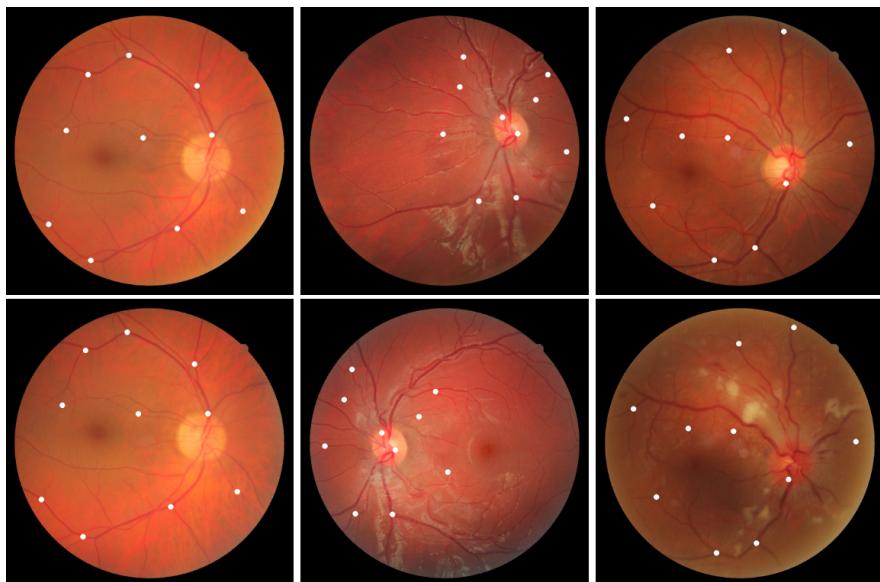
4.4.1 FIRE

Está composto por 134 pares de imaxes de retinas, con un tamaño de 2912×2912 pixels e un FOV de $45^\circ \times 45^\circ$. Están clasificadas en 3 categorías según o grado de superposición e a presenza de diferencias anatómicas: S, P e A. [6]

Categoría	Nº de pares de imaxes	Superposición (%)	Diferenzas Visuais
<i>S</i>	71	> 75	Non
<i>P</i>	49	< 75	Non
<i>A</i>	14	> 75	Si

Figura 4.9: Clasificación dos pares de imaxes en categorías.

Inclúe 10 puntos de referencia para cada imaxe, que se utilizan para a avaliación do rexistro, así como unha máscara por cada imaxe que indica a localización dos píxeles con información de cor.

Figura 4.10: Exemplo de imaxes do conxunto de datos FIRE [6] cos puntos de control indicados. De esquerda a dereita, categorías *S*, *P*, *A*.

4.4.2 RFMID

O conxunto de datos RFMID [9] proporciona 3200 imaxes de fondo de ollo en cor con resolución 1712x1712, etiquetadas según se teñen algunha anomalía ou non. Tamén proporciona etiquetas para 45 diferentes anomalías anotadas por expertos.

Para utilizarlo neste traballo, seleccionamos unha submostra e xeramos transformacións aleatorias. Gardamos as imaxes orixinais e as transformadas así como as matrices de transformación asociadas para a posterior avaliación. Tamén se divide entre transformacións de cor e de xeometría.

4.4.3 Diferencias entre os datasets

Unha vantaxe de utilizar dous conxutos de datos diferentes é que cada un deles ten características únicas que permiten avaliar o modelo en diferentes contextos. A principal diferencia é que RMiFD é un conxunto de datos sintético, no cal non introducimos diferenzas de cor e sempre teñen unha superposición do 100%, polo que o único que se avalia é a capacidade do modelo para realizar os rexistros xeométricos. Polo contrario, FIRE é un conxunto de datos real, no cal existen cambios na iluminación, contraste, superposición e demais diferencias visuais, polo que se avalia a capacidade do modelo para realizar rexistros en condicións moito más adversas.

4.5 Diseño de Experimentos

O deseño de experimentos é un proceso sistemático que busca determinar a influencia de diferentes factores sobre un resultado específico. Neste caso, o obxectivo é avaliar como diferentes parámetros afectan á calidade do rexistro de imaxes.

O custo computacional é un factor moi importante a ter en conta, xa que cada combinación de parámetros require un adestramento completo da rede por cada parella de imaxes, o que implica un alto custo enerxético e de tempo. Por exemplo, para probar unha combinación de parámetros sobre FIRE haberá que entrenar unha rede por cada parella de imaxes das 134 do conxunto de datos. A un tempo de adestramento de 3 minutos por parella, o adestramento completo levaría mais de 6 horas por cada combinación de parámetros, cunha pegada de memoria de arredor de 5 GB de VRAM. O custo temporal e de memoria dependen de varios factores, sendo os mais relevantes a regularización empregada, a resolución da imaxe, o tamaño do batch size e a función de activación. En concreto, SIREN ten un custo computacional moito maior que ReLU, requerindo arredor do doble de tempo e memoria para adestrar a rede.

Debido ao gran número de factores a ter en conta, adoptouse un enfoque de experimentación en fases.

Inicialmente realizáronse experimentos iniciais para identificar os rangos de parámetros más prometedores utilizando unha submostras representativa de 14 parellas de imaxes de cada categoría do dataset FIRE, co obxectivo: Reducir o espazo de parámetros para as fases posteriores. Nesta fase avaliouuse a métrica de loss, a resolución da imaxe, a regularización empregada e o tamaño do batch size.

Baseándose nos resultados da primeira fase, realizouse unha experimentación más exhaustiva para tentar mellorar o rendemento do rexistro, centrándose na estratexia de mostraxe, a inicialización dos pesos da rede e un axuste dinámico do batch size.

4.5.1 Metodoloxías Desenrroladas

Para este traballo tivéronse que desenrrolar varias metodoloxías específicas.

Algunhas delas inclúen a implementación de diferentes tipos de estratexias mostraxe.

Na estratexia de mostraxe intelixente, calcúlase unha máscara de probabilidade para cada imaxe, que se utiliza para seleccionar os puntos que se pasan á rede. Para calcular esta máscara, extráense mediante operadores de Sobel os vasos sanguíneos e mediante umbralización o disco óptico, que son as zonas onde se espera que haxa máis información, e dáselles maiores probabilidades de ser seleccionadas. Tamén se implementou unha estratexia de mostraxe ponderada, onde se seleccionan puntos aleatorios, pero con maior probabilidade de que caian nas zonas de interese (vasos sanguíneos e disco óptico).

Posteriormente introduciuse unha estratexia de mostraxe uniforme, onde se seleccionan un número fixo de puntos en cada imaxe, asegurando que están distribuídos uniformemente por toda a imaxe. É unha estratexia similar ao mostraxe aleatorio, pero garantindo que se cubre a maior parte posible da imaxe. Isto é relevante en experimentos con batch sizes pequenos onde unha mostraxe aleatoria non ten por que cubrir toda a imaxe de forma uniforme. Para implementalo empregouse unha distribución baseada na grella de Fibonacci (Fibonacci lattice), que permite repartir os puntos de maneira uniforme sobre a superficie circular da retina. A posición de cada punto calcúlase en coordenadas polares, asignando a cada punto un radio proporcional á raíz cadrada do seu índice dividido polo número total de puntos, e un ángulo proporcional ao índice multiplicado por 2π e dividido polo cadrado do número áureo (φ^2):

$$r_i = \sqrt{\frac{i}{N}}, \quad \theta_i = 2\pi \frac{i}{\varphi^2}$$

onde i é o índice do punto ($i = 1, \dots, N$), N é o número total de puntos e φ é o número áureo. Deste xeito, conséguese unha cobertura uniforme e eficiente da rexión de interese, evitando agrupamentos ou zonas baleiras. Na figura 4.11 pódense observar os diferentes tipos de mostraxe utilizados.

Outra metodoloxía desenrolada foi a implementación da lotería de inicialización, que consiste en probar diferentes inicializacions dos pesos da rede para determinar cal é mais beneficiosa; e o axuste dinámico do batch size, que consiste en aumentar o tamaño do batch size ao longo do adestramento. Para esta última, divídense os epochs en diferentes fases, onde cada fase fai uso dun tamaño de batch size diferente.

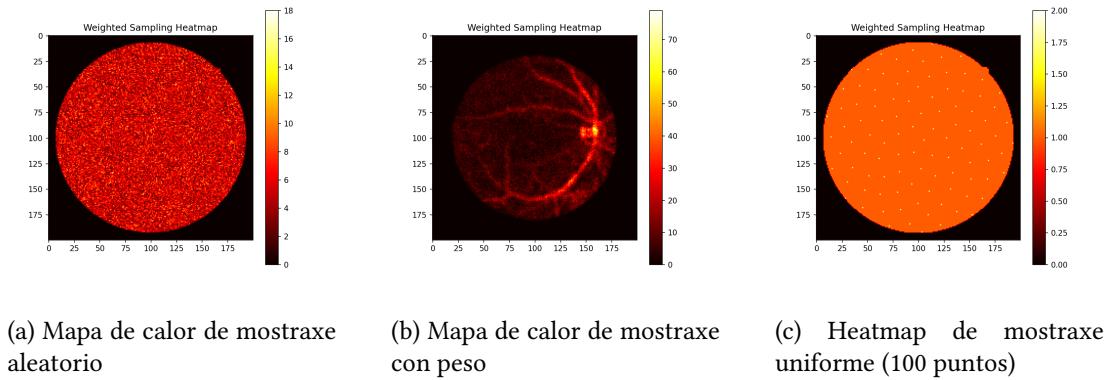


Figura 4.11: Mapas de calor de mostraxe

4.6 Métodos de Avaliación

A avaliación do rendemento do sistema de rexistro constitúe un aspecto fundamental para determinar a eficacia das modificacións implementadas. O proceso de avaliação divídese en dous enfoques complementarios: a avaliação cuantitativa, que emprega métricas numéricas obxectivas, e a avaliação cualitativa, que analiza os resultados de forma visual para detectar artefactos ou deformacións non desexadas que poidan escapar ás métricas numéricas.

Ambas avaliações son necesarias para obter unha visión completa da calidade do rexistro, xa que a avaliação cuantitativa pode non ser suficiente para detectar problemas visuais que non se reflictan nas métricas.

4.6.1 Avaliación Cuantitativa

Utilizamos como método de evaluación cuantitativa o proposto por FIRE [6] xerando un gráfico onde o eixo x representa o valor do límite de erro e o eixo y mostra a porcentaxe de pares de imaxes que foron rexistrados con éxito para cada límite de erro.

O erro de rexistro calcúlase mediante a distancia euclidiana media entre os puntos correspondentes nas imaxes fixa e móbil:

$$E = \frac{1}{N} \sum_{i=1}^N \|p_i^{\text{fixo}} - T(p_i^{\text{móvel}})\|$$

Figura 4.12: Cálculo do erro de rexistro mediante a distancia euclidiana.

onde N é o número de puntos de referencia, p son as coordenadas dos puntos e T é a transformación aplicada.

Cando o erro de rexistro entre un par de imaxes está por debaixo do limiar, considérase que o rexistro foi exitoso e viceversa. Isto dá lugar a unha curva monótona e continua que reflicte a relación entre a taxa de éxito e a precisión obxectivo, evitando así a necesidade de establecer un limiar arbitrario. Estes gráficos utilizanse para ilustrar a precisión do rexistro tanto para casos individuais (onde se utilizan o porcentaxe de parellas de puntos rexistrados con éxito) como para o conxunto completo de datos. Esta métrica facilita a comparación entre distintos métodos competidores e permite seleccionar o máis axeitado segundo a precisión desexada.

Ademais, en FIRE a avaliación segmentarase nas 3 categorías de imaxes (S, P e A) para analizar o rendemento do rexistro en cada unha delas, xa que cada categoría presenta diferentes desafíos e características.

Mentres que FIRE xa provee os puntos de referencia para a avaliación, RFMID non o fai. Polo tanto, para RFMID, utilizamos o mesmo método de avaliación, pero xerando os puntos manualmente de forma que cubran o interior da máscara da imaxe fixa (separados por 50 píxeles entre si).

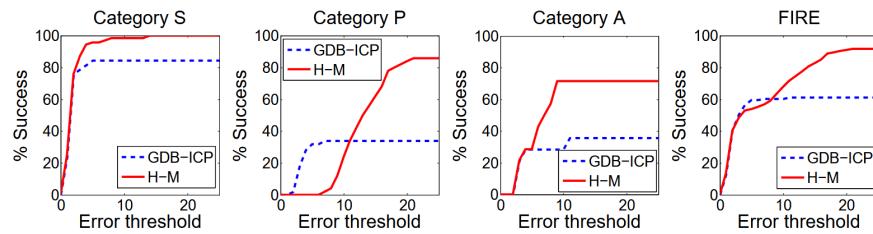


Figura 4.13: Gráfico de avaliação FIRE, [6]

No caso de RFMiD, dividiremos o conxunto de datos en varias categorías dependendo da dificultade do rexistro, que se calcula mediante a norma de Frobenius dunha matriz $A \in \mathbb{R}^{m \times n}$. Esta é unha xeralización da distancia euclidiana aplicada a matrices, onde as imaxes con transformacións mais grandes considéranse mais difíciles.

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$$

Figura 4.14: Norma de Frobenius dunha matriz $A \in \mathbb{R}^{m \times n}$, onde a_{ij} son os elementos da matriz A .

Nalgúns casos temén utilizaremos a distancia media entre os puntos correspondentes como métrica complementaria para avaliar a calidade do rexistro, xa que a taxa de éxito pode non ser suficiente para detectar os cambios.

4.6.2 Avaliación Cualitativa

No caso deste traballo, a avaliação cualitativa cobra gran importancia, xa que na cuantitativa só se está a comparar sobre un número reducido de puntos en cada parexa de imaxes. A avaliação visual permite detectar problemas que non se reflectan nas métricas cuantitativas, como artefactos visuais ou deformacións non desexadas, especialmente en rexistros que teñen deformacións locais que poden non coincidir con ningún punto.

No caso do dataset FIRE [6], a avaliação visual é especialmente relevante, xa que tan só se proporcionan 10 puntos de referencia por imaxe, que poden non ser suficientes para avaliar a calidade do rexistro en moitas zonas da imaxe. Xa que en RFMID [9] utilizanse puntos de referencia xerados manualmente que cubren toda a imaxe, a avaliação visual é algo menos relevante, xa que é máis probable que unha deformación local incorrecta sexa detectada por algún punto e se vexa reflexado nas métricas.

Co obxectivo de identificar facilmente os distintos artefactos visuais ou transformación non realista, utilizanse diferentes ferramentas como a composición de imaxes, a visualización dos vectores de desprazamento e a comparación de imaxes antes e despois do rexistro. Na figura 4.15 pódense observar algúns exemplos de estas.

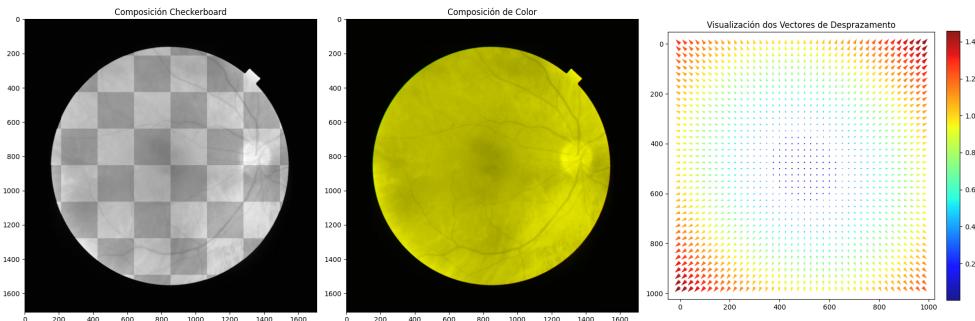


Figura 4.15: Exemplos de avaliação visual: (a) Composición de imaxes en checkerboard, (b) Composición de imaxes por cor, (c) Visualización dos vectores de desprazamento.

Capítulo 5

Experimentos e resultados

NESTE capítulo presentaranse os experimentos realizados e os resultados obtidos. Para iso, comezarase presentando unha vista xeral do proceso de experimentación, seguido dos propios experimentos realizados, para finalmente analizar os resultados obtidos en conxunto e as conclusóns que se poden extraer deles.

5.1 Vista Xeral

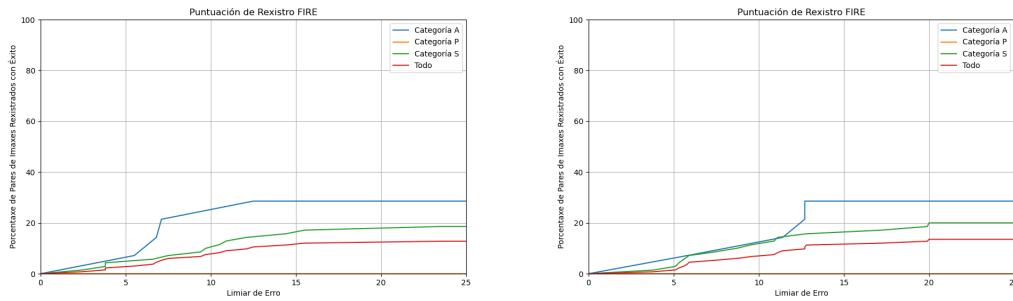
O obxectivo do traballo é determinar se as redes implícitas son aptas para a tarefa de rexistro de retinas. A comparación principal céntrase na función de activación empregada (SIREN ou ReLU), sobre os databases FIRE e RFMID.

Experimentaráse sobre as seguintes variables:

- Función de perda
- Resolución da imaxe
- Regularización
- Batch size
- Estratexias de mostraxe
- Inicialización
- Axuste dinámico do batch size

Debido ás diferencias entre as funcións de activación, é posible que cada unha require unha configuración diferente para obter os mellores resultados. Por exemplo, o sesgo que SIREN ten cara sinais de alta frecuencia fará que o proceso de regularización sexa mais relevante para evitar o sobreaxuste.

A avaliación de FIRE completa, desglosada por categoría, pódese ver nas figuras 5.1a e 5.1b. A menos que especificado de outro xeito, utilizarase un learning rate de 0.0001, batch size de 10000 puntos e 1500 epochs. Estos determináronse a partir dos utilizados orixinalmente por IDIR e ca análises cualitativa dos resultados obtidos en experimentos preliminares.



(a) Métrica FIRE ca función de activación ReLU (b) Métrica FIRE ca función de activación SIREN

Figura 5.1: Métricas dataset FIRE

Obsérvase que a categoría P resulta imposible de rexistrar, teorízase que debido ao baixo grado de superposición entre as imaxes (<75%) que impide que a rede aprenda a transformación adecuada. Nas categorías S e A os resultados son algo mellores, con taxas de éxito de arredor do 20%.

Co obxectivo de mellorar o rendemento da rede nesta tarefa, realizáronse experimentos adicionais para determinar os parámetros mais relevantes e optimizalos.

5.1.1 Descripción dos experimentos

Experimentos iniciais: Nesta parte realizaranxe experimentos para determinar uns valores aceptables para os parámetros da rede no contexto da imaxe oftalmolóxica, así como determinar a súa influenza no rendemento da rede.

- Función de perda:** Debido ás características únicas das imaxes de retina, con variabilidade en iluminación e contraste, é crucial determinar que función de perda é máis robusta para esta tarefa. Comparáronse funcións baseadas en píxeles (MSE, L1) con funcións baseadas en características estruturais (NCC, SSIM) para determinar cal captura mellor as correspondencias entre imaxes retinianas.
- Resolución da imaxe:** As imaxes de retina poden ter resolucóns de ata 2160×2160 píxeles, significativamente maiores que as imaxes de pulmón utilizadas orixinalmente por IDIR (512×512). É necesario determinar se unha maior resolución mellora o rendemento ou se introduce ruído que perxudica o rexistro.

- **Regularización:** SIREN ten un sesgo inherente cara sinais de alta frecuencia, o que pode provocar sobreaxuste. Avalíase o impacto de diferentes termos de regularización (jacobiana, hiperelástica, enerxía de flexión) para determinar os valores óptimos que eviten deformacións non realistas.
- **Batch size:** A densidade de puntos mostrados á rede pode ser crucial para o éxito do rexistro. Batch sizes maiores proporcionan máis información por iteración, pero a un maior custo computacional. Investígase o equilibrio óptimo entre eficiencia e rendemento.

Estratexias de mostraxe: As imaxes de retina teñen zonas con diferentes cantidades de información estrutural (vasos sanguíneos, disco óptico vs. fondo uniforme). Compáranse estratexias de mostraxe aleatorio, uniforme e ponderado por contido para determinar se priorizar certas rexións mellora o rexistro.

Inicialización: A natureza non convexa da función de perda pode facer que diferentes inicializacions converxan a mínimos locais distintos. Implementouse unha lotería de inicialización para seleccionar a inicialización máis prometedora baseándose no loss inicial.

Axuste dinámico do batch size: Teorízase que a rede podería beneficiarse de aprender primeiro transformacións globais con batch sizes pequenos e despois refinar con batch sizes maiores para capturar detalles locais.

5.2 Exemplos de rexistro

Diferentes exemplos de rexistro, tanto exitosos como fallidos, pódense observar na figura 5.2. A primeira imaxe corresponde ca imaxe fixa, a segunda corresponde ca imaxe rexistrada, a terceira ca imaxe móbil e a cuarta o campo de deformación aplicado a unha grella cadrada.

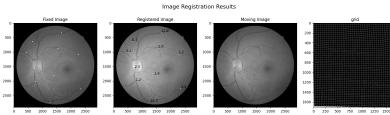
Pódense observar os puntos de control, sendo os brancos os da imaxe fixa, os verdes os da imaxe móbil e os azuis os desprazados pola rede.

5.3 Función de loss

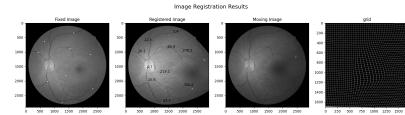
5.3.1 Planteamento

As funcións de perda valoradas para este traballo xa forón explicadas na sección 4.2.1.2.

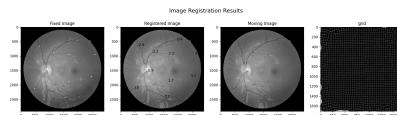
Para determinar cal é a función de perda mais adecuada para a tarefa de rexistro de retinas, realizáronse experimentos comparando o rendemento de cada unha sobre unha mostra de imaxes dos databases de FIRE e RFMID. Xa que a rede non é capaz de rexistrar con éxito a gran parte das imaxes nestas condicións, tomaráse a distancia media de todos os puntos como métrica de comparación.



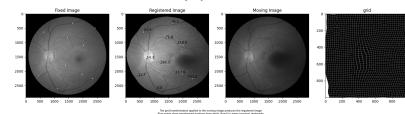
(a) Rexistro exitoso dunha parella de imaxes do dataset FIRE ca función de activación ReLU



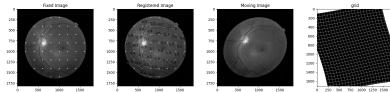
(b) Rexistro fallido dunha parella de imaxes do dataset FIRE ca función de activación ReLU



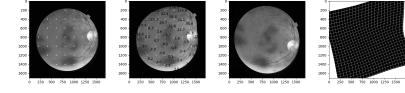
(c) Rexistro exitoso dunha parella de imaxes do dataset FIRE ca función de activación SIREN



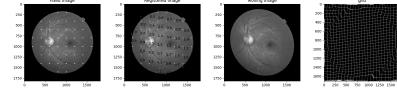
(d) Rexistro fallido dunha parella de imaxes do dataset FIRE ca función de activación SIREN



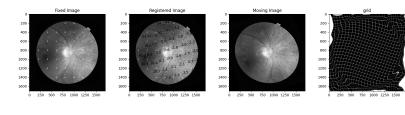
(e) Rexistro exitoso dunha parella de imaxes do dataset RFMID ca función de activación ReLU



(f) Rexistro fallido dunha parella de imaxes do dataset RFMID ca función de activación ReLU



(g) Rexistro exitoso dunha parella de imaxes do dataset RFMID ca función de activación SIREN



(h) Rexistro fallido dunha parella de imaxes do dataset RFMID ca función de activación SIREN

Figura 5.2: Exemplos de rexistro: combinacións de dataset (FIRE/RFMID), función de activación (relu/SIREN) e éxito.

5.3.2 Resultados

Presentase na figura 5.3 a comparación entre as diferentes funcións de perda.

5.3.3 Discusión

Obsérvase como as métricas que teñen en conta a estrutura da imaxe (NCC, SSIM) tenden a dar mellores resultados que aquelas que non o fan (MSE, Huber, Smooth L1) co dataset de FIRE, mentres que con RFMID ocurre ó contrario. Isto pode deberse a que as imaxes reais de retina teñen unha maior variabilidade na iluminación e contraste, polo que as métricas que non teñen en conta a estrutura da imaxe serán menos robustas a estas diferenzas. No caso de RFMID, ao ser imaxes sintéticas, a variabilidade na iluminación e contraste é nula, o que explica os mellores resultados das métricas que non teñen en conta a estrutura da imaxe. Da

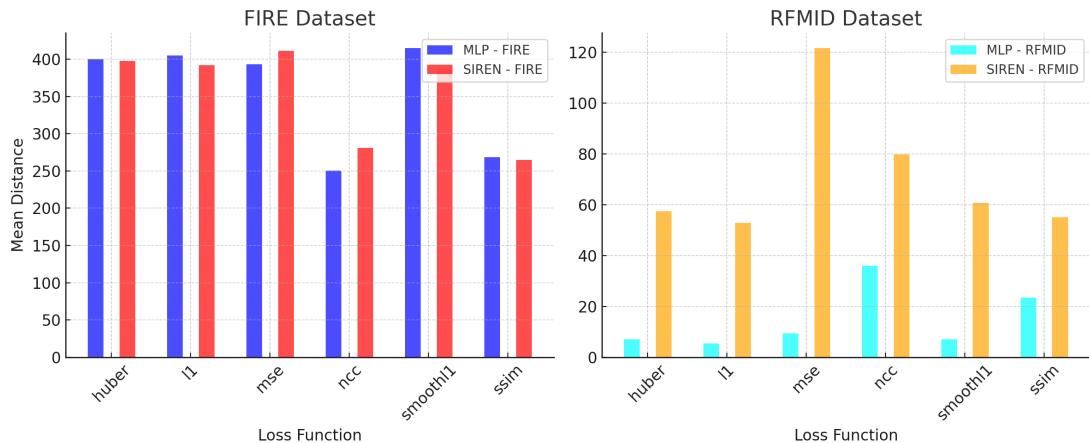


Figura 5.3: Comparación de diferentes funcións de perda sobre imaxes de FIRE e RFMID

mesma forma, a función de activación Relu tende a producir funcións predominantemente lineares, o que se adapta mellor ás transformacións realizadas no dataset RFMID.

SSIM é menos robusta ao ruído e sensible o tamaño das seccións utilizadas, así como computacionalmente mais costosa. Ademais, ten outro custo engadido xa que non é posible calcular SSIM tan só comparando os puntos mostrados xa que utiliza xanelas deslizantes para evaluar luminancia, contraste e estrutura. Para utilizala é necesario reconstruir a imaxen en cada iteracion o que ten un alto custo computacional. No caso de non reconstruir a imaxe e utilizar os puntos mostrados directamente, esta métrica funciona igualmente mais con resultados lixeiramente peores, xa que perde toda a súa capacidade de capturar variacións locais de luminancia, contraste e estrutura, o que se tradúce nunha función de perda global sen consideraciós locais.

5.3.4 Conclusíons

En base aos resultados obtidos, pódense extraer as seguintes conclusíons:

- Para o dataset FIRE, que contén imaxes reais de retina con variabilidade en iluminación e contraste, as funcións de perda baseadas en características estruturais como NCC e SSIM proporcionan resultados significativamente mellores.
- Para o dataset RFMID, que contén imaxes con tan só variación xeométrica, as funcións de perda baseadas en píxeles como L1 e Huber ofrecen mellores resultados.
- Obsérvase unha diferenza sistemática entre os modelos Relu e SIREN, sendo os primeiros más efectivos para o dataset RFMID, mentres que ambos mostran rendementos comparables para FIRE.

5.4 Resolución da imaxe

5.4.1 Planteamento

Para determinar cal é a resolución mais adecuada, realizáronse experimentos comparando o rendemento de cada unha sobre unha mostra de imaxes dos databases de FIRE e RFMID. Debido a que a rede non é capaz de rexistrar con éxito a gran parte das imaxes, tomaráse a distancia media de todos os puntos como métrica de comparación.

A resolución da imaxe inflúe de forma directa no resto de parámetros da rede. Por exemplo, un batch size de 1000 puntos nunha imaxe de 256x256 é unha densidade de puntos moito maior que nunha imaxe de 1024x1024.

Ademais, a resolución da imaxe tamén inflúe na capacidade da rede para aprender as transformacións, xa que a información que recibe é mais detallada. Isto pode ser beneficioso se estos detalles conteñen información relevante para a tarefa de rexistro, pero tamén podería ser perxudicial se conteñen unha gran parte de ruido.

O tamaño das imaxes tamén é unha das principais diferencias entre as imaxes de retina e as de pulmóns utilizadas orixinalmente por IDIR, tendo estas últimas de 512x512 mentres que as imaxes dos ollos contan con resolucóns de ata 2160x2160.

5.4.2 Resultados

Preséntase na figura 5.4 a comparación entre as diferentes resolucóns.

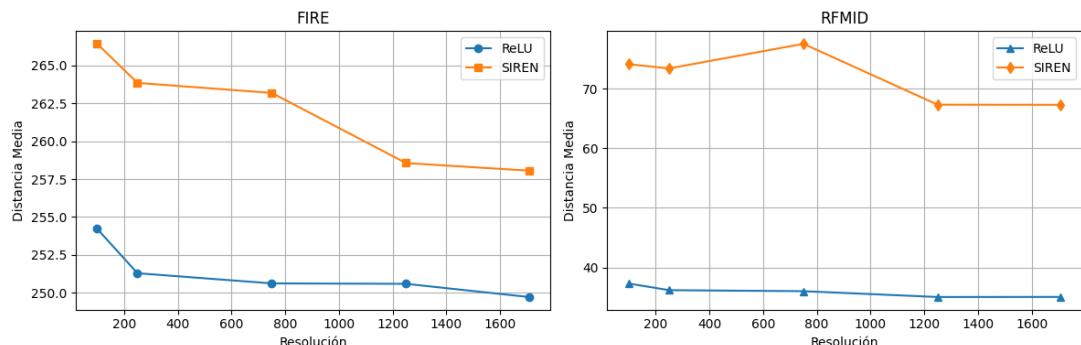


Figura 5.4: Comparación de diferentes resolucóns de perda sobre imaxes de FIRE e RFMID. Menor distancia media é mellor.

5.4.3 Discusión

Pódese observar como unha maior resolución tende a dar lixeiramente mellores resultados, pero a un custo computacional maior. Isto pode deberse á precisón ca que se fai a evaluación

mais que a unha mellor capacidade da rede para aprender as transformacións, xa que as diferencias son moi pequenas e consistentes entre os diferentes parellas de imaxes. Isto suxire que a resolución non ten un impacto significativo no rendemento da rede, e que a maioría da información relevante para a tarefa de rexistro xa está capturada en resolucións inferiores.

5.4.4 Conclusións

Baseándonos nos resultados obtidos, podemos concluír que:

1. Resolucións inferiores a 100×100 non capturan suficientes detalles das estruturas vasculares retinianas para realizar un rexistro preciso, especialmente en imaxes reais do dataset FIRE.
2. Aumentar a resolución por encima de 1250×1250 non aporta beneficios significativos.

Para os experimentos subseguientes, adoptarase unha resolución estándar de 1000×1000 píxeles, que demostrou proporcionar un bo balance entre rendemento e eficiencia computacional.

5.5 Regularización

5.5.1 Planteamento

Para determinar cal é a cantidade de regularización óptima, realizáronse experimentos comparando o rendemento de cada unha sobre unha mostra de imaxes dos databases de FIRE e RFMID cas diferentes funcións de activación e diferentes grados de regularización.

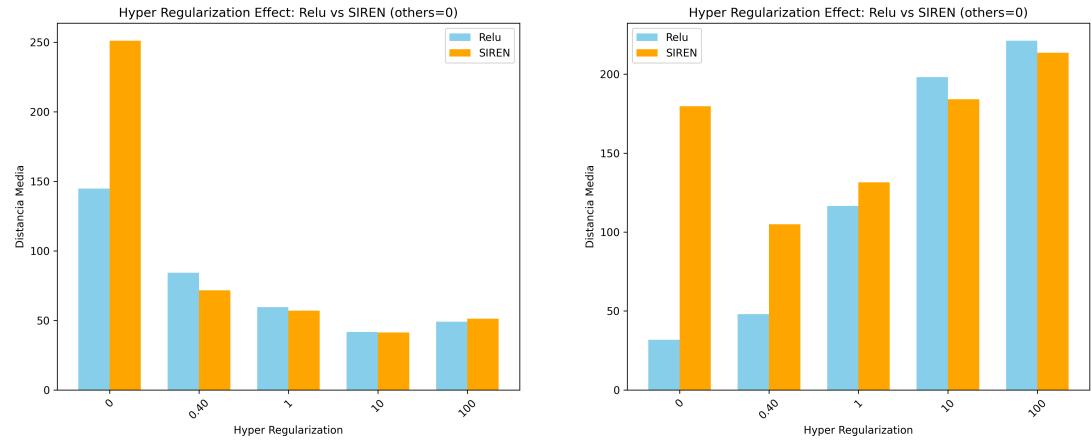
O proceso de regularización axuda a rede a evitar o sobreaxuste, modificando o termo de loss para penalizar as transformacións pouco realistas. As técnicas de regularización valoradas, que xa forón explicadas en detalle na sección 4.2.1.3

Os valores utilizados para cada tipo de regularización axustaronse a partir dos utilizados orixinalmente por IDIR e comparado o impacto de cada un deles sobre a función de perda, xa que a escala de cada un deles é diferente.

No anexo A.1 detállase unha búsqueda mais completa para explorar as relacións entre os diferentes tipos de regularización. Neste apartado só se presentarán os resultados dos experimentos realizados coa regularización hiperelástica, que se considera a mais relevante para nesta tarefa.

5.5.2 Resultados

A comparación entre os diferentes valores de regularización hiperelástica preséntase na figura 5.5.



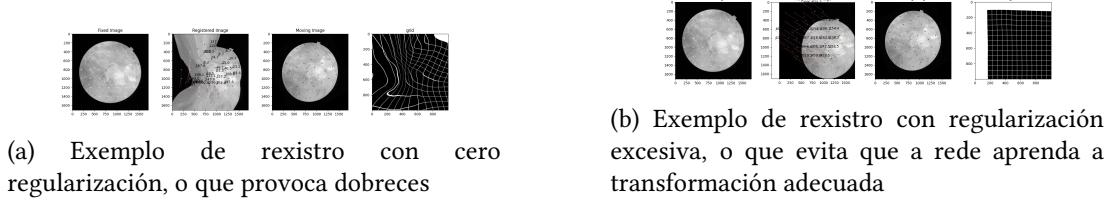
(a) Comparación de regularización hiperelástica en FIRE

(b) Comparación de regularización hiperelástica en RFMID

Figura 5.5: Comparación do impacto da regularización hiperelástica sobre os datasets FIRE e RFMID para modelos ReLU e SIREN

5.5.3 Discusión

Os resultados amosan que a regularización ten un impacto significativo no rendemento da rede. Tanto a ausencia de regularización como a regularización excesiva resultan en rendemento deficiente. Na figura 5.6 pódense observar exemplos de rexistros con ambos problemas.



(a) Exemplo de rexistro con cero regularización, o que provoca dobrices

(b) Exemplo de rexistro con regularización excesiva, o que evita que a rede aprenda a transformación adecuada

Figura 5.6: Exemplos de rexistro con ausencia e exceso de regularización

Nos resultados obsérvase que ReLU segue a dar mellores resultados que SIREN no dataset RFMID, mentres que no dataset FIRE ambos parecen ter un rendemento similar.

A regularización óptima depende do tipo de rexistro que se está a realizar. Os rexistros de transformacións lineais (RFMID) beneficianse de pouca ou ningunha regularización, mentres que os rexistros de transformacións non lineais (FIRE) e con pouca superposición beneficianse de regularizacións mais elevadas. Isto suxire que a regularización é mais relevante onde a rede ten que aprender transformacións mais complexas, xa que evita que caia en mínimos locais non desexados.

5.5.4 Conclusións

En base aos resultados, conclúese que a regularización é un compoñente indispensable para o rexistro de retinas con redes implícitas. O seu valor óptimo non é universal, senón que depende directamente da complexidade da transformación a aprender. Para deformacións sinxelas e lineais como as de RFMiD, unha regularización mínima é suficiente, pero para os desafíos presentes en FIRE, cunha maior non linearidade, un termo de regularización robusto é crucial para guiar a rede cara a solucións fisicamente plausibles e evitar o sobreaxuste. Confirmase tamén que os modelos SIREN, pola súa maior capacidade para representar detalles de alta frecuencia, son más sensibles á regularización e requieren valores xeralmente más altos que os modelos ReLU para previr artefactos. A elección do coeficiente de regularización debe considerarse unha decisión fundamental, adaptada tanto á natureza do problema de rexistro como á arquitectura da rede empregada.

5.6 Batch size

5.6.1 Planteamento

Ao longo dos experimentos realizados, o análisis cualitativo revelou que o batch size é un dos parámetros que máis impacto ten no rendemento da rede.

De agora en adiante dividimos o conxunto de datos de RFMID en varios subconxuntos según a dificultade da transformación, como detallado na sección 4.6.1.

Desta forma podemos comparar o rendemento da rede en diferentes subconxuntos de imaxes, e determinar se o rendemento da rede é consistente entre eles.

Nos experimentos co dataset FIRE, decidiuse limitarse á categoría S, xa que é a que maior número de exemplos tén e ten un maior grao de superposición entre as imaxes, o que facilita a tarefa de rexistro. Ademais, xa que a rede si que é capaz de rexistrar correctamente as imaxes dos subconxuntos mais sinxelos, utilizaremos a métrica de FIRE para medir o porcentaxe de imaxes rexistradas correctamente.

5.6.2 Resultados

Nas figuras 5.7a e 5.7b pódense observar os resultados da experimentación co dataset RFMID a distintas dificultades e con distintos tamaños de batch.

Con esta nova división do dataset, tamén se realizou a avaliación polo método de avaliación de FIRE, que se pode ver na figuras 5.8a e 5.8b.

Nas figuras 5.9a e 5.9b mostranse os resultados da experimentación co dataset FIRE.

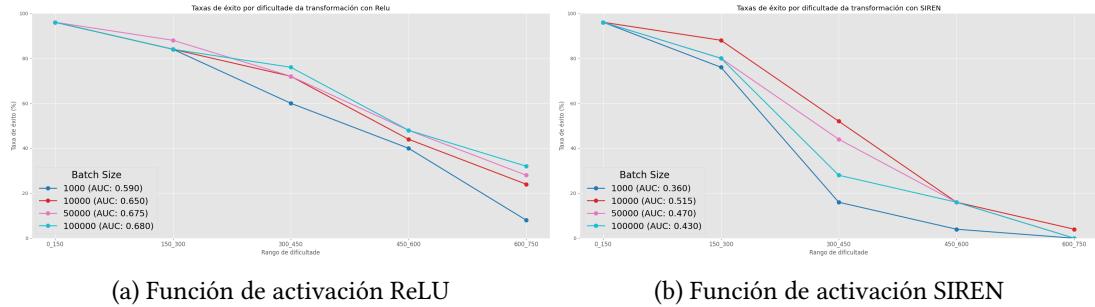


Figura 5.7: Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID

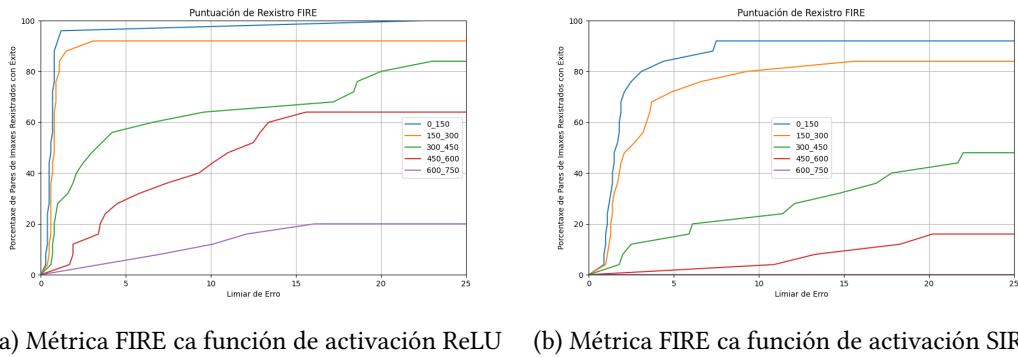


Figura 5.8: Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset FIRE

5.6.3 Discusión

Obsérvase que as redes ca función de activación ReLU tenden a ter un rendemento moito mellor que as ca función de activación SIREN. Isto pode explicarse xa que as deformacións artificiais que se aplican nas imaxes do dataset RFMID son lineais, e a función de activación ReLU é adecuada para este tipo de transformacións.

Tamén parece que o batch size é relevante, especialmente o cambio entre 1000 e 10000, mentres que batch sizes maiores (50000, 100000) non parecen ter tanto impacto, áinda que si un maior custo computacional.

Mentres que a rede é capaz de rexistrar correctamente consistentemente as imaxes do subconxunto mais sinxelos (0-150, 150-300), o rendemento decae notablemente para transformacións mais complexas (300+). Isto é mais notable cando se utiliza a función de activación SIREN, que ten dificultades incluso con transformacións de complexidade media, mentres que con ReLU decae de forma lineal.

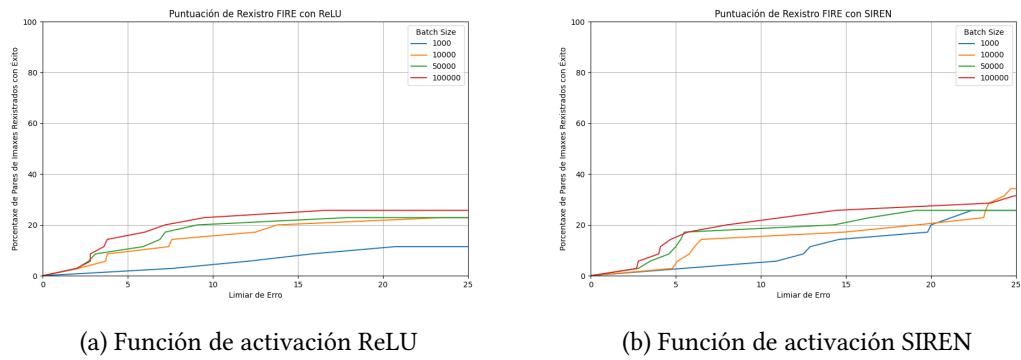


Figura 5.9: Comparación do rendemento da rede con diferentes batch sizes sobre imaxes da categoría S do dataset FIRE

5.6.4 Conclusins

O principal factor limitador do rendemento da rede é o tamaño e complexidade das transformacións que tenta aprender. Un batch size maior parece axudar, pero non é suficiente para rexistrar correctamente as imaxes con transformacións más difíciles.

5.7 Estratexias de mostraxe

Orixinalmente IDIR utiliza unha estratexia de mostraxe aleatoria para seleccionar os puntos que se pasan á rede en cada iteración. Mientras que esta estratexia parece suficiente para o rexistro de pulmóns, no caso das imaxes de retina isto non ten porque ser así. Isto débese a que as imaxes de retina conteñen seccións con moita mais información que outras, frente os CTs de pulmóns onde o sinal é mais uniforme. Por exemplo, as seccións que conteñen vasos sanguíneos ou o disco óptico probablemente teñan mais información relevante para a tarefa de rexistro, frente outras seccións como o fondo da retina. Ademais, as retinografías teñen desprazamentos moito maiores e menor superposición entre cada parella, polo que a rede ten que aprender transformacións mais complexas.

5.7.1 Plantexamento

Plantexáronse novas estratexias de mostraxe, explicadas en detalle no apartado ??, cás cales preténdese mellorar o rendemento da rede ao proporcionarlle mais información relevante para a tarefa de rexistro. As estratexias de mostraxe comparadas son as seguintes:

- Mostraxe aleatoria: Selección aleatoria de puntos da imaxe.
 - Mostraxe uniforme: Selección de puntos uniformemente espazados na imaxe. Especialmente relevante cando se usan batch sizes pequenos, xa que permite garantir que se amosan

puntos de toda a imaxe.

- Mostraxe intelixente: Selección de puntos baseándose na información do gradiente da imaxe, priorizando as áreas con maior variación.
- Mostraxe ponderada: Punto intermedio entre a mostraxe aleatoria e a intelixente, onde se seleccionan puntos aleatoriamente pero con maior probabilidade nas áreas de maior interese.

5.7.2 Resultados

Os resultados das diferentes estratexias de mostraxe sobre o dataset RFMID preséntanse na figura 5.10.

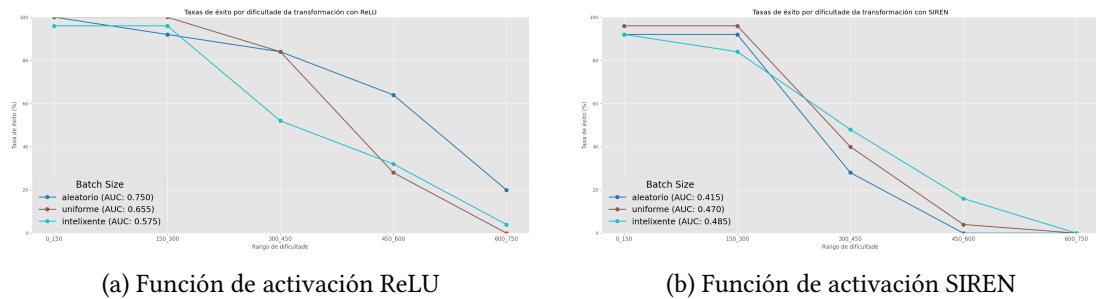


Figura 5.10: Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset RFMID

5.7.3 Discusión

A hipótese da estratexia de mostraxe intelixente non parece ser axeitada, con resultados similares á estratexia aleatoria. O mesmo ocorre coa estratexia uniforme.

Igual que en experimentos anteriores, a función de activación ReLU parece dar mellores resultados que SIREN con RFMID, especialmente con maiores dificultades de transformación.

5.7.4 Conclusións

Conclúese que, en contra da hipótese inicial, as estratexias de mostraxe implementadas, como a ponderada polo contido da imaxe, non aportan unha mellora significativa no rendemento do rexistro en comparación coa mostraxe aleatoria estándar. Isto sugere que a información relevante para a deformación está o suficientemente ben distribuída como para que unha mostraxe aleatoria sexa capaz de capturar os puntos necesarios para a converxencia, sempre que o batch size sexa adecuado.

5.8 Inicialización

5.8.1 Planteamento

É posible que a inicialización da rede sexa un factor clave, e que certos desprazamentos iniciais provoque que que a rede sexa incapaz de aprender a transformación correcta, ou que lle custe moito mais aprenderla.

Para testar esta hipótese implementouse unha lotería de inicialización, onde se utiliza o loss no epoch 0 para determinar a inicialización da rede máis beneficiosa sobre a que seguir entrenando.

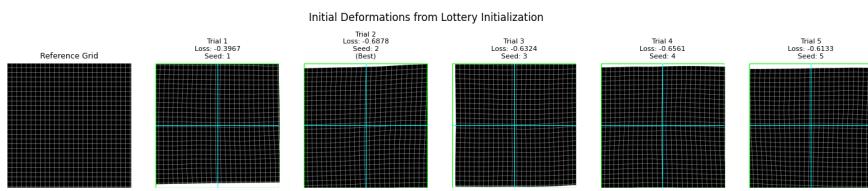


Figura 5.11: Exemplos das diferentes inicializacions ca función de activación RELU

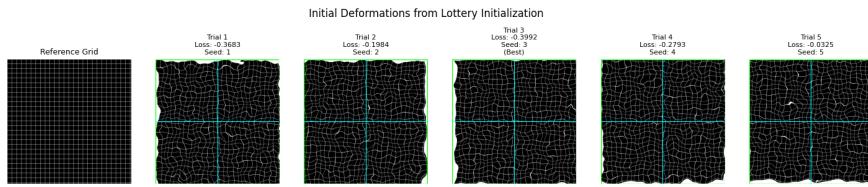


Figura 5.12: Exemplos das diferentes inicializacions ca función de activación SIREN

5.8.2 Resultados

Na figura 5.13 amósanse os resultados de diferentes valores da lotería de inicialización.

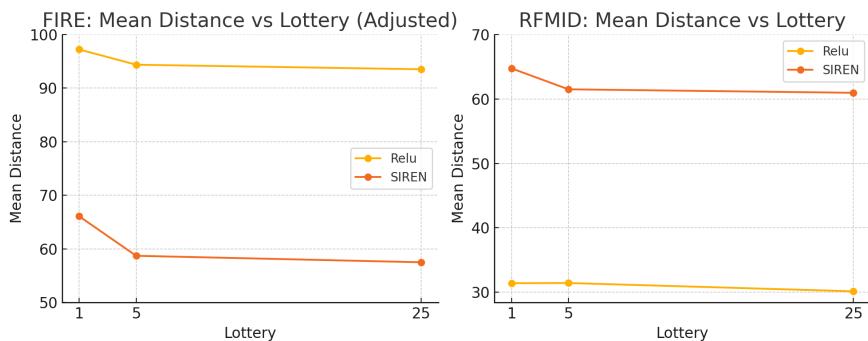


Figura 5.13: Resultados da lotería de inicialización

5.8.3 Discusión

Obsérvase que a lotería de inicialización si que provoca unha mellora, aínda que non moi significativa, e non se beneficia particularmente de utilizar mais de 5 inicializacions. É posible que fose mellor esperar ata unha iteración algo mais avanzado para determinar a inicialización, xa que no epoch 0 non hai ningunha seguridade de que non sexa un mínimo local, pero isto tamén implicaría un maior custo computacional.

5.8.4 Conclusóns

Unha posible mellora á lotería de inicialización sería utilizar un número maior de epochs antes de determinar a inicialización gañadora, xa que o loss inicial non é necesariamente representativo do rendemento final da rede.

5.9 Axuste dinámico do batch size

5.9.1 Planteamento

Teorízase que a rede pode beneficiarse de dividir o proceso de rexistro en diferentes fases, onde inicialmente utilizase un batch size reducido para aprender a transformación global, e posteriormente aumentase o batch size para aprender as transformacións locais. Para isto utilizaremos a estratexia de mostraxe uniforme, que permite asegurar que se cubre toda a imaxe de ca mesma densidade, o que é mais importante con batch sizes pequenos. O learning rate modifícase de forma proporcional para manter a relación entre este e o batch size.

5.9.2 Resultados

Os resultados de utilizar diferentes números de fases pódense observar na figura 5.14.

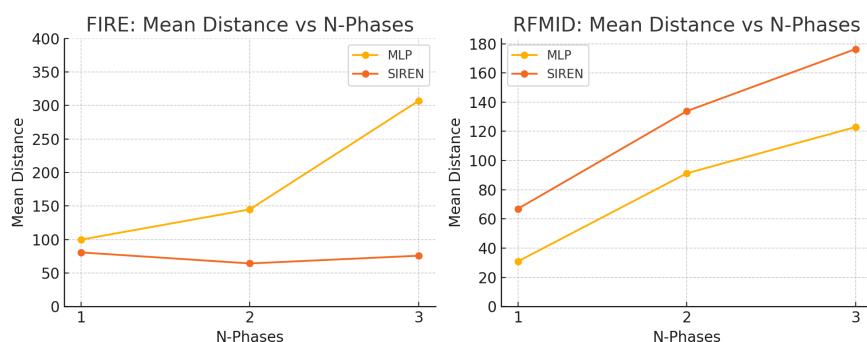


Figura 5.14: Resultados de usar distinto número de fases

5.9.3 Discusión

Os resultados presentados na figura 5.14 mostran unha tendencia clara e contraria á hipótese inicial, xa que o rendemento da rede empeora progresivamente a medida que se incrementa o número de fases. A estratexia de comenzar cun batch size pequeno para aprender a transformación global antes de refinar os detalles cun batch size maior resulta ser contraproducente.

Unha posible explicación é que a idea de que un lote pequeno favorece a aprendizaxe global é incorrecta neste contexto. Un batch size reducido ofrece unha estimación moi ruidosa e pouco representativa, o que pode levar o adestramento por un camiño inestable e impedir que a rede converxa cara a unha boa solución global na fase inicial. En cambio, un batch size grande e constante, como se validou na sección 5.6, proporciona dende o principio información suficiente tanto a nivel global como local, permitindo que a rede aprenda ambas as transformacións simultaneamente de forma máis estable e eficaz.

Polo tanto, a estratexia de dividir o adestramento en fases non só non aporta beneficios, senón que resulta prexudicial ao introducir inestabilidade nas etapas cruciais da aprendizaxe. Este experimento reforza a conclusión de que un batch size grande é fundamental para o éxito do rexistro con esta metodoloxía.

5.9.4 Conclusións

Conclúese que a estratexia de axuste dinámico do batch size é ineficaz e prexudicial para esta tarefa. A inestabilidade introducida nas fases iniciais do adestramento anula calquera posible beneficio teórico de aprender as transformacións por etapas. A rede demostra ser máis robusta e eficaz cando se adestra cun batch size grande e constante dende o principio, confirmando que este enfoque é superior para aprender simultaneamente as características globais e locais da deformación.

5.10 Comparativa e resumo de resultados

Tras a realización dos experimentos descritos anteriormente, pódense extraer as seguintes conclusións:

5.10.1 Rendemento por dataset

Dataset FIRE: O rendemento no dataset FIRE, que contén imaxes reais de retina, é significativamente mais baixo que no dataset RFMID. A categoría P resulta imposible de rexistrar debido ao baixo grado de superposición (<75%). As categorías S e A mostran taxas de éxito ao redor do 20%, sendo a categoría S lixeiramente superior.

Dataset RFMID: O rendemento no dataset RFMID é considerablemente mellor, especialmente para transformacións de baixa complexidade (norma de Frobenius 0-150), onde se alcanza un éxito do case 100%. O rendemento decae progresivamente ca complexidade da transformación.

5.10.2 Comparación de funcións de activación

Os resultados mostran unha clara diferenciación entre as funcións de activación segundo o tipo de dataset:

ReLU: Mostra un rendemento superior no dataset RFMID, debido á súa capacidade natural para aprender transformacións lineais.

SIREN: Aínda que teoricamente máis adecuada para aprender transformacións complexas, ten dificultades para representalas adecuadamente. Tende a transformacións locais e pouco realistas e se non se regulariza adecuadamente, converxe a mínimos locais non desexados.

5.10.3 Impacto dos parámetros principais

Función de perda: As métricas baseadas en características estruturais (NCC, SSIM) son superiores para imaxes reais con variabilidade de iluminación (FIRE), mentres que as métricas baseadas en píxeles (L1, MSE) son más efectivas para imaxes sen variabilidade (RFMID).

Regularización: A regularización hiperelástica é a mais relevante para evitar deformacións non realistas. O valor óptimo depende da parexa de imaxes concreta a rexistrar. SIREN require valores máis altos que ReLU xa que ten unha maior tendencia ao sobreaxuste.

Batch size: Constitúe un dos parámetros más críticos. Batch sizes maiores (10000-50000) proporcionan mellores resultados que batch sizes pequenos (1000), aínda que o beneficio decrece para valores moi altos (>50000).

Resolución: Non se observan beneficios significativos por encima de 1000×1000 píxeles, suxerindo que a información relevante para o rexistro xa está capturada en resolucións moderadas.

5.10.4 Limitacións identificadas

Complexidade das transformacións: O factor limitador principal é a complexidade das transformacións. Tanto ReLU como SIREN mostran dificultades ca transformacións de alta complexidade, independentemente da optimización de parámetros.

Superposición de imaxes: A baixa superposición entre imaxes (categoría P de FIRE) dificulta enormemente o rexistro, suxerindo que as redes implícitas requieren un mínimo de información compartida.

Estratexias de mostraxe: Contrario ao esperado, as estratexias intelixentes de mostraxe (ponderadas por contido) non mostran beneficios sobre o mostraxe aleatorio, suxerindo que a información relevante está distribuída de forma más uniforme do esperado.

5.10.5 Conclusións xerais

Os experimentos demostran que as redes implícitas son aplicables ao rexistro de imaxes de retina, pero con limitacións importantes. O rendemento depende criticamente da complexidade das transformacións e do grado de superposición entre imaxes. Aínda que os resultados son prometedores para casos de complexidade baixa a moderada, o rexistro de imaxes ca transformacións complexas ou baixa superposición segue sendo un desafío que require investigación adicional.

Capítulo 6

Conclusións

En conclusión, o proxecto de investigación realizado consistiu na adaptación do framework IDIR para o rexistro de retinografías. En especial valoramos o uso da función de activación SIREN, proposta como alternativa á función ReLU para mellorar a representación das deformacións.

A aliñación de retinografías é un problema relevante xa que é un proceso laborioso para os expertos, mais con moita uilidade clínica. A etapa inicial da revisión do estado da arte revelou que xa existían varios traballos previos que abordaban este problema, sendo os más exitosos os baseados en métodos iterativos. Actualmente os métodos de aprendizaxe profunda son unha alternativa prometedora que está gañando prominencia no campo. Particularmente, o uso de representacións implícitas para esta tarefa é un enfoque innovador que xa foi aplicado en outros campos da imaxe médica con bós resultados.

Para avaliar a efectividade do método proposto, escolléronse dous conxuntos de datos de retinografías: FIRE, que permite a avaliación do método en imaxes reais, e RFMID, sobre o que se efectuaron transformacións artificiais para simular diferentes escenarios de aliñación.

Durante a fase de experimentación exploráronse diferentes combinacións de hiperparámetros (loss, regularización, resolución, batch size...) e introducíronse diferentes técnicas para tentar mellorar a converxencia do modelo, como diferentes esquemas de mostraxe, inicialización, e técnicas de axuste dinámico do batch size.

Algunhas das dificultades atopadas durante o desenvolvemento do proxecto foron: a falta de documentación sobre o funcionamento do código orixinal, que dificultou a súa adaptación ao novo dominio; o deseño do proceso de avaliación, no cal foi complexo atopar visualizacións que permitisen interpretar os resultados facilmente; e o tempo de cómputo que requerían algúns experimentos, que requirió a implementación de optimizacións para facilitar a experimentación.

Os resultados obtidos amosan que esta arquitectura non é a máis adecuada para a tarefa de rexistro de retinografías.

Si que se obteñen bós resultados no dataset RFMID, que se basea en imaxes con transformacións lineais sintéticas, onde a función de activación ReLU tende a obter mellores resultados ca

CAPÍTULO 6. CONCLUSIÓNS

SIREN, xa que está mellor preparada para representar as transformacións lineais globais que se producen entre estas imaxes. Obsérvase tamén que o tamaño da transformación ten un impacto significativo no rendemento, xa que as imaxes de maior tamaño presentan un maior erro de rexistro.

No dataset FIRE, que contén imaxes reais, os resultados son peores ca no dataset RFMID, especialmente nas parellas de imaxes que presentan grandes deformacións ou baixo nivel de superposición. A función de activación SIREN obtén mellores resultados aquí, xa que é capaz de representar mellor as deformacións non lineais e locais que se producen entre as imaxes.

Estas diferencias no rendemento destacan a importancia da elección da función de activación en función da natureza específica das transformacións esperadas. A fase de experimentación tamén revelou que a regularización é un factor fundamental, especialmente na función de activación SIREN, onde a ausencia de regularización leva a un sobreaxuste significativo e a un rendemento moi pobre.

Cabe destacar que o método presentado neste traballo guía a optimización con tan só a métrica de NCC, que depende únicamente das intensidades dos píxeles, e que en rexitros con moito desprazamento ou deformacións complexas a topografía de función de perda será pouco convexa e con múltiples mínimos locais, o que dificulta a converxencia do modelo. Polo contrario, métodos como REMPE [37] que obteñen resultados moito mellores (rexistra exitosamente a totalidade da categoría S de FIRE) fan uso de información adicional que lles permite establecer correspondencias globais entre as imaxes.

Unha observación relevante é a diferenza entre o rendemento entre o conxunto de datos sintético (RFMID) e o conxunto de datos real (FIRE). Esta brecha demostra a dificultade de aplicar modelos adestrados en datos sintéticos a imaxes reais.

Todos estes achados responden aos obxectivos propostos no inicio do proxecto, onde adaptamos o framework IDIR para o rexistro de retinografías, exploramos a función de activación SIREN e avaliamos o rendemento do modelo en diferentes condicións.

Capítulo 7

Traballo futuro

EXISTEN varias liñas de traballo futuro que se poden seguir para mellorar o sistema actual. Os resultados obtidos neste traballo, aínda que demostran a viabilidade de adaptar o framework IDIR para o aliñamento de imaxes oftalmolóxicas 2D, tamén revelan limitacións á hora de acadar a precisión e robustez desexables. As seguintes liñas de traballo futuro considéranse prometedoras para superar estes desafíos e avanzar no campo:

7.1 Arquitecturas alternativas

Unha liña relevante de traballo futuro é a exploración de arquitecturas alternativas. Mentre que os perceptróns multicapa (MLPs) son considerados aproximadores universais [76] (son capaces de aproximar calquera función continua dada unha cantidade suficiente de neuronas), é posible que a arquitectura utilizada de 3 capas con 256 neuronas por capa non sexa o suficientemente grande para capturar as complexidades das transformacións entre as retinografías.

Unha opción sería aumentar o número de capas ou neuronas por capa. Outra sería implementar o uso de codificación posicional, que parece ser útil para a tarefa de rexistro [77].

Outra idea moi intersante é o uso de restriccóns de consistencia cíclicas, propostas por Van Harten et al. no contexto de rexistro de imaxes médicas [78]. Consiste en entrenar dúas redes á vez que estiman as transformacións directas e inversas, facendo que estas se regularicen mutuamente e estabilizando a optimización.

7.2 Invertibilidade

Unha dirección interesante para o traballo futuro é a exploración de métodos que garantan a invertibilidade das transformacións aprendidas pola rede. A rede IDIR actual non garante a invertibilidade das transformacións aprendidas, o que significa que non é posíbel aplicar a transformación inversa de maneira fiable.

Grazas aos termos de regularización utilizados durante o adestramento son poucos os casos nos que o determinante jacobiano é negativo (o que indicaría que a transformación non é invertible).

Aproximación como a de i-RevNet [79] ou aqueles baseados en campos vectoriais de velocidade [66] permiten garantir a invertibilidade das transformacións aprendidas, o que podería mellorar a precisión e a robustez do rexistro e funcionaría como un mecanismo de regulación implícita.

7.3 Enfoque híbrido

Outra liña de traballo futuro é a exploración de enfoques híbridos que combinen o rexistro baseado en redes neuronais con técnicas tradicionais de rexistro. Unha posibilidade sería utilizar o rexistro tradicional para proporcionar un rexistro inicial robusto, que despois podería ser refinado por unha rede neuronal.

Así mesmo, poderíanse explorar con máis profundidade o preprocesado das imaxes, xa que é inexistente no método actual pero podería ser útil para mellorar a calidade das imaxes e facilitar o rexistro.

Apéndices

Apéndice A

Material adicional

A.1 Anexo regularization

A.1.1 Figuras experimentos de regularización

Resultados Os resultados da experimentación extendida da regularización, realizados sobre os datasets FIRE e RFMID, preséntanse nas figuras [A.1](#).

Discusión Os resultados amosan que as interaccións entre os diferentes termos de regularización e as funcións de activación son complexas e moi dependentes da parella de imaxes concreta a rexistrar.

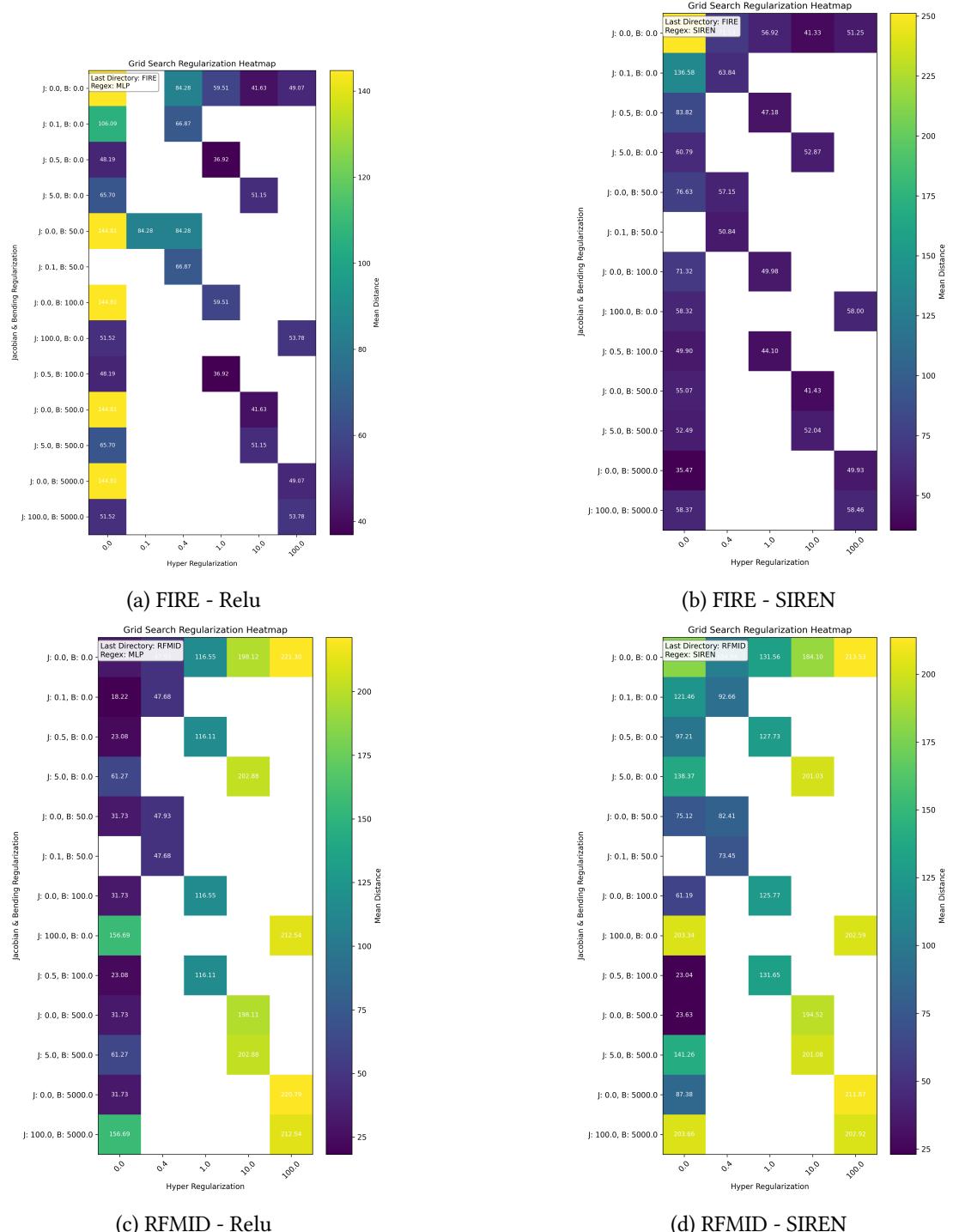


Figura A.1: Mapa de calor cos resultados de diferentes combinacións de termos de regularización e funcións de activación sobre os datasets FIRE e RFMID

Glosario

- 4DCT** Siglas en inglés de *Four-Dimensional Computed Tomography*. La Tomografía Computarizada del tórax en cuatro dimensiones es una técnica de imagen médica que permite capturar información dinámica del tórax, enfocándose principalmente en los pulmones. La cuarta dimensión se refiere al tiempo, lo que permite capturar el movimiento de las estructuras a lo largo del ciclo respiratorio, generando así una serie de imágenes 3D sincronizadas con las fases de la respiración. . 23
- BBF** Siglas en inglés de *Best Bin First*. Algoritmo de busca de vecinos en espazos de alta dimensión, empleado na comparación de descriptores.. 9
- CNN** Siglas en inglés de *Convolutional Neural Network*. Tipo de rede neuronal artificial que utiliza capas convolucionales que aplican filtros para extraer diferentes características, permitiendo o recoñecemento de patrones complexos. . 1, 11
- DFV** Siglas en inglés de *Deformation Vector Field*. Consiste nunha matriz que contén, para cada punto, un vector que indica a magnitud e dirección do desprazamento a realizar.. 7
- FBR** Siglas en inglés de *Feature-Based Registration*. Técnica de rexistro de imaxes baseada na identificación e correspondencia de características salientables, como puntos, liñas ou bordes, presentes nas imaxes. . 8
- FOV** Siglas en inglés de *Field Of View*. Ángulo ou área visible a través dun sensor ou dispositivo óptico. Determina a rexión do espazo que pode captar ou observar un sistema de visión.. 33
- GAN** Siglas en inglés de *Generative Adversarial Network*. Tipo de rede neuronal artificial composta por dúas redes que compiten entre si: un xerador, que intenta crear datos falsos realistas, e un discriminador, que tenta distinguir entre datos reais e falsos. . 11

GDB-ICP Siglas en inglés de *Generalized Dual-Bootstrap Iterative Closest Point*. Algoritmo de aliñamento iterativo de puntos. Funciona buscando correspondencias entre puntos e refinando a transformación ata conseguir o mellor axuste posible.. [9](#)

GMM Siglas en inglés de *Gaussian Mixture Models*. Modelo estatístico que representa a distribución de datos como unha combinación de varias gaussianas. Funciona axustando os parámetros das gaussianas para modelar a estrutura dos datos e identificar subconjuntos ou clusters.. [9](#)

IBR Siglas en inglés de *Image-Based Registration*. Técnica de rexistro de imaxes baseada na comparación directa dos valores de intensidade dos píxeles ou voxels das imaxes a aliñar. . [8](#)

IGRT Siglas en inglés de *Image-Guided Radiation Therapy*. Técnica de radioterapia que utiliza imaxes médicas obtidas inmediatamente antes ou durante o tratamento para localizar con precisión o tumor e adaptar a administración da radiación, mellorando a exactitude e minimizando o dano aos tecidos sans circundantes. . [5](#)

NTK Siglas en inglés de *Neural Tangent Kernel*. Ferramenta matemática para analizar o comportamento de redes neuronais durante o adestramento. Funciona aproximando a evolución dos parámetros da rede mediante un kernel que describe a súa dinámica de aprendizaxe. . [32](#)

PIIFD Siglas de *Partial Intensity Invariant Feature Descriptor*. É un descriptor de características que xera representación robustas de contornos en imaxes, mantendo certa invariancia frente a cambios parciales de intensidade e transformacións xeométricas.. [9](#)

Bibliografía

- [1] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” 2020. [Online]. Available: <https://arxiv.org/abs/2006.09661>
- [2] J. M. Wolterink, J. C. Zwienenberg, and C. Brune, “Implicit neural representations for deformable image registration,” in *Medical Imaging with Deep Learning 2022*, 2022.
- [3] R. Ferreruela, “La visión y el ojo,” *Apunts Educació Física y Deportes*, 2007. [En línea]. Disponible en: <https://www.redalyc.org/articulo.oa?id=551656954002>
- [4] V. B. Sivaraman, M. Imran, Q. Wei, P. Muralidharan, M. R. Tamplin, I. M. . Grumbach, R. H. Kardon, J.-K. Wang, Y. Zhou, and W. Shao, “Retinaregnet: A zero-shot approach for retinal image registration,” 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2404.16017>
- [5] “Scopus,” Retrieved from <https://www.scopus.com/>, 2024, accessed on 10-12-2024.
- [6] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, S. Douma, and A. A. Argyros, “Fire: Fundus image registration dataset,” *Modeling and Artificial Intelligence in Ophthalmology*, vol. 1, no. 4, p. 16–28, Jul. 2017. [En línea]. Disponible en: <https://www.maio-journal.com/index.php/MAIO/article/view/42>
- [7] F. Altaf, S. M. S. Islam, N. Akhtar, and N. K. Janjua, “Going deep in medical image analysis: Concepts, methods, challenges, and future directions,” *IEEE Access*, vol. 7, pp. 99 540–99 572, 2019.
- [8] S. Bharati, M. R. H. Mondal, P. Podder, and V. B. S. Prasath, “Deep learning for medical image registration: A comprehensive review,” 2022. [En línea]. Disponible en: <https://arxiv.org/abs/2204.11341>
- [9] S. Pachade, P. Porwal, D. Thulkar, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, L. Giancardo, G. Quellec, and F. Mériadeau, “Retinal fundus multi-disease image

- dataset (rfmid): A dataset for multi-disease detection research,” *Data*, vol. 6, no. 2, 2021. [En línea]. Disponible en: <https://www.mdpi.com/2306-5729/6/2/14>
- [10] J. J. Kanski and B. Bowling, *Clinical ophthalmology: a systematic approach*. Elsevier Health Sciences, 2011.
- [11] N. H. H. y Jamilet Navarro Vivó y Alina Honan González y Belkis Ortega Ruiz, “Importancia del diagnóstico precoz del glaucoma.” *Revista Médica Electrónica*, vol. 28, no. 1, pp. 34–37, 2014. [En línea]. Disponible en: <https://revmedicaelectronica.sld.cu/index.php/rme/article/view/259>
- [12] J. V. Forrester, A. D. Dick, P. G. McMenamin, F. Roberts, and E. Pearlman, “Chapter 5 - physiology of vision and the visual system,” in *The Eye (Fourth Edition)*, 4th ed., J. V. Forrester, A. D. Dick, P. G. McMenamin, F. Roberts, and E. Pearlman, Eds. W.B. Saunders, 2016, pp. 269–337.e2. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780702055546000058>
- [13] ——, “Chapter 1 - anatomy of the eye and orbit,” in *The Eye (Fourth Edition)*, 4th ed., J. V. Forrester, A. D. Dick, P. G. McMenamin, F. Roberts, and E. Pearlman, Eds. W.B. Saunders, 2016, pp. 1–102.e2. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780702055546000010>
- [14] T. Ilginis, J. Clarke, and P. J. Patel, “Ophthalmic imaging.” *British medical bulletin*, vol. 111, no. 1, 2014.
- [15] M. Fogel-Levin, S. R. Sadda, P. J. Rosenfeld, N. Waheed, G. Querques, B. K. Freund, and D. Sarraf, “Advanced retinal imaging and applications for clinical practice: A consensus review,” *Survey of Ophthalmology*, vol. 67, no. 5, pp. 1373–1390, 2022. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0039625722000352>
- [16] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, “Retinal image registration as a tool for supporting clinical applications,” *Computer Methods and Programs in Biomedicine*, vol. 199, p. 105900, 2021. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0169260720317338>
- [17] A. A. Goshtasby, *Theory and applications of image registration*. John Wiley & Sons, 2017.
- [18] Y. Wang, Y. Long, S. H. Fan, and Q. Dou, “Neural rendering for stereo 3d reconstruction of deformable tissues in robotic surgery,” 2022. [En línea]. Disponible en: <https://arxiv.org/abs/2206.15255>

- [19] J. A. Bogovic, P. Hanslovsky, A. Wong, and S. Saalfeld, “Robust registration of calcium images by learned contrast synthesis,” in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 1123–1126.
- [20] K. Han, S. sun, X. Yan, C. You, H. Tang, J. Naushad, H. Ma, D. Kong, and X. Xie, “Diffeomorphic image registration with neural velocity field,” 2022. [En línea]. Disponible en: <https://arxiv.org/abs/2202.12498>
- [21] H. Xiao, X. Teng, C. Liu, T. Li, G. Ren, R. Yang, D. Shen, and J. Cai, “A review of deep learning-based three-dimensional medical image registration methods,” *Quantitative Imaging in Medicine and Surgery*, vol. 11, 12 2021.
- [22] B. Zitová and J. Flusser, “Image registration methods: a survey,” *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0262885603001379>
- [23] Y. Liu, B. Yu, T. Chen, Y. Gu, B. Du, Y. Xu, and J. Cheng, “Progressive retinal image registration via global and local deformable transformations,” 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2409.01068>
- [24] X. Papademetris, A. P. Jackowski, R. T. Schultz, L. H. Staib, and J. S. Duncan, “Integrated intensity and point-feature nonrigid registration,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*, C. Barillot, D. R. Haynor, and P. Hellier, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 763–770.
- [25] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004. [En línea]. Disponible en: <https://api.semanticscholar.org/CorpusID:174065>
- [26] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008, similarity Matching in Computer Vision and Multimedia. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1077314207001555>
- [27] S. Leutenegger, M. Chli, and R. Y. Siegwart, “Brisk: Binary robust invariant scalable keypoints,” in *2011 International Conference on Computer Vision*, 2011, pp. 2548–2555.
- [28] A. Alahi, R. Ortiz, and P. Vandergheynst, “Freak: Fast retina keypoint,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 510–517.
- [29] M. Muja and D. G. Lowe, “Scalable nearest neighbor algorithms for high dimensional data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2227–2240, 2014.

- [30] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, p. 381–395, Jun. 1981. [En línea]. Disponible en: <https://doi.org/10.1145/358669.358692>
- [31] C. Stewart, C.-L. Tsai, and B. Roysam, “The dual-bootstrap iterative closest point algorithm with application to retinal image registration,” *IEEE transactions on medical imaging*, vol. 22, pp. 1379–94, 12 2003.
- [32] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, “A partial intensity invariant feature descriptor for multimodal retinal image registration,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1707–1718, 2010.
- [33] C. G. Harris and M. J. Stephens, “A combined corner and edge detector,” in *Alvey Vision Conference*, 1988. [En línea]. Disponible en: <https://api.semanticscholar.org/CorpusID:1694378>
- [34] J. Beis and D. Lowe, “Shape indexing using approximate nearest-neighbour search in high-dimensional spaces,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997, pp. 1000–1006.
- [35] Z. Ghassabi, J. Shanbezadeh, A. Sedaghat, and E. Fatemizadeh, “An efficient approach for robust multimodal retinal image registration based on ur-sift features and piifd descriptors,” *EURASIP Journal on Image and Video Processing*, vol. 1, pp. 1–16, 12 2013.
- [36] C. Liu, J. Ma, Y. Ma, and J. Huang, “Retinal image registration via feature-guided gaussian mixture model,” *J. Opt. Soc. Am. A*, vol. 33, no. 7, pp. 1267–1276, Jul 2016. [En línea]. Disponible en: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-33-7-1267>
- [37] C. Hernandez-Matas, X. Zabulis, and A. Argyros, “Rempe: Registration of retinal images through eye modelling and pose estimation,” *IEEE Journal of Biomedical and Health Informatics*, 2020.
- [38] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of ICNN’95 - International Conference on Neural Networks*, vol. 4, 1995, pp. 1942–1948 vol.4.
- [39] B. C. Lowekamp, D. T. Chen, L. Ibanez, and D. Blezek, “The design of simpleitk,” *Frontiers in Neuroinformatics*, vol. 7, 2013. [En línea]. Disponible en: <https://www.frontiersin.org/journals/neuroinformatics/articles/10.3389/fninf.2013.00045>
- [40] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, “elastix: A toolbox for intensity-based medical image registration,” *IEEE Transactions on*

- Medical Imaging*, vol. 29, pp. 196–205, 2010. [En línea]. Disponible en: <https://api.semanticscholar.org/CorpusID:15850341>
- [41] N. J. Tustison, P. A. Cook, A. J. Holbrook, H. J. Johnson, J. Muschelli, G. A. Devenyi, J. T. Duda, S. R. Das, N. C. Cullen, D. L. Gillen, M. A. Yassa, J. R. Stone, J. C. Gee, and B. B. Avants, “The ANTsX ecosystem for quantitative biological and medical imaging,” *Scientific Reports*, vol. 11, no. 1, p. 9068, Apr. 2021. [En línea]. Disponible en: <https://doi.org/10.1038/s41598-021-87564-6>
- [42] Q. Nie, X. Zhang, Y. Hu, M. Gong, and J. Liu, “Medical image registration and its application in retinal images: A review,” 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2403.16502>
- [43] S. Czolbe, P. Pegios, O. Krause, and A. Feragen, “Semantic similarity metrics for image registration,” *Medical Image Analysis*, vol. 87, p. 102830, 2023. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1361841523000907>
- [44] M. Lafarge, P. Moeskops, M. Veta, J. Pluim, and K. Eppenhof, “Deformable image registration using convolutional neural networks,” p. 27, 03 2018.
- [45] D. DeTone, T. Malisiewicz, and A. Rabinovich, “Superpoint: Self-supervised interest point detection and description,” 2018. [En línea]. Disponible en: <https://arxiv.org/abs/1712.07629>
- [46] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, “Superglue: Learning feature matching with graph neural networks,” 2020. [En línea]. Disponible en: <https://arxiv.org/abs/1911.11763>
- [47] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “Voxelmorph: A learning framework for deformable medical image registration,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, p. 1788–1800, Aug. 2019. [En línea]. Disponible en: <http://dx.doi.org/10.1109/TMI.2019.2897538>
- [48] Y. Ma, D. Niu, J. Zhang, X. Zhao, B. Yang, and C. Zhang, “Unsupervised deformable image registration network for 3d medical images,” *Applied Intelligence*, vol. 52, no. 1, p. 766–779, Jan. 2022. [En línea]. Disponible en: <https://doi.org/10.1007/s10489-021-02196-7>
- [49] R. Jena, P. Chaudhari, and J. C. Gee, “Deep implicit optimization enables robust learnable features for deformable image registration,” *Medical Image Analysis*, vol. 103, p. 103577, Jul. 2025. [En línea]. Disponible en: <http://dx.doi.org/10.1016/j.media.2025.103577>
- [50] D. Mahapatra, “Gan based medical image registration,” 2019. [En línea]. Disponible en: <https://arxiv.org/abs/1805.02369>

- [51] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. A. Hamprecht, Y. Bengio, and A. Courville, “On the spectral bias of neural networks,” 2019. [En línea]. Disponible en: <https://arxiv.org/abs/1806.08734>
- [52] A. Essakine, Y. Cheng, C.-W. Cheng, L. Zhang, Z. Deng, L. Zhu, C.-B. Schönlieb, and A. I. Aviles-Rivero, “Where do we stand with implicit neural representations? a technical and performance survey,” 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2411.03688>
- [53] S. Ramasinghe and S. Lucey, “Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps,” 2022. [En línea]. Disponible en: <https://arxiv.org/abs/2111.15135>
- [54] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. G. Baraniuk, “Wire: Wavelet implicit neural representations,” 2023. [En línea]. Disponible en: <https://arxiv.org/abs/2301.05187>
- [55] Z. Yu and Y. Tang, “Neural trajectory model: Implicit neural trajectory representation for trajectories generation,” 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2402.01254>
- [56] P. Reddy, Z. Zhang, M. Fisher, H. Jin, Z. Wang, and N. J. Mitra, “A multi-implicit neural representation for fonts,” 2022. [En línea]. Disponible en: <https://arxiv.org/abs/2106.06866>
- [57] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” 2020. [En línea]. Disponible en: <https://arxiv.org/abs/2003.08934>
- [58] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, “Occupancy networks: Learning 3d reconstruction in function space,” 2019. [En línea]. Disponible en: <https://arxiv.org/abs/1812.03828>
- [59] Q. Wu, Y. Li, L. Xu, R. Feng, H. Wei, Q. Yang, B. Yu, X. Liu, J. Yu, and Y. Zhang, “Irem: High-resolution magnetic resonance (mr) image reconstruction via implicit neural representation,” 2021. [En línea]. Disponible en: <https://arxiv.org/abs/2106.15097>
- [60] A. Molaei, A. Aminimehr, A. Tavakoli, A. Kazerouni, B. Azad, R. Azad, and D. Merhof, “Implicit neural representation in medical imaging: A comparative survey,” 2023. [En línea]. Disponible en: <https://arxiv.org/abs/2307.16142>
- [61] L. Shen, J. Pauly, and L. Xing, “Nerp: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction,” 2023. [En línea]. Disponible en: <https://arxiv.org/abs/2108.10991>

- [62] M. Tenzer, E. Tung, Z. Rasheed, and K. Shafique, “Generating trajectories from implicit neural models,” in *2024 25th IEEE International Conference on Mobile Data Management (MDM)*, 2024, pp. 129–138.
- [63] H. Zhang, L. Zhu, J. Shen, and A. Song, “Implicit neural field guidance for teleoperated robot-assisted surgery,” in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 6866–6872.
- [64] Y. Velikova, M. F. Azampour, W. Simson, M. Esposito, and N. Navab, “Implicit neural representations for breathing-compensated volume reconstruction in robotic ultrasound,” 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2311.04999>
- [65] E. Dupont, H. Loya, M. Alizadeh, A. Golinski, Y. W. Teh, and A. Doucet, “COIN++: data agnostic neural compression,” *CoRR*, vol. abs/2201.12904, 2022. [En línea]. Disponible en: <https://arxiv.org/abs/2201.12904>
- [66] S. Sun, K. Han, C. You, H. Tang, D. Kong, J. Naushad, X. Yan, H. Ma, P. Khosravi, J. S. Duncan, and X. Xie, “Medical image registration via neural fields,” 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2206.03111>
- [67] Y. Wu, T. Z. Jiahao, J. Wang, P. A. Yushkevich, J. C. Gee, and M. A. Hsieh, “Deformable image registration using neural odes,” *CoRR*, vol. abs/2108.03443, 2021. [En línea]. Disponible en: <https://arxiv.org/abs/2108.03443>
- [68] L. Ziyin, T. Hartwig, and M. Ueda, “Neural networks fail to learn periodic functions and how to fix it,” 2020. [En línea]. Disponible en: <https://arxiv.org/abs/2006.08195>
- [69] F. de Avila Belbute-Peres and J. Z. Kolter, “Simple initialization and parametrization of sinusoidal networks via their kernel bandwidth,” 2022. [En línea]. Disponible en: <https://arxiv.org/abs/2211.14503>
- [70] M. Li, X. Chen, X. Li, B. Ma, and P. Vitanyi, “The similarity metric,” *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3250–3264, 2004.
- [71] G. Palubinskas, “Image similarity/distance measures: what is really behind mse and ssim?” *International Journal of Image and Data Fusion*, vol. 8, no. 1, pp. 32–53, 2017. [En línea]. Disponible en: <https://doi.org/10.1080/19479832.2016.1273259>
- [72] M. Burger, J. Modersitzki, and L. Ruthotto, “A hyperelastic regularization energy for image registration,” *SIAM Journal on Scientific Computing*, vol. 35, no. 1, pp. B132–B148, 2013. [En línea]. Disponible en: <https://doi.org/10.1137/110835955>

- [73] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, “Nonrigid registration using free-form deformations: Application to breast mr images,” *Medical Imaging, IEEE Transactions on*, vol. 18, pp. 712–721, 09 1999.
- [74] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, “Accurate, large minibatch sgd: Training imagenet in 1 hour,” 2018. [En línea]. Disponible en: <https://arxiv.org/abs/1706.02677>
- [75] S. Jianlin. (2024, Nov) How does the learning rate change when batch size increases? [En línea]. Disponible en: <https://kexue.fm/archives/10542>
- [76] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/0893608089900208>
- [77] T. Müller, A. Evans, C. Schied, and A. Keller, “Instant neural graphics primitives with a multiresolution hash encoding,” *ACM Trans. Graph.*, vol. 41, no. 4, pp. 102:1–102:15, Jul. 2022. [En línea]. Disponible en: <https://doi.org/10.1145/3528223.3530127>
- [78] L. D. van Harten, J. Stoker, and I. Išgum, “Robust deformable image registration using cycle-consistent implicit representations,” *IEEE Transactions on Medical Imaging*, vol. 43, no. 2, p. 784–793, Feb. 2024. [En línea]. Disponible en: <http://dx.doi.org/10.1109/TMI.2023.3321425>
- [79] J.-H. Jacobsen, A. Smeulders, and E. Oyallon, “i-revnet: Deep invertible networks,” 2018. [En línea]. Disponible en: <https://arxiv.org/abs/1802.07088>