



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, maio de 2025.

Dedicatoria

Agradecimentos

 Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

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 [1]), que pode axudar a eliminar o sesgo cara sinais de baixa frecuencia e mapear mellor deformación pequenas e detalladas.

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	Estructura	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	Estado da arte	7
2.2.2	En retinas	10
2.3	Representación Neuronais Implícitas	11
2.4	IDIR	14
2.5	Traballo proposto	22
3	Metodoloxía e planificación	23
3.1	Metodoloxía do desenvolvemento	23
3.2	Planificación do proxecto	23
3.3	Recursos utilizados	23
3.3.1	Software	23
3.3.2	Hardware	24
3.3.3	Estimación de custos	24
4	Traballo Realizado	25
4.1	Vista Xeral	25
4.2	Conxuntos de datos	25
4.2.1	FIRE	25

4.2.2	RFMID	26
4.2.3	Diferencias entre os datasets	27
4.3	Métodos de Avaliación	27
4.3.1	Avaliación Cuantitativa	28
4.3.2	Avaliación Cualitativa	28
4.4	Proceso de Rexistro	29
5	Experimentos e resultados	30
5.1	Vista Xeral	30
5.2	Experimentos	30
5.2.1	Experimentos iniciais	30
5.2.2	Batch size	37
5.2.3	Estratexias de mostraxe	37
5.2.4	Inicialización	39
5.2.5	Exemplos de rexistro	39
6	Conclusións	50
7	Traballo futuro	51
7.1	Instant Neural Graphics Primitives	51
7.2	Invertibilidade	51
Glosario		54
Bibliografía		56

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	Categorías de rexistro	6
2.3	Visualización do campo de vectores de deformación (DFV). Á esquerda, representación mediante frechas. Á dereita, esta deformación aplicada a unha cuadrícula.	7
2.4	Proceso de rexistro de imaxes iterativo	8
2.5	Comparación de publicacións ó longo do tempo que relacionadas co rexistro de imaxes. Datos extraídos de Scopus [4], 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)"	12
2.6	Exemplo de rexistro de imaxes de retina [5]	12
2.7	Inicialización primera capa	16
2.8	Inicialización seguintes capas	16
2.9	Regulizador Jacobiano	19
2.10	Regulizador Hiperelástico.	20
2.11	Regulizador Bending Energy	20
2.12	Regulizador Jacobiano en 2D	20
2.13	Regulizador Hiperelástico en 2D	21
2.14	Regulizador Bending Energy en 2D	21
4.1	Clasificación dos pares de imaxes en categorías.	26
4.2	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>	27
4.3	Gráfico de avaliación FIRE, [6]	28
5.1	Comparación de diferentes funcións de perda sobre imaxes de FIRE e RFMID .	32
5.2	Mapa de calor cos resultados de diferentes combinacións de termos de regularización e funcións de activación sobre os datasets FIRE e RFMID	41

5.3	Reusltados lr cos datasets FIRE e RFMID cas funcións de activación SIREN e Relu	42
5.4	Mapa de calor cos resultados de diferentes combinacións de batch size e learning rate con unha mostra de imaxes de RMIFD ca función de activación ReLU	42
5.5	Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación ReLU (learning rate = 1e-5)	43
5.6	Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación ReLU (learning rate = 5e-5)	43
5.7	Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación ReLU (learning rate = 1e-4)	44
5.8	Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación SIREN (learning rate = 1e-5)	44
5.9	Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación SIREN (learning rate = 5e-5)	45
5.10	Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación SIREN (learning rate = 1e-4)	45
5.11	Heatmaps de mostraxe	45
5.12	Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset RFMID	46
5.13	Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset FIRE	46
5.14	Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset RFMID ca función de activación RELU	47
5.15	Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset RFMID ca función de activación SIREN	47
5.16	Exemplos das diferentes inicializacións ca función de activación RELU	47
5.17	Exemplos das diferentes inicializacións ca función de activación SIREN	47
5.18	Resultados de usar distinto número de fases	48
5.19	Resultados da lotería de inicialización	48
5.20	Exemplos de rexistro: combinacións de dataset (FIRE/RFMID), función de activatción (relu/SIREN) e éxito.	49

Índice de Táboas

3.1	Comparativa entre os servidores utilizados	24
3.2	Estimación de custos dos recursos humanos	24
4.1	Comparación de Resultados entre IDIR y Replicación	26
5.1	Distancias medias segundo función de perda, tipo de rede e datasets (FIRE vs. RFMID)	32
5.2	Distancias medias para o dataset FIRE ca función de activación Relu	34
5.3	Distancias medias para o dataset FIRE ca función de activación SIREN	34
5.4	Distancias medias para o dataset RFMID ca función de activación Relu	34
5.5	Distancias medias para o dataset RMIFD ca función de activación SIREN	34

Capítulo 1

Introducción

Neste primer capítulo expónense as motivacións e obxetivos 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. Esta problema é axeitada para ser automatizada, xa que é non require dun xuízo clínico, senón que se basea na comparación de características visuais das imaxes.

Xa existen diversas técnicas para realizar aliñamento de imaxes automático, especialmente ca chegada do deep learning 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 datasets 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 Estructura

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 obxetivos 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 discutiranse os resultados e compararanse cos resultados de outros métodos.
- **Capítulo 5: Experimentos e resultados:** neste capítulo presentaranse os experimentos realizados e os resultados obtidos.
- **Capítulo 6: Conclusíons:** neste capítulo presentaranse 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. 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.

[3], [13]

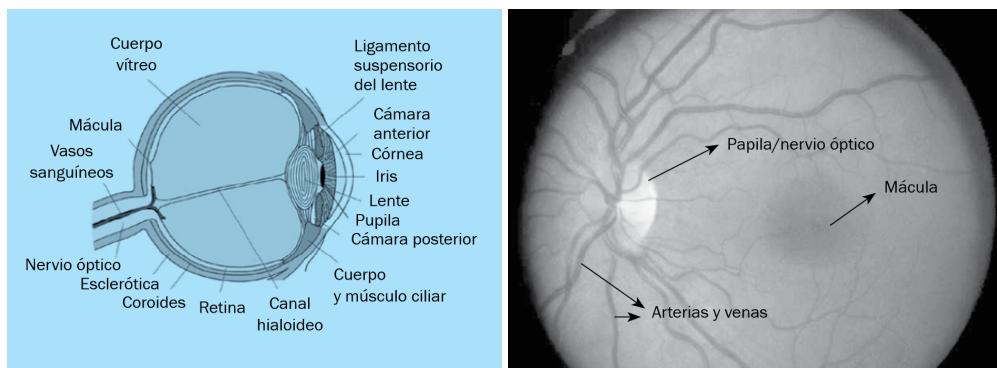


Figura 2.1: Imaxes do ollo humano, extraídas de [3]. Á esquerda, vista lateral do ollo anotada. Á dereita, retinografía do ollo anotada.

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.

Por exemplo, no caso do stitching de fotografías panorámicas, o rexistro de imaxes permite identificar correspondencias entre puntos característicos en múltiples tomas solapadas e axustar a súa posición relativa nun marco común. Esta etapa é necesaria para a posterior fusión das imaxes, de modo que as distintas vistas se aliñen con precisión, producindo un resultado final continuo e sen irregularidades visuais.

O rexistro de imaxes ten utilidade en moitos campos diferentes como a imaxe satelital, xeografía, robótica... [17] mais o campo da imaxe médica é dos mais interesantes pola súa aplicación práctica e é o que se aborda neste traballo.

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. As imaxes poden variar a nivel temporal, espacial, de dimensión ou de modalidade.

Tamén é fundamental para a intervención guiada por imaxe (telecirurxia, (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.

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

[20], [8] 2.2

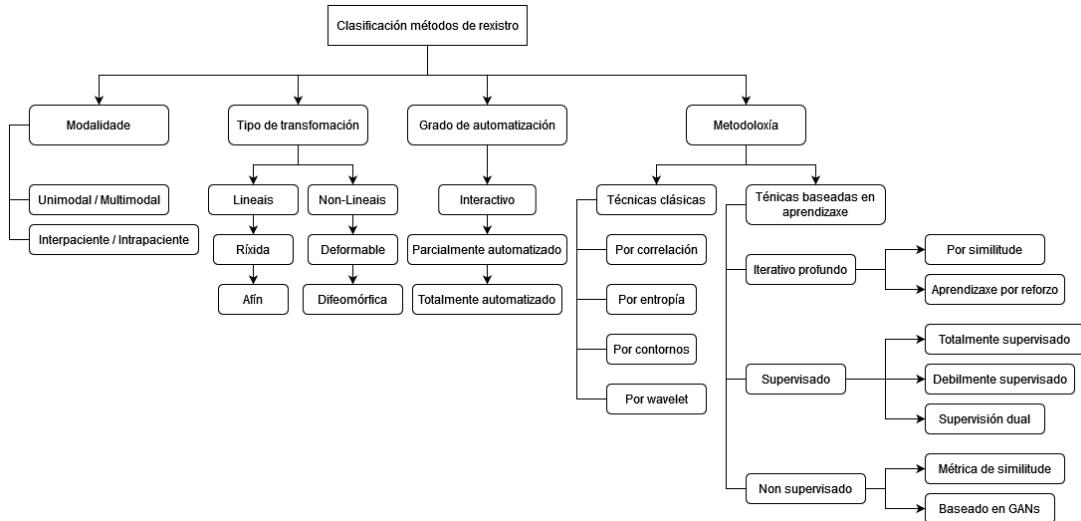


Figura 2.2: Categorías de rexistro

O rexistro tamén se pode diferenciar entre simétrico e asimétrico, dependendo de se a transformación se calcula para unha imaxe ou para ambas.

No caso de traballar con dúas imaxes de forma asimétrica, a imaxe de referencia denomínase imaxe fixa (f) e a imaxe que se quere rexistrar imaxe móvil (m). Dependendo do tipo de transformación utilizada esta pode ser clasificada en ríxida, afín ou deformable. A ríxida tan só permite rotación e traslación, mentres que a afín permite ademais escalado e cizallamento. Ámbas transformacións poden ser representadas por unha matriz de 2 dimensións xa que son deformacións lineais. Ao contrario, a transformación deformable é non lineal, polo que require dunha dimensión adicional ás da imaxe a rexistrar (unha imaxe de 2d require unha matriz 3d).

Esta matriz denominase campo de vectores de deformación (DFVs), e permite representar deformacións locais na imaxe, facendoa moito mais flexible para representar transformacións complexas 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.

Dentro da deformables poden distinguirse entre transformacións difeomórficas e non difeomórficas. As transformacións difeomórficas son aquellas que son continuas, invertibles e diferenciables en todo o seu dominio. Se non tén esta característica, non se pode garantir que a información da imaxe móvil se manteña intacta tras a transformación. Por iso as transformacións difeomórficas son preferidas en moitos casos [21],

Existen diversos métodos para realizar aliñamento de imaxes, que poden estar automatizados

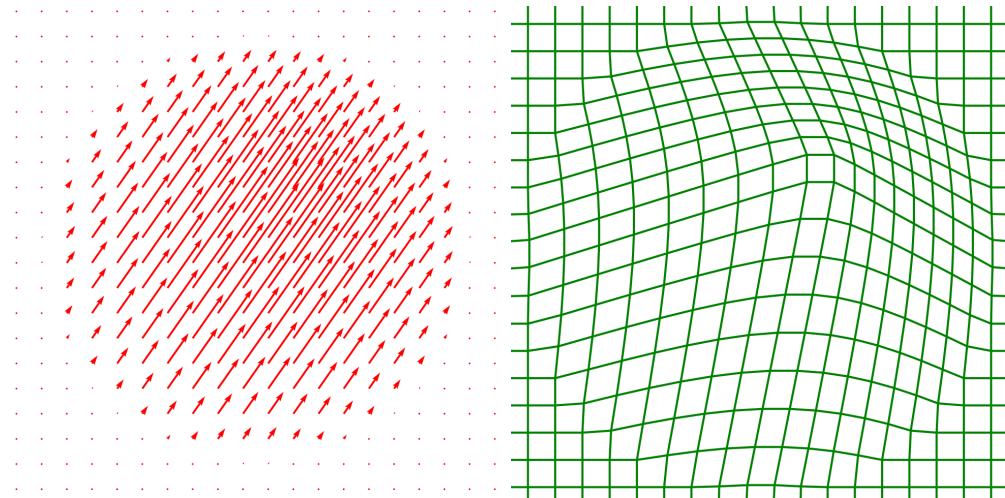


Figura 2.3: Visualización do campo de vectores de deformación (DFV). Á esquerda, representación mediante frechas. Á dereita, esta deformación aplicada a unha cuadrícula.

en maior ou menor medida, sendo moitos deles híbridos. [20]

2.2.1 Estado da arte

Pese á gran cantidade de avances que está a ocorrer no campo do aprendizaxe profunda, os métodos clásicos de rexistro de imaxes seguen a ser o estado da arte na maioría de casos, principalmente debido á importancia da precisión e a robustez en imaxe médica.

2.2.1.1 Métodos clásicos

Tradicionalmente impregáronse métodos iterativos baseados na extracción de características seguido dun proceso de optimización entre cada par de imaxes. Isto require recoñecer as características de interese en cada imaxe e utilizar unha métrica de similitude para determinar a calidade do rexistro. O proceso consiste en calcular o grado de semellanza entre as imaxes e actualizar os parámetros da transformación de forma iterativa utilizando algún mecanismo de optimización ata que se cumpran os criterios de terminación. O resultado final pode ser os parámetros da transformación ou a imaxe fusionada. A figura 2.4 mostra un diagrama do proceso de rexistro iterativo.

A principal desventaxa destes métodos é a súa lentitude, xa que requieren varias iteracións para converxer. Non obstante, son moi precisos e robustos, polo que áinda se empregan en moitos casos.

No proceso de extracción de características destacan algoritmos de SIFT [22], SURF [23], BRISK [24] ou FREAK [25].

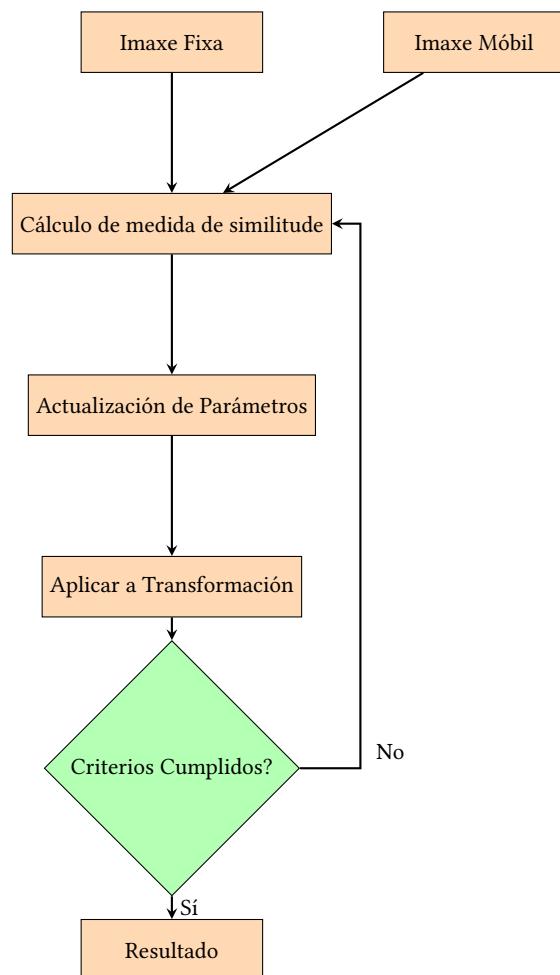


Figura 2.4: Proceso de rexistro de imaxes iterativo

Para encontrar a correspondencia entre os puntos característicos das imaxes, empreganse algoritmos como RANSAC [26] ou FLANN [27].

Tamén existen múltiples programas que facilitan o rexistro de imaxes, como SimpleITK [28], Elastix [29] ou ANTs [30].

2.2.1.2 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. Estos métodos tenden a ser mais rápidos que os métodos convencionais, a custo de algo de precisión. Ademais, estos métodos requiren 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.

Unha aproximación común é empregar estes conxuntos de datos para optimizar unha CNN que, dadas dúas imaxes novas e non vistas, predice o DFV correspondente. [31]

Durante o proceso de entrenamento, a rede ten acceso aos DFVs ca deformación correcta, ou pódense obter indirectamente a través da optimización dunha métrica de similitude de imaxes.

Existen moitas extensións a esta aproximación, como o uso de múltiples etapas ou o uso de redes adversarias durante o entrenamento.

Tamén se propuxeron métodos híbridos que combinan a optimización iterativa ca aprendizaxe profunda, entrenando unha CNN nova por cada parella de imaxes. Desta forma conséguese evitar a necesidade de grandes conxuntos de datos para o adestramento.

Os métodos de aprendizaxe profunda poden ser clasificados en dous tipos según se requiran de DFVs anotados ou non na etapa de entrenamento: supervisados (requíren de DFVs anotados) e non supervisados (non requíren de DFVs anotados). [32]

Existe un gran interés polos métodos baseados en aprendizaxe profundo, como se reflexa no crecente número de publicacións no campo 2.5.

Métodos Supervisados Existen dúas subcategorías según o grado de supervisión utilizado na etapa de entrenamento: totalmente supervisados ou débilmente supervisados. O rexistro totalmente supervisado fai uso de DVF_s de referencia para supervisar o proceso de aprendizaxe. O termo de perda adoita basearse na discrepancia entre os DVF_s de referencia e os DVF_s predichos.

No lugar dos DVF_s de referencia, o rexistro débilmente supervisado pode utilizar outras etiquetas de referencia implícitas. Estas etiquetas non se basan en datos explícitos como os DFVs, senón que utilizan información indirecta para guiar o proceso de rexistro, como a semellanza entre as imaxes ou restricións 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] ...

Métodos Non Supervisados Un dos maiores retos para entrenar redes efectivas con imaxes médicas é a recolección de datos anotados de calidade para o adestramento [7]. A creación de conxuntos de DFVs anotados é un proceso laborioso e costoso, que normalmente sólo pódese executado por especialistas, polo que os métodos de rexistro non supervisados son de gran interese. Xa que a imaxe fixa e a imaxe móvil xa conteñen toda a información necesaria para un rexistro correcto, os métodos non supervisados parecen mais adecuados para a tarefa de rexistro. 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.

[33] é un dos frameworks mais utilizados no rexistro de imaxes non supervisado facendo uso de CNNs.

GANs As CNN (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 móvil 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 [34], onde se propón un modelo de rexistro de imaxes baseado en GANs que demostrou ser efectivo tanto en retinas como con resonancias magnéticas cardiovasculares.

Posteriormente [35] propuxo SymReg-GAN, baseado en GANs para o rexistro de imaxe simétrico, abordando os problemas inherentes ao rexistro asimétrico debido á súa natureza direccional. O método proposto impón a simetría asegurando que a mapeo de imaxe A a B e de volta a A sexa consistente, e está deseñado para escenarios multimodais.

2.2.2 En retinas

Os métodos que funcionan ben en moitos dominios de imaxe médica (cerebro, pulmóns, etc) adoitan requirir de axustes para funcionar en retinas, polo que en imaxe oftalmolóxica hai un estado do arte paralelo. O rexistro mediante landmarks segue a ser moi usado neste contexto, e hay especial interese por métodos moi robustos que permitan o rexistro e fusión de imaxes de diferentes modalidades, por exemplo entre imaxes de fondo do ollo e anxiografías.

[GDB-ICP](#) é un dos métodos tradicionalmente mais utilizados, proposto por [36], consiste en comezar cunha ou máis estimacións iniciais que só son precisas en pequenas rexións da imaxe, chamadas rexións bootstrap. En cada rexión bootstrap, o algoritmo itera sobre os seguintes pasos:

- 1: refina a estimación da transformación usando restricións só dentro da rexión bootstrap;
- 2: expande a rexión bootstrap; 3: comproba se se pode usar un modelo de transformación de orde superior, deténdose cando a rexión se expande para cubrir a superposición entre imaxes.

[37] propón un framework nomeado Harris-PIIFD no cal comezan por detectar os puntos de esquina como candidatos a puntos de control utilizando Harris [38]. Posteriormente, introducen o algoritmo [PIIFD](#) para describir os puntos característicos, e aplícase o algoritmo [PIIFD](#) [39] para identificar coincidencias entre pares de imaxes. Finalmente, refinanse as coincidencias e elimínanse as incorrectas, e escóllese o tipo de transformación a aplicar en función do número de coincidencias. A transformación más sinxela (ríxida) require polo menos dous pares de puntos de control. Se o número de coincidencias é maior ou igual a tres pero menor que seis, aplícase a transformación afín; e se o número é maior ou igual a seis, aplícase a transformación polinomial de segunda orde.

Varias melloras a este método foron propostas, como [40], que introduce o algoritmo [UR-SIFT-PIIFD](#), que combina UR-SIFT con Harris-PIIFD para extraer puntos invariantes á escala e obter unha maior robustez. [41] propón un método baseado en [GMM](#) para a extracción de puntos de control, que demostra ser mais robusto que os métodos anteriores, especialmente en imaxes con mala calidade ou pouca solapación. [42] constrúen sobre os modelos anteriores e utilizan [SURF](#) para a extracción de puntos de control e [PIIFD](#) para a descripción dos mesmos, logo presentan [MFSP](#), un método non ríxido que combina a preservación de estrutura cos puntos característicos para o rexistro das imaxes. A transformación final efectúase mediante [TPS](#) [43].

[REMPE](#) [44] propón a estimación simultánea da pose das cámaras que adquiriron as imaxes e da forma e pose do ollo. Utiliza un modelo elipsoidal para o ollo e estima a posición das cámaras inicialmente mediante RANSAC, seguido de una variante de [PSO](#) [45] para refinala. Versións anteriores deste método utilizaron modelos esféricos en lugar de elipsoidais, e [SURF](#) en lugar de SIFT [46].

2.6

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

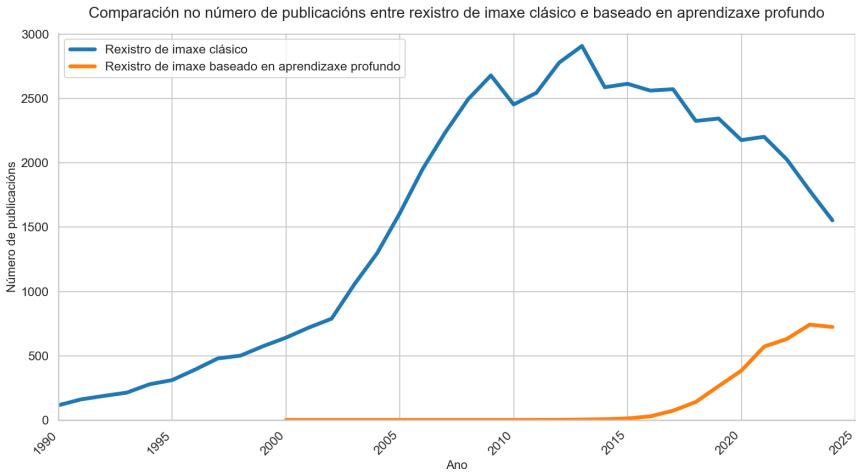


Figura 2.5: Comparación de publicacóns ó longo do tempo que relacionadas co rexistro de imaxes. Datos extraídos de Scopus [4], realizando as consultas: "TITLE-ABS-KEY(image AND registration) AND NOT(deep AND learning)" e "TITLE-ABS-KEY(image AND registration) AND (deep AND learning)"

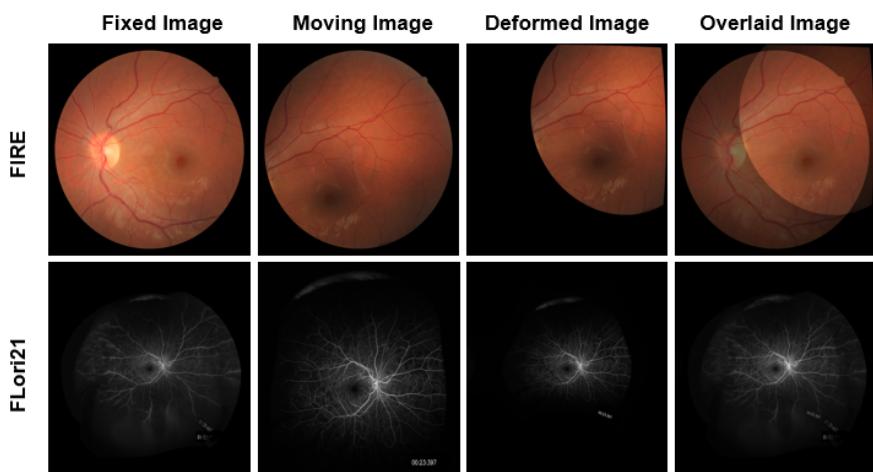


Figura 2.6: Exemplo de rexistro de imaxes de retina [5]

entrada é dividido en celdas e cada celda é asignada un valor (por exemplo nubes de puntos, matrices de píxeles ou véxoles...). 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 obtéñense 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 mellores resultados [47], polo que moita investigación diríxese a atopar alternativas que melloren a representación do sinal. [48]

Por exemplo SIREN [1], sobre a que profundizaremos mais adiante. [49] propón as funcións de activación gaussianas como alternativa a SIREN, e argumenta que poden obter mellores representacións e mais robustas. [50] achega unha nova función de activación baseada en wavelets, que demostra ser especialmente útil para a representación de imaxes.

As representacións implícitas poden ser clasificadas en dúas categorías: generalizables e sobreajustadas [51]. As representacións sobreajustadas céntranse en reproducir con precisión unha única sinal, mentres que as representacións generalizables poden modelar varias.

As INR son utilizadas en todo tipo de campos, dende xeración de imaxes [52], pasando por reconstrucción de obxectos [53] [54] ou modelado de sinais complejas [55].

As representacións implícitas están a recibir cada vez máis atención da comunidade médica [56], e son especialmente útiles para as tarefas de imaxe inversa, que requieren a reconstrucción de representacións correctas a partir de datos incompletos ou ruidosos. No caso de [57], propuxeron unha representación implícita para a reconstrucción de imaxes de resonancia magnética a partir de datos incompletos facendo uso de redes implícitas, e obtiveron resultados comparables a métodos tradicionais.

[53] propuxeron unha representación implícita para a representación de escenas 3D, optimizando unha función volumétrica continua que modela a densidade de volume e a radiancia emitida

en cada punto do espazo. Utilizando 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. Demostrando que as representacións implícitas están capacitadas para modelar escenas 3D complexas con alta fidelidade visual.

As representacións implícitas teñen bastante potencial no campo de planificación de traxectorias, como demostran [51] e [58], que propóñen o uso de INRs para modelar entornos e planificar traxectorias para un ou varios axentes. 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 comparable ás obtidas é adecuada para a maioria das aplicacións.

Mais especificamente, en [59] utilizan este tipo de representacións para garantir a seguridade do paciente durante a cirurxía teleoperada e optimizar a traxectoria do robot para evitar colisións co paciente, neste caso na boca e gorxa. 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.

Outro uso interesante das representacións neuronais implícitas é a compresión de imaxes. COIN [60] converter os datos en representacións neuronais implícitas, funcións que mapean coordenadas a valores RGB, logrando unha compresión eficiente e unha redución significativa do tempo de codificación en moitas modalidades.

[61] ten outros casos de uso interesantes.

Tamén se empregan en segmentación, compresión e síntesis de imaxes.

No caso do aliñamento de imaxes, buscase optimizar a función que mapea cada localización da imaxe fixa a unha localización da móbil.

2.4 IDIR

IDIR (Implicit Deformable Image Registration) é un método de aliñamento de imaxes baseado en redes neuronais. A súa principal diferenza frente a unha rede convolucional tradicional é que, en lugar de predicir a transformación entre imaxes, optimízase unha rede para esta mesma represente esta transformación.

O que [2] propón é optimizar directamente o DFV facendo uso dunha representación

implícita, de forma que a deformación está representada nos propios pesos dunha MLP.

[62] e [63] propuxeron métodos de rexistro de imaxes similares de forma independente, baseados en Neural ODE (ODE-Nets)[64], unha familia de modelos de aprendizaxe profundo que trata a rede como un sistema continuo en lugar de unha secuencia de capas discretas.

2.4.0.1 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 muestreados aleatoriamente do espazo de coordenadas dentro da máscara. O térmico de perda é a 'normalized cross-correlation' entre os valores dos píxeles muestreados na imaxe fixa e os correspondentes da imaxe móbil. Utilizan Adam de optimizador, cun learning rate de 0.0001.

2.4.0.2 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 desvantaxas.

As ReLUs teñen un sesgo cara a sinais de baixa frecuencia [47], o que significa que o modelo pode ter dificultades para representar pequenas deformacións locais no rexistro de imaxes.

[65] 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 [53] 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 [53] non utiliza unha función de activación periódica, mais para a representación adecuada de zonas de alta frecuencia utilizaron positional encoding, 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 2.7: 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 2.8: 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 [66] implementan unha versión simplificada de SIREN para facilitar o estudo destas, e propoñen melloras proceso de inicialización. Unha delas e utilizar a distribución Kaiming (He) en lugar da uniforme. Tamén propoñen un método para escoller un valor de w apropriado según o problema a resolver.

2.4.0.3 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 aliñamento 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. [67]

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. [68]

Regulizador Hiperelástico en 2D 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 $\mu_y, \mu_{\hat{y}}$ son as medias, $\sigma_y, \sigma_{\hat{y}}$ son as desviacións estándar, $\sigma_{y\hat{y}}$ é a covarianza, e C_1, C_2 son constantes para evitar divisións entre cero. [68]

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>

2.4.0.4 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 vóxeles, 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 ocurrir 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 desvíen de 1, tentando preservar áreas locais e evitar estiramentos ou dobreces extremas.

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

Figura 2.9: Regulizador Jacobiano

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

- Hyperelastic regularizer Tamén se pódense engadir restriccións ao DVF con este termo proposto por [69]. 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.

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

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 2.10: Regulizador Hiperelástico.

- Bending energy penalty Pódese impoñer a suavidade da deformación empregando esta penalización proposta en [70], 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. ...

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

Figura 2.11: Regulizador Bending Energy

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 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 2.12: Regulizador Jacobiano en 2D

- Hyperelastic regularizer:
- Bending energy penalty:

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

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

Figura 2.13: Regulizador Hiperelástico en 2D

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

Figura 2.14: Regulizador Bending Energy en 2D

2.4.0.5 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 e móbil, L_{reg} é un termo de regularización na transformación Φ , e α é un termo de ponderación.

$$\hat{\Phi} = \underset{\Phi}{\text{Arg min}} L_{data}(M \circ \Phi, F) + \alpha L_{reg}(\Phi) \quad (2.1)$$

A principal innovación que introduce [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 INR dentro dun MLP 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 calquier tamaño, ao contrario dunha CNN tradicional que ten que ser adaptada para cada resolución.

Segundo, facelo desta forma permite aproveitar as capacidades de librarías 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 eficientes 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, entrenarase 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.

2.5 Traballo proposto

Baseándose no framework introducido por [2], propónse modificalo para adaptalo a imaxes médicas de retina.

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 desenrrolo

Ao ser un proxecto de investigación, a metodoloxía de traballo mais adecuada é unha metodoloxía iterativa e incremental.

Os requisitos evolucionan ao longo do proxecto, polo que é vital planificar para poder adaptarse a estes cambios.

3.2 Planificación do proxecto

Implementar IDIR orixinal e replicar resultados.

Implementar adaptacións necesarias para o seu uso en imaxes oftalmolóxicas. - Imaxes de pulmón son 4d, imaxes oftalmolóxicas son 2d (ignoramos color non aporta moita información)

Implmentar evaluación para FIRE e RFMID.

Experimentar para mellorar resultados.

Escribir memoria.

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 desenrrolo

de redes neuronais. (version y CUDA). Tamén se empregan librerías de apoio como NumPy (para traballar con matrices), Matplotlib (visualización), OpenCV ou scikit-learn (manxo de imaxes). (versions y links).

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

3.3.2 Hardware

O proxecto foi desenvolvido 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 imáxenes na súa resolución orixinal foi necesario emplegar o segundo debido ás limitacións de memoria da GPU.

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

Táboa 3.1: Comparativa entre os servidores utilizados

3.3.3 Estimación de custos

Os costos do hardware son ignorados xa que xa estaba disponible antes da realización do proxecto. Os costos dos recursos humanos calcúlanse para un estudiante e doués tutores, resultando nun custo estimado de *PLACEHOLDER*.

Recurso	Coste por hora	Horas estimadas	Coste total
Estudiante			€
Titor 1			€
Titor 2			€

Táboa 3.2: Estimación de custos dos recursos humanos

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

Capítulo 4

Traballo Realizado

NESTE apartado presentarase o traballo realizado, comezando por unha vista xeral do proceso, seguido dunha explicación dos diferentes módulos desenvolvidos e a súa interacción, así como os conxuntos de datos empregados. Finalmente, presentaranse os resultados obtidos acompañados dunha análise dos mesmos.

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.

Inicialmente replicáronse os resultados obtidos por Wolterink et al [2], que se mostran na táboa seguinte:

4.2 Conxuntos de datos

4.2.1 FIRE

[6] 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.

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.

Táboa 4.1: Comparación de Resultados entre IDIR y Replicación

Scan	IDIR Original / 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)

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 4.1: Clasificación dos pares de imaxes en categorías.

4.2.2 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 utilizalo 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.

...preguntar David



Figura 4.2: Exemplo de imaxes do conxunto de datos FIRE [6] cos puntos de control indicados. De esquerda a dereita, categorías *S*, *P*, *A*.

4.2.3 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.

4.3 Métodos de Avaliación

A evaluación divídese en dous tipos: a avaliação cualitativa, na cal analízanse os resultados de forma visual, e a avaliação cuantitativa, na cal se utilizan métricas numéricas para comparar os resultados en base a un criterio obxetivo.

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.

4.3.1 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 error de rexistro calcúlase ca distancia media entre os puntos correspondentes na imaxe fixa e móbil (c_j, r_j). Cando o erro de rexistro entre un par de imaxes está por debaixo do límite, 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.

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).

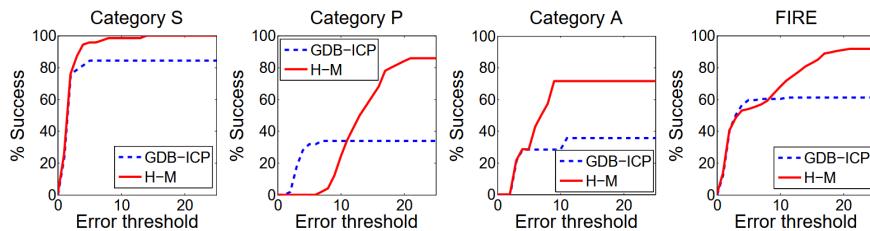


Figura 4.3: Gráfico de avaliação FIRE, [6]

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

4.3.2 Avaliación Cualitativa

No caso deste traballo, a avaliación cualitativa cobra gran importancia, xa que na cuantitativa so se está a comparar sobre un número reducido de puntos en cada parexa de imaxes. A avaliación 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 avaliación visual é especialmente relevante, xa que tan

só se proporcionan 10 puntos de referencia por imaxe, que non son suficientes para avaliar a calidade do rexistro en moitas zonas da imaxe. Xa que en RFMID [9] uilízanse puntos de referencia xerados manualmente que cubren toda a imaxe, a avaliación visual é algo menos relevante, xa que é mais probable que unha deformación local incorrecta sexa detectada por algún punto e se vexa reflexado nas métricas.

4.4 Proceso de Rexistro

Inicializase a rede, ca arquitectura modificada nas capas de entrada e saída (orixinalmente [3, 256, 256, 256, 3], agora [2, 256, 256, 256, 2]) 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 entrenamento por cada epoch repítese o seguinte proceso:

Se muestrean "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, aplicase 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 propágase cara atrás para actualizar os pesos da rede.

O proceso de muestreo pode ser aleatorio ou utilizando unha estratexia de muestreo específica. Mientras que para o caso dos pulmóns a estratexia aleatoria era suficiente, para o fondo de ollo é necesario unha estratexia mais intelixente xa que gran parte da imaxe non contén información relevante. As imaxes de rertinas conteñen a maior parte de información relevante nos arredores do disco óptico e vasos sanguíneos, polo que é necesario muestrear máis puntos nesta zona. Neste caso se implementouse unha estrategia na que probabilidade de que una coordenada sexa muestreada é determinada por un mapa de pesos que indica a importancia de cada punto. Este mapa de pesos é calculado ao inicia a partir da imaxen fixa, aplicando operadores de Sobel de 10 píxeles de tamaño para computar os gradientes e a súa magnitud, e posteriormente utilizando unha kernel gaussiano moi agresivo (201, 201) para suavizar o resultado.

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. Aínda que partimos do traballo de IDIR, a tarefa de rexistro de retinas presenta diferenzas importantes respecto ao rexistro de pulmóns. Por este motivo, non podemos asumir que os parámetros óptimos para pulmóns sexan tamén os mellores para retinas.

Ao longo dos experimentos, a comparación principal céntrase na función de activación empregada (SIREN ou ReLU). Debido ao gran espazo de búsqueda que implica probar todas as configuracións posibles, os experimentos iniciais centraránse en fixar unha unha serie de parámetros con valores razonables para poder experimentar só con aqueles que teñan un impacto máis significativo.

Ademais, 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.

5.2 Experimentos

5.2.1 Experimentos iniciais

Inicialmente tentaremos determinar uns valores aceptables para varios dos parámetros da rede. Isto é relevante xa que moitos destes parámetros son dependentes uns de outros.

A menos que especificado de outro xeito, os parámetros utilizados para os experimentos son os seguintes:

- Función de perda: NCC
- Regularización:
 - Bending energy penalty: 10.0
 - Jacobian regularizer: 0.05
 - Hyperelastic regularizer: 0.25
- Learning rate: 0.0001
- Batch size: 10000 puntos
- Epochs: 1500
- Optimizador: Adam
- Resolución: 1000x1000

5.2.1.1 Función de loss

Planteamento A función de loss é un dos aspectos más importantes á hora de entrenar unha rede neuronal. As funcíons de perda valoradas para este traballo xa forón explicadas en [2.4.0.3](#).

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, tomaráse a distancia media de todos os puntos como métrica de comparación.

Resultados Os resultados obtidos son os seguintes: [5.1](#), [5.1](#)

Discusión Obsérvase como as métricas que teñen en conta a estructura 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 estructura 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 estructura da imaxe.

NCC tende a ser mais robusta a cambios uniformes na intensidade global, mentres que SSIM tende a ser mellor con cambios locais. SSIM tamén é menos robusta ao ruído, sensible

Loss Function	FIRE Dataset		RFMID Dataset	
	Relu	SIREN	Relu	SIREN
huber	399.86	397.45	7.13	57.31
l1	404.83	391.87	5.42	52.91
mse	392.94	410.97	9.50	121.54
ncc	250.59	281.03	36.04	79.84
smoothl1	414.79	387.71	7.01	60.74
ssim	268.72	264.98	23.34	55.07

Táboa 5.1: Distancias medias segundo función de perda, tipo de rede e datasets (FIRE vs. RFMID)

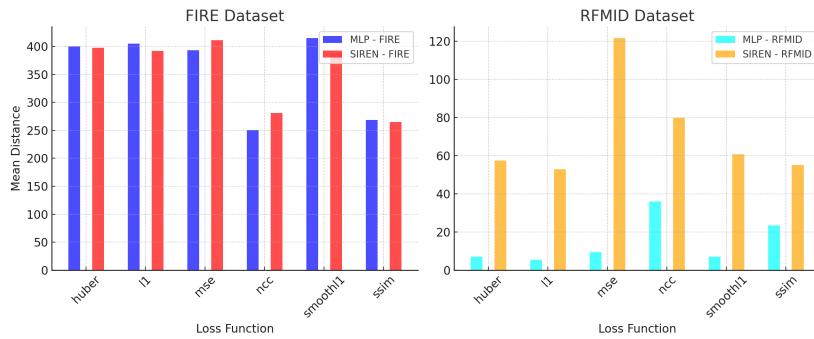


Figura 5.1: Comparación de diferentes funcións de perda sobre imaxes de FIRE e RFMID

o tamaño das seccións utilizadas e 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.

Conclusións En base aos resultados obtidos, pódense extraer varias conclusións relevantes:

- 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. En particular, NCC mostra o menor erro

medio (250.59 para Relu e 281.03 para SIREN). 2. Para o dataset RFMID, que contén imaxes con tan só variación xeométrica, as funcións de perda baseadas en píxeles como L1 e Smooth L1 ofrecen mellores resultados. Concretamente, L1 presenta o menor erro medio para Relu (5.42) e resultados competitivos para SIREN (52.91). 3. 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. Isto débese a que Relu tende a producir funcións predominantemente lineares, o que se adapta mellor ás transformacións realizadas no dataset RFMID. 4. SSIM, a pesar de ser teoricamente robusta a cambios locais, non mostra unha vantaxe significativa sobre NCC.

A elección de NCC como función de perda estándar baséase tanto na súa robustez empírica coma na súa consistencia co obxectivo de alinear imaxes reais de retina, onde a variabilidade en iluminación e contraste é un factor importante.

5.2.1.2 Resolución da imaxe

Planteamento A resolución da imaxe é un aspecto clave xa que 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 moi maior que nunha imaxe de 512x512.

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 perjudicial 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 resolucións de ata 2160x2160.

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.

Resultados 5.2, 5.3, 5.4, 5.5

Discusión Pódese observar como unha maior resolución tende a dar lixeiramente mellores resultados, pero a un custo computacional maior. Isto pode deberse mais á precisión 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

Resolution	Mean Distance
100	254.22
250	251.29
750	250.62
1250	250.59
1708	249.72

Táboa 5.2: Distancias medias para o dataset FIRE ca función de activación Relu

Resolution	Mean Distance
100	266.43
250	263.85
750	263.19
1250	258.56
1708	258.06

Táboa 5.3: Distancias medias para o dataset FIRE ca función de activación SIREN

Resolution	Mean Distance
100	37.29
250	36.18
750	36.01
1250	35.03
1708	35.04

Táboa 5.4: Distancias medias para o dataset RFMID ca función de activación Relu

Resolution	Mean Distance
100	68.12
250	73.42
750	77.55
1250	67.33
1708	67.31

Táboa 5.5: Distancias medias para o dataset RMIFD ca función de activación SIREN

a maioría da información relevante para a tarefa de rexistro xa está capturada en resolucións inferiores.

Conclusións Baseándonos nos resultados obtidos, podemos concluir 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.
3. O comportamento respecto á resolución é consistente para ambos tipos de modelos (Relu e SIREN) e para ambos datasets (FIRE e RFMID), o que suxire que estas conclusións son xeneralizables.

Para os experimentos subseguientes, adoptarase unha resolución estándar de 1000×1000 píxeles, que demostrou proporcionar o mellor rendemento global para a tarefa de rexistro de retinas.

5.2.1.3 Regularización

Planteamento O proceso de regularización axuda 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 en 2.4.0.4, son as seguintes:

- Regularizador do Jacobiano: penaliza as desviacións do determinante da matriz Jacobiana respecto a 1, limitando expansións ou compresións locais excesivas.
- Regularizador hiperelástico: engade termos basados na enerxía de deformación, controlando a extensión e a expansión de superficie e garantindo transformacion suaves y difeomórficas.
- Penalización da enerxía de flexión: mide a magnitude das segundas derivadas do campo de deformación, promovendo que a superficie resultante sea o mais realista posible e reduciendo oscilacións de alta frecuencia.

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.

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. Na traballo orixinal de IDIR, chegaban a usar 13 pasadas debido a que traballan en 3D.

Ademais, pese a que os diferentes termos de regularización valoran diferentes aspectos, cabe ter en conta que tamén coinciden nalgúns das propiedades que valoran. Por exemplo, o regularizador hiperelástico pode considerarse un caso mais xeral que inclúe indirectamente penalizaciones de volumen (Jacobian) e de suavidade (parte de la energía de flexión).

Para determinar cal é a regularización óptima, realizáronse experimentos comparando o rendemento de cada unha sobre unha mostra de imaxes dos datasets de FIRE e RFMID cas

diferentes funcións de activación. Destes experimentos obterase un mapa de calor que mostra a distancia media entre os puntos correspondentes para cada combinación de parámetros.

Resultados 5.2

Discusión Os resultados amosan que a regularización ten un impacto significativo no rendemento da rede.

Conclusíons

5.2.1.4 Learning rate

Planteamento 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 magnitude 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 heurísiticas mais comúns para relacionar o learning rate e o batch size é a regla de escalado linear [71]. 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 mais grandes temos unha mellor aproximación do gradiente real, é posible utilizar un learning rate maior sen que a rede diverxa.

Para determinar cal é o learning rate óptimo, realizáronse experimentos comparando o rendemento de cada un sobre unha mostra de imaxes dos databases de FIRE e RFMID cas diferentes funcións de activación.

Resultados 5.3, 5.4

Discusión Valores de learning rate moi altos (0.001 and 0.005) son contraproducentes, xa que a rede diverxe rapidamente.

A dependencia entre o learning rate e o batch size confirmase. Xeralmente un learning rate baixo ($1.0e-05$, $1.0e-06$) parece requerir de batch sizes maiores e viceversa, o cal é consistente co que se esperaba.

Tamén observase que batch sizes maiores tenden a dar melhores resultados.

Conclusíons

5.2.2 Batch size

Planteamento Ao longo dos experimentos realizados, o análisis cualitativo revelou que o batch size é un dos parámetros que más impacto ten no rendemento da rede. Unha vez determinados uns valores aceptables nos parámetros anteriores, realizáronse experimentos para determinar cal era o batch size óptimo.

Neste caso dividimos o conxunto de datos de RFMID en varios subconxuntos según a dificultade da transformación, medida mediante a norma de Frobenius. A norma de Frobenius dunha matriz $A \in \mathbb{R}^{m \times n}$ defíñese como:

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

onde a_{ij} son os elementos da matriz A . Esta é unha xeralización da distancia euclidiana aplicada a matrices, onde as imaxes con transformacións mais grandes considéranse mais difíciles.

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

Resultados 5.5, 5.6, 5.7 5.10, 5.8, 5.9

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, aínda que si un maior custo computacional.

Conclusións Interesa determinar cal é o punto de inflexión onde o aumento do batch size non compensa o aumento do rendemento.

5.2.3 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.

Planteamento Para solucionar isto, propúxose unha estratexia de mostraxe mais intellixente, onde se calcula 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.

Posteriormente tamén se introducíu unha estratexia de mostraxe uniforme, onde se seleccionan un número fixo de puntos en cada 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.

Así mesmo implementouse un scheduling do batch size, ca idea de utilizar un batch size pequeno nos epochs iniciais para que a rede aprenda unha transformación global a grandes rasgos, e aumentar o número de puntos conforme avanza o entrenamento para que a rede aprenda as transformacións locais. A estratexia de mostraxe uniforme é a mais axeitada para este caso, especialmente cando se utiliza un batch size pequeno. O learning rate modificaráse de forma proporcional para manter a relación entre ambos.

Na figura 5.11 pódense observar os diferentes tipos de mostraxe utilizados.

Resultados 5.12 5.13 5.14 5.15

Discusión A hipótese da estratexia de mostraxe intellixente non parece ser a mais axeitada, con resultados peores ca estratexia aleatoria. A estratexia uniforme tampouco parece ser a mais adecuada, xa que non se obtén resultados significativamente mellores ca a aleatoria.

Conclusións

5.2.4 Inicialización

Planteamento É posible que a inicialización da rede sexa un factor mais importante que a estratexia de mostraxe. É posible que certas inicializacíons coloquen os desprazamentos de forma 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 mais beneficiosa. É posible que fore mellor esperar ata un epoch algo mais avanzado para determinar a inicialización, xa que no epoch 0 non hay ningunha seguridade de que non sexa un mínimo local, pero isto tamén implicaría un maior custo computacional.

Resultados 5.16, 5.17

5.18, 5.19

Discusión Obsérvase que varias o batch size ao longo do proceso de optimización non ten os resultados esperados, empeorando o rendemento da rede. Tamén se observa que a lotería de inicialización si que provoca unha mellora, áinda que non moi significativa, e non se beneficia particularmente de utilizar mais de 5 inicializacíons.

Conclusións 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. Poderíase tamén determinar como de sensible é a rede a diferentes inicializacíons.

5.2.5 Exemplos de rexistro

5.20

Estes exemplos ilustran o resultado do rexistro de diferentes imaxes, mostrando tanto rexistros exitosos como fallidos. A primeira imaxe corresponde ca imaxe fixa, a segunda corresponde ca imaxe rexistrada, a terceira ca imaxe móvil e a cuarta os campos de deformación aplicado a unha gralla cadrada.

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

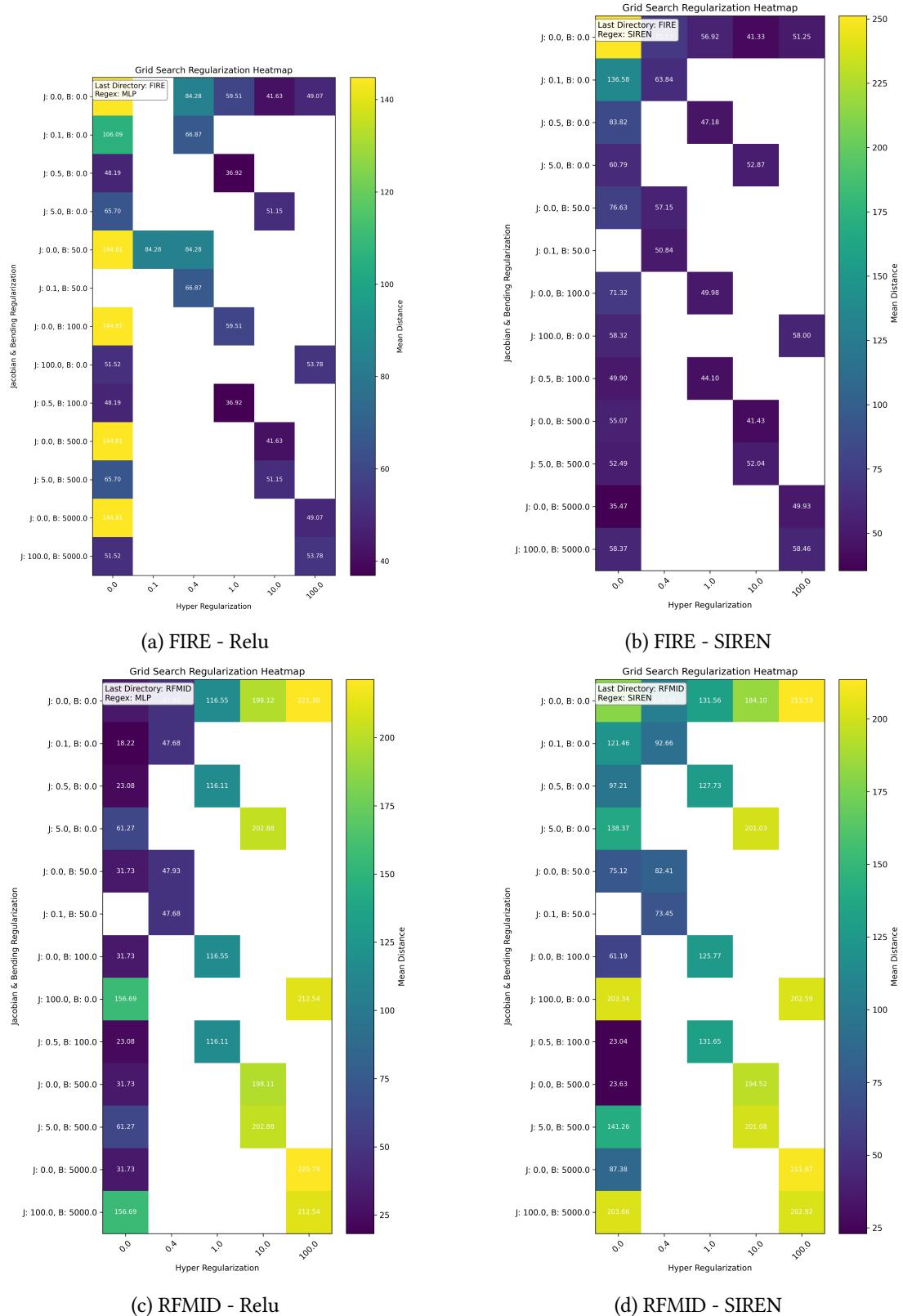


Figura 5.2: Mapa de calor cos resultados de diferentes combinacóns de termos de regularización e funcións de activación sobre os datasets FIRE e RFMID

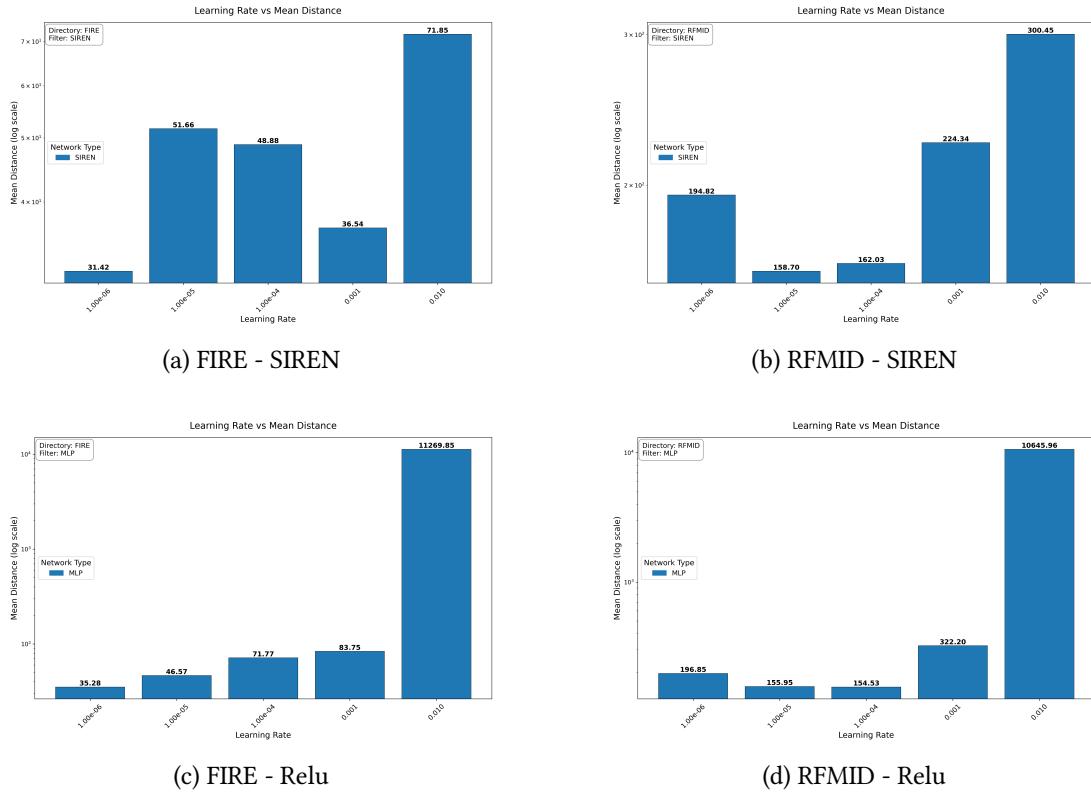


Figura 5.3: Reusltados lr cos datasets FIRE e RFMID cas funcións de activación SIREN e ReLU.

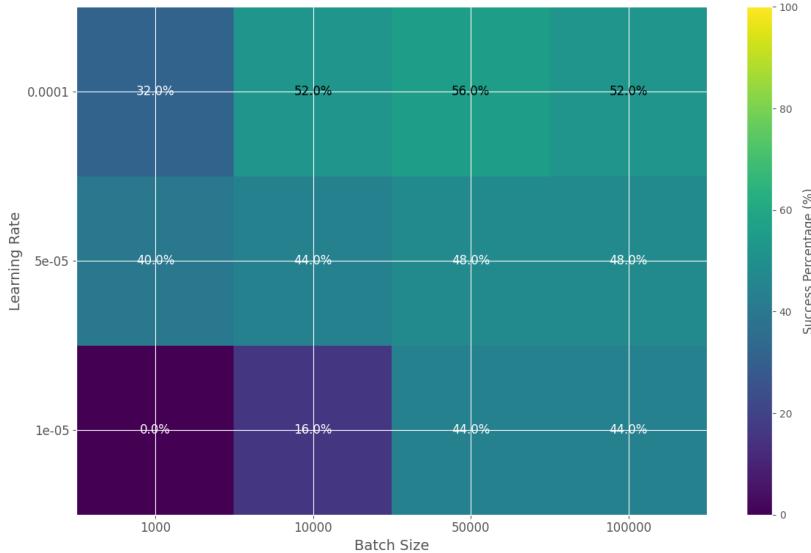


Figura 5.4: Mapa de calor cos resultados de diferentes combinacións de batch size e learning rate con unha mostra de imaxes de RMIFD ca función de activación ReLU

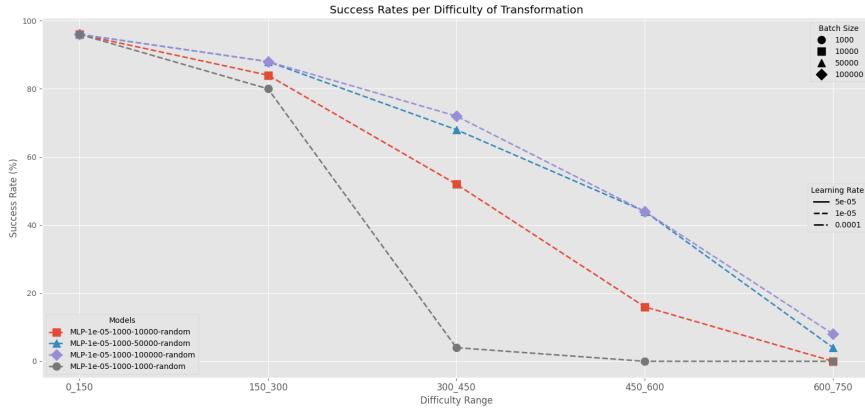


Figura 5.5: Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación ReLU (learning rate = 1e-5)

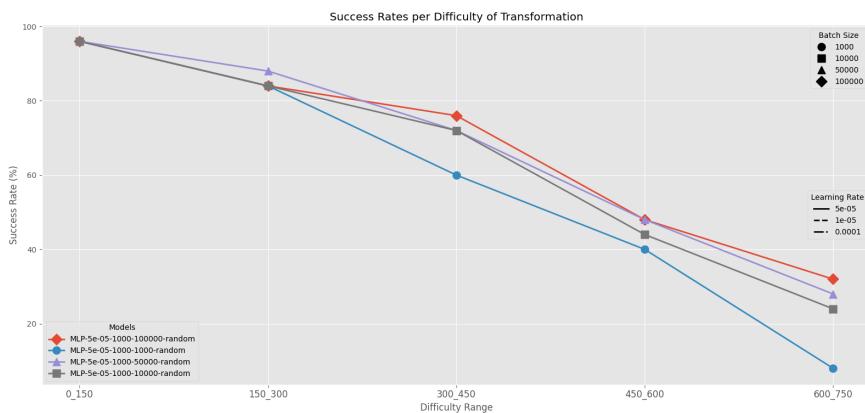


Figura 5.6: Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación ReLU (learning rate = 5e-5)

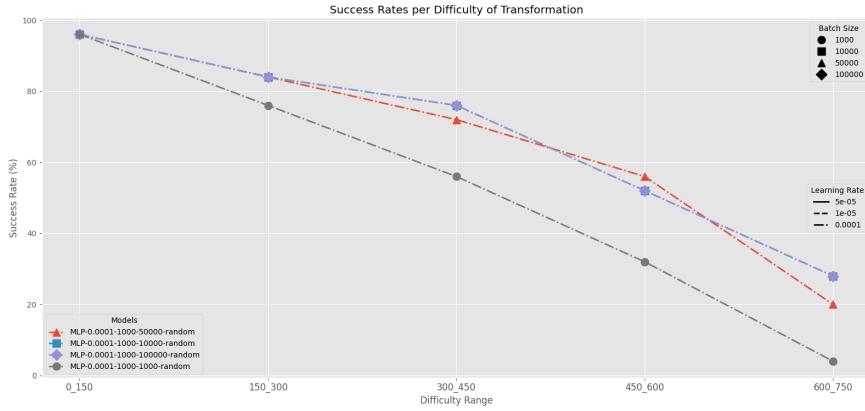


Figura 5.7: Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación ReLU (learning rate = 1e-4)

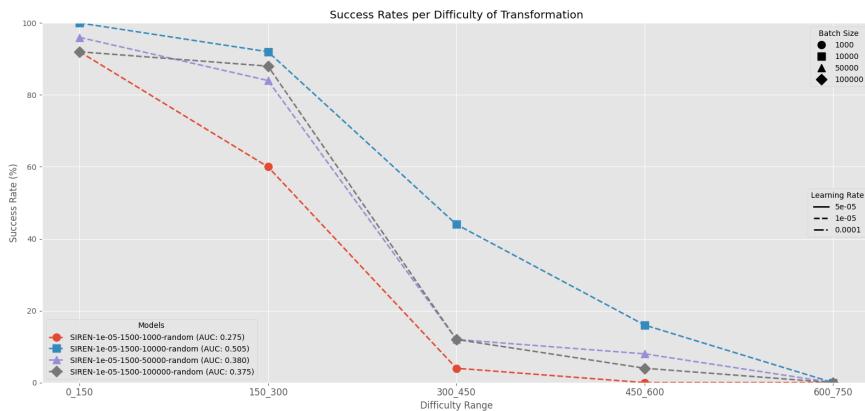


Figura 5.8: Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación SIREN (learning rate = 1e-5)

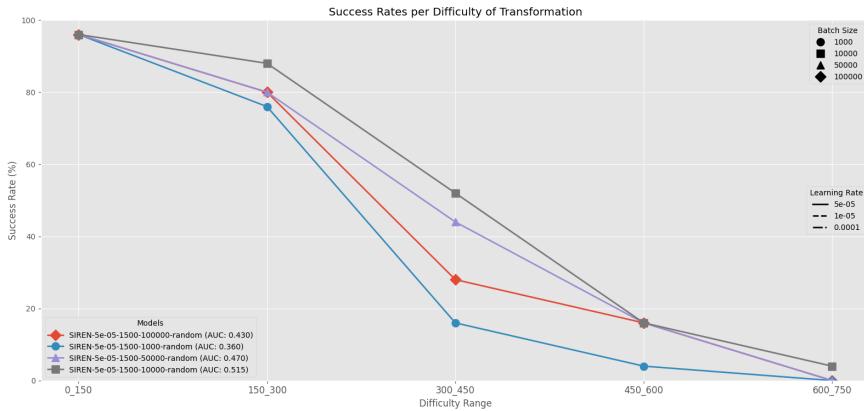


Figura 5.9: Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación SIREN (learning rate = 5e-5)

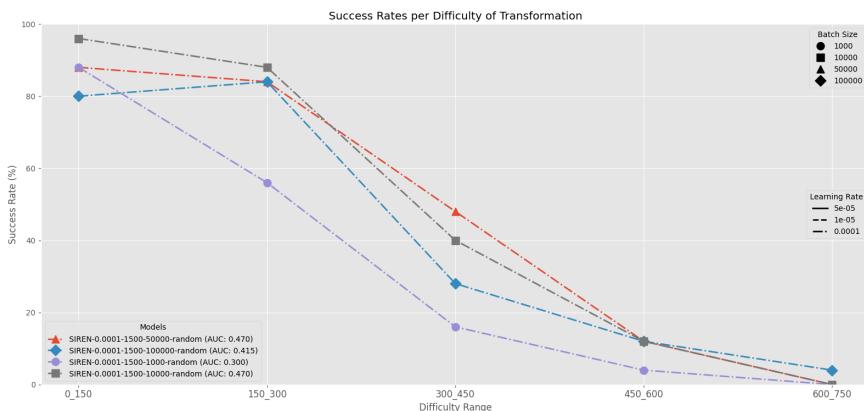


Figura 5.10: Comparación do rendemento da rede con diferentes batch sizes sobre imaxes do dataset RFMID ca función de activación SIREN (learning rate = 1e-4)

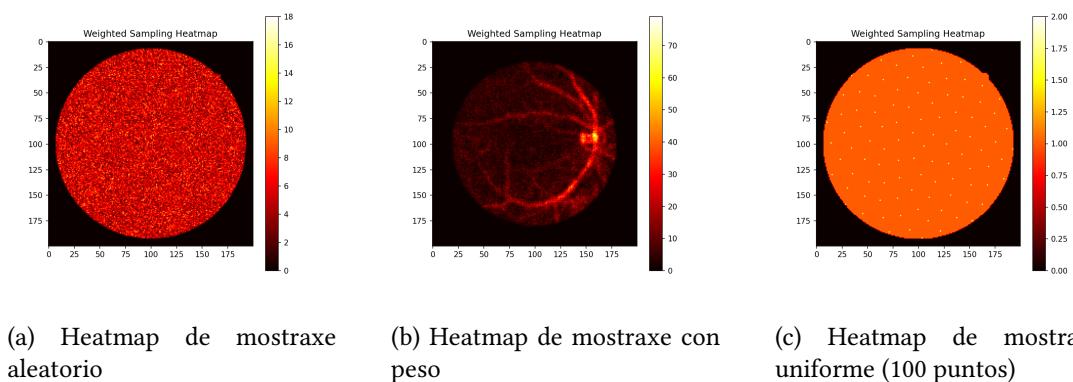


Figura 5.11: Heatmaps de mostraxe

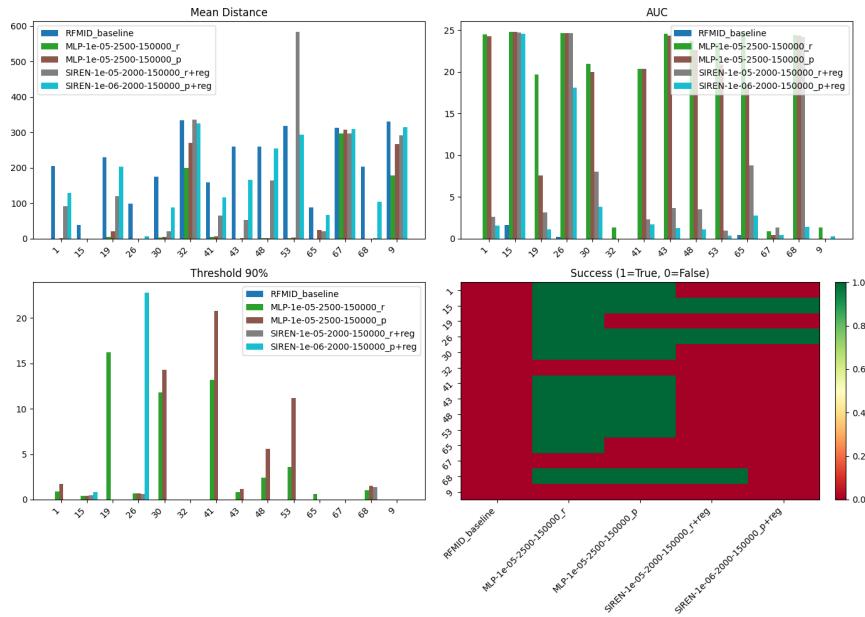


Figura 5.12: Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset RFMID

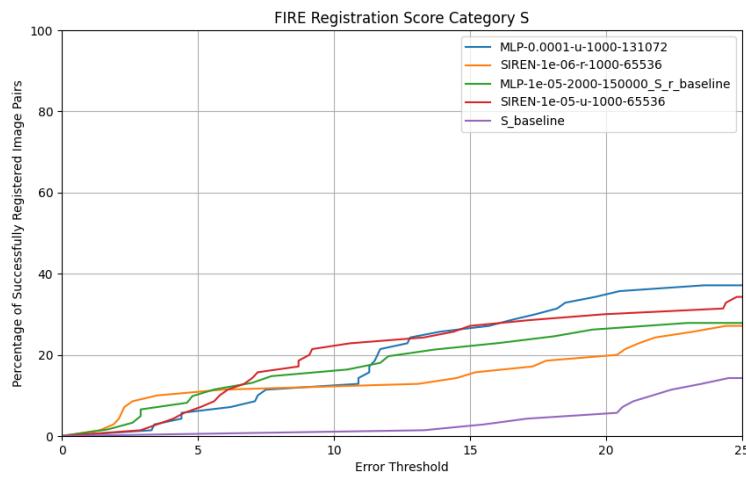


Figura 5.13: Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset FIRE

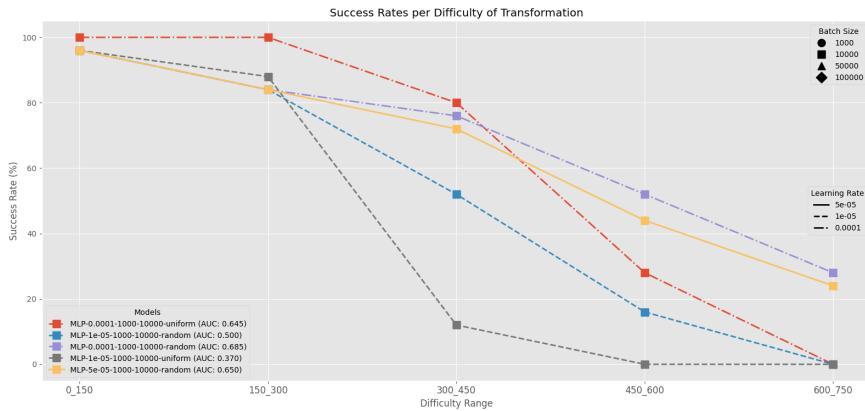


Figura 5.14: Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset RFMID ca función de activación RELU

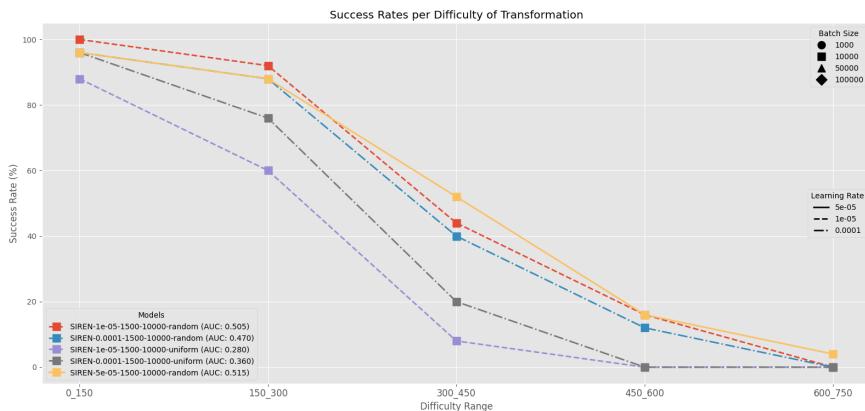


Figura 5.15: Comparación das diferentes estratexias de mostraxe sobre imaxes do dataset RFMID ca función de activación SIREN



Figura 5.16: Exemplos das diferentes inicializacions ca función de activación RELU



Figura 5.17: Exemplos das diferentes inicializacions ca función de activación SIREN

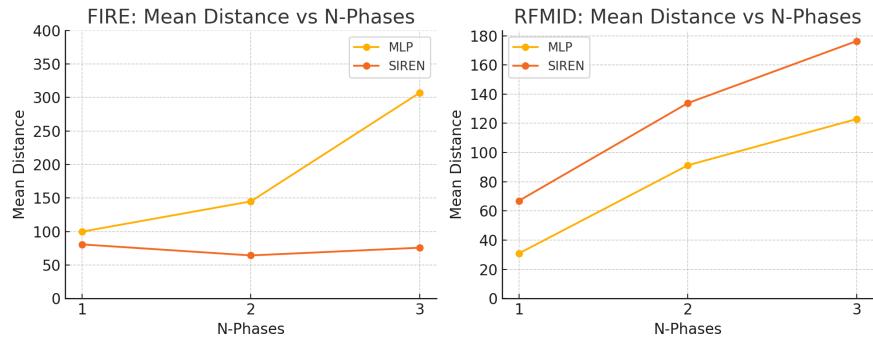


Figura 5.18: Resultados de usar distinto número de fases

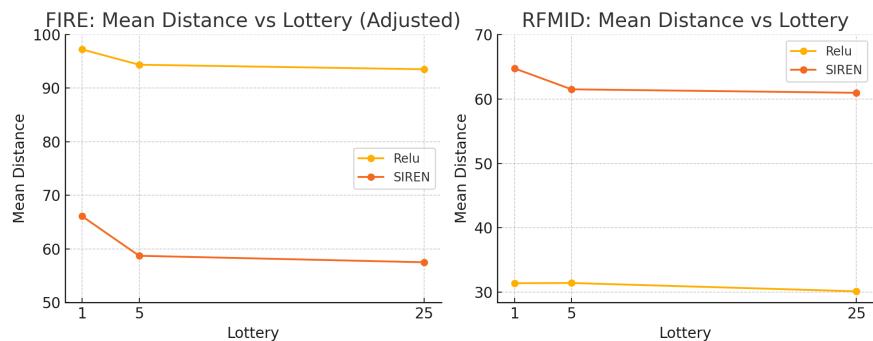
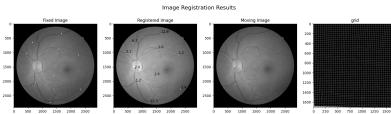
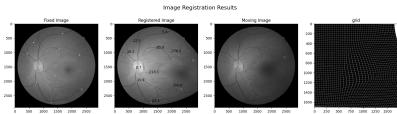


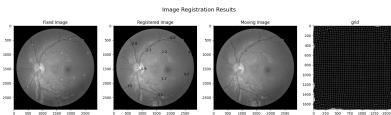
Figura 5.19: Resultados da lotería de inicialización



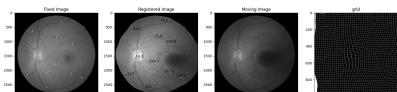
(a) Rexistro exitoso dunha parella de imaxes do dataset FIRE ca función de activación ReLU



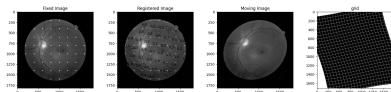
(b) Rexistro fallido dunha parella de imaxes do dataset FIRE ca función de activación ReLU



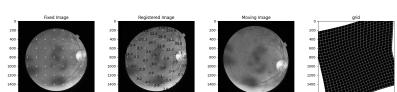
(c) Rexistro exitoso dunha parella de imaxes do dataset FIRE ca función de activación SIREN



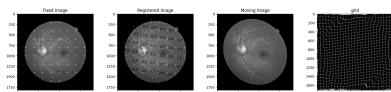
(d) Rexistro fallido dunha parella de imaxes do dataset FIRE ca función de activación SIREN



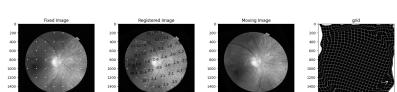
(e) Rexistro exitoso dunha parella de imaxes do dataset RFMID ca función de activación ReLU



(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 5.20: Exemplos de rexistro: combinacíons de dataset (FIRE/RFMID), función de activación (relu/SIREN) e éxito.

Capítulo 6

Conclusións

Capítulo 7

Traballo futuro

EXISTEN varias liñas de traballo futuro que se poden seguir para mellorar o sistema. A continuación, descríbense algunas das que teñen maior potencial:

7.1 Instant Neural Graphics Primitives

Introducidas por [72], propoñen encodear os inputs da rede a un espacio dimensional superior.

Encodear os inputs da rede é unha técnica que xa se emprega en moitas ocasións (one-hot encodings, transformers...) Eles utilizan 'sparse parametric encodings' utilizando unha tabla de hashes de múltiples resolucións, que tamén tén parámetros entrenables e fai parte do traballo de aprendizaxe da rede. Isto permítelles un entrenamento e inferencia moito mais rápido que outros métodos, sen ter que sacrificar en rendemento.

[73] aplica estas ideas á tarefa de rexistro, con moi bós resultados. Notablemente, resuelven o 'sliding boundary problem', que se refiere ás complicación de modelar o movemento relativo entre diferentes estructuras. No caso da imaxe pulmonar, surxe cuando os lóbulos dos pulmóns se deslizan entre sí durante la respiración.

7.2 Invertibilidade

Os resultados de esta rede non teñen ningunha garantía de ser difeomórficos, mais debido 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).

É posible garantir a invertibilidade da transformación facendo uso de redes invertibles [74]. Tamén con campos de velocidade neuronais [62] ...

Apéndices

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. . 25
- CNN** Siglas en inglés de *Convolutional Neural Network*. Tipo de red . . 1
- CNN** Siglas en inglés de *Generative Adversarial Network*. Tipo de red . . 10
- DFV** Siglas en inglés de *Campo de Vectores de Deformación..* 6
- FOV** Siglas en inglés de *Field Of View.* . 25
- GDB-ICP** Siglas en inglés de *Generalized Dual-Bootstrap Iterative Closest Point.* . 11
- GMM** Siglas en inglés de *Gaussian Mixture Models.* . 11
- IGRT** Siglas en inglés de *Image-guided radiation therapy.* . 5
- MFSP** Siglas en inglés de *Mixture Feature and Structural Preservation Feature Points Registration.* . 11
- NTK** Siglas en inglés de *Neural Tangent Kernel.* . 22
- PIFD** Siglas en inglés de *Best Bin First.* . 11
- PIIFD** Siglas en inglés de *Partial Intensity Invariant Feature Descriptor.* . 11
- PSO** Siglas en inglés de *Particle Swarm Optimization.* . 11

REMPE Siglas en inglés de *Registration of Retinal Images Through Eye Modelling and Pose Estimation.* . [11](#)

SURF Siglas en inglés de *Speeded-Up Robust Features.* . [11](#)

TPS Siglas en inglés de *Thin Plated Spline.* . [11](#)

UR-SIFT-PIIFD Siglas en inglés de *Uniform Robust Scale Invariant Feature Transform e Harris-Partial Intensity Invariant Feature Descriptor.* . [11](#)

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] “Scopus,” Retrieved from <https://www.scopus.com/>, 2024, accessed on 10-12-2024.
- [5] 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>
- [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] 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.
- [21] 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>
- [22] 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>
- [23] 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>
- [24] 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.
- [25] 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.
- [26] 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>
- [27] 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.
- [28] 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>
- [29] 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>
- [30] 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>
- [31] M. Lafarge, P. Moeskops, M. Veta, J. Pluim, and K. Eppenhof, “Deformable image registration using convolutional neural networks,” p. 27, 03 2018.
- [32] 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>
- [33] 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>
- [34] D. Mahapatra, “Gan based medical image registration,” 2019. [En línea]. Disponible en: <https://arxiv.org/abs/1805.02369>
- [35] Y. Zheng, X. Sui, Y. Jiang, T. Che, S. Zhang, J. Yang, and H. Li, “Symreg-gan: Symmetric image registration with generative adversarial networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5631–5646, 2022.
- [36] 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.
- [37] 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.
- [38] 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>
- [39] 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.

- [40] 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.
- [41] 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 liña]. Disponible en: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-33-7-1267>
- [42] H. Tang, A. Pan, Y. Yang, K. Yang, Y. Luo, S. Zhang, and S. Ong, “Retinal image registration based on robust non-rigid point matching method,” *Journal of Medical Imaging and Health Informatics*, vol. 8, pp. 240–249, 02 2018.
- [43] F. Bookstein, “Principal warps: thin-plate splines and the decomposition of deformations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.
- [44] 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.
- [45] 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.
- [46] C. Hernandez-Matas, X. Zabulis, and A. A. Argyros, “Retinal image registration through simultaneous camera pose and eye shape estimation,” in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3247–3251.
- [47] 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 liña]. Disponible en: <https://arxiv.org/abs/1806.08734>
- [48] 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 liña]. Disponible en: <https://arxiv.org/abs/2411.03688>
- [49] S. Ramasinghe and S. Lucey, “Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps,” 2022. [En liña]. Disponible en: <https://arxiv.org/abs/2111.15135>
- [50] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. G. Baraniuk, “Wire: Wavelet implicit neural representations,” 2023. [En liña]. Disponible en: <https://arxiv.org/abs/2301.05187>

- [51] 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>
- [52] 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>
- [53] 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>
- [54] 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>
- [55] 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>
- [56] 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>
- [57] 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>
- [58] 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.
- [59] 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.
- [60] 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>
- [61] 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>

- [62] 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>
- [63] 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>
- [64] T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, “Neural ordinary differential equations,” *CoRR*, vol. abs/1806.07366, 2018. [En línea]. Disponible en: <http://arxiv.org/abs/1806.07366>
- [65] 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>
- [66] 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>
- [67] 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.
- [68] 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>
- [69] 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>
- [70] 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.
- [71] 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>
- [72] 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>

-
- [73] X. Li, F. Zhang, M. Li, D. Weber, A. Lomax, J. Buhmann, and Y. Zhang, “Neural graphics primitives-based deformable image registration for on-the-fly motion extraction,” 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2402.05568>
 - [74] J.-H. Jacobsen, A. Smeulders, and E. Oyallon, “i-revnet: Deep invertible networks,” 2018. [En línea]. Disponible en: <https://arxiv.org/abs/1802.07088>