



TRABALLO FIN DE GRAO
GRAO EN ENXEÑARÍA INFORMÁTICA
MENCIÓN EN COMPUTACIÓN



Aliñamento de imaxes oftalmolóxicas usando representacións neuronais implícitas

Estudiante: Mateo Amado Ares

Dirección: José Rouco Maseda

Jorge Novo Buján

A Coruña, decembro de 2024.

Dedicatoria

Agradecementos

Lorem ipsum dolor sit amet, consectetur 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, consectetur 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. David Rivas Villar

Resumo

O aliñamento da imaxe oftalmolóxica é un campo moi relevante. Aliñar imaxes médicas é útil para, entre outras cousas, revisar o avance dunha enfermidade ao longo do tempo ou fusionar diferentes modalidades. O caso dos ollos é de particular importancia xa que permiten a observación in-vivo de tecido neuronal e vasos sanguíneos. Aliñar as imaxes manualmente é un proceso tedioso e complexo, polo que automatizar este proceso é moi beneficioso.

Neste traballo explórase o uso de redes de representación implícita, onde se parametriza a imaxe como unha función continua coas coordenadas como entrada e o valor do pixel como saída, como unha alternativa para o aliñamento de imaxes. Estas aportan vantaxes fronte 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, adoitan empregar unha función de activación sinusoidal (SIREN), que pode axudar a eliminar o sesgo cara sinais de baixa frecuencia e mapear mellor deformación pequenas e detalladas.

Adaptando o traballo realizado por [1], valorarase se este método é apto para a tarefa de aliñamento de imaxes oftalmolóxicas e como se compara con métodos convencionais.

Abstract

Ophthalmic image alignment is a highly relevant field. Aligning medical images is useful for, among other things, reviewing the progression of a disease over time or merging different modalities. The case of eyes is particularly important as they allow in-vivo observation of neuronal tissue and blood vessels. Manually aligning images is a tedious and complex process, so automating this process is beneficial.

This work explores the use of implicit representation networks, where the image is parameterized as a continuous function with coordinates as input and pixel value as output. This provides advantages over traditional discrete representations such as resolution independence and the ability to dispense with large databases since they are trained through an optimization process for each group of images. Furthermore, instead of using standard activation functions like RELU, they typically employ a sinusoidal activation function (SIREN), which can help eliminate bias towards low-frequency signals and better map small and detailed deformations.

Based on the work done by [1], this study will evaluate whether this method is suitable for the task of aligning ophthalmic images and how it compares to conventional methods.

Palabras chave:

- 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 oftalmolóxica	4
2.1.3	Retinografía	4
2.2	Rexistro de imaxes	4
2.2.1	Estado da arte	6
2.2.2	En retinas	8
2.3	Representación Neuronaís Implícitas	9
2.4	IDIR	10
2.5	Traballo proposto	13
3	Metodoloxía e planificación	14
3.1	Metodoloxía do desenrrolo	14
3.2	Planificación do proxecto	14
3.3	Recursos utilizados	14
3.3.1	Software	14
3.3.2	Hardware	15
3.3.3	Estimación de custos	15
4	Traballo Realizado	16
4.1	Vista Xeral	16
4.2	Conxuntos de datos	16

4.2.1	FIRE	16
4.2.2	RFMID	16
4.3	Avaliación	16
4.4	Discusión	16
5	Experimentos e resultados	17
6	Conclusións	18
7	Traballo futuro	19
7.1	Instant Neural Graphics Primitives	19
7.2	Invertibilidade	19
A	Positional Encoding	21
	Bibliografía	23

Índice de Figuras

2.1	Proceso de rexistro de imaxes iterativo	7
-----	---	---

Índice de Táboas

3.1	Comparativa entre os servidores utilizados	15
3.2	Estimación de custos dos recursos humanos	15

Capítulo 1

Introdución

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

1.1 Motivación

A oftalmoloxía vólase 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 foron tomadas en instantes separados no tempo, é preciso aliñalas para poder comparalas de xeito efectivo. O aliñamento de imaxes é un proceso que se leva a cabo para poder comparar imaxes dun mesmo paciente tomadas en distintos momentos, ou para comparar imaxes de diferentes pacientes. Consiste en deformar dúas ou máis imaxes de forma que as características de interese se atopen na mesma posición (superpostas). Este é un proceso tedioso e propenso a erros, polo que calquer mellora nel é de gran interese para os profesionais da saúde. Esta tarefa é axeitada para ser automatizada, xa que é non require dun xuízo clínico, senón que se basea na comparación de características visuais das imaxes. A automatización deste proceso permitiría aos profesionais da saúde dedicar máis tempo a tarefas máis relevantes.

Xa existen diversas técnicas para realizar aliñamento de imaxes automático, especialmente ca chegada do deep learning á visión por computador para imaxes médicas, onde unha aproximación habitual é o uso de redes neuronais convolucionais. Estes métodos, aínda que efectivos, teñen limitacións. Unha delas é que requiren 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 máis rápidos. Adaptando o traballo realizado por [1], preténdese aplicar redes de representación implícita para o aliñamento de imaxes oftalmolóxicas.

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 [1] para aplicalo a imaxes oftalmolóxicas.
- Comparar o rendemento do método proposto co de métodos automáticos convencionais nos datasets de FIRE [2] e RFMID [3].
- Analizar a influencia de distintos parámetros no rendemento do método proposto, 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: Introducción:** 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 discutiránse 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: Conclusións:** neste capítulo presentaranse as conclusións do traballo e as súas implicacións.
- **Capítulo 7: Traballo futuro:** neste capítulo propoñeránse liñas de traballo futuro.

Capítulo 2

Contexto

NESTE apartado introdúcese o contexto relevante a este traballo que provee os conceptos básicos necesarios para a súa comprensión. Para elo descríbese 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 que se encarga do estudo e tratamento do ollo e os seus trastornos. O ollo humano é un dos órganos dos que mais dependemos e maior cantidade de información sensorial aporta, e en consecuencia tamén é un dos mais complexos do noso corpo. Así mesmo, é unha das rexións que proporciona máis datos sobre o estado de saúde do paciente, xa que permite observar directamente os vasos sanguíneos e o tecido neuronal "in-vivo". Isto permite a detección temprana de enfermidades, que poden ser diagnosticadas mediante a observación da retina.

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. O ollo humano está composto por varias estruturas, cada unha cunha función específica que permite a percepción visual. Entre elas destacan:

(imagen ilustrativa do ollo con partes señaladas)

- **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, expándese 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 osíxeno necesarios á retina e eliminan os seus residuos metabólicos.

...

2.1.2 Imaxe oftalmolóxica

Existen diversas modalidades de imaxe médica que permiten observar o ollo, cada unha con diferentes propiedades e aplicacións. As mais utilizadas son a fotografía de fondo de ollo (retinografía), a tomografía de coherencia óptica (OCT) e a angiografía con fluoresceína.

2.1.3 Retinografía

Este traballo céntrase na retinografía xa que é a mais común. Isto é débese en gran parte á súa accesibilidade, requirindo equipo máis 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.

Para realizala utilízase unha 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.

(imagen retina con partes señaladas)

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 alínealas 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.

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

O rexistro de imaxes ten utilidade en moitos campos diferentes como a imaxe satelital, xeografía, robótica... [4] 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 corregir os movementos involuntarios do paciente durante a adquisición de imaxes, como no caso da respiración en imaxes de pulmóns.

Tamén é fundamental para a intervención guiada por imaxe (telecirurxía, radioterapia guiada por imaxe (IGRT)...) que non podería funcionar sen a utilización axeitada de técnicas de rexistro de imaxes. [5]

Ata recentemente, gran parte do traballo de rexistro facíase de forma manual por expertos, 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. Existen diversos métodos para realizar aliñamento de imaxes, que poden estar automatizados en maior ou menor medida, sendo gran parte deles híbridos.

No caso de traballar con dúas imaxes, a imaxe de referencia denomínase imaxe fixa (f) e a imaxe que se quere rexistrar imaxe móbil (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 denomínase campo de vectores de deformación (DFV), e permite representar deformacións locais na imaxe, facendoa moito mais flexible para representar transformacións complexas e detalladas. ...más de DFV ...

Dentro da deformables poden distinguirse entre transformacións difeomórficas e non

difeomórficas.

As transformacións difeomórficas son aquelas que son continuas, invertibles e diferenciáveis en todo o seu dominio. Se non téñ esta característica, non se pode garantir que a información da imaxe móbil se manteña intacta tras a transformación. Por iso as transformacións difeomórficas son preferidas en moitos casos [6],

2.2.1 Estado da arte

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 2.1 mostra un diagrama do proceso de rexistro iterativo.

A principal desvantaxe destes métodos é a súa lentitude, xa que requiren de moitas iteracións para converxer. Non obstante, son moi precisos e robustos, polo que aínda se empregan en moitos casos.

máis de rexistro con landmarks...

Métodos de aprendizaxe profunda

Ca chegada dos métodos de aprendizaxe profunda á imaxe médica, comezaron a empregarse redes neuronais para realizar o aliñamento de imaxes. Estes métodos tenden a ser máis rápidos que os métodos convencionais, a custo de algo de precisión. Ademais, estes métodos requiren dunha gran cantidade de datos para ser adestrados, o que pode ser unha desvantaxe 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.

Durante o proceso de entrenamento, a rede ten acceso aos DFVs da 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 coa aprendizaxe profunda, entrenando unha CNN nova por cada parella de imaxes. Desta forma conséguese

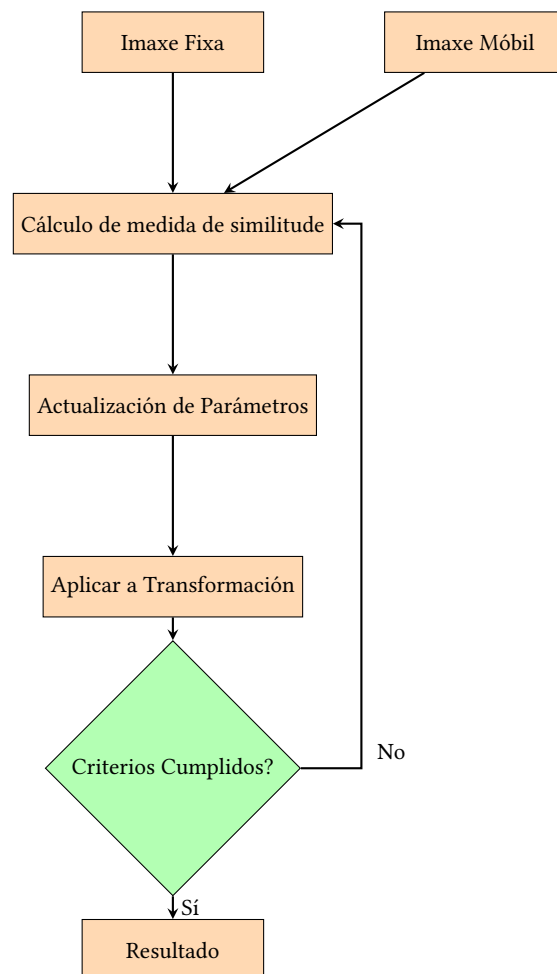


Figura 2.1: Proceso de rexistro de imaxes iterativo

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).

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 DVFs de referencia para supervisar o proceso de aprendizaxe. O termo de perda adoita basearse na discrepancia entre os DVFs de referencia e os DVFs predichos.

No lugar dos DVFs de referencia, o rexistro débilmente supervisado pode utilizar outras etiquetas de referencia implícitas. Estas etiquetas non se basan en datos explícitos como os DVFs, 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. [7] ...

Métodos Non Supervisados Un dos maiores retos para traballar de forma automática con imaxes médicas é a recolección de datos anotados de calidade para o adestramento [8]. A creación de conxuntos de DFVs anotados é un proceso laborioso e costoso, que normalmente sólo pódese executar por especialistas, polo que os métodos de rexistro non supervisados son de gran interese. Xa que a imaxe fixa e a imaxe móbil xa conteñen toda a información necesaria para un rexistro correcto, os métodos non supervisados parecen máis 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.

[9] é un dos métodos máis recoñecidos de rexistro de imaxes non supervisado facendo uso de CNNs.

O rexistro non supervisado de imaxes médicas utilizando GANs é unha subcategoría desta técnica.

2.2.2 En retinas

Os métodos que funcionan ben en moitos dominios de imaxe médica (cerebro, pulmóns, etc) suelen requirir de axustes para funcionar en retinas, polo que en imaxe oftalmolóxica hai un estado do arte paralelo. Rexistro con landmarks que é moi usado tanto en imaxe natural como en retinas.

2.3 Representación Neuronais Implícitas

A representación de coñecemento é un dos problemas máis importantes na área da computación, e as redes profundas son unha das ferramentas máis útiles, especialmente no campo da visión por computador. Tradicionalmente empréganse representacións discretas, onde o espazo de entrada é dividido en celdas e cada celda é asignada un valor (por exemplo nubes de puntos, matrices de píxeles ou vóxeles...). Unha das principais desvantaxas 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 neuronales. 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 sinais.

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 diferenciáveis, 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 [10], polo que moita investigación diríxese a atopar alternativas que melloren a representación do sinal. [11]

Por exemplo SIREN [12], sobre a que profundizaremos mais adiante. [13] propón as funcións de activación gaussianas como alternativa a SIREN, e argumenta que poden obter mellores representacións e mais robustas. [14] 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 [15]. 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 [16], pasando por reconstrucción de obxectos [17] [18] ou modelado de sinais complexas [19].

As representacións implícitas están a recibir cada vez máis atención da comunidade médica [20], e son especialmente útiles para as tarefas de imaxe inversa, que requiren a reconstrucción de representacións correctas a partir de datos incompletos ou ruidosos. No caso de [21],

propuxeron unha representación implícita para a reconstrución de imaxes de resonancia magnética a partir de datos incompletos facendo uso de redes implícitas, e obtiveron resultados comparables a métodos tradicionais.

[17] 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 [15] e [22], que propoñen o uso de INRs para modelar entornos e planificar traxectorias para un ou varios axentes. A principal vantaxe de facelo desta forma fronte 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 maioría das aplicacións.

Mais especificamente, en [23] 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 reconstrución 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.

[24] and 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 fronte 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 [1] propón é optimizar directamente o DFV facendo uso dunha representación implícita, de forma que a deformación está representada nos propios pesos dunha MLP.

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 entrenamiento (2500 en total), 10000 puntos son muestreados aleatoriamente do espazo de coordenadas dentro da máscara. O término 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.

Función de activación

Unha elección estándar para a función de activación é ReLU: $\sigma(x) = \max(0, x)$. Non obstante, para redes de representación implícita como ca que estamos traballando, esta ten unha serie de desventajas.

Sen embargo, as ReLUs teñen un sesgo cara a sinais de baixa frecuencia [10], o que significa que o modelo pode ter dificultades para representar pequenas deformacións locais no rexistro de imaxes.

Existen varias formas de superar este sesgo, como preprocesar as coordenadas de entrada con funcións de activación periódicas [17] ou substituír a función de activación ReLU por unha función de activación periódica [12]. Neste traballo escollemos a segunda opción, utilizando unha función de activación periódica para obter un modelo de tipo SIREN, $\sigma(x) = \sin(x)$. 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.

[17] non utiliza unha función de activación periódica, mais para a representación adecuada de zonas de alta frecuencia utilizaron positional encoding para incorporálas de forma implícita na rede con bós resultados. [A](#)

Termos de regularización

Debido a que el registro de imágenes deformables es un problema mal planteado (ill-posed problem**), es común regularizar el DVF para evitar deformaciones poco realistas. Los métodos de registro basados en redes neuronales convolucionales (CNN) representan los DVF como muestras en una cuadrícula de vóxeles, y por lo tanto, solo pueden aproximar gradientes espaciales mediante esquemas de diferencias finitas. Esto conlleva errores de discretización y pérdidas de precisión.

Facendo uso de representacións implícitas, todas as operacións son diferenciables, e os gradientes poden ser computados de forma analítica en lugar de ter que aproximalos. 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 varias veces. 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 indica que está a ocorrer folding e a transformación non será invertible. ...
- **Hyperelastic regularizer** Tamén se pódenn engadir restricións ao DVF con este termo proposto por [25]. 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 e está controlado pola matriz do Jacobiano da transformación. A matriz de cofactores e o determinante da matriz do Jacobiano da transformación controlan a área e o volume respectivamente, penalizando o crecemento e a contracción por igual. ...
- **Bending energy penalty** Pódese impoñer a suavidade do DVF empregando esta penalización proposta en (Rueckert et al., 1999). Require que as segundas derivadas do DVF sexan pequenas en todo o dominio, polo que non pode ser utilizado nunha rede que utilice ReLU como función de activación (a segunda derivada de unha ReLU é sempre igual a 0). ...

Método

Sendo o obxectivo encontrar unha transformación espacial óptima entre a imaxe fixa e a imaxe móbil, é necesario obter a función de deformación $\Phi(x) = u(x) + x$ que mapea cada coordenada x na imaxe fixa a unha coordenada na imaxe móbil, 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.

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

A aportación clave é que a transformación Φ está implícitamente representada nunha 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, estes 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 calquer tamaño, ao contrario dunha CNN tradicional que ten que ser adaptada para cada resolución.

Segundo, facelo desta forma permite aproveitar as capacidades de librerías como PyTorch para calcular os gradientes da transformación respecto das coordenadas. Isto permite obter gradientes máis 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. Neural Tangent Kernel (NTK) es un concepto que 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 precindindo así da necesidade de grandes conxuntos de datos para o seu adestramento.

2.5 Traballo proposto

Baseándose no framework introducido por [1], propónse modificalo para adaptalo a imaxes médicas de retina.

Metodoloxía e planificación

NESTA sección explícase a metodoloxía de traballo empregada para o desenvolvemento do proxecto, así como a planificación do mesmo. Ademais, descríbense 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 adptacións necesarias para o seu uso en imaxes oftalmolóxicas. - Imaxes de pulmón son 4d, imaxes oftalmolóxicas son 2d (ignoramos color non aporta moita información)

- Implmentar avaliación 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, decidiúse empregar moito do mesmo software ca o traballo orixinal para facilitar a implementación e reproducibilidade. O mais relvante 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 desenrolado nun ordenador portátil conectado por ssh a un servidor con GPU. Utilizáronse dous servidores diferentes, un montado por min¹ e outro facilitado polo grupo de investigación VARPA (Visión Artificial y Reconocimiento de Patrones).

A gran parte dos experimentos foron realizado no primeiro, mais para poder executar o proxecto cas imáxenes na súa resolución orixinal foi necesario empregar 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 custos do hardware son ignorados xa que xa estaba disponible antes da realización do proxecto. Os custos dos recursos humanos calcúlanse para un estudante e doús tutores, resultando nun coste 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>

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. Finalmete, presentaranse os resultados obtidos acompañados dunha análise dos mesmos.

4.1 Vista Xeral

4.2 Conxuntos de datos

4.2.1 FIRE

4.2.2 RFMID

4.3 Avaliación

4.4 Discusión

Experimentos e resultados

NESTE capítulo presentarase os experimentos realizados e os resultados obtidos. Para iso, comezase presentando unha vista xeral do proceso de experimentación, A continuación, presentarase os resultados obtidos cas diferentes configuracións probadas.

Capítulo 6

Conclusiones

Traballo futuro

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

7.1 Instant Neural Graphics Primitives

Introducidas por [26], 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 téñen 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.

[27] aplicao estas ideas á tarefa de rexistro, con moi bós resultados. Notablemente, resuelven o 'sliding boundary problem', que se refire ás complicacións de modelar o movemento relativo entre diferentes estruturas. No caso da imaxe pulmonar, surxe cando os lóbulos dos pulmóns se deslizan entre si durante a 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 [28]. Tamén con campos de velocidade neuronais [29] ...

Apéndices

Positional Encoding

EXEMPLO de capítulo con formato de apéndice, onde se pode incluír material adicional que non teña cabida no corpo principal do documento.

Bibliografía

- [1] J. M. Wolterink, J. C. Zwienenberg, and C. Brune, “Implicit neural representations for deformable image registration,” in *Medical Imaging with Deep Learning 2022*, 2022.
- [2] 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>
- [3] S. Pachade, P. Porwal, D. Thulkar, M. Kokare, G. Deshmukh, V. Sahasrabuddhe, L. Giancardo, G. Quelled, and F. Mériaudeau, “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>
- [4] A. A. Goshtasby, *Theory and applications of image registration*. John Wiley & Sons, 2017.
- [5] 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>
- [6] 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>
- [7] 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>
- [8] 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.

- [9] 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>
- [10] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. A. Hamprecht, Y. Bengio, and A. Courville, “On the spectral bias of neural networks,” 2019. [En línea]. Disponible en: <https://arxiv.org/abs/1806.08734>
- [11] A. Essakine, Y. Cheng, C.-W. Cheng, L. Zhang, Z. Deng, L. Zhu, C.-B. Schönlieb, and A. I. Aviles-Rivero, “Where do we stand with implicit neural representations? a technical and performance survey,” 2024. [En línea]. Disponible en: <https://arxiv.org/abs/2411.03688>
- [12] V. Sitzmann, J. N. P. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, “Implicit neural representations with periodic activation functions,” 2020. [En línea]. Disponible en: <https://arxiv.org/abs/2006.09661>
- [13] S. Ramasinghe and S. Lucey, “Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps,” 2022. [En línea]. Disponible en: <https://arxiv.org/abs/2111.15135>
- [14] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. G. Baraniuk, “Wire: Wavelet implicit neural representations,” 2023. [En línea]. Disponible en: <https://arxiv.org/abs/2301.05187>
- [15] 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>
- [16] 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>
- [17] 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>
- [18] 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>
- [19] 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>

- [20] 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>
- [21] 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>
- [22] 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.
- [23] 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.
- [24] 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>
- [25] 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>
- [26] 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>
- [27] 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>
- [28] J.-H. Jacobsen, A. Smeulders, and E. Oyallon, “i-revnet: Deep invertible networks,” 2018. [En línea]. Disponible en: <https://arxiv.org/abs/1802.07088>
- [29] 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>