



UNIVERSIDADE ESTADUAL PAULISTA

"JÚLIO DE MESQUITA FILHO"

Instituto de Geociências e Ciências Exatas

Programa de Pós-Graduação em Ciência da Computação

Uemerson Pinheiro Junior

**Registro Não-Rígido de Imagens Médicas
usando Block-Based Principal Component
Analysis como Camada de Pooling**

Rio Claro - SP

2024

UNIVERSIDADE ESTADUAL PAULISTA
"Júlio de Mesquita Filho"
Instituto de Geociências e Ciências Exatas
Câmpus de Rio Claro

Uemerson Pinheiro Junior

**Registro Não-Rígido de Imagens Médicas usando
Block-Based Principal Component Analysis como Camada
de Pooling**

Dissertação de Mestrado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista "Júlio de Mesquita Filho", como parte dos requisitos para obtenção do título de Mestre no Programa de Pós-Graduação em Ciência da Computação.

Orientador: Prof. Dr. Denis Henrique Pinheiro Salvadeo

Rio Claro - SP
2024

P654r	Pinheiro Junior, Uemerson Registro não-rígido de imagens médicas usando block-based principal component analysis como camada de pooling / Uemerson Pinheiro Junior. -- Rio Claro, 2024 92 p. : il., tabs. Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Instituto de Geociências e Ciências Exatas, Rio Claro Orientador: Denis Henrique Pinheiro Salvadeo 1. Registro Não-Rígido de Imagem. 2. Imagens Médicas. 3. Aprendizado Profundo. 4. Análise de Componentes Principais. 5. Block-Based Principal Component Analysis. I. Título.
-------	--

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade Estadual Paulista (UNESP), Instituto de Geociências e Ciências Exatas, Rio Claro. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA
"Júlio de Mesquita Filho"
Instituto de Geociências e Ciências Exatas
Câmpus de Rio Claro

Uemerson Pinheiro Junior

Registro Não-Rígido de Imagens Médicas usando Block-Based Principal Component Analysis como Camada de Pooling

Dissertação de Mestrado apresentada ao Instituto de Geociências e Ciências Exatas do Câmpus de Rio Claro, da Universidade Estadual Paulista "Júlio de Mesquita Filho", como parte dos requisitos para obtenção do título de Mestre no Programa de Pós-Graduação em Ciência da Computação.

Comissão Examinadora

Prof. Dr. Denis Henrique Pinheiro Salvadeo

IGCE / UNESP / Rio Claro (SP)

Orientador

Prof. Dr. Wallace Correa de Oliveira Casaca

IBILCE / UNESP / São José do Rio Preto (SP)

Prof. Dr. Marcelo Zanchetta do Nascimento

FACOM / UFU / Uberlândia (MG)

Conceito: Aprovado.

Rio Claro (SP), 04 de março de 2024

Agradeço a Deus e a Jesus, luz e guia em minha jornada, à minha amada família, pilar de amor incondicional, e aos meus queridos amigos, fontes constantes de alegrias.

Agradecimentos

Agradeço primeiramente a Deus e a Jesus por me permitirem chegar até aqui, concedendo-me força, sabedoria e perseverança ao longo desta jornada acadêmica. Sou imensamente grato por permitirem a realização deste sonho de concluir o mestrado.

Expresso minha profunda gratidão aos meus pais, Uemerson Pinheiro e Gleicy Lane Oliveira Beirigo Pinheiro, e ao meu irmão Simon Oliveira Beirigo Pinheiro, que sempre estiveram ao meu lado, oferecendo amor incondicional, apoio emocional e encorajamento nos momentos desafiadores.

Gostaria de expressar meus agradecimentos ao meu orientador, Denis Henrique Pinheiro Salvadeo, pela paciência, confiança e por me proporcionar o privilégio de aprender com ele. Sou muito grato pelo cuidado que ele teve comigo, onde, mesmo sem me conhecer previamente, entrou em contato, perguntando se estava tudo bem e quais motivos de não estar participando das aulas ministradas por ele, nas quais tinha me inscrito. Consequentemente, essa interação culminou em eu me tornar seu orientando e realizar este trabalho.

Agradeço aos meus queridos amigos, Cryfort Stone, Gabriel Ferreira e Guilherme Ferreira, a quem considero como verdadeiros irmãos. Também expresso minha gratidão aos meus amigos Emanuel Silva Araujo, Leonardo Fonseca Ohashi, Teddy Roberto e Tércio Borges Ribeiro, bem como aos amigos que fiz no grupo PRISMA (Pattern Recognition, Image and Signal Processing, and Multimedia Applications), em especial, o Lucas de Brito Silva, que se tornou um grande amigo. Vocês foram pilares fundamentais em minha vida, proporcionando apoio, amizade e momentos memoráveis. Cada um de vocês contribuiu de maneira única para enriquecer minha jornada, e sou grato por ter amigos tão especiais ao meu lado.

Agradeço sinceramente a todos os membros da comissão avaliadora por seu tempo e valiosa contribuição a este momento especial em minha jornada acadêmica.

E por fim, serei eternamente grato a todos que, de alguma forma, contribuíram e apoiaram meu caminho. Vocês foram fundamentais e são parte essencial da minha jornada e da realização deste sonho.

"...Amem uns aos outros. Como eu os amei, vocês devem amar uns aos outros."

João 13:34 - Nova Versão Internacional

Resumo

O registro não-rígido de imagem é fundamental em muitas tarefas de análise em imagens médicas. O registro não-rígido visa estabelecer correspondências espaciais com o intuito de minimizar as diferenças entre a imagem fixa e a imagem móvel. As técnicas convencionais de registro de imagens, devido à abordagem iterativa, são lentas, mesmo se realizada com as melhores técnicas e com as melhores GPUs. Além disso, possui dificuldades com convergência ou estagnação prematura, principalmente com imagens multimodais. Desta forma, recentemente, foi introduzida na literatura a abordagem de registro de imagem com aprendizado profundo visando resolver os problemas de lentidão, de convergência ou estagnação prematura dos métodos convencionais. Algumas destas novas abordagens são baseadas na rede U-Net, utilizando camadas clássicas de *pooling*, tal como o *Max Pooling*, que não consideram a relação espacial para realizar uma compressão mais representativa dos dados. Por outro lado, outras técnicas perdem a relação espacial dos dados, como o uso da análise de componentes principais (PCA) global como *pooling*. Assim, o presente trabalho tem o objetivo de investigar a utilização da técnica PCA baseada em Blocos (Block-based PCA, i.e., BPCA) como técnica de *pooling* para ambas as etapas de subamostragem e reconstrução de uma U-Net aplicada ao problema de registro de imagens médicas. A fim de analisar a efetividade desse método, foram realizados experimentos nos conjuntos de dados OASIS e IXI, que são conjuntos de dados de imagens 3D de ressonância magnética ponderadas em T1 do cérebro. Os experimentos revelaram que o BPCA superou o *Max Pooling* no conjunto de dados IXI e apresentou desempenho equivalente no conjunto de dados de validação do OASIS, apresentando valores inferiores apenas no conjunto de testes do OASIS. Além disso, a combinação dos dois métodos, em que o BPCA é utilizado nas três camadas iniciais de *pooling* e, na última camada, o *Max Pooling*, superou os resultados dos outros experimentos na métrica *Dice*. Assim, com os resultados qualitativos e quantitativos, foi demonstrado que o método BPCA é uma alternativa viável para o registro não-rígido de imagens.

Palavras-chave: Registro Não-Rígido de Imagem; Imagens Médicas; Aprendizado Profundo; Análise de Componentes Principais; Block-Based Principal Component Analysis.

Abstract

The non-rigid image registration is fundamental in many tasks of analysis in medical images. Non-rigid registration aims to establish spatial correspondences with the purpose of minimizing differences between the fixed and moving images. Conventional image registration techniques, due to their iterative approach, are slow, even when implemented with the best techniques and GPUs. Moreover, they encounter difficulties with convergence or premature stagnation, particularly with multimodal images. Thus, a recent introduction in the literature is the deep learning-based image registration approach, aiming to address the issues of slowness, convergence, or premature stagnation encountered by conventional methods. Some of these new approaches are based on the U-Net architecture, utilizing classical pooling layers such as Max Pooling, which do not consider spatial relationships for a more representative data compression. On the other hand, other techniques lose the spatial relationship of the data, such as the use of global Principal Component Analysis (PCA) as pooling. Therefore, the present study aims to investigate the use of Block-based PCA (BPCA) as a pooling technique for both downsampling and upsampling stages of a U-Net applied to the medical image registration problem. In order to analyze the effectiveness of this method, experiments were conducted on the OASIS and IXI datasets, which are datasets of 3D T1-weighted brain magnetic resonance images. The experiments revealed that BPCA outperformed Max Pooling in the IXI dataset and showed equivalent performance in the OASIS validation dataset, presenting lower values only in the OASIS test dataset. Furthermore, the combination of both methods, where BPCA is used in the initial three pooling layers and Max Pooling is applied in the final layer, surpassed the results of other experiments in the Dice metric. Thus, with qualitative and quantitative results, it was demonstrated that the BPCA method is a viable alternative for non-rigid image registration.

Keywords: Non-rigid Image Registration; Medical Images; Deep Learning; Principal Component Analysis; Block-Based Principal Component Analysis.

Lista de ilustrações

Figura 1 – <i>Perceptron</i> de uma camada, possuindo as entradas que realizam o estímulo inicial, uma função de ativação que é responsável por realizar o aprendizado e uma saída que é o resultado predito.	22
Figura 2 – Exemplo de uma rede <i>Multilayer Perceptron</i> (MLP), com duas camadas intermediárias.	23
Figura 3 – Arquitetura típica de uma Rede Neural Convolucional (CNN).	26
Figura 4 – O mapa de características resultante da operação convolucional é representado por $I * K$, onde I representa uma imagem e K um <i>kernel</i> (filtro). . .	27
Figura 5 – Representação da camada de <i>pooling</i> utilizando a função de sumarização <i>max pooling</i>	27
Figura 6 – Representação da arquitetura da rede FCN para realizar segmentação de imagens.	28
Figura 7 – Representação das etapas de <i>max-pooling</i> , <i>max-locations</i> e de <i>unpooling</i> .	29
Figura 8 – Representação da arquitetura da U-Net	29
Figura 9 – Exemplo de uma arquitetura de Redes Adversárias Generativas típica para reconhecer e gerar dígitos escritos à mão.	30
Figura 10 – Fluxograma da estrutura de registro de imagens em algoritmos de otimização.	34
Figura 11 – Processo de registro de imagem na abordagem por Métricas de Similaridade Profunda.	36
Figura 12 – Processo de treinamento da abordagem de Registro Supervisionado de Ponta a Ponta.	36
Figura 13 – Processo de treinamento da abordagem por Aprendizado Profundo por Reforço.	37
Figura 14 – Processo de treinamento do paradigma de Registro Não Supervisionado de Ponta a Ponta.	38
Figura 15 – Processo de registro de imagem por meio do paradigma de Registro Fracamente Semi-Supervisionado de Ponta a Ponta.	39
Figura 16 – Gráfico de barras representando as principais métricas utilizadas para avaliar os registros de imagens, de acordo com o número de publicações em que foram utilizadas.	41
Figura 17 – Visão geral do método proposto por Balakrishnan et al. (2019).	43

Figura 18 – Arquitetura convolucional U-Net proposta pelo <i>framework VoxelMorph</i> . Cada retângulo representa um volume 3D gerado a partir de convoluções 3D aplicadas no volume anterior. O tamanho dos filtros convolucionais estão dentro dos retângulos e abaixo deles é apresentada a resolução espacial de cada volume em relação à sua entrada. As setas representam as <i>skip connections</i> que concatenam características do codificador e decodificador.	44
Figura 19 – Arquitetura das redes (a) geradoras e (b) discriminadoras utilizadas em (MAHAPATRA et al., 2018).	45
Figura 20 – Representação esquemática do DLIR <i>framework</i> .	46
Figura 21 – Fluxo geral do registro de imagem da LungRegNet.	47
Figura 22 – Arquitetura da rede proposta em (HU et al., 2020).	48
Figura 23 – Arquitetura da rede multi-resolução proposta.	50
Figura 24 – Diagrama da rede U-Net do <i>framework VoxelMorph</i> com adaptações a fim de utilizar o método BPCA como camada de <i>pooling</i> .	54
Figura 25 – Diagrama de blocos do método BPCA.	58
Figura 26 – Resultado do <i>Max Pooling</i> e BPCA aplicados em diferentes resoluções em uma imagem.	59
Figura 27 – Exemplos de imagens contidas no conjunto de dados OASIS.	60
Figura 28 – Exemplos de imagens contidas no conjunto de dados IXI.	61
Figura 29 – <i>Dice Score</i> obtido ao longo do treinamento no conjunto de validação do conjunto de dados OASIS.	64
Figura 30 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do OASIS em uma visão sagital.	67
Figura 31 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do OASIS em visão sagital.	68
Figura 32 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do OASIS em uma visão coronal.	69
Figura 33 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do OASIS em visão coronal.	70
Figura 34 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do OASIS em uma visão axial.	71
Figura 35 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do OASIS em visão axial.	72
Figura 36 – <i>Dice Score</i> obtido ao longo do treinamento no conjunto de validação do conjunto de dados IXI.	73

Figura 37 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do IXI em uma visão sagital.	76
Figura 38 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do IXI em visão sagital.	77
Figura 39 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do IXI em uma visão coronal.	78
Figura 40 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do IXI em visão coronal.	79
Figura 41 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do IXI em uma visão axial.	80
Figura 42 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do IXI em visão axial.	81

Lista de tabelas

Tabela 1 – Conjuntos de dados públicos disponíveis divididos entre tipos de órgãos, registros e modalidades.	51
Tabela 2 – Tempo de treinamento de cada experimento no conjunto de dados OASIS, expressos em segundos e em horas. O número destacado em negrito indica o melhor tempo.	65
Tabela 3 – Resultados fornecidos pelos organizadores do desafio do conjunto de dados de teste do OASIS, os números destacados em negrito indicam as melhores classificações.	65
Tabela 4 – Resultados fornecidos pelos organizadores do desafio do conjunto de dados de validação do OASIS, as melhores classificações estão destacadas em negrito.	66
Tabela 5 – Tempo de treinamento de cada experimento no conjunto de dados IXI, expressos em segundos e em horas. O número destacado em negrito indica o melhor tempo.	74
Tabela 6 – Resultados da média do <i>Dice Score</i> , a porcentagem de $ J_\phi \leq 0$ e o tempo médio gasto para realizar o registro na GPU em segundos, referentes a cada experimento realizado. As melhores classificações estão destacadas em negrito.	74

Lista de abreviaturas e siglas

ANTs	<i>Advanced Normalization Tools</i>
BPCA	<i>Block-Based Principal Component Analysis</i>
CC	<i>Cross-correlation</i>
CNN	<i>Convolutional Neural Network</i>
CPU	<i>Central Processing Unit</i>
DBT	<i>Digital Breast Tomosynthesis</i>
DFV	<i>Deformation Vector Field</i>
DDF	<i>Dense Displacement Field</i>
DIR	<i>Deformable Image Registration</i>
DLIR	<i>Deep Learning Image Registration framework</i>
DRL	<i>Deep Reinforcement Learning</i>
DSC	<i>Dice Similarity Coefficient</i>
DSM	<i>Deep Similarity Metrics</i>
FCN	<i>Fully Convolutional Network</i>
GAN	<i>Generative Adversarial Network</i>
GC	<i>Gain Coefficient</i>
GPU	<i>Graphics Processing Unit</i>
GRU	<i>Gated Recurrent Units</i>
MI	<i>Mutual Information</i>
MLP	<i>Multilayer Perceptron</i>
MSE	<i>Mean Squared Error</i>
PCA	<i>Principal Component Analysis</i>
ReLU	<i>Rectified Linear Unit</i>

RIR	<i>Rigid Image Registration</i>
RM	Ressonância Magnética
RNA	Redes Neurais Artificiais
SE2ER	<i>Supervised End-to-End Registration</i>
SNN	<i>Spiking Neural Networks</i>
SSIM	<i>Structural Similarity Index Measure</i>
SVD	<i>Singular Value Decomposition</i>
SyN	<i>Symmetric Normalization</i>
TC	Tomografia Computadorizada
TRE	<i>Target Registration Error</i>
UE2ER	<i>Unsupervised End-to-End Registration</i>
WSE2ER	<i>Weakly-Semi-Supervised End-to-End Registration</i>
XAI	<i>Explainable Artificial Intelligence</i>

Sumário

1	INTRODUÇÃO	17
1.1	Objetivos	19
1.1.1	Objetivos Específicos	19
1.2	Principais Contribuições do Trabalho	19
1.3	Estrutura do Texto	20
2	REDES NEURAIIS E APRENDIZADO PROFUNDO	21
2.1	Redes Neurais Artificiais	21
2.2	Aprendizado Profundo	24
2.2.1	Rede Neural Convolucional	25
2.2.2	Rede Completamente Convolucional	28
2.2.3	U-Net	29
2.2.4	Redes Adversárias Generativas	30
2.3	Considerações Finais do Capítulo	31
3	REGISTRO DE IMAGENS	32
3.1	Conceitos Iniciais	32
3.2	Registro de imagens com aprendizado profundo	33
3.3	Métricas utilizadas em registro de imagens	40
3.4	Trabalhos correlacionados	42
3.5	Conjunto de dados públicos disponíveis	51
3.6	Considerações Finais do Capítulo	53
4	UMA PROPOSTA METODOLÓGICA PARA O REGISTRO NÃO- RÍGIDO DE IMAGENS MÉDICAS	54
4.1	Metodologia	54
4.2	Principal Component Analysis	56
4.3	Block-Based Principal Component Analysis	57
4.4	Conjuntos de dados	59
4.5	Métricas	61
4.6	Experimentos	62
4.7	Configurações	62
5	RESULTADOS	64
5.1	Conjunto de dados OASIS	64
5.2	Conjunto de dados IXI	73

5.3	Considerações Finais do Capítulo	82
6	CONCLUSÕES	83
	REFERÊNCIAS	85

1 Introdução

Nos últimos anos, com o crescimento dos recursos computacionais e o surgimento do aprendizado profundo (do inglês, *Deep Learning*), houve avanços significativos na performance de algoritmos para várias tarefas de visão computacional, incluindo registro de imagens. O registro não-rígido, também conhecido como registro de imagem deformável (DIR, do inglês *Deformable Image Registration*) tem sido um componente central de várias aplicações médicas nas últimas décadas, como em aplicações clínicas envolvendo diagnóstico de doenças e monitoramento, procedimentos cirúrgicos e radioterapia (CHEN et al., 2020; MANSILLA; MILONE; FERRANTE, 2020).

Imagens médicas podem ser definidas como imagens que possibilitam visualizar partes (geralmente internas, mas também externas) do corpo de um organismo, permitindo visualizar estruturas anatômicas desses organismos. As imagens médicas são utilizadas em pesquisas e diagnósticos médicos. Existem diversos tipos de modalidades dessas imagens. A modalidade é o que determina qual foi a técnica utilizada para construir a imagem. Alguns tipos de modalidades são: ressonância magnética, radiografia, tomografia computadorizada, tomografia computadorizada por emissão de fóton único e tomografia por emissão de pósitrons (NIELSEN, 2003).

As imagens médicas são utilizadas pelos profissionais da saúde em geral para auxiliar no diagnóstico e tratamento de pacientes. As análises dessas imagens são feitas pelos especialistas que utilizam principalmente sua interpretação visual e experiência, resultando em uma avaliação subjetiva, além de uma avaliação normalmente requerer muito foco e, conseqüentemente, o tempo desses profissionais (PATIL; DEORE, 2013). Essas limitações por parte dos especialistas podem resultar em diagnósticos divergentes ou com um tempo mais longo, prejudicando os pacientes. Desta forma, o ideal é que as análises de imagens médicas sejam realizadas por sistemas e algoritmos de máquinas precisos e eficientes, em conjunto com o especialista para corroborar ou confirmar o diagnóstico (PATIL; DEORE, 2013; KER et al., 2017).

Os autores em Litjens et al. (2017) demonstram o uso do aprendizado profundo em imagens médicas, em tarefas de classificação de exames, diagnósticos ou lesões, detecção, segmentação, registro de imagens e entre outras, sendo aplicadas em várias áreas como neurologia, oftalmologia, mastologia, cardiologia, entre outras.

Na maior parte dos procedimentos médicos, há uma série de casos em que algumas imagens médicas precisam ser adquiridas para diagnóstico, prognóstico, tratamento e acompanhamento. As imagens médicas podem ser obtidas com variações em termos temporais, espaciais, dimensionais ou modalidade, sendo inevitável a necessidade de alinhamento para essas imagens tiradas em tempos e condições diferentes. O registro de imagem não-rígido em

aplicações médicas tem como objetivo encontrar uma transformação espacial que distorça a imagem móvel a fim de que a mesma fique o mais próximo possível da imagem de referência, também conhecida como imagem fixa (GOSHTASBY, 2017). O registro não-rígido é considerado um problema mal posto, pois, não há uma verdade absoluta disponível para a deformação desejada (CHEN et al., 2020).

Um dos problemas no registro de imagem convencional é que o mesmo realiza o registro por meio de algoritmos de otimização iterativa. Resumidamente, em cada iteração, é utilizada uma medida de similaridade para alcançar um melhor alinhamento. O algoritmo termina de ser executado quando é verificado que não há possibilidade de realizar um registro superior ao melhor registro realizado ou quando algum outro critério de parada seja alcançado.

A tarefa de registro de imagem utilizando um algoritmo convencional é tipicamente computacionalmente intensiva e demorada, mesmo com as melhores implementações e com os melhores *hardwares*, devido à sua execução iterativa (CHEN et al., 2020), o que inviabiliza o uso prático dos mesmos em operações clínicas em que é preciso obter o resultado em tempo real. Um desafio adicional surge na necessidade de realizar o registro para cada par de imagens novas, implicando na resolução repetida do problema de otimização sempre que novas entradas são apresentadas (CHEN et al., 2022). Além disso, a imprecisão das medidas de similaridade, especialmente no registro multimodal, faz com que os algoritmos de registro convencionais fiquem presos nos ótimos locais, impedindo-os de encontrar o ótimo global, que é o objetivo desejado, resultando em uma perda de eficiência e causando convergência ou estagnação prematura (BOVEIRI et al., 2020).

Outro problema é a disponibilidade dos conjuntos de dados de imagens médicas. Ainda é muito escasso esse tipo de conjunto de dados quando comparados ao conjunto de dados de imagens gerais, devido à sensibilidade dos dados e ser necessário diagnósticos ou opiniões de especialistas para se obter as anotações desses dados (TAVARES, 2018).

Por outro lado, a abordagem de utilizar o aprendizado profundo veio para tentar resolver os problemas de lentidão, de convergência ou estagnação prematura, com espaço ainda para ser explorado no contexto de registro de imagens médicas. A questão de falta de disponibilidade dos conjuntos de dados de imagens médicas com anotações pode ser resolvida com a aprendizagem profunda não supervisionada, na qual as redes desse tipo de aprendizagem não necessitam de conjuntos de dados com anotações para o seu treinamento.

Além disso, na literatura de registro de imagens usando aprendizado profundo, uma das arquiteturas de redes neurais mais utilizadas é a U-Net, que se utiliza de camadas de *pooling*, especialmente para etapa de compressão dos mapas de atributos. As técnicas de *pooling* comumente empregadas nestes tipos de rede, tal como o *Max Pooling* e o *PCA global*, não consideram simultaneamente a relação espacial e a manutenção da estrutura espacial da imagem após a compressão, de tal forma que o resultado da compressão pode trazer muitas perdas de representatividade da estrutura geral da imagem, o que pode prejudicar o processo

de registro em encontrar pontos fortes de similaridade. Por sua vez, a técnica de extração de atributos PCA baseada em blocos (BPCA, do inglês *Block-Based Principal Component Analysis*) (SALVADEO et al., 2011) é capaz de atender estes critérios. Assim, vislumbrou-se a questão a ser investigada nesta pesquisa: A utilização de BPCA como camada de *pooling* seria adequada para o registro não-rígido de imagens médicas e qual a sua efetividade?

1.1 Objetivos

O presente trabalho tem como objetivo geral adaptar a rede U-Net do *framework VoxelMorph* (BALAKRISHNAN et al., 2019) para utilizar o método estatístico BPCA (SALVADEO et al., 2011) para comprimir e reconstruir as imagens nos caminhos de subamostragem (*downsampling*) e reconstrução (*upsampling*), visando melhorar o desempenho do registro do *framework*.

1.1.1 Objetivos Específicos

A fim de atingir o objetivo geral foram levantados os seguintes objetivos específicos:

- Adaptar a rede U-Net do *framework VoxelMorph* para utilizar o método estatístico BPCA como camada de *pooling*;
- Verificar o desempenho dos registros do *framework* com a a rede U-Net adaptada para utilizar o método estatístico BPCA como camada de *pooling* e também sem nenhuma adaptação, e comparar o desempenho de ambos os experimentos;
- Misturar os métodos de *pooling Max Pooling* e o BPCA, alternando também a ordem desta mistura e comparar o desempenho destes experimentos;
- Adaptar o caminho de reconstrução do *framework* para utilizar o BPCA inverso como *unpooling*, isto é, reverter o BPCA no caminho da reconstrução, e comparar o desempenho do registro com os outros experimentos;
- Realizar os experimentos propostos nos conjunto de dados escolhidos.

1.2 Principais Contribuições do Trabalho

Este trabalho propõe uma abordagem inovadora para o registro de imagens, introduzindo o uso da técnica de *pooling Block-Based Principal Component Analysis*. Além disso, o trabalho explora combinações de técnicas de *pooling*, como BPCA e *Max Pooling*. Adicionalmente, o trabalho introduz uma ideia inovadora ao propor a reversão do BPCA no caminho de reconstrução da rede U-net, denominado como BPCA inverso.

Por fim, com o objetivo de estimular colaborações, facilitar pesquisas futuras na área e promover o avanço do conhecimento na comunidade científica, todo o código fonte utilizado neste trabalho está disponível como *open-source*¹.

1.3 Estrutura do Texto

No Capítulo 1, a introdução apresentou a contextualização, o problema e o objetivo geral deste trabalho. Por sua vez, o restante do texto está organizado da seguinte forma. No Capítulo 2, sobre redes neurais e aprendizado profundo, é apresentado o referencial teórico do presente trabalho, que foi útil como fonte de pesquisa para o desenvolvimento do mesmo. No Capítulo 3, sobre o registro de imagens, é aprofundado o tema fundamental deste trabalho, detalhando também as métricas utilizadas para avaliar o desempenho do registro, os trabalhos correlacionados e os conjuntos de dados disponíveis para o registro de imagens médicas. No Capítulo 4, é apresentada a metodologia do trabalho, contendo a proposta, os experimentos a serem realizados, os conjuntos de dados, as métricas, experimentos e as configurações utilizadas. No Capítulo 5, penúltimo capítulo, os resultados dos experimentos realizados são expostos, bem como é feita uma análise destes resultados nas considerações finais do capítulo. E no Capítulo 6, último capítulo, são feitas as conclusões deste trabalho.

¹ Repositório Github do projeto - <https://github.com/Uemerson/masters-degree-in-computer-science-from-unesp>

2 Redes Neurais e Aprendizado Profundo

Neste capítulo é apresentado o referencial teórico deste trabalho, abordando as principais arquiteturas de redes neurais com aprendizado profundo encontradas na literatura de registro de imagens.

2.1 Redes Neurais Artificiais

Redes Neurais Artificiais (RNA) são modelos computacionais inspirados na estrutura neural de seres vivos, sendo capazes de realizar as seguintes operações: aprendizado, associação, generalização e abstração, além de adquirir e manter o conhecimento baseado em informações. As RNAs podem ser definidas como um conjunto de unidades de processamento, definidas por neurônios artificiais, altamente interligadas por uma grande quantidade de interconexões, denominadas sinapses, efetuando operações simples, e transmitindo seus resultados às unidades de processamento vizinhas (SILVA; SPATTI; FLAUZINO, 2010; CUNHA, 2010).

O principal componente das redes neurais artificiais é o neurônio artificial, que foi introduzido pela primeira vez por McCulloch e Pitts (1943). Utilizando os conceitos propostos por McCulloch e Pitts (1943) de neurônio artificial, Rosenblatt (1958) desenvolveu o primeiro modelo treinável de um neurônio artificial, denominado *Perceptron*. O *Perceptron* é constituído de apenas uma camada, conforme ilustrado na Figura 1, e em termos matemáticos pode ser descrito conforme a Equação 1:

$$y = g \left(\sum_{i=1}^n x_i w_i + b \right), \quad (1)$$

em que $x_1, x_2 \dots x_n$ representam os sinais das entradas formando o axônio, e $w_1, w_2 \dots w_n$ representam os pesos sinápticos formando as sinapses, b representa o viés (bias), g é a função de ativação que modela a saída no neurônio. O somatório da multiplicação das entradas com os pesos é somado com o valor do viés e o resultado deste cálculo é utilizado pela função de ativação para ativar ou não o neurônio. As funções de ativações comumente usadas são: sigmóide, tangente hiperbólica e a ReLU (do inglês, *Rectified Linear Unit*), descritas em termos matemáticos da Equação 2 até a Equação 4 (MARCON, 2020; FERREIRA, 2019), respectivamente:

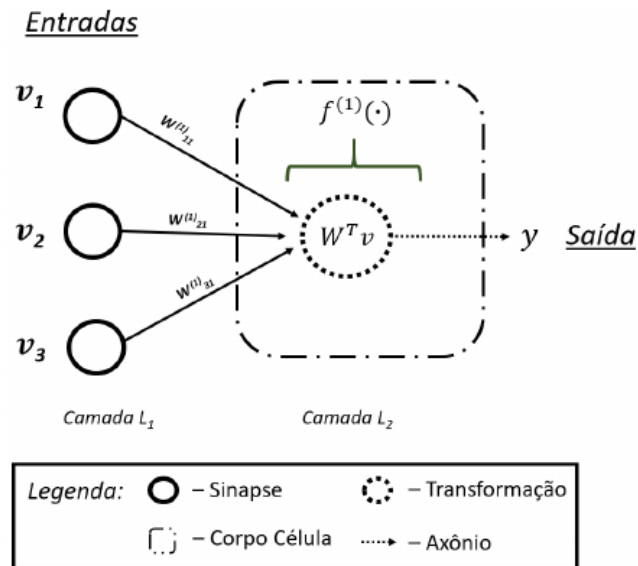
$$\phi(x) = \frac{1}{1 + e^{-x}}, \quad (2)$$

$$\phi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \text{ e} \quad (3)$$

$$\phi(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} = \max(0, x). \quad (4)$$

Nas Equações 2 a 4, a variável x é a entrada para a função de ativação. Atualmente, a função de ativação mais utilizada é a ReLU, devido à essa função transformar as entradas negativas em zero, simplificando e acelerando os cálculos e o treinamento. Ao transformar os valores negativos em zero, é possível evitar o problema da dissipação do gradiente e fornecer às redes, após serem treinadas, características esparsas (KRISTIADI; HEIN; HENNIG, 2020; LIN; SHEN, 2018; KER et al., 2017).

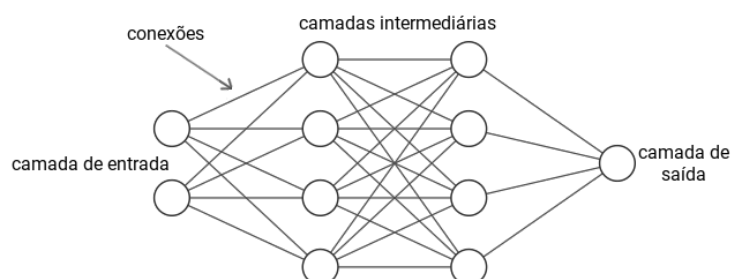
Figura 1 – *Perceptron* de uma camada, possuindo as entradas que realizam o estímulo inicial, uma função de ativação que é responsável por realizar o aprendizado e uma saída que é o resultado predito.



Fonte: (ANDRADE, 2020)

O *Perceptron* com apenas uma camada é utilizado para solucionar problemas lineares. Ao se adicionar mais camadas intermediárias, é possível solucionar também problemas não lineares. A arquitetura com múltiplas camadas intermediárias ou também nomeadas camadas internas ou escondidas (do inglês, *Hidden Layers*) é conhecida como *Multilayer Perceptron* (MLP) ou *Perceptron* de Múltiplas Camadas (ROSENBLATT, 1961). As MLPs também são conhecidas como *feedforward neural network* e são a estrutura base do aprendizado profundo (GOODFELLOW; BENGIO; COURVILLE, 2016; HAO, 2019). Cada camada possui diversos neurônios, sendo interligados, porém os neurônios de uma mesma camada geralmente não são interligados entre si, conforme ilustrado na Figura 2.

Figura 2 – Exemplo de uma rede *Multilayer Perceptron* (MLP), com duas camadas intermediárias.



Fonte: Adaptado de (CUNHA, 2010)

As camadas são conectadas por meio de pesos sinápticos e sinais de saídas. Os sinais de saída de cada camada são o resultado do processamento dos sinais de entrada da camada pela função de transferência não linear ou função de ativação. Cada camada sucessiva utiliza a saída da camada anterior como sinal de entrada. Se fosse utilizada uma função de transferência linear, então a MLP seria capaz de modelar apenas funções lineares (GARDNER; DORLING, 1998).

O treinamento da MLP tem como objetivo encontrar a combinação de pesos sinápticos que resultam no menor erro possível. O algoritmo de retropropagação de erro (do inglês, *error backpropagation*) (RUMELHART; HINTON; WILLIAMS, 1986) é bem conhecido e utilizado para realizar o treinamento da MLP (HAYKIN, 2007). As redes MLP foram impulsionadas somente depois da criação do algoritmo de *backpropagation*.

No primeiro passo, o algoritmo *backpropagation* propaga os sinais de entrada através da rede, camada por camada, iniciando pela camada de entrada indo até a camada de saída. No final da propagação é obtido como resposta real da rede um conjunto de saídas. No segundo passo, o sinal de erro (em geral, calculado com base na diferença da resposta real da rede com a resposta desejada) é propagado através da rede, contra a direção das conexões sinápticas, ou seja, de trás para frente (direção das saídas para as entradas), ajustando os pesos sinápticos de modo que a resposta da rede se mova para mais perto da resposta desejada (SANTOS; PORSANI; HIRATA, 2010; HAYKIN, 2007). A ideia básica é que pesos sejam ajustados com base na sua contribuição para o erro.

As arquiteturas de redes neurais são projetadas para seguir algum tipo de treinamento, de acordo com o problema em questão. Os três principais tipos de aprendizado de máquina podem ser categorizados em: aprendizado supervisionado (do inglês, *supervised learning*), aprendizado semi-supervisionado (do inglês, *semi-supervised learning*) e aprendizado não supervisionado (do inglês, *unsupervised learning*) (HAO, 2019).

O aprendizado supervisionado necessita que o conjunto de dados de treinamento estejam rotulados com sua classe correspondente ou saída desejada, desta forma, os algoritmos de

aprendizado supervisionado são guiados a aprenderem um conjunto de padrões com base no conhecimento fornecido pelos dados e seus rótulos. No aprendizado não supervisionado não há necessidade de um conjunto de dados com rótulos, pois os algoritmos de aprendizado não supervisionado aprendem os padrões a partir de uma caracterização de similaridade contida no conjunto de dados, ou seja, não há necessidade de guiar o aprendizado para esses algoritmos. Os algoritmos de aprendizado semi-supervisionado utilizam tanto do aprendizado supervisionado quanto do aprendizado não supervisionado, ou seja, aprendem a partir de exemplos rotulados e não rotulados, sendo bastante útil quando existe um pequeno número disponível de exemplos no conjunto de dados rotulados (SANCHES, 2003).

2.2 Aprendizado Profundo

O aprendizado de máquina (do inglês, *machine learning*) vem sendo muito utilizado no mundo contemporâneo, estando presente cada vez mais no dia a dia, podendo ser utilizado em tarefas como: identificar objetos em imagens, transformar áudio em texto, buscar recomendações relevantes em plataformas de *streaming* ou em sites de vendas utilizando informações dos usuários, filtrar conteúdos específicos nas redes sociais, entre outras tarefas (LECUN; BENGIO; HINTON, 2015).

O aprendizado profundo (do inglês, *deep learning*) é uma das subáreas do campo de aprendizado de máquina, surgindo de pesquisas relacionadas especialmente a redes neurais artificiais. O aprendizado profundo pertence à categoria de algoritmos das redes neurais artificiais, pois utiliza conceitos derivados das redes neurais artificiais (OISHI; YAGAWA, 2017; BUDIARTO et al., 2018), que tem uma base de modelagem em estruturas de dados de grafos.

O crescente aumento do poder computacional obtido por meio das unidades de processamento gráfico (GPUs - *Graphics Processing Units*) e a disponibilidade de grandes volumes de dados impulsionaram o aprendizado profundo, tornando a sua utilização mais ativa. O aprendizado profundo provou ser uma solução robusta para diversas tarefas como segmentação de imagens, classificação de imagens, processamento de linguagem natural, entre outras (BIZOPOULOS; KOUTSOURIS, 2018).

O aprendizado profundo no campo de visão computacional se tornou popular quando um trabalho utilizando uma abordagem de aprendizado profundo com base em uma rede neural convolucional (LECUN et al., 1998; KRIZHEVSKY; SUTSKEVER; HINTON, 2012) ganhou de forma esmagadora no final de 2012 a competição mundial de visão computacional conhecida como *ImageNet Classification* (SUZUKI, 2017), passando a ser dominante em quase todas as abordagens para tarefas de reconhecimento e detecção (LECUN; BENGIO; HINTON, 2015).

Um dos benefícios que o aprendizado profundo tem quando comparado com os métodos tradicionais de aprendizado de máquina é não precisar de ter muito conhecimento do domínio

do problema a ser resolvido. Além disso, as redes neurais profundas são melhores para problemas que escalam, pois, ao aumentar o conjunto de dados, as redes neurais profundas geralmente aumentam a precisão nos resultados obtidos (BIZOPOULOS; KOUTSOURIS, 2018).

Existem algumas limitações que os métodos de aprendizado profundo possuem quando comparados com outros métodos de aprendizado de máquina. Em geral, ainda não é possível abstrair o processo das redes neurais profundas, ou seja, não se sabe justificar como as redes neurais funcionam, esse problema é conhecido como caixa preta (do inglês, *black box*). No entanto, com o objetivo de superar o desafio da caixa preta, a área de pesquisa em Inteligência Artificial Explicável (XAI, do inglês *Explainable Artificial Intelligence*) tem como objetivo desenvolver modelos de inteligência artificial que sejam facilmente interpretáveis e compreensíveis para os seres humanos (REDDY; KUMAR, 2023). Muitos métodos de aprendizado de máquina superam os métodos de redes neurais profundas quando os conjuntos de dados são escassos. Desta forma, é necessário um grande conjunto de dados para treinar as redes neurais profundas, em geral. Outro problema encontrado é selecionar a rede neural profunda certa para cada problema, pois existem inúmeras arquiteturas de redes neurais profundas disponíveis, de tal modo que selecionar a apropriada para resolver um problema específico é uma tarefa árdua. E, por fim, as redes neurais profundas necessitam de um grande poder computacional para serem treinadas, porém não é necessário muito poder computacional após serem treinadas (MAMOSHINA et al., 2016).

A seguir, são introduzidas algumas arquiteturas de redes neurais profundas, que dentre outras tarefas são utilizadas também em tarefas de visão computacional e processamento digital de imagens, incluindo o registro de imagem.

2.2.1 Rede Neural Convolucional

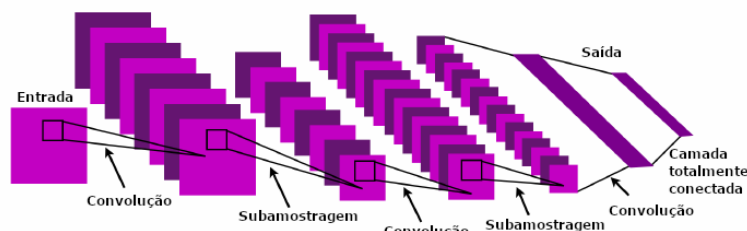
A Rede Neural Convolucional (CNN ou ConvNet, do inglês *Convolutional Neural Network*) introduzida por Lecun et al. (1998) e suas derivações têm sido aplicadas com sucesso em aplicações de detecção, segmentação e reconhecimento de objetos em imagens.

A CNN é projetada para processar dados de entrada que possuem a topologia em forma de grade, ou seja, dados que vem em forma de matrizes. Alguns exemplos de dados de matrizes são: 1D para sinais e sequências, incluindo linguagem, 2D para imagens ou espectrogramas de áudio e 3D para vídeo ou imagens volumétricas (LECUN; BENGIO; HINTON, 2015; GOODFELLOW; BENGIO; COURVILLE, 2016).

Em geral, a arquitetura da CNN é composta de diversas camadas. Essas camadas podem ser divididas em três tipos, sendo: o primeiro tipo, a camada convolucional; o segundo, a camada de subamostragem (*pooling*) e; o terceiro tipo, a camada totalmente conectada. Na Figura 3, onde está ilustrada a arquitetura de uma CNN. É possível visualizar esses tipos de camadas. Cada camada é responsável por extrair características dos dados de entrada. A saída

de uma camada é utilizada como entrada para a camada seguinte. A ideia é ir abstraindo a representação das características de acordo com a progressão através das camadas (MARCON, 2020; LECUN; BENGIO; HINTON, 2015).

Figura 3 – Arquitetura típica de uma Rede Neural Convolucional (CNN).

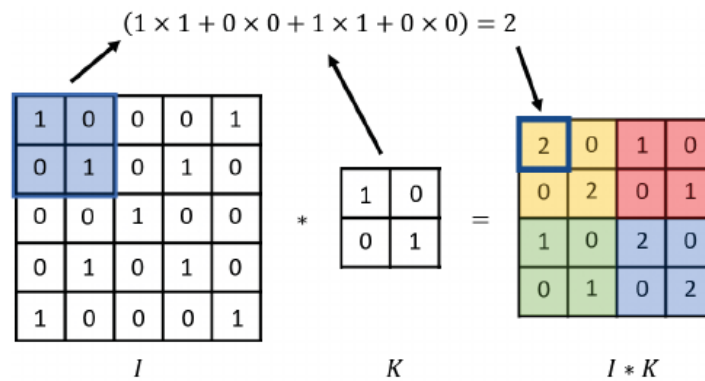


Fonte: Adaptado de (LECUN; KAVUKCUOGLU; FARABET, 2010)

A camada convolucional é composta de filtros treináveis. Os filtros possuem uma área pequena (por exemplo, um tamanho de 3x3, 5x5 ou 7x7 de *pixels*) e percorrem os dados de entrada. Dada uma imagem como entrada, esses pequenos filtros vão percorrendo a imagem e detectando as características mais marcantes dela, obtendo um mapa de ativação (ou mapa de características). Cada neurônio de cada camada está conectado apenas com os neurônios próximos da camada anterior. As camadas compartilham os pesos entre seus neurônios conduzindo os filtros a aprenderem os padrões frequentes que ocorrem em qualquer parte da imagem, por exemplo, uma borda da imagem (MARCON, 2020; HAFEMANN, 2014).

A operação convolucional consiste em deslizar os filtros sobre a entrada. Dada uma imagem como exemplo de entrada, para cada local que o filtro está passando, os *pixels* da imagem definidos sobrepostos pela janela do filtro são multiplicados com os pesos correspondentes contidos no filtro e, no final, é feita uma somatória do resultado das multiplicações, resultando em uma saída de mapa de características. Essa operação é ilustrada de forma visual na Figura 4 (MARCON, 2020).

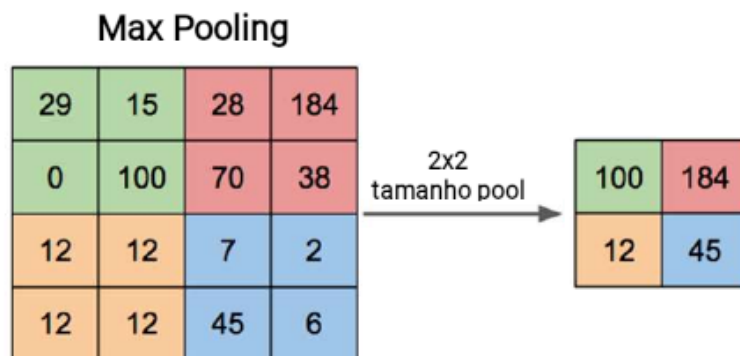
Figura 4 – O mapa de características resultante da operação convolucional é representado por $I * K$, onde I representa uma imagem e K um *kernel* (filtro).



Fonte: Adaptado de (MARCON, 2020)

A camada de subamostragem (*pooling*) é utilizada para simplificar saídas da camada anterior. Observa-se que, no exemplo da Figura 5, a área da camada de *pooling* é de tamanho 2x2. Esta área vai ser utilizada para percorrer por toda a saída da camada anterior resumindo a informação daquela área em um valor único. Uma das formas de sumarizar a saída da camada anterior é utilizar o método de ativação máxima (*max-pooling*), no qual pega-se apenas o valor máximo dentro da vizinhança da área da camada de *pooling*. Existem outras funções de *pooling*, por exemplo, funções que usam a média, mediana, norma L2 de uma região retangular e entre outras, porém a mais utilizada é a *max-pooling* (MARCON, 2020; LECUN; BENGIO; HINTON, 2015).

Figura 5 – Representação da camada de *pooling* utilizando a função de sumarização *max pooling*.



Fonte: Adaptado de (YANI et al., 2019)

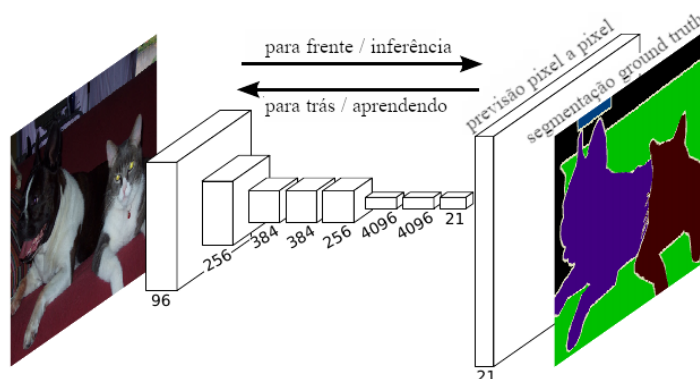
Ao finalizar os processos de convolução e *pooling*, no final da rede, é adicionada uma ou mais camadas totalmente conectadas (normalmente, correspondente a uma MLP tradicional), que utiliza a saída da camada anterior como entrada e tem como função realizar a classificação dos padrões de entrada por exemplo (FERREIRA, 2019).

2.2.2 Rede Completamente Convolutacional

A Rede Completamente Convolutacional (FCN, do inglês *Fully Convolutional Network*), introduzida por Long, Shelhamer e Darrell (2015), é muito similar à CNN, porém, a arquitetura da FCN não utiliza a camada totalmente conectada, que é geralmente utilizada para realizar a classificação. Na FCN, a camada totalmente conectada é substituída por uma camada convolutacional, permitindo que a rede resulte em uma saída do tamanho da imagem de entrada. Essa camada convolutacional é utilizada para tarefas como classificar, segmentar e entre outras *pixel por pixel* da imagem de entrada (SKEIKA, 2019; FERREIRA, 2019).

A arquitetura da rede FCN é composta por diversas camadas convolucionais, que assim como na CNN produzirão diferentes mapas de características de diferentes profundidades. No final da rede fica a camada convolutacional que irá realizar a predição, conforme ilustrado na Figura 6.

Figura 6 – Representação da arquitetura da rede FCN para realizar segmentação de imagens.



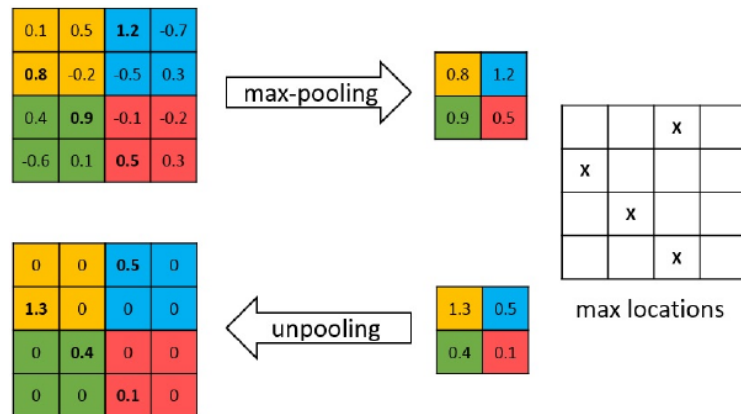
Fonte: Adaptado de (LONG; SHELHAMER; DARRELL, 2015)

Observa-se na Figura 6 que, no final da rede FCN, existe uma camada convolutacional denominada predição por *pixels*. Essa camada é responsável por realizar a predição *pixel por pixel*, ou seja, para cada *pixel* da imagem é atribuída a sua respectiva classe. Conforme a Figura 6, o tamanho final da predição por *pixels* foi determinado pela quantidade de classes distintas contidas no conjunto de dados (SKEIKA, 2019).

A FCN utiliza dois caminhos para a redução da dimensão da entrada e a expansão da mesma no final da rede, denominados *downsampling* e *upsampling*, respectivamente. No caminho de *downsampling*, a dimensão da entrada é reduzida por meio da etapa de *pooling*, enquanto no caminho de *upsampling*, a dimensão da entrada que foi reduzida ao longo das camadas é expandida novamente (SKEIKA, 2019).

Não é possível reverter completamente a operação de *max-pooling* realizada durante a etapa de *pooling*, portanto, o resultado da reversão desta operação na etapa de *unpooling* é aproximado (ZEILER; FERGUS, 2014).

Figura 7 – Representação das etapas de *max-pooling*, *max-locations* e de *unpooling*



Fonte: (DAVID; NETANYAHU, 2016)

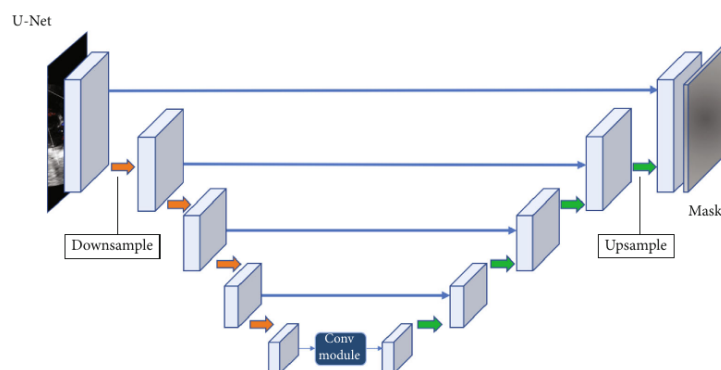
Conforme ilustrados na Figura 7 durante a operação de *max-pooling* (caminho de *downsampling*), as localizações dos máximos (*max-locations*) são armazenadas para serem usadas durante o *unpooling* (caminho de *upsampling*), onde os valores aproximados são restaurados para os seus locais exatos e os locais restantes (vizinhos) são preenchidos com zero (DAVID; NETANYAHU, 2016).

2.2.3 U-Net

A U-Net é uma Rede Neural estendida das FCNs e foi proposta por Ronneberger, Fischer e Brox (2015), desenvolvida com o objetivo de realizar com precisão a segmentação de imagens biológicas com um conjunto de dados de treino pequeno (SKEIKA, 2019).

Conforme ilustrada na Figura 8, a arquitetura da U-Net possui dois caminhos: o caminho de contração, representado na figura pelo lado esquerdo (*downsample*), no qual a entrada tem sua dimensão reduzida; e o caminho de expansão simétrico, representado na figura pelo lado direito (*upsample*), onde a entrada tem sua dimensão recuperada.

Figura 8 – Representação da arquitetura da U-Net



Fonte: (LIU et al., 2021)

O caminho de contração tem a mesma arquitetura tradicional de uma CNN, com várias aplicações repetidas de convoluções, com cada uma acompanhada por uma função de ativação ReLU e subamostragem com a operação de *max-pooling*, reduzindo a dimensão da entrada. Em cada etapa de redução da entrada (*pooling*) é duplicado o número de canais de características. No caminho de expansão, em cada etapa é seguido por *upsampling* do mapa de características e deconvolução, que reduz pela metade o número de canais de características. No fim, é feita uma concatenação com o mapa de características correspondente do caminho de contração (FERREIRA, 2019; LIU et al., 2021).

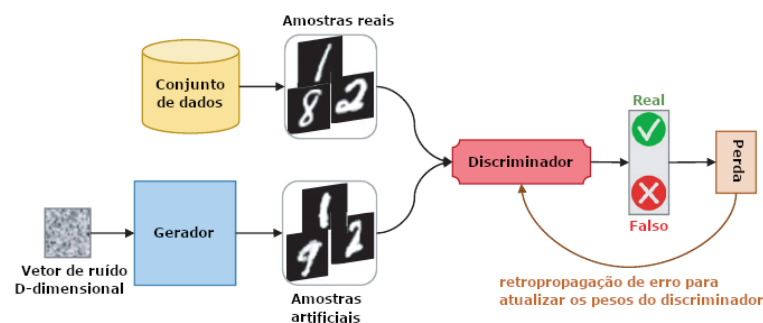
A U-Net combina os mapas de características do caminho de contração (codificador) com seus mapas de características correspondentes simétricos do caminho de expansão (decodificador) em todas as etapas, formando uma estrutura em forma de "U". Concatenar as etapas permite que o decodificador para cada etapa aprenda as características relevantes perdidas ao se agrupar no codificador e também é o que diferencia a U-Net das outras arquiteturas FCN (FERREIRA, 2019).

2.2.4 Redes Adversárias Generativas

As Redes Adversárias Generativas (GANs, do inglês *Generative Adversarial Networks*) foram propostas por Goodfellow et al. (2014) e é um modelo composto por duas redes neurais artificiais, sendo uma rede neural geradora e uma rede neural discriminadora, que ficam competindo entre si.

A rede neural geradora tem como objetivo gerar novas amostras de dados com base nos dados reais de treinamento, de tal modo que as amostras falsas geradas tentam ser o mais próximo possível dos dados verdadeiros. Enquanto a rede neural discriminadora tem como objetivo aprender a diferenciar as amostras de dados geradas pela rede geradora das amostras de dados reais (FERREIRA, 2019; KER et al., 2017).

Figura 9 – Exemplo de uma arquitetura de Redes Adversárias Generativas típica para reconhecer e gerar dígitos escritos à mão.



Fonte: Adaptado de (LOYOLA-GONZALEZ, 2019)

Observa-se na Figura 9 que a rede neural discriminadora tenta identificar entre os

exemplos reais e os exemplos gerados pela rede neural geradora quais exemplos são verdadeiros e quais são falsos. Neste exemplo, tanto a rede rede neural geradora quanto a rede neural discriminadora são treinadas utilizando o *backpropagation* (LOYOLA-GONZALEZ, 2019).

Durante o treinamento, os pesos de ambas as redes são atualizados para que o desempenho delas melhore. Desta forma, a rede geradora é atualizada para que ela gere exemplos que consigam enganar a rede discriminadora, enquanto que a rede discriminadora é atualizada para melhorar a capacidade da mesma de distinguir os exemplos falsos dos exemplos verdadeiros (FERREIRA, 2019).

A convergência se dá quando a rede discriminadora não é capaz de distinguir o que é uma amostra gerada ou uma amostra real. Desta forma, o discriminador pode ser descartado ou não, podendo ser utilizado para classificar os exemplos, e a rede geradora está pronta para ser utilizada como geradora de exemplos (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.3 Considerações Finais do Capítulo

Neste capítulo foram apresentados os conceitos de redes neurais artificiais e do aprendizado profundo. Após foi apresentado em detalhes a CNN, FCN, a U-Net e, por fim, as GANs. No próximo capítulo serão abordadas técnicas de registro de imagens com aprendizado profundo, que se utilizam nas arquiteturas de redes neurais apresentadas aqui.

3 Registro de imagens

Neste capítulo são apresentados os conceitos, processos e os tipos de métodos de registro de imagem, assim como, apresenta o registro de imagens com aprendizado profundo, expondo também algumas métricas utilizadas no registro de imagem. Por fim, são apresentados alguns trabalhos correlacionados e conjuntos de dados públicos disponíveis para o registro de imagem.

3.1 Conceitos Iniciais

É inevitável a falta de alinhamento entre imagens adquiridas de pontos de vista diferentes, em tempos diferentes, em condições diferentes ou de modalidades diferentes. Um exemplo de imagens com essas características são as imagens médicas, capturadas para diagnóstico, prognóstico, tratamento e acompanhamento. Desta forma, para resolver esse problema é utilizada uma técnica de registro de imagens.

O registro de imagens é um processo de alinhamento de imagens com base em um sistema de coordenadas geométricas semelhantes (BOVEIRI et al., 2020). No processo de registro de imagem, um par de imagens é mapeado (também conhecido como registro de pares) ou um conjunto de imagens é mapeado (conhecido como registro de grupo). Durante o mapeamento, é detectada uma transformação geométrica de coordenadas em comum entre as imagens, a fim de realizar um alinhamento ideal (CHEN et al., 2020).

O processo de registro de duas imagens sem a utilização de aprendizado profundo pode ser definido como um problema de otimização (LIU et al., 2021), podendo ser expresso pela Equação 5:

$$\hat{T} = \arg \min_T S(F, T(M)), \quad (5)$$

em que F denota a imagem fixa de referência e M denota a imagem em movimento (desalinhada) a ser registrada. T denota a transformação geométrica desejada que mapeia as coordenadas em comum entre F e M . Por sua vez, S representa uma função de custo que mede o quanto a imagem fixa e a imagem em movimento mapeada por T estão desalinhadas. O registro de imagens acontece de forma iterativa, melhorando a estimativa de T de tal forma que a função de custo S seja minimizada (CHEN et al., 2020; VOS et al., 2019).

Com base na transformação geométrica desejada, os métodos de registro de imagem podem ser classificados como registro rígido, registro afim e registro não-rígido, também conhecido como registro de imagem deformável (CHEN et al., 2020; FU et al., 2020a).

O registro rígido considera apenas a translação ao longo das coordenadas x , y e z e a rotação em torno dos eixos x , y e z . Desta forma, esse tipo de transformação pode ser modelada usando 6 parâmetros ou graus de liberdade, por meio da combinação dos 3 parâmetros de rotação e os 3 parâmetros de translação. O registro afim adiciona 6 novos graus de liberdade, sendo eles: o alongamento (escala) ao longo dos eixos x , y e z e a inclinação ou distorção (cisalhamento) nos planos xy , yz , e xz (BOVEIRI et al., 2020; CRUM; HARTKENS; HILL, 2004).

Quando a imagem apresenta deformações elásticas é necessário aplicar o registro não-rígido para realizar o alinhamento ideal. O registro rígido possibilita uma quantidade baixa de graus de liberdade, enquanto o registro não-rígido possibilita um número maior de deformações (RODRIGUES, 2010; WRANGSJÖ; PETTERSSON; KNUTSSON, 2005).

Enquanto os métodos de registro rígido e afim lidam apenas com diferenças globais como translação, rotação e escala, o registro não-rígido lida com as diferenças locais entre as imagens. Diferenças adquiridas por meio da deformação de *scanner*, movimento dos pacientes, intervenções cirúrgicas ou anatomias diferentes (FORNEFETT; ROHR; STIEHL, 2001).

No registro rígido (RIR, do inglês *Rigid Image Registration*), todos os *pixels* da imagem são transladados e/ou rotacionados uniformemente de forma que toda relação *pixel a pixel* permanece a mesma antes e depois da transformação. No registro não-rígido, a relação *pixel a pixel* não se mantém a mesma após a transformação (OH; KIM, 2017).

A seguir, é apresentado o funcionamento do registro de imagens utilizando o aprendizado profundo, bem como suas vantagens sobre os métodos tradicionais de registro de imagens.

3.2 Registro de imagens com aprendizado profundo

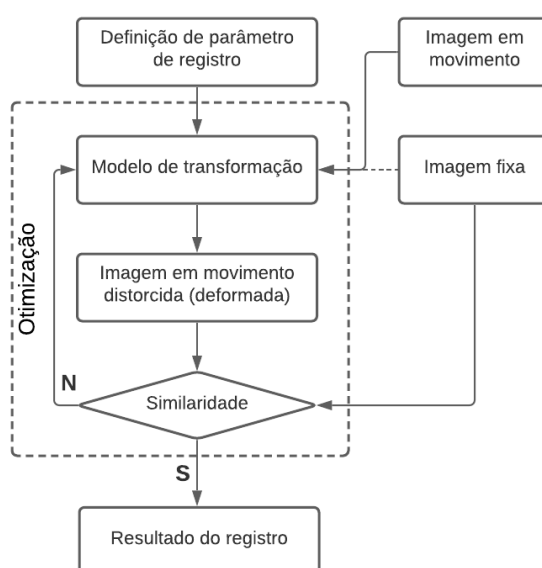
Até recentemente, o registro de imagens médicas ainda era feito de forma manual por especialistas, tornando essa tarefa bastante desafiadora e a qualidade dos registros altamente dependente da experiência desses profissionais. Para resolver esses problemas, que clinicamente são altamente prejudiciais, o registro automático foi desenvolvido (HASKINS; KRUGER; YAN, 2020).

Ainda que os métodos de registro automático tenham sido amplamente utilizados antes, durante e depois do surgimento das redes neurais profundas, após o surgimento delas, vários métodos de registro de imagens utilizando aprendizado profundo tem sido propostos, atingindo os melhores desempenhos (HASKINS; KRUGER; YAN, 2020).

O processo de registro de imagem em geral tem os seguintes passos. O primeiro passo é escolher um modelo de transformação (rígido, afim ou não-rígido) adequado ao problema e inicializar os parâmetros no modelo de transformação escolhido. O segundo passo é utilizar o modelo de transformação para distorcer a imagem em movimento. No terceiro passo ocorre a

avaliação da dissimilaridade entre a imagem em movimento distorcida com a imagem fixa. No quarto passo é utilizado um algoritmo de otimização adequado para atualizar os parâmetros no modelo de transformação, por meio da otimização da função de custo, que é formulada utilizando a métrica de dissimilaridade. É realizada uma iteração do segundo passo até o quarto passo enquanto um critério de convergência apropriado não for atendido (CHEN et al., 2020). Normalmente, os algoritmos convencionais de registro de imagens são compostos por três componentes distintos: o modelo de transformação, a métrica de similaridade e um algoritmo de otimização, conforme é ilustrado na Figura 10.

Figura 10 – Fluxograma da estrutura de registro de imagens em algoritmos de otimização.



Fonte: Adaptado de (CHEN et al., 2020)

Um dos problemas no processo iterativo, que os algoritmos convencionais de registro de imagens usam, é a lentidão, chegando a levar dezenas de minutos, mesmo com uma implementação eficiente nas melhores GPUs contemporâneas. Desta forma, esses algoritmos convencionais são normalmente intensivos, requerindo um poder de processamento computacional muito alto. É desvantajoso utilizar esses algoritmos em operações clínicas que acontecem em tempo real, onde qualquer perda de tempo prolongada não é bem vista (BOVEIRI et al., 2020; CHEN et al., 2020).

Outro problema dos algoritmos convencionais de registro de imagens que usam métodos de otimização é o fato deles serem ineficientes por causa da convergência ou estagnação prematura, devido às medidas de similaridade terem muitos ótimos locais em torno do ótimo global, especialmente em imagens de diferentes modalidades (registro de imagens multimodais) (CHEN et al., 2020).

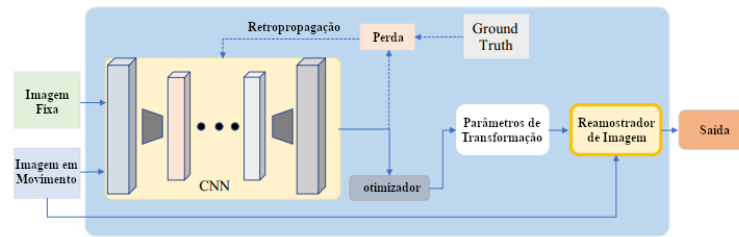
No registro de imagens com aprendizado profundo, diferente dos métodos tradicionais de registro iterativos, se realiza o registro em apenas um passo, deixando o custo principalmente para a etapa de treinamento dos modelos. Ou seja, o processo de otimização ocorre durante o treinamento, não sendo necessário repeti-lo a cada conjunto de imagens diferentes. Desta forma, é possível realizar o registro de imagens em tempo real, sendo uma das primeiras vantagens de se utilizar o aprendizado profundo para esta tarefa. Além disso, as técnicas de registro com aprendizado profundo não sofrem de convergência ou estagnação (BOVEIRI et al., 2020).

Ainda, o aprendizado profundo eliminou as tarefas de escolher, reduzir, selecionar, e normalizar características dos métodos convencionais. Sendo essas tarefas essenciais para se obter o melhor desempenho no problema de registro (BOVEIRI et al., 2020).

Segundo Boveiri et al. (2020), o registro de imagens médicas com aprendizado profundo pode ser dividido em cinco abordagens, sendo elas: Métricas de Similaridade Profunda (DSM, do inglês *Deep Similarity Metrics*), Registro Supervisionado de Ponta a Ponta (SE2ER, do inglês *Supervised End-to-End Registration*), Aprendizado Profundo por Reforço (DRL, do inglês *Deep Reinforcement Learning*), Registro Não Supervisionado de Ponta a Ponta (UE2ER, do inglês *Unsupervised End-to-End Registration*), Registro Fracamente Semi-Supervisionado de Ponta a Ponta (WSE2ER, do inglês *Weakly-Semi-Supervised End-to-End Registration*). Este trabalho adota uma abordagem de Registro Não Supervisionado de Ponta a Ponta, visto que o *framework Voxelmorph*, utilizado para auxiliar no processo de registro de imagem não-rígida, segue esta abordagem de registro.

A abordagem por Métricas de Similaridade Profunda (BOVEIRI et al., 2020) utiliza diferentes tipos de redes neurais profundas para aprenderem as métricas de similaridade por meio de um grande conjunto de dados com anotações (*ground-truth*). Após o treinamento, as redes são aptas em modelar de forma precisa e significativa as diferenças estruturais entre as entradas de pares de imagens ou pares de *patches*. Essa abordagem combina o uso das redes neurais profundas e as abordagens de otimização iterativas convencionais. Desta forma, a rede treinada fornece frequentemente as métricas aprendidas para os algoritmos de otimização convencionais, a fim de produzir os parâmetros de transformação. Esse processo está ilustrado na Figura 11. Esta abordagem foi capaz de superar as métricas dos métodos convencionais, principalmente no registro multimodal, no qual os métodos convencionais não foram muito eficazes. As desvantagens desta abordagem são a necessidade de um grande conjunto de dados, o que é difícil para imagens médicas que não existem muitos conjuntos de dados públicos disponíveis, e a dependência de abordagens convencionais baseadas em iterações, o que torna essa abordagem lenta e inutilizável para uso clínico (BOVEIRI et al., 2020; HASKINS; KRUGER; YAN, 2020).

Figura 11 – Processo de registro de imagem na abordagem por Métricas de Similaridade Profunda.

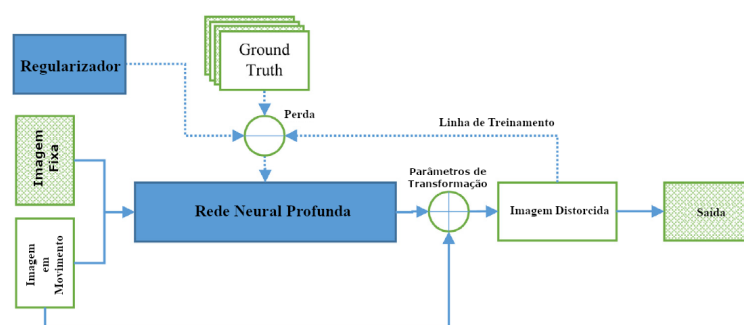


Fonte: Adaptado de (HASKINS; KRUGER; YAN, 2020)

Na Figura 11, observa-se que o fluxo da abordagem de registro de imagem por Métricas de Similaridade Profunda inicia-se com as entradas da imagem fixa e da imagem em movimento na rede. Depois, a rede gera um parâmetro de transformação que é então enviado para um algoritmo de otimização (similar ao convencional) que avalia esse parâmetro. Ao verificar que o parâmetro pode ser melhorado, o processo inicia novamente desde o começo, as iterações acabam quando o parâmetro não poder ser mais melhorado. Quando a iteração acaba o melhor parâmetro de transformação gerado é utilizado para realizar o registro de imagens (HASKINS; KRUGER; YAN, 2020; BOVEIRI et al., 2020).

O processo de treinamento do Registro Supervisionado de Ponta a Ponta (BOVEIRI et al., 2020) é bastante similar ao da abordagem por Métricas de Similaridade Profunda, porém não utiliza nenhum método de otimização iterativo convencional para realizar o registro de imagem, conforme é ilustrado na Figura 12. Desta forma, ao eliminar a abordagem convencional baseada em iteração, o registro pode ser feito em um único passo, permitindo utilizar o registro com essa abordagem em ambiente clínico e ser realizado em tempo real. Um problema do Registro Supervisionado de Ponta a Ponta é a necessidade de um grande conjunto de dados com anotações (*ground-truth*) para o treinamento das redes. Para o registro afim e rígido, as técnicas predominantes são CNN e U-Net (BOVEIRI et al., 2020).

Figura 12 – Processo de treinamento da abordagem de Registro Supervisionado de Ponta a Ponta.

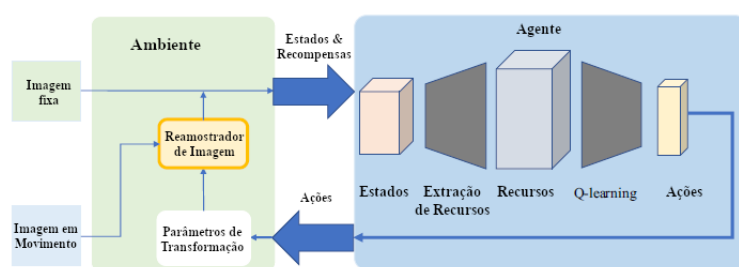


Fonte: Adaptado de (BOVEIRI et al., 2020)

Observa-se na Figura 12 que durante o treinamento a abordagem por Registro Supervisionado de Ponta a Ponta utiliza conjunto de dados com anotações para produzir parâmetros de transformação em um único passo. A imagem fixa e em movimento são enviadas para a rede como entrada. Após, a rede gera um parâmetro de transformação usado para distorcer a imagem em movimento. A imagem distorcida é utilizada para melhorar a rede ao ser comparada sua similaridade com as anotações verdadeiras. Após a rede ser treinada, o registro de imagens é realizado em um único passo, ou seja, sem a necessidade de repetir o processo de otimização para cada conjunto de imagens diferente (BOVEIRI et al., 2020).

No paradigma de registro de imagem por Aprendizado Profundo por Reforço (BOVEIRI et al., 2020), o aprendizado das redes treinadas se dão por reforço, ou seja, é fornecida uma recompensa ou uma punição para cada ação de registro realizada, com o objetivo de maximizar o parâmetro de transformação gerado pelas redes. O processo de registro por reforço é ilustrado na Figura 13. Um dos desafios desta abordagem é lidar com campos de deformação de alta resolução e ser um pouco mais lento quando comparado com método de registro Supervisionado de Ponta a Ponta. Essa abordagem ainda é mais rápida que os métodos tradicionais no registro de imagens médicas (HASKINS; KRUGER; YAN, 2020; CHEN et al., 2020).

Figura 13 – Processo de treinamento da abordagem por Aprendizado Profundo por Reforço.



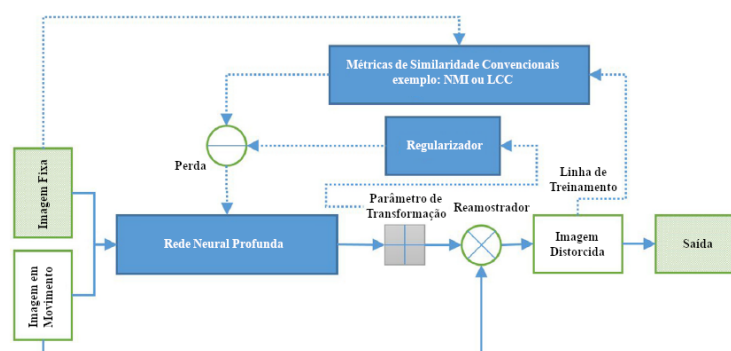
Fonte: Adaptado de (HASKINS; KRUGER; YAN, 2020)

Na Figura 13 observa-se a necessidade um agente para a abordagem de registro de imagens por Aprendizado Profundo por Reforço. O agente aprende a mapear os estados em ações segundo as recompensas que vai recebendo do ambiente. A recompensa é relativa ao desempenho obtido do registro por meio das ações fornecidas pelo agente. Desta forma, as ações geradas pelo agente são fornecidas e melhoradas em cada iteração, com base nas recompensas recebidas pelo ambiente (HASKINS; KRUGER; YAN, 2020; BOVEIRI et al., 2020).

Os conjuntos de dados médicos, especialmente para a tarefa de registro, geralmente são pequenos, o que os tornam inapropriados para o aprendizado profundo baseado em grandes volumes de dados. Essa limitação foi um grande motivador para o surgimento do Registro Não Supervisionado de Ponta a Ponta (BOVEIRI et al., 2020), onde diferentes redes profundas são treinadas sem a necessidade de um conjunto de dados com anotações (*ground-truth*) e realizam o registro em um único passo. Técnicas de aumento de dados (do inglês, *data augmentation*) podem ser utilizadas para aumentar o conjunto de dados de treinamento, e o

processo de aprendizagem é guiado por meio de uma medida ou uma combinação de medidas de similaridade tradicionais, que são empregadas como função de custo (BOVEIRI et al., 2020). Não são necessários conjuntos de dados com anotações, o que evita outro problema, que é o modelo treinado ser dependente da qualidade das anotações. A abordagem de registro não supervisionado é ilustrada na Figura 14. As medidas utilizadas como função de custo para treinar as redes para o registro multimodal são ineficientes, portanto essa abordagem não supervisionada herda essa ineficiência (BOVEIRI et al., 2020; CHEN et al., 2020).

Figura 14 – Processo de treinamento do paradigma de Registro Não Supervisionado de Ponta a Ponta.



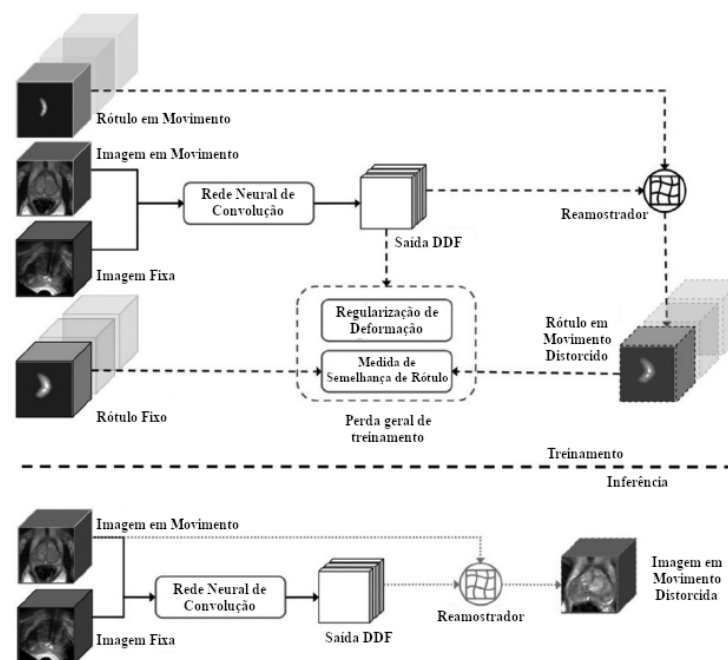
Fonte: Adaptado de (BOVEIRI et al., 2020)

Observa-se na Figura 14 que o fluxo do processo de Registro Não Supervisionado de Ponta a Ponta inicia-se com a imagem fixa e em movimento sendo enviadas para a rede como uma entrada. Então a rede gera uma transformação de parâmetros utilizada no registro da imagem em movimento, que é utilizada para gerar a imagem deformada. Em seguida, a imagem deformada é utilizada em uma métrica de similaridade e o resultado é utilizado para melhorar a rede durante o treinamento, que consequentemente irá melhorar a transformação de parâmetro. Depois que a rede é treinada, o registro de imagens acontece de uma vez só, sem precisar repetir o processo de otimização para cada conjunto de imagens diferente (BOVEIRI et al., 2020).

Segundo (BOVEIRI et al., 2020), o Registro Fracamente Semi-Supervisionado de Ponta a Ponta pode ser considerado como a melhor abordagem atualmente. Existem duas principais categorias diferentes para esta abordagem. Na primeira categoria é utilizado um conjunto de dados pequeno, mas totalmente anotado, com o máximo de pontos de referência possível e cada um desses pontos recebe um rótulo de classe diferente. A rede treinada aprende a detectar os pontos de referência em qualquer par de imagens usadas como entrada e realizar o registro. Pode ser utilizada também a medida de Erro de Registro do Alvo (TRE, do inglês *Target Registration Error*) (BOVEIRI et al., 2020) como função de custo para treinar as redes, porém não é uma tarefa simples. Ao detectar os pontos de referência, os modelos se tornam mais eficientes e aumentam a sua precisão. Na segunda categoria são utilizadas

as GANs, onde o gerador utiliza as entradas da imagem fixa e a imagem em movimento e tenta produzir parâmetros de transformação de modo que a imagem móvel, ao ser distorcida com o parâmetros de transformação produzido, não seja detectada (discriminada) como uma falsa pelo discriminador usando a anotação (*ground-truth*). As GANs também permitem o treinamento com um pequeno conjunto de dados. O processo de registro de imagens por meio desta abordagem é ilustrado na Figura 15. Essa abordagem mescla os pontos fortes das abordagens de registro supervisionado e não supervisionado, ao mesmo tempo que evita as deficiências das mesmas (BOVEIRI et al., 2020; CHEN et al., 2020).

Figura 15 – Processo de registro de imagem por meio do paradigma de Registro Fracamente Semi-Supervisionado de Ponta a Ponta.



Fonte: Adaptado de (HU et al., 2018)

Na Figura 15, a parte superior representa o fluxo de treinamento na abordagem de Registro Fracamente Semi-Supervisionado de Ponta a Ponta. Observa-se que, no treinamento, é utilizado um conjunto de dados com anotações. O fluxo é similar ao da abordagem de Registro Não Supervisionado de Ponta a Ponta com a principal diferença que a rede usa a similaridade entre as anotações da imagem fixa e em movimento para melhorar a regularização de deformação, desta forma, melhorando o campo de deslocamento denso (DDF, do inglês *Dense Displacement Field*), utilizado para deformar a imagem em movimento. Na parte inferior, após a rede ser treinada, é necessário apenas um par de imagens, onde a rede prevê o DDF sem a necessidade de imagens com anotações (HU et al., 2018).

Segundo Chen et al. (2020), o foco das pesquisas futuras serão em precisão, generalização, deformação realista e suave. Ainda, os autores esperam que com o aumento dos conjuntos

de dados públicos multimodais, o foco de pesquisas utilizando aprendizado profundo para o registro com esse tipo de conjunto de dados aumente.

Por sua vez, os autores em Boveiri et al. (2020) acreditam que as próximas tendências e contribuições futuras estarão relacionadas à evolução dos campos de visão computacional e aprendizado de máquina. Os novos modelos advindos desta evolução, por exemplo, os Modelos de Disparos Neurais (SNN, do inglês *Spiking Neural Networks*) (MAASS, 1997) e Unidades Recorrentes (GRU, do inglês *Gated Recurrent Units*) (CHO et al., 2014), possuem alto potencial de contribuição para as pesquisas futuras.

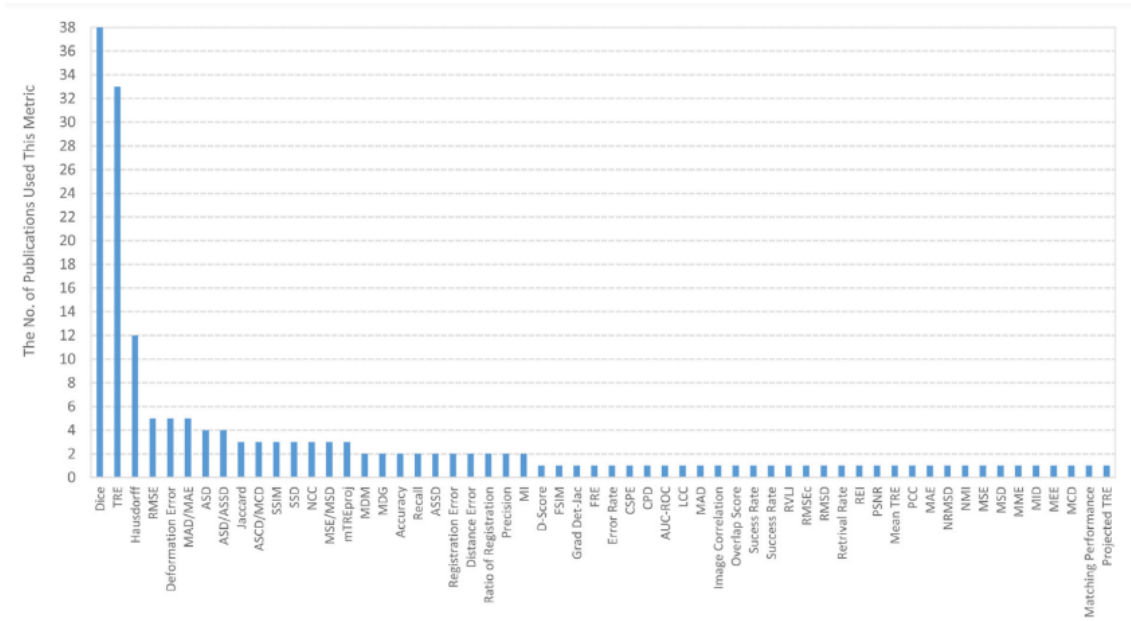
A seguir, são descritos os principais métodos encontrados na literatura para medir de maneira quantitativa o resultado do processo de registro de imagens.

3.3 Métricas utilizadas em registro de imagens

Ao realizar o processo de registro, seja utilizando técnicas tradicionais ou utilizando aprendizado profundo, aplicar métricas que quantifiquem e qualifiquem o procedimento realizado são importantes. Na literatura, observa-se o uso das seguintes métricas para avaliar o procedimento de registro: Correlação Cruzada (CC, do inglês *Cross-correlation*), Informação mútua (MI, do inglês *Mutual Information*), Erro de Registro do Alvo (TRE, do inglês *Target Registration Error*) e Coeficiente de Similaridade Dice (DSC, do inglês *Dice Similarity Coefficient*).

No trabalho de revisão de registro de imagens usando aprendizado profundo dos autores Boveiri et al. (2020), no capítulo da revisão da literatura, as métricas mais utilizadas em relação ao número de publicações são representadas em um gráfico de barras, conforme ilustrado na Figura 16.

Figura 16 – Gráfico de barras representando as principais métricas utilizadas para avaliar os registros de imagens, de acordo com o número de publicações em que foram utilizadas.



Fonte: (BOVEIRI et al., 2020)

Observa-se na Figura 16 que a métrica mais utilizada é Dice, logo em seguida vem o TRE. Liu et al. (2021) reforça que as duas métricas mais utilizadas para avaliar o desempenho do registro de imagens são Dice e MSE.

O Coeficiente de Similaridade Dice (DICE, 1945) é uma métrica utilizada para quantificar a similaridade entre duas regiões, sendo representada matematicamente na Equação 6:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}, \quad (6)$$

em que A representa a primeira região, que podem ser as máscaras da imagem fixa rotulada com a verdade fundamental (do inglês, *ground-truth*) no caso de registro de imagens. E B representa a região predita pela rede, que podem ser as máscaras preditas da imagem em movimento (imagem registrada). Cada *pixel* da primeira região é comparado com apenas o *pixel* correspondente na mesma posição da segunda região. A interseção corresponde aos *pixels* simultaneamente iguais em ambas regiões. O valor resultante da Equação 6 é entre 0 e 1, onde 0 indica que não há nenhuma correspondência entre as regiões (sem sobreposição) e 1 indica total correspondência entre as regiões (TAVARES, 2018; PATIL; DEORE, 2013; BOVEIRI et al., 2020).

A métrica de Erro de Registro do Alvo utiliza pontos anatômicos conhecidos entre duas imagens para quantificar a precisão na correspondência entre esses pontos, representado matematicamente na Equação 7 (BOVEIRI et al., 2020):

$$TRE = \sum_n^i |l_i^A - l_i^B|, \quad (7)$$

em que $l_1^A, l_2^A \dots l_n^A$ são os pontos de referências da imagem A e $l_1^B, l_2^B \dots l_n^B$ são os pontos de referências da imagem B . O TRE é uma das métricas de desempenho mais bem aceita, tendo como desvantagem a necessidade de determinar os pontos de referência manualmente por um especialista, porém é possível utilizar técnicas para automatizar ou semi automatizar os pontos de referência correspondentes (BOVEIRI et al., 2020; HOU et al., 2011).

A métrica do Erro Quadrático Médio (MSE, do inglês *Mean Squared Error*) é utilizada para verificar a diferença simétrica entre a imagem de referência e a imagem registrada (alvo), representada matematicamente na Equação 8 (CHEN et al., 2020):

$$MSE = \frac{1}{NM} \sum_x^M \sum_y^N [I(x, y) - I'(x, y)]^2, \quad (8)$$

em que $I(x, y)$ é a imagem de referência e $I'(x, y)$ é a imagem alvo, enquanto M e N são as dimensões das imagens. Quanto menor o valor resultante da equação significa um erro menor de similaridade e uma similaridade maior entre as imagens (JOSHI, 2012).

Na próxima seção são descritos trabalhos correlacionados com o tema registro de imagens utilizando aprendizado profundo.

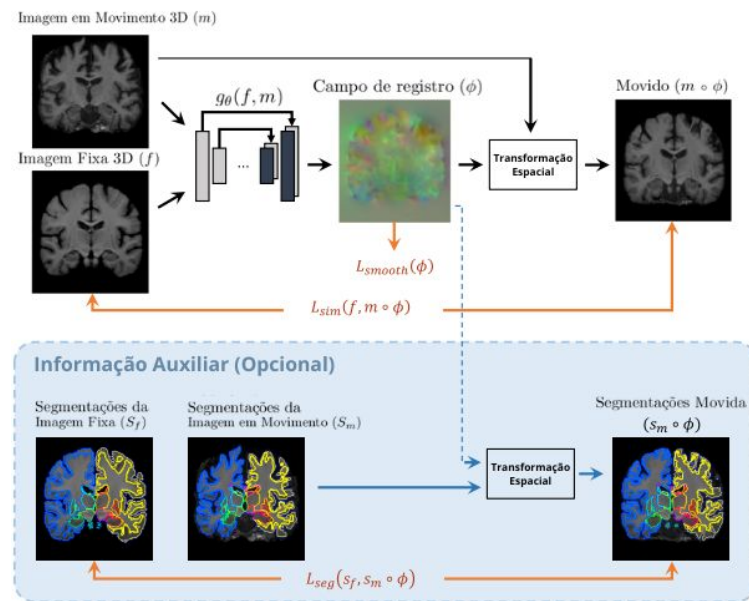
3.4 Trabalhos correlacionados

O trabalho Balakrishnan et al. (2019) propôs um *framework* de aprendizado profundo não supervisionado que utiliza uma arquitetura CNN similar à U-Net para realizar em par o registro não-rígido de imagens médicas 3D, denominado *VoxelMorph*. Foram realizados experimentos em oito conjuntos de dados públicos diferentes de ressonância magnética do cérebro: OASIS, ABIDE, ADHD200, MCIC, PPMI, HABS, FreeSurfer Buckner40 e Harvard GSP. Os resultados obtidos demonstram que a abordagem proposta atingiu resultados no registro de imagens, em termos de *Dice Score*, comparáveis aos métodos tradicionais em estado da arte *NiftyReg* e o *Symmetric Normalization* (SyN) (AVANTS et al., 2008), onde a implementação deste método foi obtida por meio dos pacotes públicos do *software* ANTs (AVANTS et al., 2011). Além disso, o método proposto reduziu o tempo de registro de horas para minutos em uma CPU e para menos de um segundo em uma GPU, em comparação com os métodos tradicionais em estado da arte.

Na Figura 17 é ilustrada uma visão geral do método proposto por Balakrishnan et al. (2019). A imagem de referência f e a imagem em movimento m são utilizadas como entrada para a rede g , utilizando um conjunto de parâmetros θ . A rede gera como resultado o campo de registro ϕ , também conhecido como campo de deformação. O destaque em azul indica o

bloco opcional que pode ser incluído durante a *pipeline* de treinamento, onde são incluídas informações auxiliares, como mapas de segmentação anatômica, elaborados por especialistas humanos ou algoritmos. A função de perda não supervisionada da rede é composta por dois componentes sendo uma métrica de similaridade $\mathcal{L}_{sim}$ e um regularizador $\mathcal{L}_{smooth}$. Foram conduzidos experimentos utilizando tanto a métrica MSE quanto a *Cross-correlation* como medidas de similaridade. Ao incorporar informações auxiliares durante o treinamento, como as segmentações da imagem fixa s_f e as segmentações da imagem em movimento s_m , um componente adicional $\mathcal{L}_{seg}$ é introduzido na função de perda para calcular o *Dice Score* entre a segmentações movida ($s_m \circ \phi$), obtida ao aplicar s_m e ϕ na função de transformação espacial, e s_f .

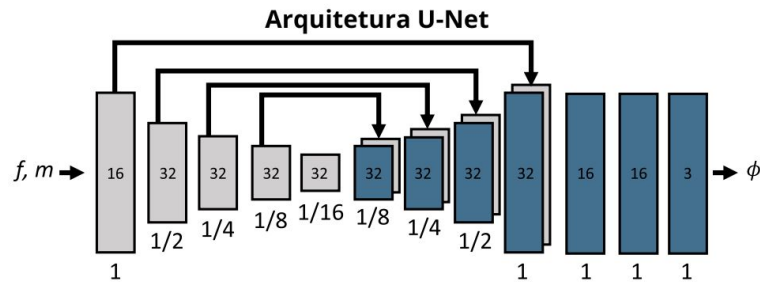
Figura 17 – Visão geral do método proposto por Balakrishnan et al. (2019).



Fonte: Adaptado de (BALAKRISHNAN et al., 2019)

Conforme ilustrado na Figura 18, a arquitetura proposta por Balakrishnan et al. (2019) é baseada na arquitetura da rede U-Net. As entradas m e f são concatenadas, resultando em uma imagem 3D de 2 canais, formando uma entrada única para a rede. Na etapa de codificação a dimensão espacial de cada camada é reduzida pela metade ao alterar o tamanho do *stride*. Na etapa de decodificação é alternado entre *upsampling*, convoluções e *skip connections* que propagam características aprendidas na etapa de codificação diretamente para as camadas que geram o registro. Cada camada convolucional é seguida por uma camada de ativação *LeakyReLU*. As camadas convolucionais extraem as características necessárias para estimar ϕ , que é aplicado junto com m em uma função de transformação espacial baseada em uma *Spatial Transformer Networks* (JADERBERG et al., 2015), obtendo a imagem movida ($m \circ \phi$).

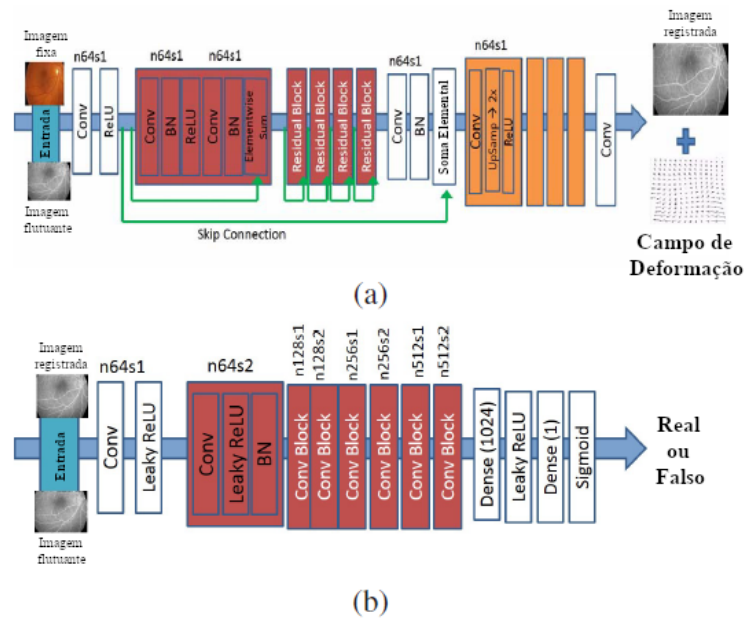
Figura 18 – Arquitetura convolucional U-Net proposta pelo *framework VoxelMorph*. Cada retângulo representa um volume 3D gerado a partir de convoluções 3D aplicadas no volume anterior. O tamanho dos filtros convolucionais estão dentro dos retângulos e abaixo deles é apresentada a resolução espacial de cada volume em relação à sua entrada. As setas representam as *skip connections* que concatenam características do codificador e decodificador.



Fonte: Adaptado de (BALAKRISHNAN et al., 2019)

No trabalho de Mahapatra et al. (2018) foi proposto o uso do aprendizado profundo com o uso das Redes Adversárias Generativas (GANs) para realizar o registro deformável de imagens médicas multimodais. O uso das GANs eliminou a necessidade de métodos iterativos demorados e permitiu o registro de imagem diretamente com o campo de deformação. Na Figura 19 é possível visualizar a arquitetura proposta. Os experimentos são realizados em dois conjuntos de dados diferentes. O primeiro conjunto de dados é composto de imagens coloridas do fundo da retina e imagens de angiografia de fluorescência. Já o segundo conjunto é composto de ressonâncias magnéticas cardíacas adquiridas no *Sunnybrook*. Os experimentos foram realizados em ambos os conjuntos de dados, comparando o método proposto com o método tradicional *Elastix* (KLEIN et al., 2010) e o método de aprendizado profundo DIRNet (VOS et al., 2017). Os resultados obtidos nos experimentos demonstraram que o método proposto performou melhor que o método convencional e o método de aprendizagem profunda que utiliza uma abordagem de transformação tradicional.

Figura 19 – Arquitetura das redes (a) geradoras e (b) discriminadoras utilizadas em (MAHAPATRA et al., 2018).



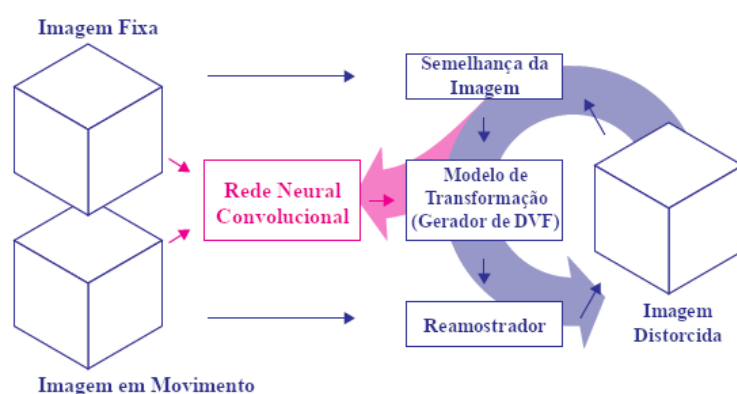
Fonte: Adaptado de (MAHAPATRA et al., 2018)

A rede geradora, Figura 19 (a), recebe a imagem fixa e a imagem em movimento (flutuante) como entrada e produz a imagem registrada e o campo de deformação. A rede geradora é formada por camadas convolucionais 3×3 , seguida por *batch normalization* e ativação ReLU. A rede discriminadora, Figura 19 (b), possui oito camadas convolucionais, onde os tamanhos dos *kernels* são progressivamente aumentados, iniciando com 64 e alcançando 512 ao longo da rede. A função de ativação utilizada pela rede é a *LeakyReLU* e o tamanho do *stride* nas convoluções é utilizado para reduzir as dimensões da imagem. Por último, são inseridas duas camadas densas, seguidas por uma função de ativação sigmóide, a fim de gerar o mapa de probabilidades. O objetivo da rede discriminadora é avaliar a similaridade da distribuição de intensidade entre a imagem registrada e a imagem fixa, e o erro entre o campo de deformação gerado e o de referência. As redes utilizam como função de perda a combinação da perda de conteúdo, que garante que a imagem gerada tenha características desejadas, e a perda *adversarial*. A perda de conteúdo é formada por três componentes: informação mútua normalizada (NMI), métrica de índice de similaridade estrutural (SSIM), e a distância L2 entre duas imagens.

O trabalho Vos et al. (2019) propôs um *framework* para o registro afim e deformável de imagens utilizando aprendizado profundo não supervisionado, denominado *Deep Learning Image Registration (DLIR) framework*. O DLIR é composto de ConvNets treinadas utilizando uma técnica similar às utilizadas nos registros de imagens convencionais baseados em intensidade, sem a necessidade de um conjunto de dados rotulados. Na Figura 20 é ilustrado o esquema do DLIR. Os experimentos foram realizados nos seguintes conjuntos de dados: tomografias

computadorizadas de tórax de baixa dose obtidas no *National Lung Screening Trial* (NLST), ressonâncias magnéticas cardíacas obtidas no *Sunnybrook*, e tomografias computadorizadas do tórax em 4D realizadas em dez momentos diferentes, obtidas no DIR-Lab. Os resultados demonstraram que o registro de imagens do método proposto é equiparável a outros métodos como por exemplo o *SimpleElastix* (MARSTAL et al., 2016). Na comparação em termos de tempo de execução do registro, o método proposto foi muito mais rápido que o convencional. As principais dificuldades encontradas foram o conjunto de dados pequeno para treinar as ConvNets, o que prejudicou o desempenho dos resultados. As limitações de *hardware* podem ter parcialmente prejudicado os resultados, pois a arquitetura escolhida para a ConvNet foi abaixo do que se é desejado devido a esses limites. Os autores ainda afirmam que o trabalho proposto pode ser estendido com outros métodos de registros tradicionais ou registros usando aprendizado profundo e, desta forma, obter resultados ainda melhores.

Figura 20 – Representação esquemática do DLIR *framework*.



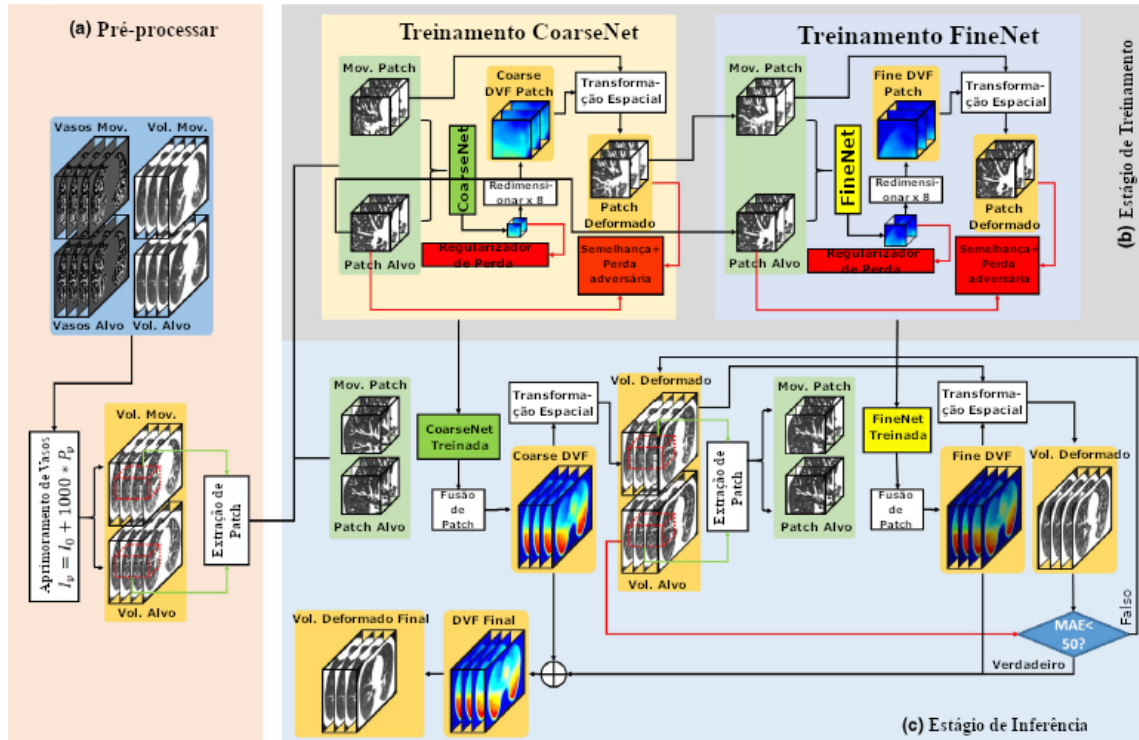
Fonte: Adaptado de (VOS et al., 2019)

Observa-se na Figura 20 que o processo de treinamento do *framework* DLIR é similar ao dos métodos convencionais de registro. Porém, ao adicionar a ConvNet, o método proposto passa a permitir o treinamento não supervisionado para o registro de imagem. Diferente dos métodos convencionais, onde a atualização da transformação de parâmetro é feita de forma iterativa usando a similaridade da imagem (representado pela seta azul grande), o DLIR utiliza a similaridade da imagem para atualizar os pesos da ConvNet usando o *backpropagation* (representado pela seta vermelha grande), permitindo o registro de imagem em um único passo.

Por sua vez, foi proposto no trabalho de Fu et al. (2020b), um método de aprendizado profundo não supervisionado para o registro de imagem deformável de tomografia computadorizada pulmonar, denominado LungRegNet. A LungRegNet é composta de duas sub-redes, CoarseNet e FineNet, sendo que ambas as redes possuem um gerador e um discriminador. A CoarseNet é responsável por prever um grande movimento pulmonar em uma imagem em escala grosseira, enquanto a FineNet prevê o movimento pulmonar local em uma imagem em escala fina. Na Figura 21 é ilustrado o fluxo geral do registro de ambas sub-redes. Os

experimentos foram realizados em dez diferentes tipos de conjuntos de dados públicos em tomografia computadorizada pulmonar 4D obtidos do DIR-Lab. Os resultados obtidos nos experimentos mostraram que o método proposto foi superior aos outros métodos de aprendizado profundo e teve um desempenho comparável com os métodos convencionais em estado da arte.

Figura 21 – Fluxo geral do registro de imagem da LungRegNet.



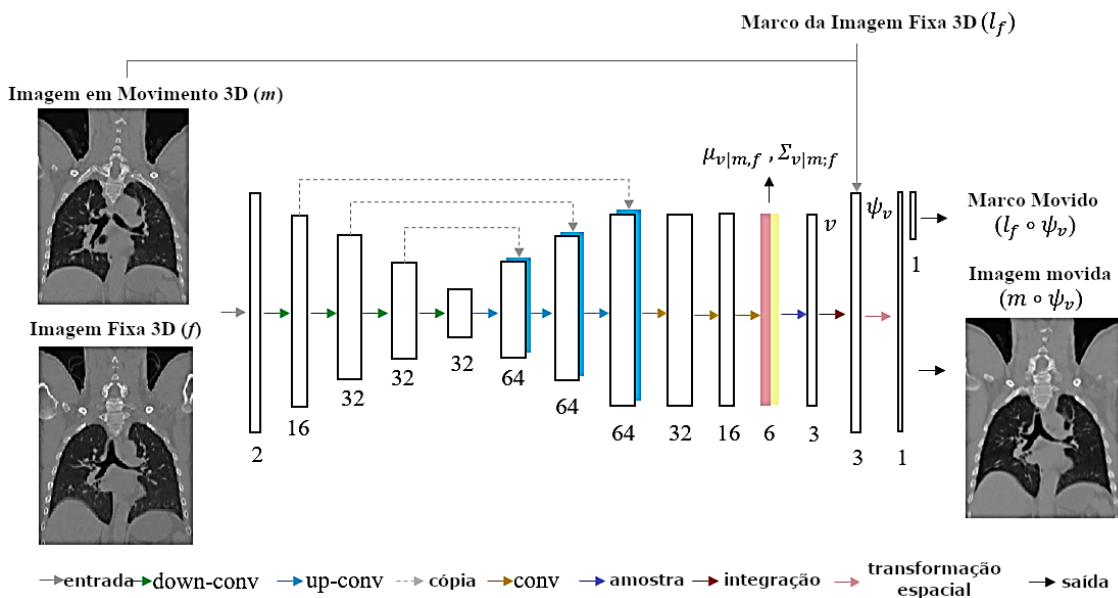
Fonte: (FU et al., 2020b)

Conforme ilustrado na Figura 21, o processo de registro da LungRegNet inicia com a etapa de pré-processamento da imagem. Nesta etapa, é feita a extração das estruturas vasculares pulmonares, aumentando o contraste da imagem de pequenas estruturas vasculares. Na próxima etapa é realizado o treinamento das redes CoarseNet e FineNet. O gerador de ambas as redes é treinado para prever o campo vetorial de deformação (DFV, do inglês *Deformation Vector Field*), que é utilizado para deformar a imagem em movimento usando uma transformação espacial. O discriminador de ambas as redes é treinado para distinguir as imagens deformadas das imagens originais e o objetivo do mesmo é regularizar a DVF, evitando que as imagens deformadas não fiquem realísticas. Na última etapa, a de inferência, os *patches* das imagens são regularizados e fundidos. Ao fundir os patches, é obtida a imagem inteira de referência e em movimento. Essas imagens são aplicadas em ambas as redes treinadas e então é feito o registro da imagem.

No trabalho de Hu et al. (2020) foi proposto um aprendizado restrito por pontos de referências (*landmarks*) utilizando uma CNN. A maioria dos métodos de registro de imagens médicas por aprendizado profundo aprendem o campo de deformação por meio da intensidade

da similaridade, ignorando a importância de alinhar os pontos de referências anatômicos, por exemplo, os pontos de ramificação das vias aéreas e vasos. A CNN utilizada foi estendida para utilizar o registro probabilístico difeomórfico, o que permitiu grandes deformações. Além disso, foi adicionada a restrição por pontos de referências para a aprendizagem, o que permitiu um registro com precisão em estruturas anatômicas finas. O método proposto é semi-supervisionado, pois precisa de pontos de referências definidos por especialistas para o treinamento da rede, e não é necessário nenhum ponto de referência para realizar o registro. Na Figura 22 é ilustrada a arquitetura da rede proposta pelos autores. Os experimentos foram realizados em quarenta imagens de tomografia computadorizada de pulmão em adultos saudáveis, sendo que em cada imagem existem onze anotações de especialistas com pontos de referências. Os resultados demonstraram que o registro de imagem do método proposto foi superior aos métodos em estado da arte tradicionais e o método de registro de imagem com aprendizado profundo *VoxelMorph*. O pacote de *software Elastix* com a métrica de informação mútua e a transformação espacial BSpline e o pacote de *software ANTs* com normalização simétrica (SyN, do inglês *Symmetric Normalization*) foram os métodos tradicionais comparados. O método de aprendizado profundo foi o *VoxelMorph*, que também foi utilizado como linha de base e estendido para criar o método proposto pelos autores. O registro de imagens limitada por pontos de referência mostrou ser eficaz e superior aos métodos comparados e ser mais rápido, além de garantir um alinhamento melhor dos principais pontos de referência anatomicamente significativos. O método proposto também pode ser aplicado em outros conjuntos de dados.

Figura 22 – Arquitetura da rede proposta em (HU et al., 2020).



Fonte: Adaptado de (HU et al., 2020)

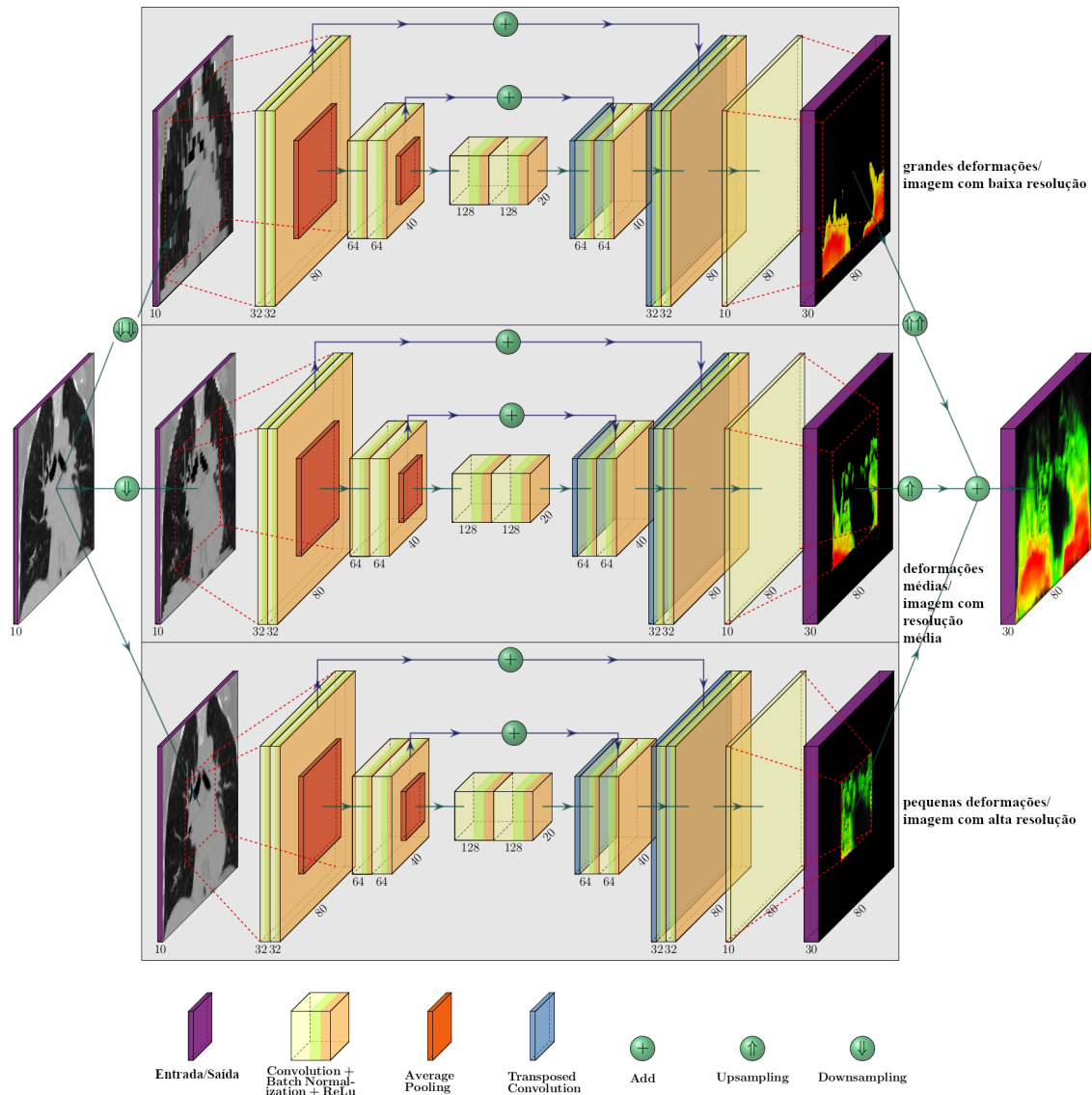
Na Figura 22 observa-se que a rede proposta pelos autores possui uma estrutura semelhante à U-Net e todas as ativações da rede são *LeakyReLU*. A imagem fixa é representada

por f , m representa a entrada da imagem em movimento e v o campo de velocidade. A rede é composta de quatro camadas convolucionais com subamostragem, três camadas convolucionais com superamostragem e três *skip connections* com operação de cópia. Na camada de transformação é realizado o registro, onde os pontos de referências da imagem fixa são utilizados para gerar o campo de deformação a ser aplicado na imagem em movimento.

No trabalho de Fechter e Baltas (2020) foi proposta uma abordagem de aprendizado profundo não supervisionado para o registro de imagens em um conjunto de dados públicos 3D e 4D contendo imagens em movimentos periódicos. Foi empregada uma U-Net combinada com uma abordagem de registro grosseiro para fino com um módulo de transformação espacial diferencial. Para o conjunto de dados em 4D é necessária uma grande quantidade de memória. Desta forma, foi necessário dividir esse conjunto de dados em pequenos *patches* não sobrepostos. A rede é treinada até a convergência com cada um desses *patches* e, por fim, são montados os *patches* de campo vetorial. Os resultados demonstraram que o método proposto conseguiu realizar o registro das imagens em movimentos periódicos de forma concorrente aos métodos em estado da arte comparados. A arquitetura do método proposto pelos autores é ilustrada na Figura 23. Os experimentos foram realizados nos seguintes conjuntos de dados: DIR-Lab, que consiste em conjunto de dados de tomografias computadorizadas do tórax em 4D realizadas em dez momentos diferentes; Popi, que é um conjunto de dados de tomografias computadorizadas do pulmão e; o conjunto de dados *Sunnybrook*, que é composto de ressonâncias magnéticas cardíacas.

O método proposto demonstrou que não foi necessário treinamento para se obter bons resultados. No registro do conjunto de dados 3D são calculados ao mesmo tempo o campo de deformação vetorial e uma aproximação do seu inverso. O método obteve bons resultados no registro, independente do local do órgão ou modalidade. Durante o registro, não foi utilizada uma imagem de referência, como é comum nas outras abordagens. Foi utilizada a imagem vizinha próxima da imagem a ser registrada. Alguns experimentos demonstram que a rede obteve resultados melhores no registro de imagens com pequenas deformações. Os métodos convencionais obtiveram resultados superiores ao método proposto, devido nos testes executados pelo método proposto serem utilizados os mesmos parâmetros em todos os conjuntos de dados que foram utilizados nos experimentos. Comparado com os outros métodos de aprendizagem profunda, o método proposto por estes autores teve um tempo de computação mais longo no registro. Essa desvantagem pode ser reduzida significativamente quando usado um pequeno conjunto de dados e com uma etapa de ajuste fino, onde o desempenho do registro não é muito reduzido. Os principais benefícios do método proposto são a sua capacidade de generalizar, o que permite ser usado em diferentes modalidades e diferentes regiões do corpo, além de não ser necessário nenhum dado de treinamento com antecedência. O método proposto pode ser aplicado em dois casos, quando o treinamento de uma rede não é possível devido à falta de conjuntos de dados ou os conjuntos de dados são muito heterogêneos (FECHTER; BALTAS, 2020).

Figura 23 – Arquitetura da rede multi-resolução proposta.



Fonte: Adaptado de (FECHTER; BALTAS, 2020)

Conforme ilustrado na Figura 23, os retângulos sombreados de cinza representam o processamento dos dados de entrada de imagens com diferentes resoluções. A U-Net é treinada com o conjunto de dados de entrada que é reduzido (*downsampled*) por duas vezes até a convergência, de forma que seja possível calcular as grandes deformações, representadas em vermelho na imagem de saída. As deformações médias e pequenas, representadas por amarelo e verde, respectivamente, são calculadas treinando a rede com uma redução (*downsampled*) simples também com o conjunto de entradas. A arquitetura da U-Net se mantém a mesma, independente da resolução da imagem.

Na próxima seção, são abordados os conjuntos de dados públicos disponíveis para realizar o registro de imagens médicas encontrados na literatura.

3.5 Conjunto de dados públicos disponíveis

Para o registro de imagens médicas, o trabalho de Chen et al. (2020) teve o esforço de sumarizar os principais conjuntos de dados públicos disponíveis, dividindo-os entre as modalidades e órgãos. Na Tabela 1 estão os conjuntos de dados divididos pelos órgãos e modalidades.

Tabela 1 – Conjuntos de dados públicos disponíveis divididos entre tipos de órgãos, registros e modalidades.

Órgão	Tipo de registro	Conjunto de dados	Modalidade
Cérebro	Monomodal	ADNI	RM
		IXI	RM
		OASIS	RM
		BRATS2015	RM
		LPBA40	RM
		IBIS	RM
		IBSR18	RM
		MGH10	RM
		CUMC12	RM
		ABIDE	RM
		ADHD200	RM
		MCIC	RM
		PPMI	RM
		HABS	RM
		Harvard GSP	RM
		FreeSurfer Buckner40	RM
		Mindboggle101	RM
		BraTS2017	RM
		BrainWeb	RM Simulada
	Multi-modal	RIRE	TM, RM Simulada
		BITE	Ultrassom, RM Simulada
		RESECT	Ultrassom, RM Simulada
Coração	Monomodal	Sunnybrook	Sequência de Imagens (Cine) de RM
		ACDC	Sequência de Imagens (Cine) de RM
Continua na próxima página			

Tabela 1 – continuação da página anterior

Órgão	Tipo de registro	Conjunto de dados	Modalidade
	Multi-modal	MM-WHS	TM, RM Simulada
Joelho	Multi-modal	OAI	RM Simulada, Raio X
Fígado	Monomodal	MICCAI 2007 <i>Grand Challenge</i>	TM
		MSD	TM
		SLIVER	TM
		LiTS	TM
Tórax	Monomodal	COPDGen	TM
		NLST	TM, Raio X
		DIR-Lab-COPDgen	TM
		DIR-Lab-4DCT	TM
		SPARE	TM, TM de Feixe Cô- nico
		POPI	TM
		LIDC-IDRI	TM
		<i>Empire 10 lung datasets</i>	TM
		<i>NIH ChestXray14 dataset</i>	Raio X
		JSRT	Raio X
		<i>Montgomery County x-ray database</i>	Raio X
		<i>Shenzhen Hospital x-ray database</i>	Raio X
Vários órgãos	Multi-modal	<i>UK Biobank Imaging Study</i>	RM
Todo o corpo	Multi-modal	<i>VISCERAL Anatomy3</i>	TM, RM

Fonte: (CHEN et al., 2020)

No trabalho de Boveiri et al. (2020) são identificados os conjuntos de dados que são frequentemente utilizados baseado no número de publicações. Segundo os autores, os conjuntos de dados frequentemente usados são: privados, ANDI, LONI, IXI, OASIS e DIRLAB. Sendo que cada um destes conjuntos de dados foram usados em mais de cinco trabalhos. Ainda, os autores listaram outros conjuntos de dados frequentemente usados: *Sunnybrook*, ACDC, MCIC, MGH10, XEF, *Harvard GSP*, HABS, PPMI, CUMC12, IBSR18, *BrainWeb*, *SmartTarget*,

ADHD200, ABIDE, TKA, e VIPS. É importante citar que os conjuntos de dados privados não estavam disponíveis publicamente até a data que o artigo foi publicado.

3.6 Considerações Finais do Capítulo

Neste capítulo foi apresentado em detalhes o processo do registro de imagens com métodos tradicionais e aprendizado profundo, também foi levantado os principais tipos de registro de imagens com aprendizado profundo. Foram apresentadas também as principais métricas utilizadas para quantificar o desempenho do registro de imagens, os trabalhos correlatos sendo úteis como fonte de pesquisa para este trabalho e, por fim, foram levantados os conjuntos de dados públicos de imagens médicas utilizados em tarefas de registro de imagens. Com base no levantamento realizado neste capítulo, este documento apresenta uma proposta para o registro de imagens médicas, que está descrita no próximo capítulo.

4 Uma proposta metodológica para o registro não-rígido de imagens médicas

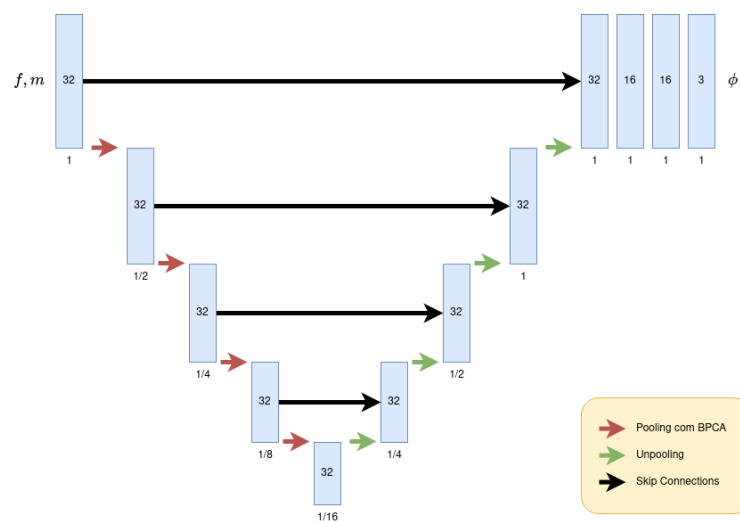
Neste capítulo são apresentados a metodologia deste trabalho, os conjuntos de dados, as métricas a serem utilizadas, os experimentos a serem realizados e as configurações utilizadas.

4.1 Metodologia

Conforme apontado no Capítulo 1, este trabalho visa investigar a utilização do método de extração de atributos Block-Based PCA (BPCA) (mais detalhes sobre o método podem ser encontrados na Seção 4.3) como uma camada de *pooling* em uma rede U-Net para fins de registro de imagens, ou seja, considerando um método estatístico para compressão, mas que mantém a relação espacial e considera toda a informação redundante na imagem.

Desde modo, considerando que o *VoxelMorph* é um *framework* de registro de imagem não-rígido baseado em uma arquitetura U-Net, além de ser um dos principais métodos de registro baseados em aprendizado profundo, a metodologia básica deste trabalho consiste em substituir sua técnica de *pooling*, que é feita por meio da convolução com subamostragem, pelo BPCA.

Figura 24 – Diagrama da rede U-Net do *framework VoxelMorph* com adaptações a fim de utilizar o método BPCA como camada de *pooling*.



Fonte: Baseado em (BALAKRISHNAN et al., 2019)

Na Figura 24, as variáveis f e m são da imagem fixa e da imagem em movimento, respectivamente, concatenadas em uma única imagem 3D de 2 canais, representando uma

única entrada para a rede. Os retângulos representam um volume 3D, que são gerados ao aplicar convoluções 3D no volume anterior. Foram aplicadas convoluções 3D tanto na etapa do codificador quanto na etapa do decodificador. Os valores dentro dos retângulos são o tamanho dos filtros convolucionais, o tamanho do *kernel size* utilizado foi de 3 e o *stride* foi de 1. Cada convolução é seguida por uma camada *LeakyReLU* com parâmetro 0,2. O campo de registro também denominado campo de deformação representado por ϕ é estimado pelas camadas de convoluções ao capturar as características hierárquicas do par de imagens de entrada.

Com exceção do tamanho dos filtros convolucionais, que foram aumentados a fim de garantir o correto funcionamento do BPCA inverso, os outros hiperparâmetros do modelo foram configurados com os valores padrão fornecidos pelo *framework*. Isso se deve ao fato de que o principal objetivo deste trabalho é investigar a performance dos registros com diferentes técnicas de *pooling*.

Na etapa do codificador foi utilizado em cada experimento diferentes métodos de *pooling* a fim de reduzir as dimensões espaciais pela metade em cada camada. Desta forma, as camadas sucessivas do codificador trabalharam com representações mais grosseiras da entrada. Os métodos utilizados se diferenciaram entre BPCA, *Max Pooling*, *pooling* direto na camada de convolução ao aumentar o stride conhecido também como convolução com subamostragem, alternância entre tipos de *poolings* sendo as três primeiras camadas com *Max Pooling* e a última com BPCA e, por fim, o inverso, onde as três primeiras camadas iniciais utilizaram o BPCA e, a camada final, o *Max Pooling*.

Na etapa do decodificador foi alternado entre *upsampling*, convoluções e *skip connections*. O método de *upsampling* utilizado nos experimentos é conhecido como vizinho mais próximo (do inglês, *Nearest Neighbor*), no qual envolve a duplicação dos *pixels* ou amostras existentes para criar novos. Somente um dos experimentos foi utilizado um método de *upsampling* diferente no qual consiste reconstruir o BPCA por meio da multiplicação matriz transposta. As *skip connections* tem a função de propagar características aprendidas durante a etapa do codificador diretamente para camadas que geram o registro.

Durante o treinamento, o *VoxelMorph* aplica o campo de deformação ϕ , estimado pela rede, por meio de uma função de transformação espacial para deformar m e obter $m \circ \phi$. Esse resultado é então utilizado em uma função de perda que compara f com $m \circ \phi$, visando aprimorar continuamente a suavidade de ϕ ao longo do processo de treinamento (BALAKRISHNAN et al., 2019).

Para a implementação desta metodologia e dos experimentos a serem realizados para a geração de resultados foram utilizadas as seguintes ferramentas: linguagem *Python*², o *framework* de aprendizado de máquina *PyTorch*³, a ferramenta *Tensorboard*⁴ para a análise das

² <https://www.python.org/>

³ <https://pytorch.org/>

⁴ <https://www.tensorflow.org/tensorboard>

métricas com a acurácia, além das bibliotecas *Pandas*⁵ para análise dos dados, o *framework* *Voxelmorph*⁶ para o registro de imagens e *Matplotlib*⁷ para visualizar as imagens e os gráficos.

Com o propósito de impulsionar o desenvolvimento acadêmico e de pesquisa, bem como viabilizar aprimoramentos futuros no âmbito do registro de imagens mediante a utilização de camadas de *pooling* personalizadas, como exemplificado pela camada de BPCA, os códigos desenvolvidos neste projeto foram disponibilizados como *software* de código aberto nos repositórios do autor⁸ no *Github*, em conformidade com a licença *Apache v2.0*⁹.

Na próxima seção, é apresentada uma análise do *Principal Component Analysis* (PCA) e os procedimentos necessários para realizar efetivamente essa técnica em uma matriz de dados.

4.2 Principal Component Analysis

O método estatístico de extração de atributos Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*) criado por Pearson (1901), é uma abordagem que tem como objetivo reduzir uma quantidade significativa da dimensionalidade dos dados, tendo a menor perda possível da informação, ou seja, preservando ao máximo a informação relevante.

A seguir, apresenta-se a representação matemática dos passos essenciais para reduzir a dimensionalidade de uma matriz de dados por meio do PCA (PAUL; SUMAN; SULTAN, 2013). O primeiro passo para realizar o PCA em uma matriz de dados é subtrair a média μ dos dados, representada matematicamente na Equação 9 (SMITH, 2002; PAUL; SUMAN; SULTAN, 2013):

$$\mu = \frac{\sum_{i=1}^n \mathbf{X}_i}{n}, \quad (9)$$

onde $\mathbf{X}$ é a matriz de mistura, que contém todos os exemplos (no caso do BPCA, cada exemplo será um bloco) utilizados para calcular o PCA, $\mathbf{X}_i$ é um exemplo do conjunto de dados, μ é o vetor média dos exemplos de $\mathbf{X}$ e n é o número de observações $\mathbf{X}_i$.

A próxima etapa é calcular a matriz de covariância $\mathbf{C}$ de $\mathbf{X}$, representada matematicamente na Equação 10 (SMITH, 2002; PAUL; SUMAN; SULTAN, 2013):

$$\mathbf{C} = \text{cov}(\mathbf{X}) = E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T], \quad (10)$$

onde E é o valor esperado.

⁵ <https://pandas.pydata.org/>

⁶ <https://github.com/voxelmorph/voxelmorph>

⁷ <https://matplotlib.org/>

⁸ Perfil do autor no *Github* - <https://github.com/Uemerson>

⁹ <https://www.apache.org/licenses/LICENSE-2.0>

O próximo passo é calcular os autovetores e autovalores da matriz de covariância C (SMITH, 2002; PAUL; SUMAN; SULTAN, 2013). Um autovetor em álgebra linear é um vetor não nulo que, quando sujeito a uma transformação linear, muda no máximo por um fator escalar. O autovetor representa os componentes principais, ou seja, as direções nas quais a variação é máxima, na matriz de covariância. Os autovalores são suas magnitudes correspondentes. Assim, o autovetor que possui o maior autovalor correspondente representa a direção de máxima variância. Na matriz de covariância C um autovetor $\vec{v}$ é representado matematicamente na Equação 11 (CASADO, 2023):

$$C\vec{v} = \lambda\vec{v} \Rightarrow (C - \lambda I)\vec{v} = 0, \quad (11)$$

onde I representa a matriz identidade, λ representa o autovalor associado ao autovetor $\vec{v}$ de C .

O próximo passo é ordenar os autovetores em ordem decrescente de seus autovalores e escolher os componentes principais mais representativos, obtendo a matriz de vetores de características Z (CASADO, 2023). E por fim calcular a matriz de projeção, representado matematicamente na Equação 12 (CASADO, 2023; SMITH, 2002):

$$W = Z^T \cdot X^T, \quad (12)$$

onde, X^T representa a matriz de dados original transposta e Z^T é a matriz com os autovetores nas colunas transpostas, de modo que os autovetores agora estão nas linhas, com o autovetor mais significativo no topo. Por meio destes passos detalhados acima os dados originais tem sua dimensionalidade reduzida preservando ao máximo as informações relevantes (CASADO, 2023; SMITH, 2002).

O PCA também pode ser obtido por meio do método de Decomposição em Valores Singulares (SVD, do inglês *Singular Value Decomposition*) (GERBRANDS, 1981). Desta forma, é possível reverter a dimensionalidade dos dados usando o SVD, sendo necessário multiplicar a matriz diagonal dos valores singulares S pela transposta da coluna de vetores singulares à direita V , representado pela expressão $S_{1:k,1:k} V_{:,1:k}^T$, onde k corresponde ao tamanho da dimensionalidade das colunas selecionadas.

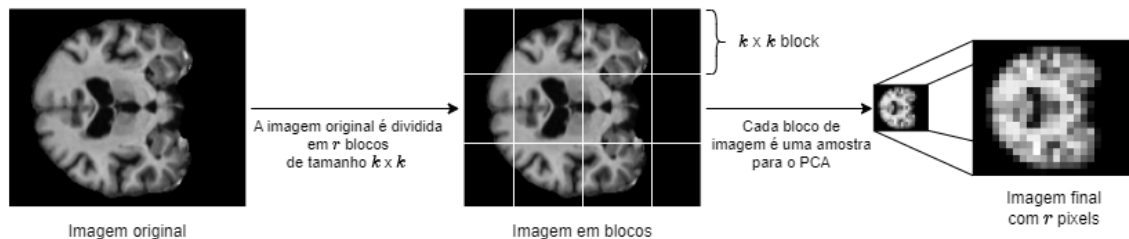
Na próxima seção é detalhado o funcionamento do *Block-Based Principal Component Analysis*.

4.3 Block-Based Principal Component Analysis

O *Block-Based Principal Component Analysis* (BPCA) (SALVADEO et al., 2011) é uma das variantes do método estatístico de extração de atributos conhecido como *Principal*

Component Analysis (PCA). O método divide uma imagem em blocos, geralmente de tamanhos iguais (4x4, 8x8 e assim por diante). No BPCA, cada bloco da imagem representa uma amostra, enquanto no PCA, cada imagem é uma amostra. Depois, cada bloco $k \times k$ é projetado nos m autovetores associados aos m maiores autovalores da matriz de covariância da mistura de blocos. Isso permite reduzir cada bloco de tamanho $k \times k$ pixels a um subconjunto menor de pixels, o que reduz as dimensões espaciais de cada imagem. Na Figura 25, é representado o diagrama de blocos do método BPCA aplicado a uma imagem de uma fatia de ressonância magnética ponderada em T1 do cérebro.

Figura 25 – Diagrama de blocos do método BPCA.



Fonte: autor

Conforme ilustrado na Figura 25, a imagem é subdividida em r blocos de tamanho $k \times k$ pixels, sendo que cada bloco se torna uma amostra para o PCA. O principal componente principal foi selecionado para fazer o mapeamento, transformando cada bloco de tamanho $k \times k$ pixels em um único pixel. Assim, é produzida uma representação final da imagem em termos de r pixels/atributos (SALVADEO et al., 2011).

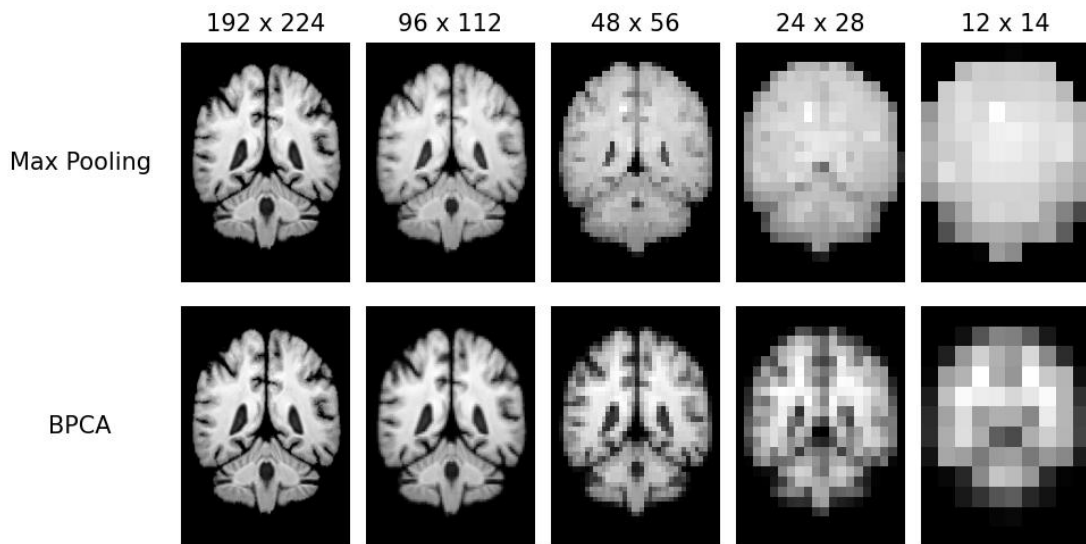
Destaca-se que o método utiliza um modelo estatístico para comprimir as imagens, preservando a estrutura da imagem e sua relação espacial, levando em consideração toda a informação contida na imagem. Estas características podem ser relevantes ao problema de registro de imagens, uma vez que há correspondência espacial entre as imagens a serem registradas.

A abordagem do BPCA nos conjuntos de dados deste trabalho se assemelha ao processo de *Max Pooling*. Consiste em percorrer todo o volume 3D com uma janela de tamanho 2x2x2 e com passo 2, extraindo blocos que servirão como amostras para o PCA. A matriz resultante dessa operação é transformada em uma matriz bidimensional para aplicação do PCA. Nos experimentos foi selecionado apenas um componente principal, transformando os blocos 2x2x2 em um valor único. Por fim, a matriz resultante do PCA é redimensionada de volta para uma forma tridimensional.

Na Figura 26, é ilustrada a aplicação dos métodos de *pooling Max Pooling* e BPCA em uma imagem em diferentes resoluções. As colunas são identificadas pelas suas respectivas resoluções, sendo a primeira coluna a imagem com sua resolução original. Cada linha corresponde aos resultados da aplicação do *Max Pooling* e BPCA em diferentes resoluções da imagem,

respectivamente. Exceto pela primeira coluna, cada resultado em uma coluna representa a aplicação do método de *pooling* à resolução anterior.

Figura 26 – Resultado do *Max Pooling* e BPCA aplicados em diferentes resoluções em uma imagem.



Fonte: autor

Nota-se na Figura 26 que a partir da coluna 3 (48×56), o BPCA mantém mais informações, como as estruturas da imagem, enquanto o *Max Pooling* perde mais informações ao comprimir as imagens.

A seguir é apresentado em detalhe os conjuntos de dados a serem utilizados nos experimentos realizados neste trabalho.

4.4 Conjuntos de dados

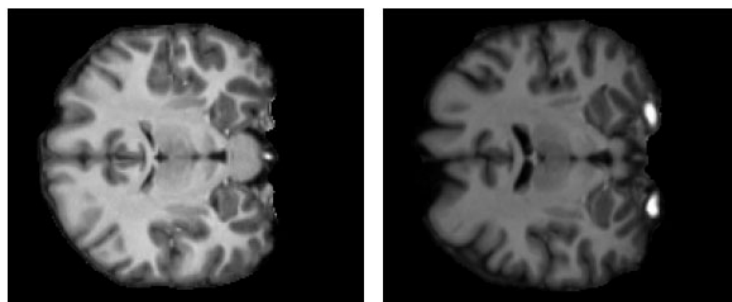
A fim de responder a questão proposta nesta pesquisa, a metodologia proposta foi avaliada em algumas bases de dados públicas de imagens médicas, que estão especificadas a seguir.

O OASIS(MARCUS et al., 2007) é um conjunto de dados público obtido a partir do desafio *Learn2Reg* de 2021 (HERING et al., 2023) e pode ser baixado no formato pkl no repositório público do *Github*¹⁰. As imagens do conjunto de dados possuem uma resolução de $160 \times 192 \times 224$. O conjunto de dados é composto de 451 exemplos de imagens de ressonância magnética ponderadas em T1 do cérebro. Adotou-se a divisão padrão proposta pelo conjunto de dados entre treinamento, validação e teste. Desta forma, 394, 19 e 38 imagens foram usadas para

¹⁰ https://github.com/junyuchen245/TransMorph_Transformer_for_Medical_Image_Registration/blob/main/OASIS/TransMorph_on_OASIS.md

treinamento, validação e teste, respectivamente. Foram utilizadas as 24 anotações de rótulos anatômicos automaticamente segmentados com *FreeSurfer*¹¹ e SAMSEG¹² (do inglês *Sequence Adaptive Multimodal SEGmentation*) disponíveis na *toolbox Neurite*¹³, para avaliação do desempenho do registro. Na Figura 27, é mostrado um exemplo de uma imagem de uma fatia do conjunto de dados OASIS.

Figura 27 – Exemplos de imagens contidas no conjunto de dados OASIS.



Fonte: autor

O conjunto de dados IXI também é um conjunto de dados público disponível no repositório público no *GitHub*¹⁴, que foi preparado para o trabalho (CHEN et al., 2022). Os exemplos deste conjunto de dados também possuem uma resolução de 160x192x224. Consiste em 576 exemplos de imagens de ressonância magnética ponderadas em T1 do cérebro. Adotou-se a distribuição padrão sugerida pelo conjunto de dados entre treinamento, validação e teste. Portanto, dentre os exemplos fornecidos, foram utilizadas 403 imagens para o processo de treinamento, 58 imagens para a validação e mais 115 imagens para o teste. Para a avaliação do desempenho do registro, foram utilizadas segmentações com 30 rótulos de estruturas anatômicas. Na Figura 28, é mostrado um exemplo de uma imagem de uma fatia do conjunto de dados IXI.

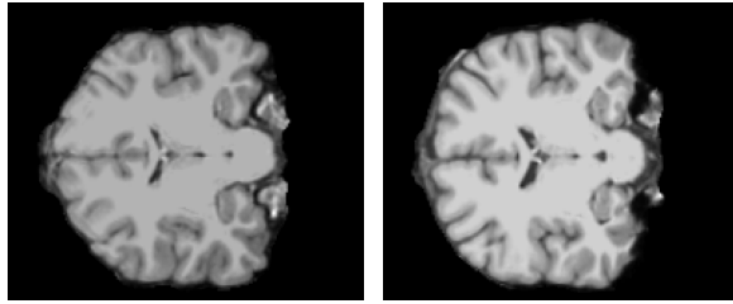
¹¹ <https://surfer.nmr.mgh.harvard.edu/>

¹² <https://surfer.nmr.mgh.harvard.edu/fswiki/Samseg>

¹³ <https://github.com/adalca/neurite>

¹⁴ https://github.com/junyuchen245/TransMorph_Transformer_for_Medical_Image_Registration/blob/main/IXI/TransMorph_on_IXI.md

Figura 28 – Exemplos de imagens contidas no conjunto de dados IXI.



Fonte: autor

A seguir são apresentadas as métricas que foram utilizadas neste trabalho para avaliar o desempenho do registro de imagens.

4.5 Métricas

A fim de comparar o desempenho do registro quantitativamente será utilizada a métrica *Dice Score* em ambos conjuntos de dados, comparando a imagem resultante do registro com a imagem alvo. Essa métrica permite avaliar a similaridade entre duas ou mais regiões, possibilitando também verificar a similaridade usando a imagem completamente. Essa e outras métricas são detalhadas na Seção 3.3.

A quantificação numérica da regularidade dos campos de deformação foi obtida por meio das porcentagens de valores não positivos no determinante da matriz Jacobiana nos campos de deformação ($|J_\phi| \leq 0$). Se $|J_\phi| > 0$ então o campo de deformação preserva as características anatômicas e evita distorções indesejáveis, ou seja, uma transformação difeomórfica (MOK; CHUNG, 2020). O registro de imagem não-rígido difeomórfico é crucial em aplicações médicas pois mantém características anatômicas, como estruturas e proporções, durante a transformação (BALAKRISHNAN et al., 2018; CHEN et al., 2022).

Os resultados do conjunto de dados de validação do OASIS são obtidos diretamente do placar de líderes do desafio, enquanto os resultados de teste são fornecidos de forma limitada diretamente pelos organizadores do desafio. As métricas fornecidas pelo desafio são *Dice* de 35 estruturas cerebrais corticais e subcorticais, a porcentagem do percentil 95 da distância de *Hausdorff* (HD95) e o desvio padrão do logaritmo do determinante de Jacobiano (SDlogJ).

A métrica HD95 compara coordenadas de pontos correspondentes tanto na imagem de referência como na imagem registrada. Assim, um valor baixo nessa métrica indica que a correspondência de pontos entre as imagens é mais precisa, ou seja, o registro com menor valor nessa métrica é superior aos que teve valores mais alto. O desvio padrão do logaritmo do determinante de Jacobiano é utilizado para avaliar a variação na deformação entre as imagens registradas (HERING et al., 2023).

Na próxima seção, são apresentados com detalhes quais os experimentos realizados neste trabalho.

4.6 Experimentos

O experimento básico a ser realizado é o registro de imagens com a alteração na rede U-Net, onde será utilizado o método estatístico BPCA como camada de *pooling*. Além disso, pretende-se realizar uma série de experimentos com algumas variações para comparação, especialmente para se avaliar parâmetros e abordagens suavemente diferentes, entre elas: a) intercalação entre métodos de *poolings* diferentes (BPCA e *Max Pooling*), sendo as três primeiras camadas de *pooling* usando o BPCA e a última usando *Max Pooling* e também o inverso, com as três primeiras camadas de *pooling* usando o *Max Pooling* e a última com o BPCA; b) utilização ou não do BPCA inverso para o *upsampling*; c) realizar o *pooling* direto na camada de convolução (convolução com subamostragem);

Os resultados obtidos no experimento base demonstrarão se o registro com o *framework VoxelMorph*, considerando as adaptações na rede U-Net para utilizar o método estatístico BPCA para comprimir no caminhos de codificação (*downsample*), trazem uma maior capacidade de representação comprimida em favor de uma melhor qualidade no registro de imagens, em comparação ao método sem nenhuma adaptação. Ao intercalar os *poolings* é possível identificar se o BPCA atua melhor nas camadas iniciais de *poolings* (representações mais finas da entrada) ou nas finais (representações mais grosseiras da entrada). O BPCA inverso é um experimento que possibilita verificar a efetividade desta técnica de *unpooling* para o BPCA e sua efetividade no registro de imagens.

Caso os resultados sejam superiores é possível afirmar que essas alterações melhoram o desempenho do registro de imagem. Já as variações propostas permitem realizar um estudo exploratório no sentido de identificar como a misturas de diferentes tipos de *pooling*, a utilização ou não do BPCA inverso e a convolução com subamostragem podem afetar o registro de imagens.

Na próxima seção são apresentadas as configurações utilizadas neste trabalho.

4.7 Configurações

No que se refere ao *hardware* utilizado para realizar os experimentos deste trabalho, foi utilizada uma máquina com 64 GB de memória RAM, um processador (CPU) Intel(R) Core(TM) i9-12900F de 12ª geração e uma placa de vídeo (GPU) NVIDIA GeForce RTX 4090. Uma vez que o propósito principal deste trabalho é avaliar o desempenho do registro utilizando diferentes métodos de *poolings*, para garantir uma avaliação imparcial, os mesmos parâmetros foram mantidos constantes em todas as sessões de treinamento. Foi utilizado como função de

perda o erro médio quadrático (MSE), juntamente com o otimizador ADAM com uma taxa de aprendizado de 10^{-4} , e o processo de treinamento foi executado por 500 épocas. A arquitetura U-Net utilizada neste trabalho consistiu em 5 camadas de convolução de codificação (com 32, 32, 32, 32 e 32 canais) e 7 camadas de convolução de decodificação (com 32, 32, 32, 32, 16, 16 e 3 canais), resultando em um total de 355.859 mil parâmetros treináveis.

No próximo capítulo serão apresentados os resultados obtidos nos experimentos realizados neste trabalho.

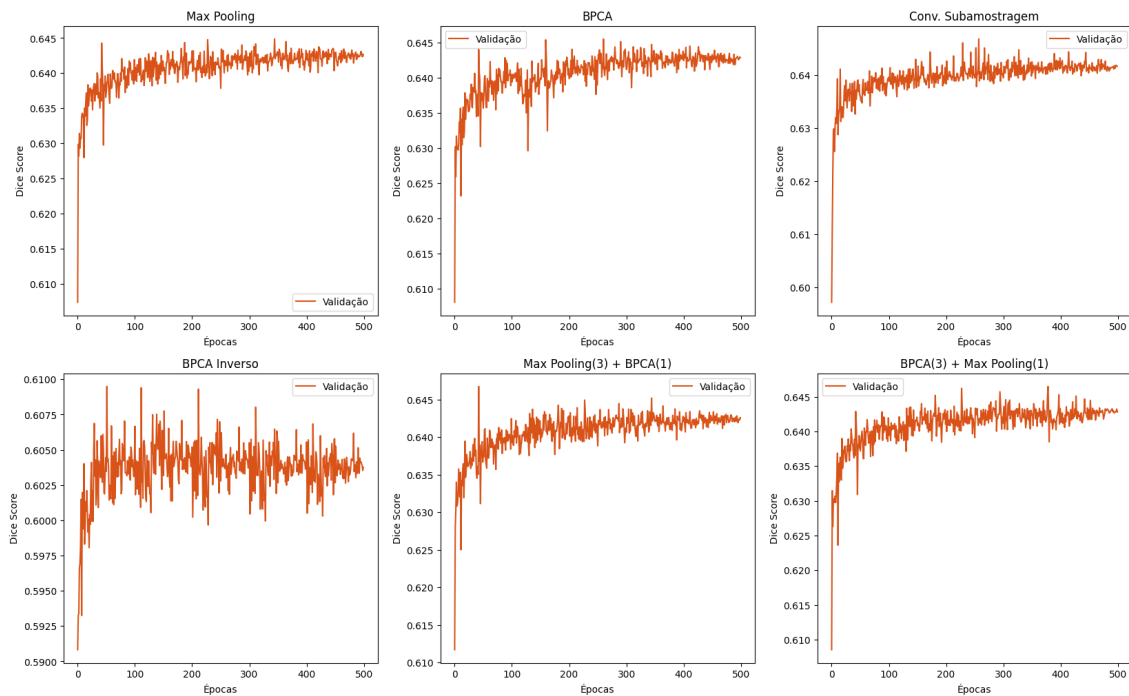
5 Resultados

Neste capítulo são apresentados os resultados obtidos deste trabalho, organizados em seções para cada conjunto de dados.

5.1 Conjunto de dados OASIS

São ilustrados na Figura 29 os gráficos da evolução da métrica *Dice Score* ao longo das 500 épocas de treinamento no conjunto de dados de validação OASIS, nos diferentes experimentos realizados.

Figura 29 – *Dice Score* obtido ao longo do treinamento no conjunto de validação do conjunto de dados OASIS.



Fonte: autor

Nota-se na Figura 29, que os valores do *Dice Score* não convergiram, indicando uma possibilidade de melhoria nos resultados obtidos, ao aumentar o número de épocas nos treinamentos. O maior *Dice Score* obtido foi no experimento utilizando a convolução com subamostragem, atingindo aproximadamente o valor de 0,6467 e o pior valor obtido foi no experimento com o BPCA inverso com 0,6094.

Na Tabela 2, encontram-se os tempos de treinamento de cada experimento, expressos em segundos e em horas.

Tabela 2 – Tempo de treinamento de cada experimento no conjunto de dados OASIS, expressos em segundos e em horas. O número destacado em negrito indica o melhor tempo.

Método de <i>Pooling</i>	Tempo em segundos	Tempo em horas
<i>Max Pooling</i>	110941	30,817
BPCA	144160	40,044
Convolução com subamostragem	92161	25,600
BPCA inverso	145447	40,402
<i>Max Pooling</i> (3) + BPCA(1)	110734	30,759
BPCA(3) + <i>Max Pooling</i> (1)	145327	40,369

Fonte: autor

Diferente do conjunto de dados IXI, o OASIS é um conjunto de dados associado a um desafio, e, desta forma, o conjunto de dados de teste com o *ground truth* não é disponibilizado publicamente. Para obter os resultados do conjunto de testes, é necessário submeter os resultados dos métodos, incluindo os campos deformáveis, aos organizadores do desafio. Os organizadores, a fim de preservar a equidade do desafio limitam o número de envios evitando ajustes excessivos ao conjunto de testes e avaliam apenas uma submissão por método. Neste trabalho, apesar das variações nos métodos de *pooling*, os organizadores decidiram tratá-los como não distintos entre si. Dessa forma, uma exceção foi feita, permitindo a avaliação de três experimentos diferentes. Os resultados dessas avaliações estão detalhadamente apresentados na Tabela 3.

Tabela 3 – Resultados fornecidos pelos organizadores do desafio do conjunto de dados de teste do OASIS, os números destacados em negrito indicam as melhores classificações.

Método de <i>Pooling</i>	<i>Dice Score</i>	HD95	SDlogJ
<i>Max Pooling</i>	0,638	3.148	0.012
BPCA	0,637	3,139	0,012
BPCA(3) + <i>Max Pooling</i> (1)	0,639	3.141	0,012

Fonte: autor

Observa-se que na Tabela 3, os resultados de *Dice Score* obtidos nos experimentos são bem próximos, porém o maior valor foi do método no registro que utilizou a mistura de *pooling* entre BPCA, nas três camadas iniciais, e o método *Max Pooling* na camada de *pooling* final. Na métrica de HD95, os três métodos também tiveram valores próximos, mas o melhor resultado foi com o registro usando o BPCA. Os resultados obtidos na métrica SDlogJ foram iguais para os três métodos diferentes.

Na Tabela 4 estão os resultados do conjunto de dados de validação do OASIS, obtidos diretamente do placar de líderes do desafio.

Tabela 4 – Resultados fornecidos pelos organizadores do desafio do conjunto de dados de validação do OASIS, as melhores classificações estão destacadas em negrito.

Método de <i>Pooling</i>	<i>Dice Score</i>	HD95	SDlogJ
<i>Max Pooling</i>	0,647 ± 0,047	3,186 ± 0,688	0,010 ± 0,001
BPCA	0,647 ± 0,046	3,153 ± 0,676	0,010 ± 0,001
Convolução com subamostragem	0,649 ± 0,046	3,143 ± 0,655	0,010 ± 0,001
BPCA inverso	0,611 ± 0,051	3,540 ± 0,710	0,008 ± 0,001
<i>Max Pooling</i> (3) + BPCA(1)	0,649 ± 0,049	3,151 ± 0,617	0,010 ± 0,001
BPCA(3) + <i>Max Pooling</i> (1)	0,649 ± 0,045	3,159 ± 0,671	0,010 ± 0,001

Fonte: autor

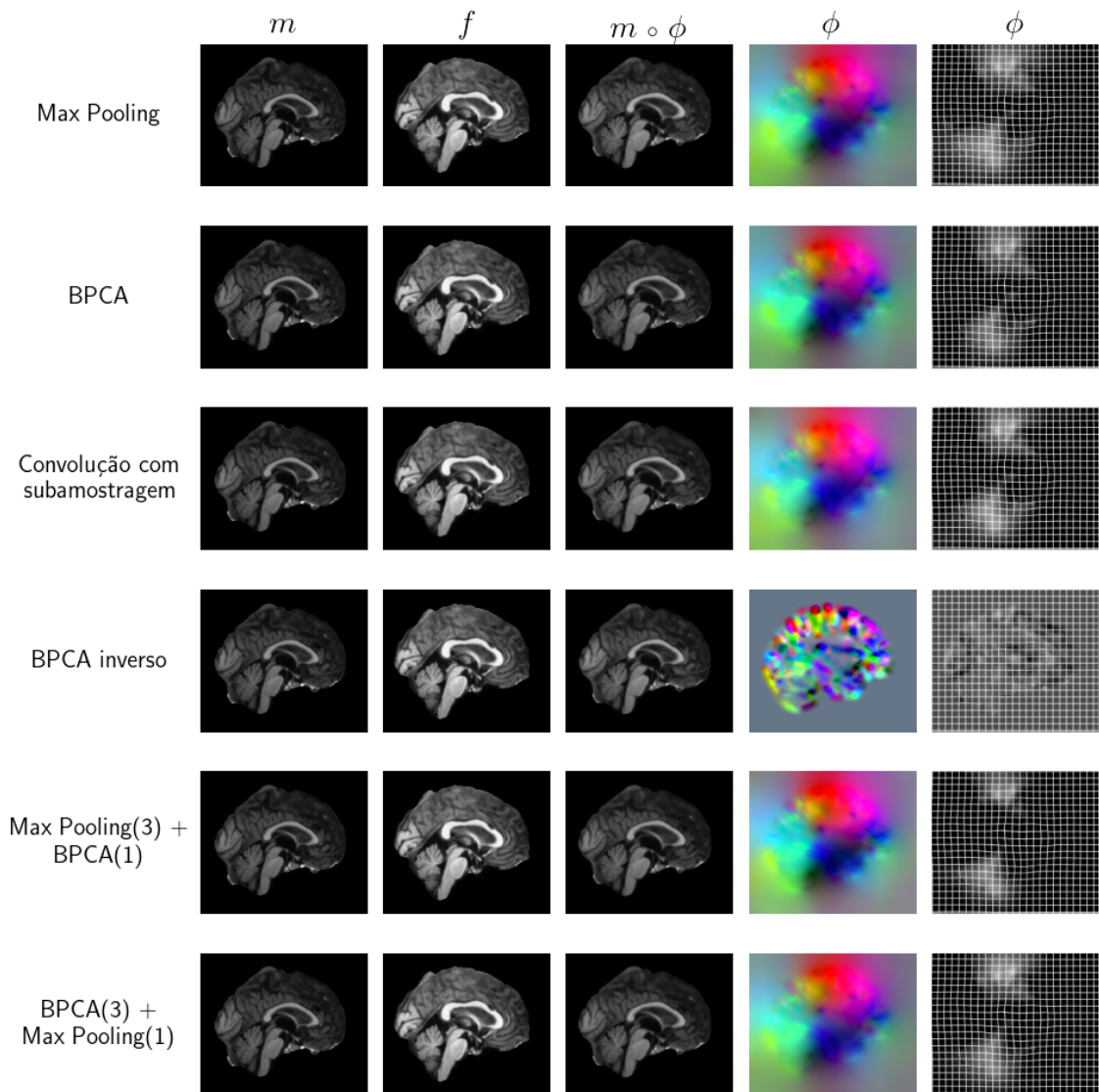
A Tabela 4 demonstra que os métodos que misturam os *pooling* e a convolução com subamostragem obtiveram o mesmo *Dice Score*, porém a mistura de técnicas de *pooling* onde o BPCA é usado nas três primeiras camadas e o *Max Pooling* foi superior por ter um desvio padrão menor que os demais. A convolução com subamostragem obteve a melhor classificação na métrica HD95. O BPCA inverso obteve a melhor classificação na métrica SDlogJ.

Não foi possível realizar o teste T (MONTGOMERY; RUNGER, 2002) para verificar se há diferenças significativas em termos estatísticos dos resultados de *Dice Score* no conjunto de dados OASIS, pois tanto os resultados do conjunto de validação quanto os do teste forneceram apenas os valores finais, sem disponibilizar todas as amostras.

Nas Figuras 30, 32 e 34, os símbolos m , f e $m \circ \phi$ representam, respectivamente, a imagem móvel, a imagem fixa ou de referência, e a imagem resultante após o processo de registro. Na coluna 4, identificada como ϕ , é apresentado o campo de deformação gerado pela rede, onde o deslocamento em cada dimensão espacial x , y e z é mapeado para cada um dos canais de cor RGB. A coluna 5, identificada como ϕ , apresenta o resultado do campo de deformação aplicado em um *grid*. Cada linha corresponde ao resultado de experimentos de registro utilizando diferentes abordagens, como *Max Pooling*, BPCA, convolução com subamostragem, BPCA inverso, três camadas de *Max Pooling* e uma de BPCA, e três camadas de BPCA e uma de *Max Pooling*, respectivamente.

A Figura 30 mostra duas fatias cerebrais sagitais de imagens de ressonância magnética extraídas de dois escaneamentos diferentes do conjunto de dados de teste do OASIS.

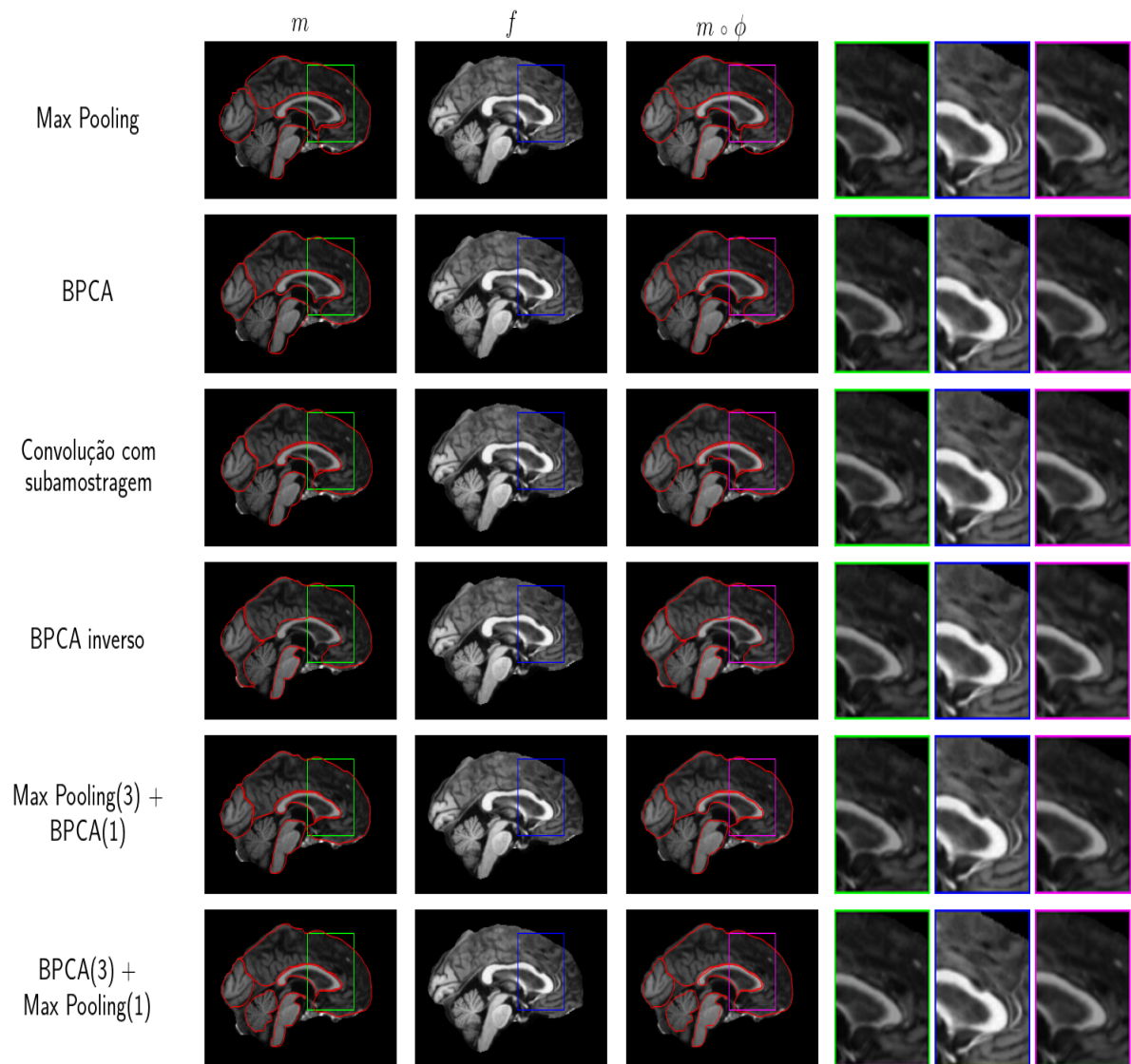
Figura 30 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do OASIS em uma visão sagital.



Fonte: autor

Na Figura 30, as diferenças entre as imagens em movimento m e as imagens registradas $m \circ \phi$ não são tão explícitas. No entanto, ao sobrepor m e $m \circ \phi$, torna-se aparente que as estruturas nas imagens $m \circ \phi$ se expandiram e que houve um clareamento em algumas regiões das imagens, destacadas com marcações em vermelho na Figura 31. Observa-se nos resultados do experimentos, que não foi gerado nenhum artefato e que visualmente os resultados são bem similares, confirmando os resultados da Tabela 3 e 4. Os campos de deformação ϕ obtidos nos experimentos são regulares/suaves, ou seja, as transições entre os deslocamentos em diferentes partes da imagem são graduais e contínuas, em vez de abruptas e irregulares como é o caso do BPCA inverso na linha 4.

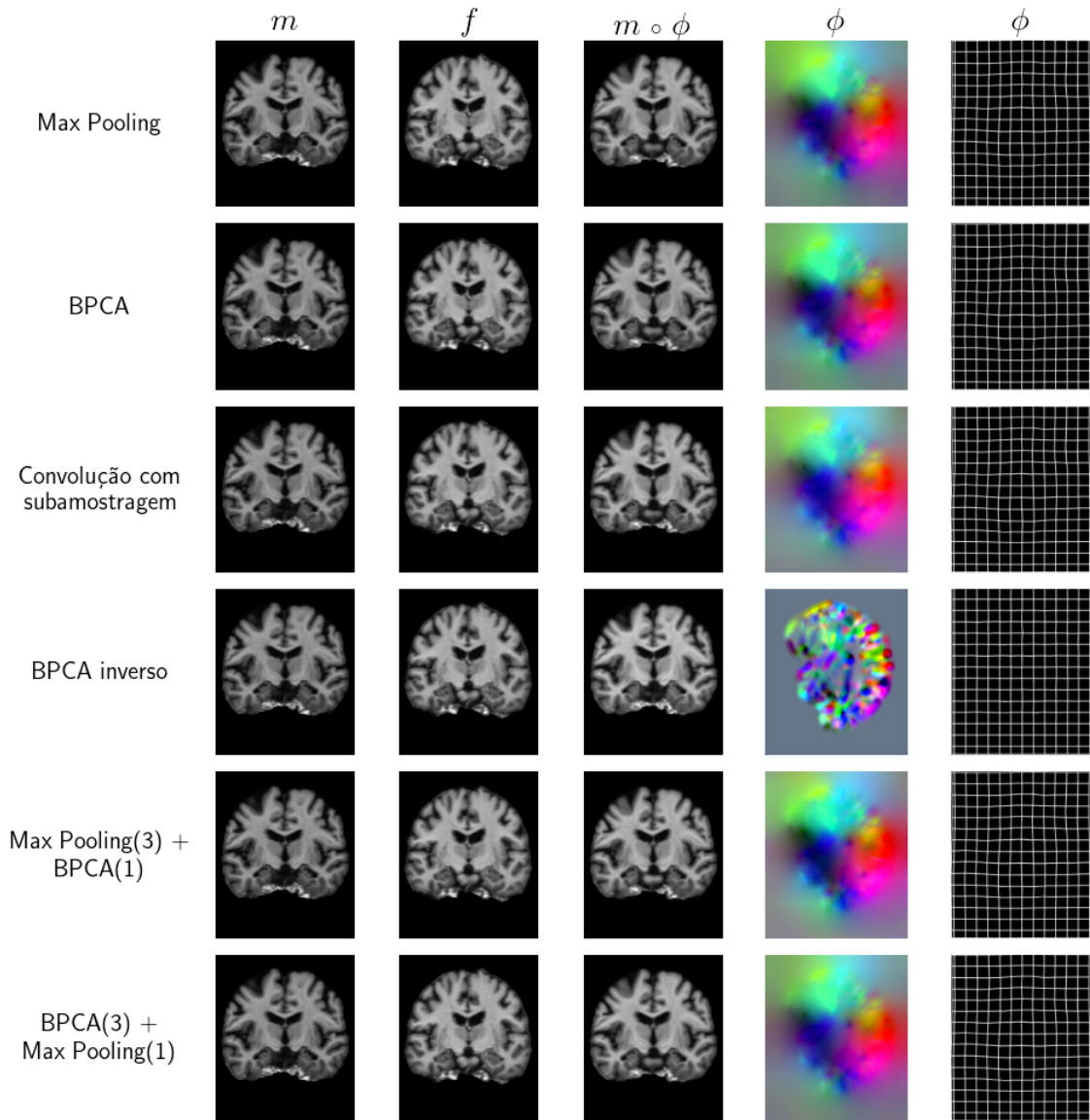
Figura 31 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do OASIS em visão sagital.



Fonte: autor

A Figura 32 mostra duas fatias cerebrais coronal de imagens de ressonância magnética extraídas de dois escaneamentos diferentes do conjunto de dados de teste do OASIS.

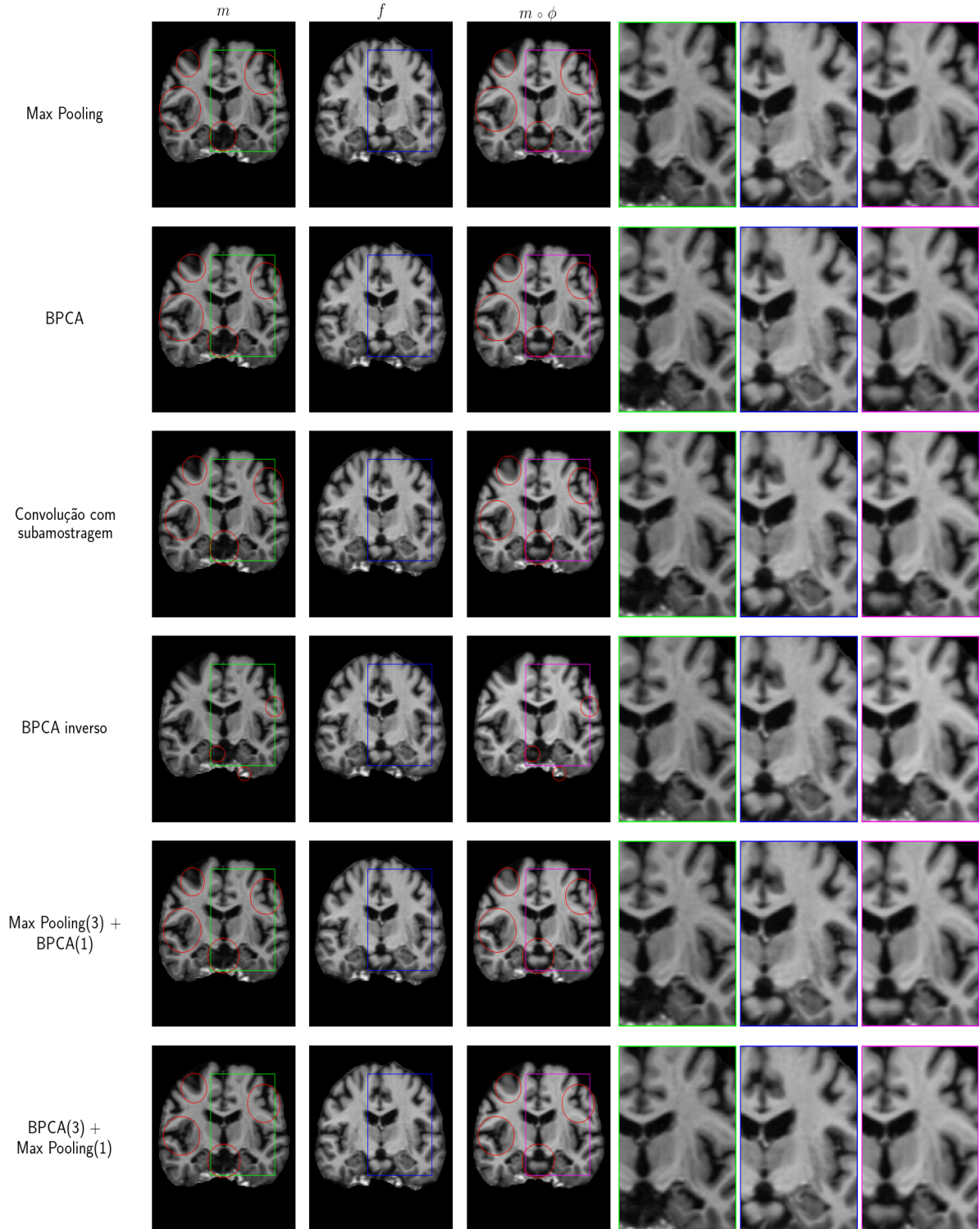
Figura 32 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do OASIS em uma visão coronal.



Fonte: autor

Os resultados das imagens em movimento m e dos registros $m \circ \phi$ na Figura 32 revelam uma diferença notável na região inferior central entre essas imagens, onde foram incorporadas estruturas da imagem fixa f . Essas diferenças estão realçadas com marcações em vermelho na Figura 33. Os registros obtidos não tiveram nenhum artefato e são visualmente parecidos, confirmando os resultados da Tabela 3 e 4. Somente o experimento na linha 4 com BPCA inverso que o campo de deformação ϕ apresentou uma deslocação irregular em cada dimensão espacial.

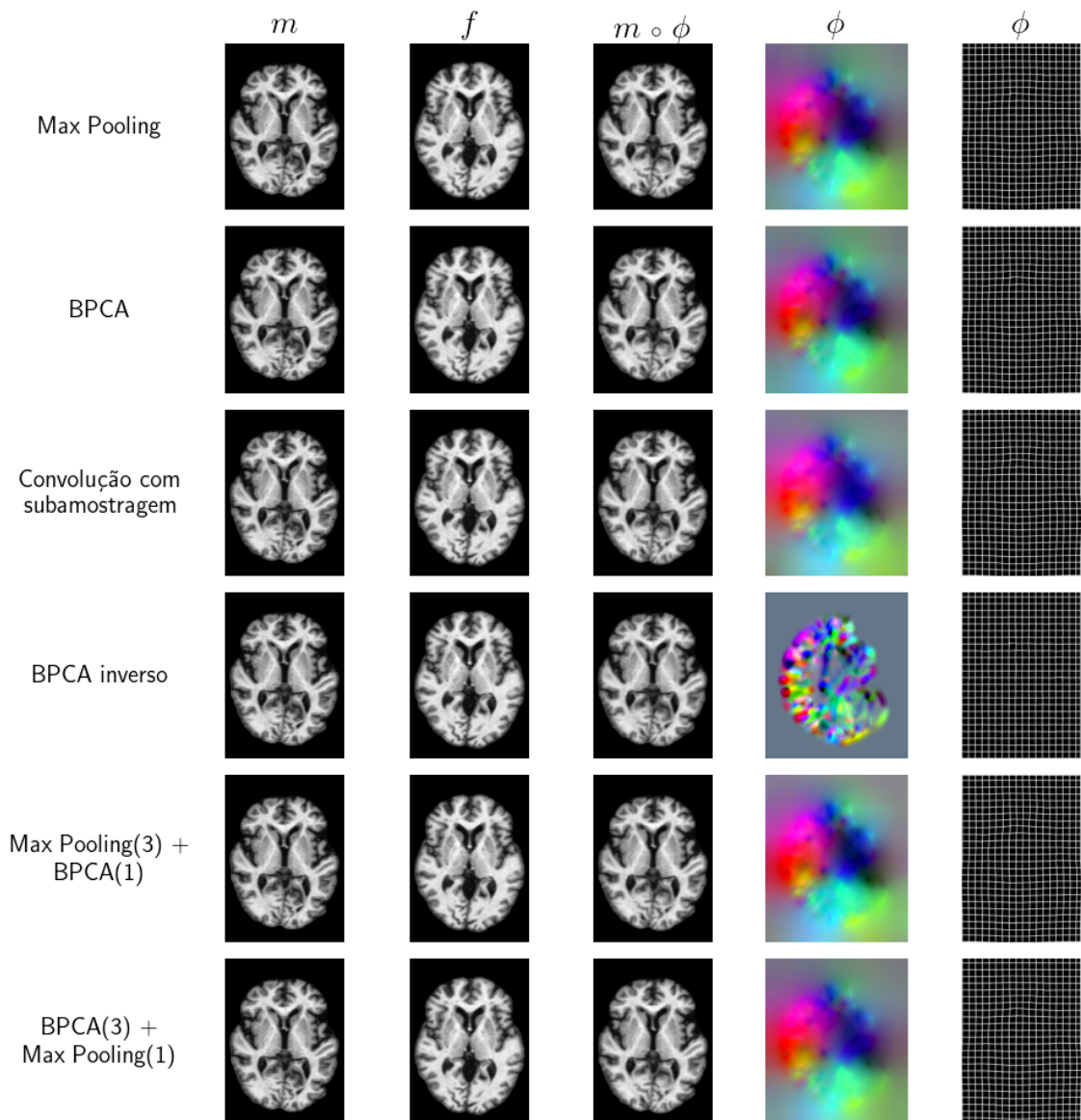
Figura 33 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do OASIS em visão coronal.



Fonte: autor

A Figura 34 mostra duas fatias cerebrais axiais de imagens de ressonância magnética extraídas de dois escaneamentos diferentes do conjunto de dados de teste do OASIS.

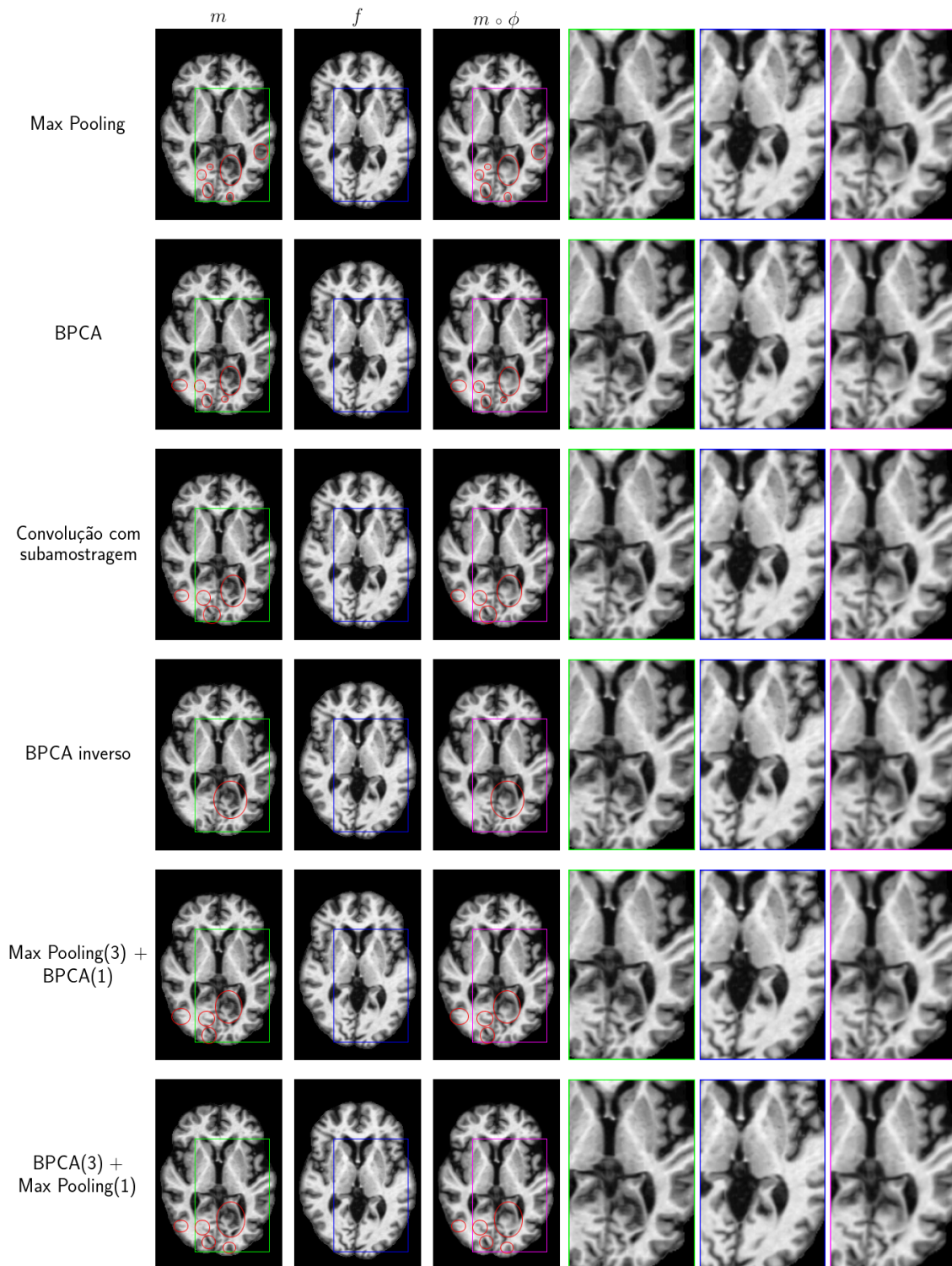
Figura 34 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do OASIS em uma visão axial.



Fonte: autor

Na Figura 34, as imagens em movimento m e as imagens fixas f possuem estruturas parecidas. Desta forma, os resultados das imagens registradas $m \circ \phi$ não apresentam diferenças tão visíveis das imagens em movimento. Porém, ao sobrepor m com $m \circ \phi$, nota-se que as estruturas de $m \circ \phi$ se comprimiram. A Figura 35 destaca as diferenças das imagens com marcações em vermelho. Os resultados $m \circ \phi$ não apresentaram nenhum artefato e são bem similares entre si o que confirma os resultados da Tabela 3 e 4. Os campos de deformação ϕ obtidos com o BPCA inverso na linha 4 mais uma vez demonstraram uma deslocação irregular nas dimensões espaciais.

Figura 35 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do OASIS em visão axial.



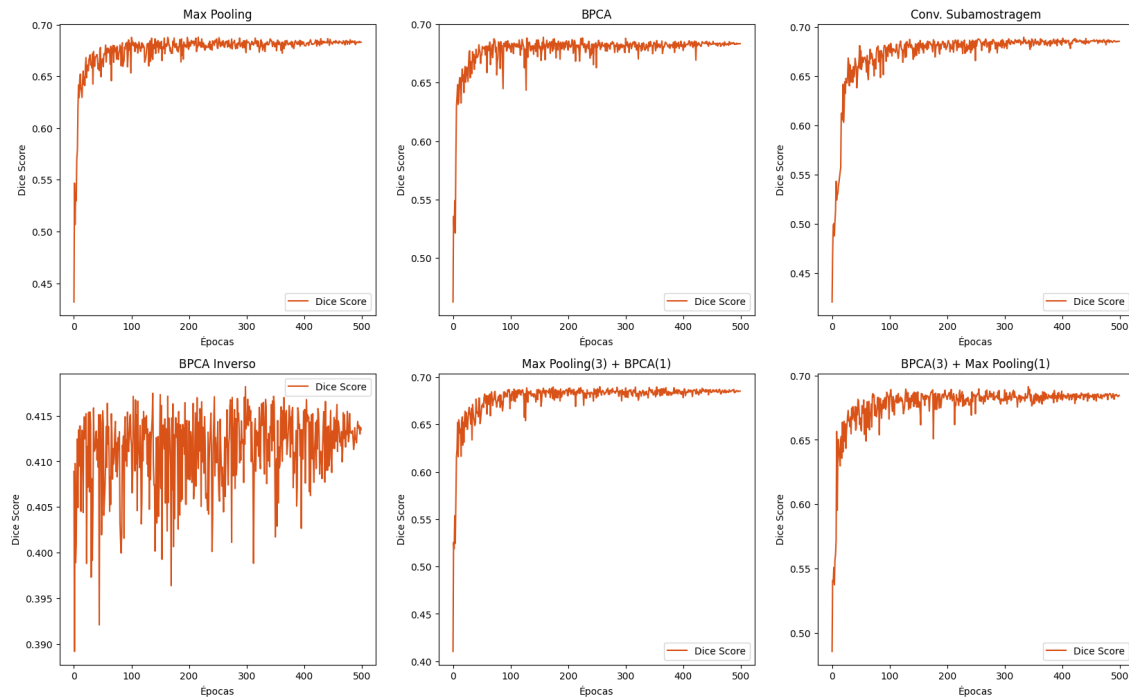
Fonte: autor

Na próxima seção é apresentado os resultados obtidos dos experimentos realizados no conjunto de dados IXI.

5.2 Conjunto de dados IXI

Os gráficos apresentados na Figura 36 exibem a evolução da métrica *Dice Score* ao longo das 500 épocas de treinamento no conjunto de dados de validação IXI, abrangendo os diversos experimentos realizados.

Figura 36 – *Dice Score* obtido ao longo do treinamento no conjunto de validação do conjunto de dados IXI.



Fonte: autor

É possível observar na Figura 36 que os resultados obtidos começaram a convergir, com exceção do experimento com o BPCA inverso, que precisa de mais épocas de treinamento para começar a convergir. O maior *Dice Score* obtido foi de 0,6912 no experimento com três camadas iniciais de *pooling* utilizando o BPCA e a final o *Max Pooling* e o menor valor alcançado foi de 0,4181 no experimento com o BPCA inverso.

Na Tabela 5, encontram-se os tempos de treinamento de cada experimento, expressos em segundos e em horas.

Tabela 5 – Tempo de treinamento de cada experimento no conjunto de dados IXI, expressos em segundos e em horas. O número destacado em negrito indica o melhor tempo.

Método de <i>Pooling</i>	Tempo em segundos	Tempo em horas
<i>Max Pooling</i>	69492	19,303
BPCA	92294	25,637
Convolução com subamostragem	60164	16,712
BPCA inverso	86155	23,932
<i>Max Pooling</i> (3) + BPCA(1)	70077	19,466
BPCA(3) + <i>Max Pooling</i> (1)	92618	25,727

Fonte: autor

Na Tabela 6 são apresentados os resultados dos experimentos conduzidos no conjunto de dados de teste do IXI. Cada linha na apresentação dos resultados corresponde a um experimento específico, identificado pela primeira coluna, enquanto as colunas subsequentes indicam as métricas empregadas para quantificar numericamente a eficácia de cada experimento realizado.

Tabela 6 – Resultados da média do *Dice Score*, a porcentagem de $|J_\phi| \leq 0$ e o tempo médio gasto para realizar o registro na GPU em segundos, referentes a cada experimento realizado. As melhores classificações estão destacadas em negrito.

Método de <i>Pooling</i>	<i>Dice Score</i>	% de $ J_\phi \leq 0$	GPU (seg)
<i>Max Pooling</i>	0,521 $\pm$ 0,028	0	0,067
BPCA	0,523 $\pm$ 0,028	<0,0001	0,132
Convolução com subamostragem	0,522 $\pm$ 0,029	0	0,046
BPCA inverso	0,299 $\pm$ 0,027	0	0,140
<i>Max Pooling</i> (3) + BPCA(1)	0,523 $\pm$ 0,028	<0,0001	0,067
BPCA(3) + <i>Max Pooling</i> (1)	0,524 $\pm$ 0,028	0	0,134

Fonte: autor

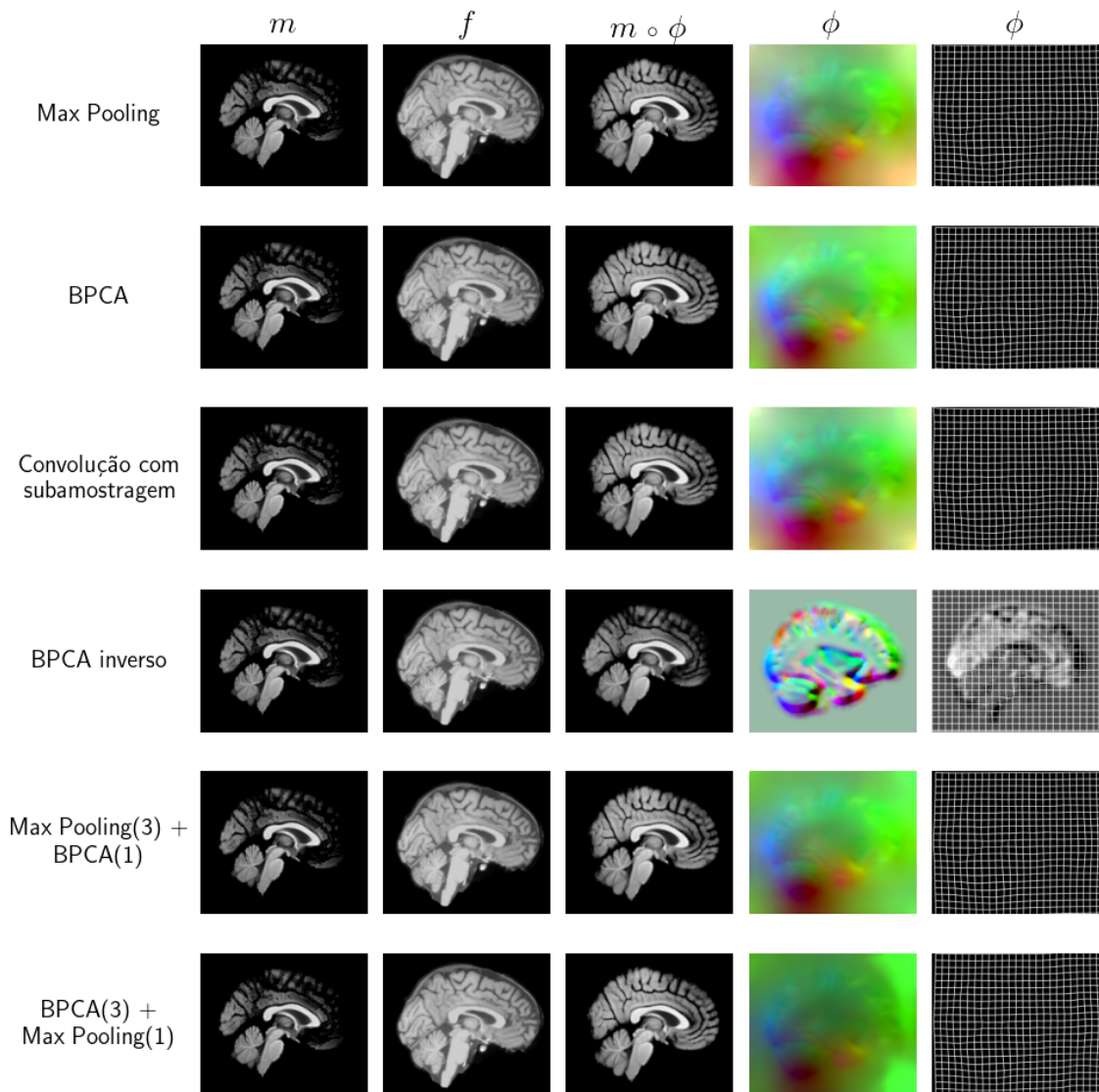
Na Tabela 6 observa-se que, os resultados obtidos com os diferentes experimentos apresentam valores médio de *Dice Score* próximos. O maior valor médio de *Dice Score*, destacado em negrito, foi obtido no registro que utilizou a mistura de *pooling* entre BPCA, nas três camadas iniciais, e o método *Max Pooling* na camada de *pooling* final. O melhor tempo médio de registro, também destacado em negrito, foi obtido no experimento sem a utilização de camada de *pooling*, ou seja, realizando a redução da entrada no caminho de codificação por meio da convolução com subamostragem.

Foi realizado o teste T, com as amostras de *Dice Score*, com um nível de significância de 0,05. Os testes realizados são entre a amostra de *Dice Score* do experimento utilizando a abordagem de três primeiras camadas de *pooling* usando o BPCA e a última usando o *Max Pooling*, que foi o que teve melhor resultado, com a amostra de *Dice Score* dos outros experimentos. Com exceção da abordagem do BPCA inverso, os resultados dos demais testes indicaram falha em rejeitar a hipótese nula. Isso sugere que, em termos estatísticos, não houve diferenças significativas nas performances observadas, exceto pela abordagem do BPCA inverso.

Nas Figuras 37, 39 e 41 os símbolos m , f e $m \circ \phi$ denotam a imagem móvel, a imagem fixa ou imagem de referência e a imagem resultante após o registro, respectivamente. É crucial enfatizar que um registro adequado resultará em uma imagem $m \circ \phi$ em que as estruturas se assemelham às da imagem fixa f . A coluna 4, identificada como ϕ , demonstra o campo de deformação gerado pela rede, onde o deslocamento em cada dimensão espacial x , y e z é mapeado para cada um dos canais de cor RGB. Na coluna 5, identificada como ϕ , é apresentado o resultado do campo de deformação aplicado em um *grid*. Em cada linha está o resultado dos experimentos de registros utilizando *Max Pooling*, BPCA, convolução com subamostragem, BPCA inverso, três camadas de *Max Pooling* e uma de BPCA e três camadas de BPCA e uma de *Max Pooling* respectivamente.

A Figura 37 mostra duas fatias cerebrais sagitais de imagens de ressonância magnética extraídas de dois escaneamentos diferentes do conjunto de dados de teste do IXI.

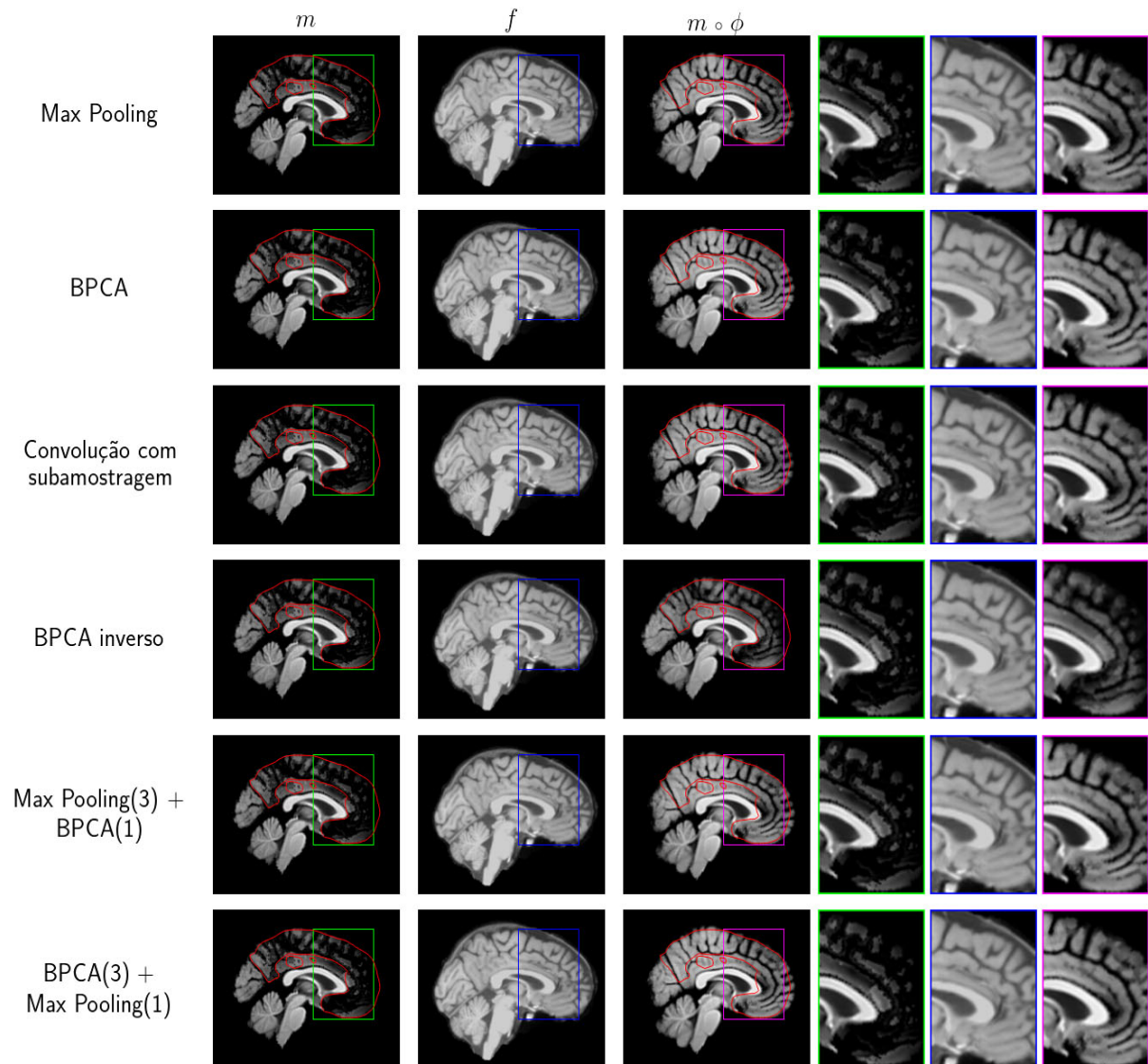
Figura 37 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do IXI em uma visão sagital.



Fonte: autor

É possível notar na Figura 37 que os resultados das imagens registradas $m \circ \phi$ possuem diferenças notáveis das imagens em movimento m , realçados com marcações em vermelho na Figura 38, como por exemplo os tons de cinzas se tornaram mais claros. Além disso, os resultados obtidos não apresentam artefatos de distorções que comprometem os resultados dos registro e parecem visualmente similares, confirmando os resultados próximos obtidos na Tabela 6. Os campos de deformações produzidos pelos experimentos mostram uma deslocação suave nas dimensões espaciais, com exceção do experimento do BPCA inverso na linha 4, que é bem mais irregular que os outros.

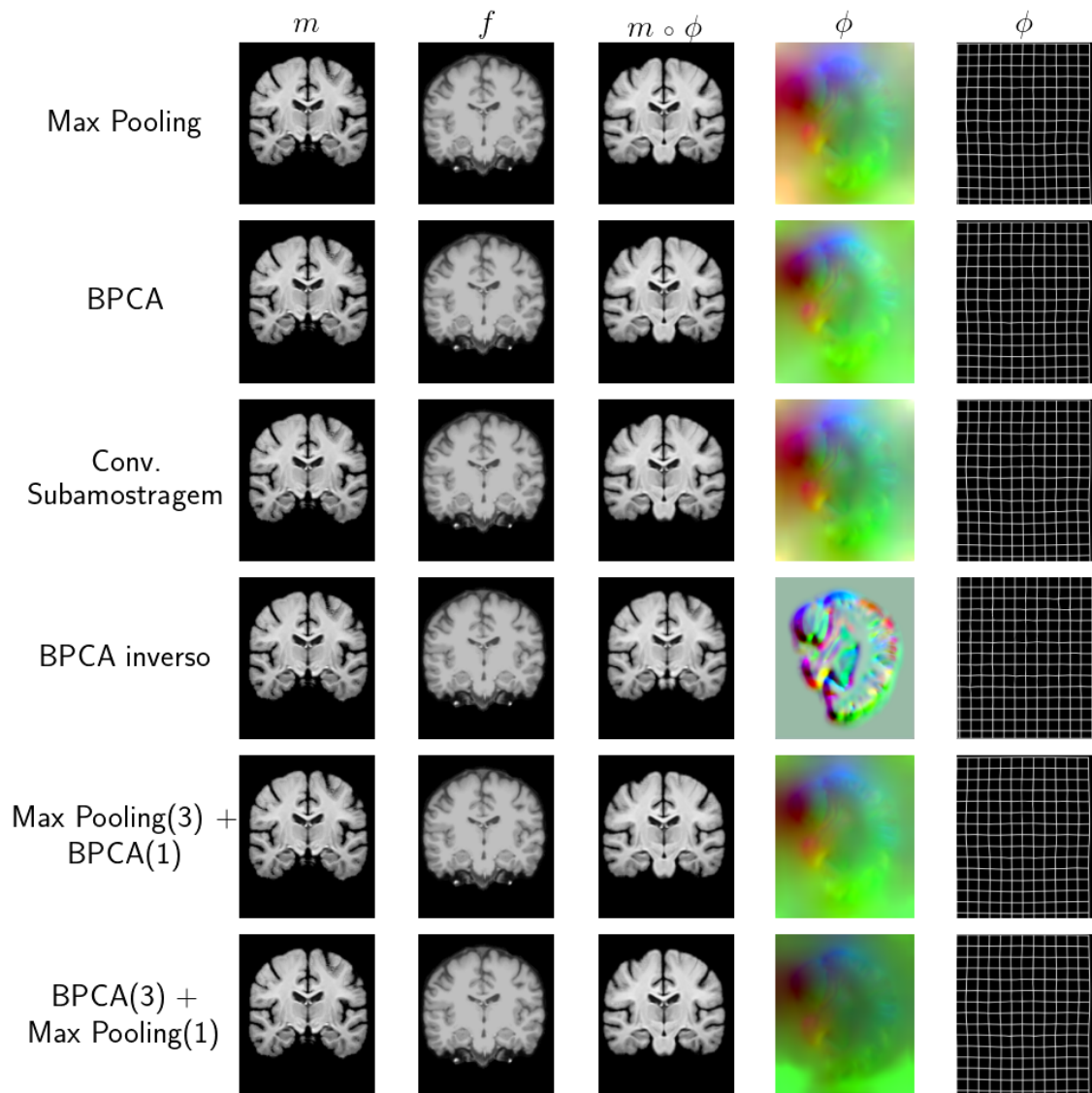
Figura 38 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do IXI em visão sagital.



Fonte: autor

A Figura 39 mostra duas fatias cerebrais coronais de imagens de ressonância magnética extraídas de dois escaneamentos diferentes do conjunto de dados de teste do IXI.

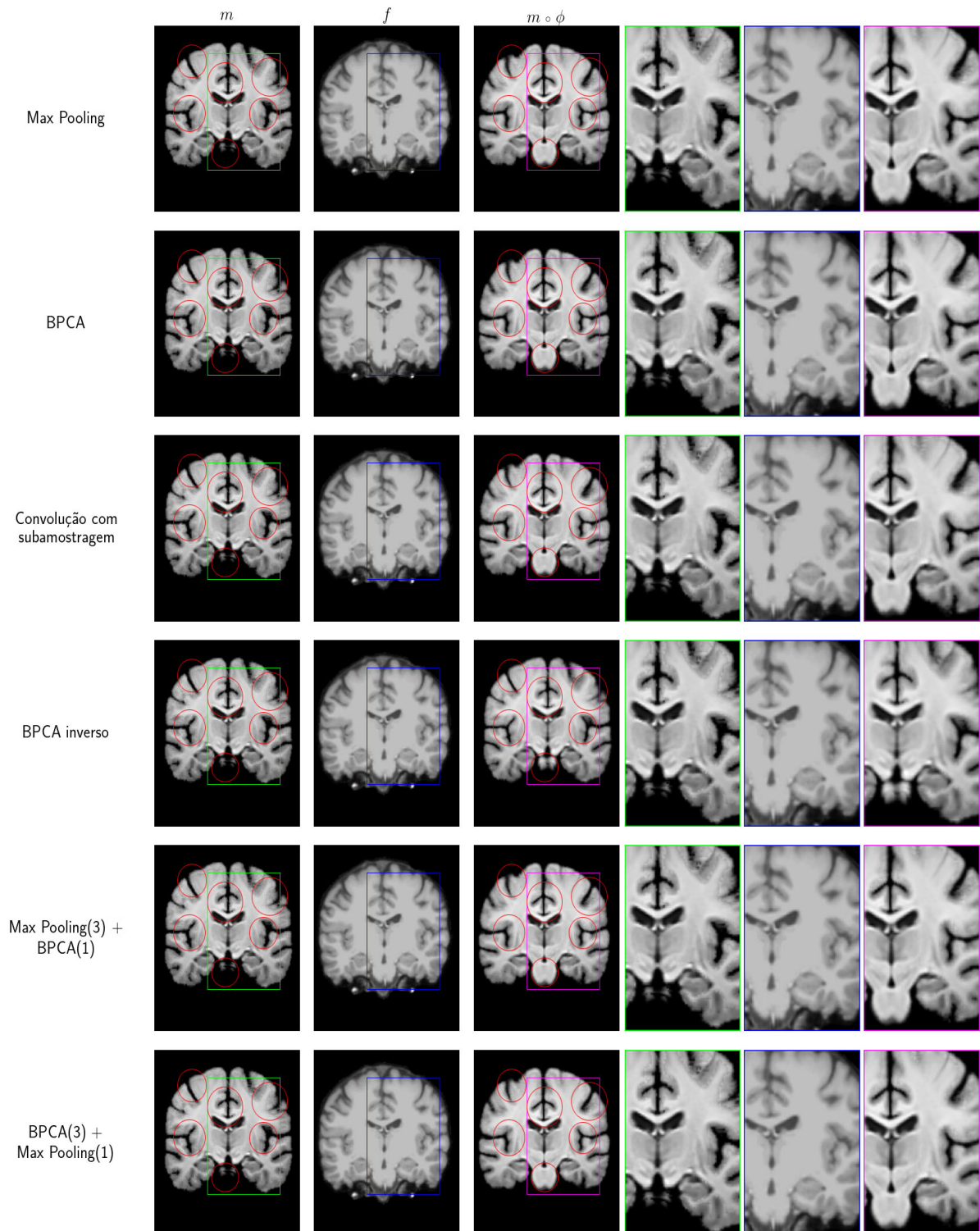
Figura 39 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do IXI em uma visão coronal.



Fonte: autor

É evidente na Figura 39 que os resultados das imagens registradas $m \circ \phi$ possuem diferenças com as imagens em movimento m , destacados com marcações em vermelho na Figura 40, como por exemplo a parte inferior das fatias terem aumentado de tamanho, com exceção do experimento do BPCA inverso na linha 4, que não teve um aumento significativo. É possível notar que os resultados não apresentaram artefatos que prejudicaram os registros mesmo com as estruturas da imagem em movimento e fixa sendo extremamente diferentes. À exceção do experimento com o BPCA inverso, os resultados apresentam uma semelhança visual, confirmando a proximidade com os resultados obtidos na Tabela 6. É possível observar que as deslocamentos do campo de deformação nas dimensões espaciais gerado pelo BPCA inverso são irregulares, enquanto dos outros experimentos são regulares.

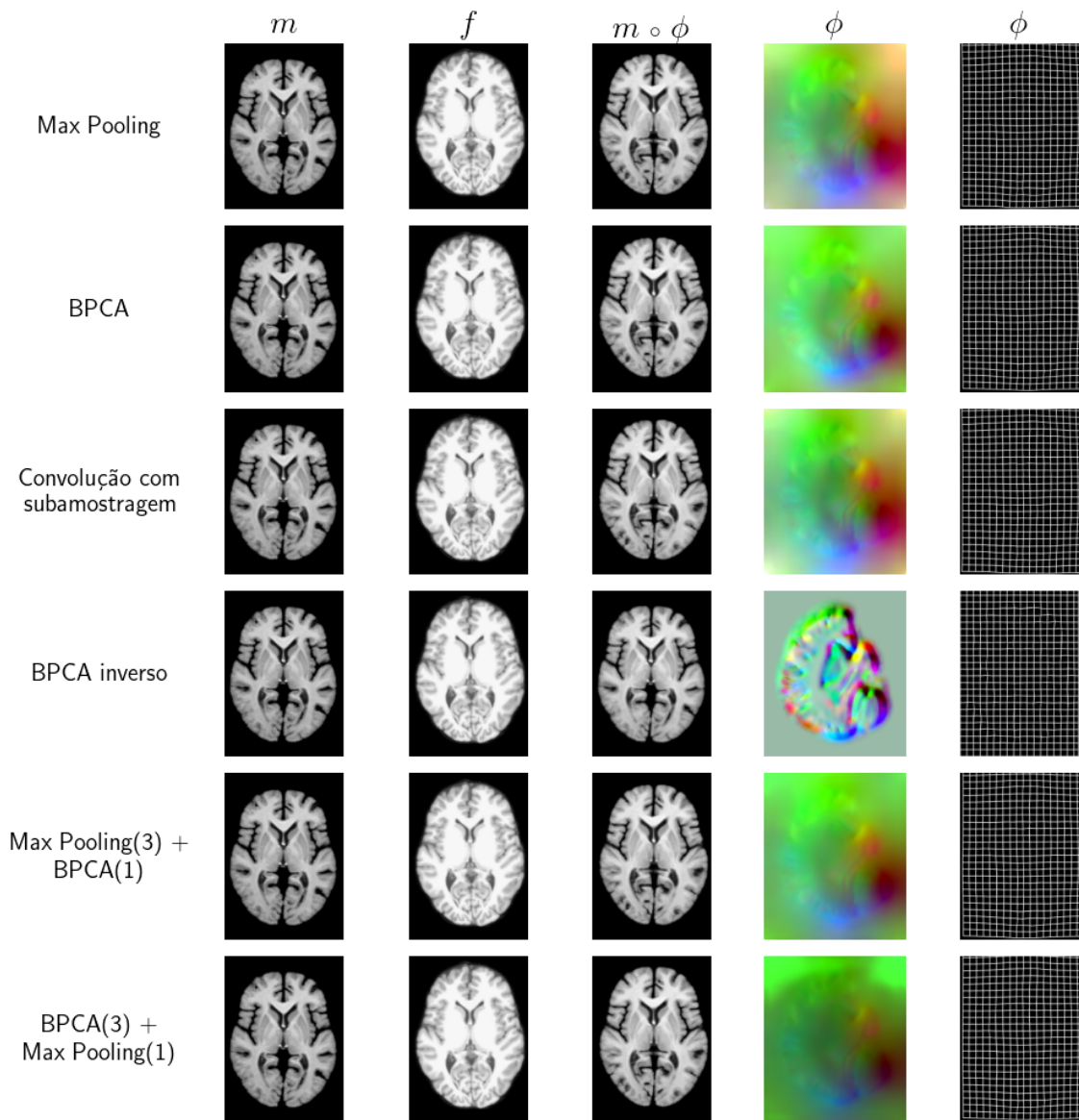
Figura 40 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do IXI em visão coronal.



Fonte: autor

A Figura 41 mostra duas fatias cerebrais axiais de imagens de ressonância magnética extraídas de dois escaneamentos diferentes do conjunto de dados de teste do IXI.

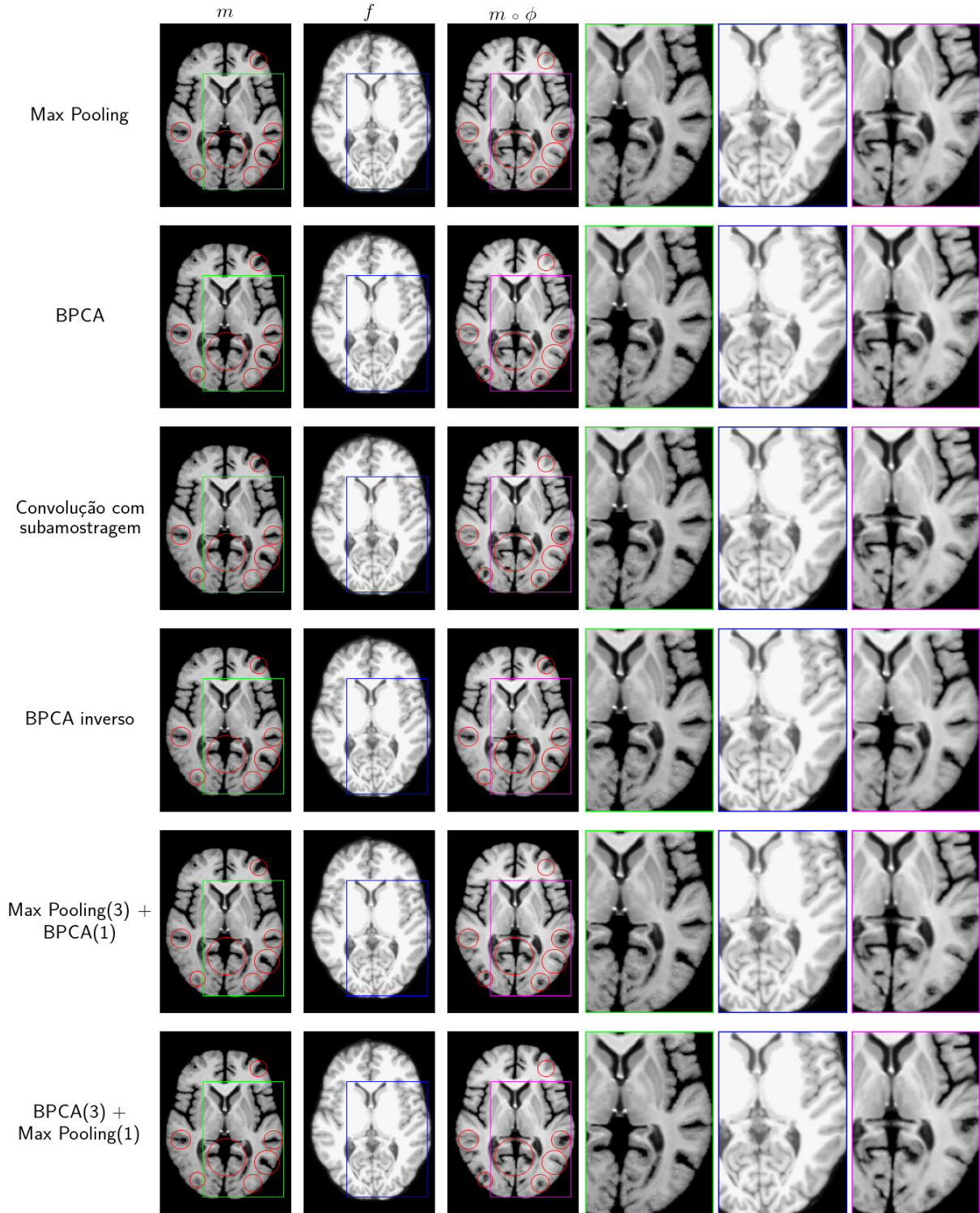
Figura 41 – Resultados dos registros obtidos nos experimentos no conjunto de dados de teste do IXI em uma visão axial.



Fonte: autor

É possível notar na Figura 41 que os resultados das imagens registradas $m \circ \phi$ não apresentam diferenças tão explícitas, devido à imagem em movimento m apresentar estruturas bastante semelhantes à imagem fixa f . Também é notável que não foi gerado nenhum artefato nos resultados dos registros e que esses resultados são bem similares visualmente confirmando novamente a proximidade com os resultados obtidos na Tabela 6. Os campos de deformações produzidos pelos experimentos mostram uma deslocação regular nas dimensões espaciais, com exceção do BPCA inverso que gerou um campo de deformação irregular.

Figura 42 – Regiões de interesse ampliadas nos retângulos verde, azul e rosa, nas últimas colunas. Marcações em vermelho destacam diferenças entre imagens nos experimentos do conjunto de dados do IXI em visão axial.



Fonte: autor

Na próxima seção, apresentaremos as considerações finais deste capítulo, discutindo os resultados apresentados até o momento.

5.3 Considerações Finais do Capítulo

Neste capítulo, apresentaram-se os resultados obtidos nos experimentos propostos. Os resultados indicam que as combinações de *poolings*, onde o BPCA foi empregado nas três camadas iniciais e o *Max Pooling* na camada final, alcançaram os melhores desempenhos na métrica *Dice Score* em ambos conjuntos de dados. Apesar desta abordagem ter tido o melhor desempenho, o teste T realizado no conjunto de dados de teste do IXI mostrou que em termos estatísticos não houve diferenças significativas nos resultados de *Dice Score* em comparação com as outras abordagens.

Os treinamentos que utilizaram a técnica de *pooling* BPCA foram notavelmente mais prolongados, devido à complexidade das operações estatísticas necessárias comparadas com a simples extração dos maiores números de uma matriz (*Max Pooling*). Além disso, é fundamental destacar que a técnica de *Max Pooling* é altamente otimizada, ao passo que a técnica com o BPCA pode ser aprimorada.

Os resultados indicam baixas porcentagens de valores não positivos no determinante da matriz Jacobiana nos campos de deformação ($|J_\phi| \leq 0$) e $SD\log J$. Isso sugere que os campos de deformação gerados são difeomórficos, ou seja, preservam características anatômicas e evitam distorções indesejáveis.

Os resultados da métrica *Dice Score* no conjunto de validação do conjunto de dados OASIS não convergiram, indicando a possibilidade de melhoria de desempenho ao aumentar o número de épocas durante o treinamento. Os resultados no conjunto de dados IXI apresentaram sinais de convergência, sugerindo que para aprimorar o desempenho seria necessário ajustar os parâmetros da rede, como, por exemplo, aumentar o tamanho da mesma.

Os resultados das imagens obtidas indicam a ausência de artefatos que pudessem impactar de forma negativa os registros. Adicionalmente, observou-se que as estruturas dos registros passaram por modificações visuais, com o intuito de se assemelharem à imagem de referência, sem comprometer a integridade global da imagem. Os campos de deformação, com exceção dos registros com o BPCA inverso, apresentaram suavidade, preservando assim as características anatômicas sem apresentar distorções indesejáveis.

No capítulo seguinte, serão discutidas as considerações finais deste trabalho, destacando os principais entendimentos alcançados ao longo da pesquisa. Além disso, serão apresentadas possíveis direções para futuras investigações que possam contribuir para o progresso do campo de estudo.

6 Conclusões

O registro não-rígido de imagens é uma tarefa desafiadora, tanto em termos computacionais quanto de precisão, especialmente quando estamos em um contexto de imagens médicas, cuja fidelidade à realidade é fundamental para um diagnóstico preciso. Por exemplo, esta tarefa não pode acabar gerando artefatos. Além disso, como não há uma verdade absoluta disponível para a deformação desejada, torna o registro não-rígido de imagens um problema de registro mal-condicionado, isto é, um problema que não tem uma solução única.

Como uma das abordagens modernas para o registro de imagem não-rígido, o uso de aprendizado profundo tem se despontado como a principal técnica, com grande potencial de melhorar a qualidade de registros. Porém, o registro com aprendizado profundo na área médica sofre com a falta de conjuntos de dados para treinamento. Além disso, o registro não-rígido ainda é uma tarefa desafiadora, mesmo utilizando o aprendizado profundo.

Por sua vez, algumas técnicas existentes, tal como o *VoxelMorph* são baseadas na definição básica de uma rede U-Net, que embutem etapas de compressão e expansão/reconstrução que não consideram fatores importantes nestas, especialmente a consideração de toda a imagem para a compressão, enquanto preservam a relação espacial. Deste modo, foi proposto utilizar o BPCA que é uma técnica que leva em consideração esses fatores, com o objetivo de verificar se uma representação mais adequada pode ser obtida e que beneficie o processo de registro de imagens.

Com base nos resultados obtidos nos conjuntos de dados OASIS e IXI, podemos concluir que a combinação de *poolings*, utilizando o BPCA nas três primeiras camadas e o *Max Pooling* na camada final, resultou em desempenho ligeiramente superior em comparação com outras técnicas de *pooling*. Isso sugere, possivelmente, que o BPCA é mais eficaz ao lidar com as representações mais finas da entrada. Nota-se que tanto o BPCA como os métodos misturados demonstraram a capacidade de produzir mudanças visuais significativas nos registros. Esses resultados apontam a viabilidade do BPCA como uma alternativa para o registro de imagens médicas, oferecendo uma abordagem diferente e promissora. Entretanto, são necessárias investigações e comparações adicionais para uma avaliação abrangente da eficácia do BPCA como camada de *pooling* no contexto do registro não-rígido de imagens médicas.

Uma das limitações do trabalho é em relação ao tempo mais prolongado de treinamento quando utilizada a técnica de *pooling* BPCA. Outra limitação é em relação aos experimentos com o BPCA serem conduzidos utilizando apenas um componente principal. Além disso, devido ao objetivo do trabalho ser comparar diferentes métodos de *pooling*, não houve preocupação em buscar hiperparâmetros diferentes a fim de obter um desempenho melhor no registro.

Em futuras pesquisas, destaca-se a importância de avaliar o impacto do aumento do número de componentes principais no método de *pooling* BPCA. Além disso, sugerem-se análises que envolvam a modificação da ordem das misturas de *pooling* entre o BPCA e o *Max Pooling*. Por exemplo, a configuração de duas camadas iniciais utilizando o BPCA, seguidas pelas finais empregando o *Max Pooling*, e vice-versa. Adicionalmente, explorar as variações alternadas entre BPCA e *Max Pooling* ao longo das camadas do método. Outra abordagem importante consiste em conduzir experimentos similares, porém incorporando a mistura do BPCA com convoluções com subamostragem. Explorar diferentes hiperparâmetros e aprimorar a implementação do BPCA são abordagens importantes a serem consideradas em trabalhos futuros. Também é importante investigar a aplicação do BPCA de forma individual para cada fatia do volume 3D, isto é, gerar os blocos e aplicar o PCA em cada fatia individualmente.

Para trabalhos futuros, considerando que o BPCA demonstrou resultados superiores em relação ao *Max Pooling* e a combinação dessas técnicas de *pooling* obteve desempenho superior em comparação com outros experimentos, é válido explorar a aplicação do BPCA como técnica de *pooling* em outras tarefas de processamento de imagens, tais como segmentação, classificação, filtragem de ruído e fusão de imagens. Ainda, utilizar outras métricas, como MSE, SSIM e GC, como aplicadas no trabalho Benvenuto et al. (2022), para avaliar de forma quantitativa os resultados do registro pode proporcionar uma análise mais abrangente da eficácia do BPCA.

Assim ao propor, a utilização de uma técnica estatística como camada de *pooling* que apresenta um balanço entre representação comprimida de maior capacidade e manutenção da informação/relação espacial da imagem em comparação com o *Max Pooling*, este estudo investigativo e exploratório trouxe uma nova linha de pensamento, especialmente no sentido de unir arquiteturas de redes neurais para aprendizado profundo e técnicas tradicionais de aprendizado de máquina e reconhecimento de padrões estatístico.

Referências

- ANDRADE, N. *Desenvolvimento de Métricas de Similaridade em Registro de Imagens Médicas Fundamentada em Redes Neurais Profundas*. 150 p. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal de São Paulo, São José dos Campos, SP, 2020.
- AVANTS, B.; EPSTEIN, C.; GROSSMAN, M.; GEE, J. Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain. *Medical Image Analysis*, v. 12, n. 1, p. 26–41, 2008.
- AVANTS, B. B.; TUSTISON, N. J.; SONG, G.; COOK, P. A.; KLEIN, A.; GEE, J. C. A Reproducible Evaluation of ANTs Similarity Metric Performance in Brain Image Registration. *NeuroImage*, v. 54, n. 3, p. 2033–2044, 2011.
- BALAKRISHNAN, G.; ZHAO, A.; SABUNCU, M. R.; DALCA, A. V.; GUTTAG, J. An Unsupervised Learning Model for Deformable Medical Image Registration. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 9252–9260.
- BALAKRISHNAN, G.; ZHAO, A.; SABUNCU, M. R.; GUTTAG, J.; DALCA, A. V. Voxelmorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, v. 38, n. 8, p. 1788–1800, 2019.
- BENVENUTO, G. A.; COLNAGO, M.; DIAS, M. A.; NEGRI, R. G.; SILVA, E. A.; CASACA, W. A Fully Unsupervised Deep Learning Framework for Non-Rigid Fundus Image Registration. *Bioengineering*, v. 9, n. 8, p. 17, 2022. ISSN 2306-5354.
- BIZOPOULOS, P.; KOUTSOURIS, D. Deep Learning in Cardiology. *IEEE Reviews in Biomedical Engineering*, IEEE, v. 12, p. 168–193, 2018.
- BOVEIRI, H. R.; KHAYAMI, R.; JAVIDAN, R.; MEHDIZADEH, A. Medical Image Registration Using Deep Neural Networks: A Comprehensive Review. *Computers & Electrical Engineering*, v. 87, p. 106767, 2020.
- BUDIHARTO, W.; GUNAWAN, A. A.; SUROSO, J. S.; CHOWANDA, A.; PATRIK, A.; UTAMA, G. Fast Object Detection for Quadcopter Drone Using Deep Learning . In: IEEE. *2018 3rd International Conference on Computer and Communication Systems (ICCCS)*. [S.l.], 2018. p. 192–195.
- CASADO, M. d. I. N. T. *Reflectance and Colorimetric Modelling of Multilayer Bodies of Translucent Dental Materials*. 189 p. Tese (Programa de Doctorado en Física y Ciencias del Espacio) — Universidad de Granada, Granada, 2023.
- CHEN, J.; FREY, E. C.; HE, Y.; SEGARS, W. P.; LI, Y.; DU, Y. Transmorph: Transformer for unsupervised medical image registration. *Medical Image Analysis*, p. 102615, 2022.
- CHEN, X.; DIAZ-PINTO, A.; RAVIKUMAR, N.; FRANGI, A. Deep Learning in Medical Image Registration. *Progress in Biomedical Engineering*, IOP Publishing, v. 3, n. 1, p. 012003, dec 2020.

CHO, K.; MERRIËNBOER, B. van; GULCEHRE, C.; BAHDANAU, D.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. Association for Computational Linguistics, Doha, Qatar, p. 1724–1734, out. 2014.

CRUM, W. R.; HARTKENS, T.; HILL, D. Non-rigid Image Registration: Theory and Practice. *The British journal of radiology*, British Institute of Radiology, v. 77, n. suppl_2, p. S140–S153, 2004.

CUNHA, G. B. d. *Redes Neurais Artificiais Aplicadas à Previsão da Incidência de Malária no Estado de Roraima*. 125 p. Tese (Doutorado em Engenharia Elétrica) — Universidade Federal de Uberlândia, Uberlândia, MG, 2010.

DAVID, O. E.; NETANYAHU, N. S. DeepPainter: Painter Classification Using Deep Convolutional Autoencoders. In: VILLA, A. E.; MASULLI, P.; RIVERO, A. J. P. (Ed.). *Artificial Neural Networks and Machine Learning – ICANN 2016*. Cham: Springer International Publishing, 2016. p. 20–28. ISBN 978-3-319-44781-0.

DICE, L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology*, JSTOR, v. 26, n. 3, p. 297–302, 1945.

FECHTER, T.; BALTAS, D. One-Shot Learning for Deformable Medical Image Registration and Periodic Motion Tracking. *IEEE Transactions on Medical Imaging*, v. 39, n. 7, p. 2506–2517, 2020.

FERREIRA, J. L. *Segmentação Automática da Próstata em Imagens de Ressonância Magnética Utilizando Redes Neurais Convolucionais, Mapa Probabilístico e Treinamento Adversário*. 67 p. Dissertação (Mestrado em Ciência da Computação) — Universidade Federal do Maranhão, São Luís, MA, 2019.

FORNEFETT, M.; ROHR, K.; STIEHL, H. Radial Basis Functions With Compact Support for Elastic Registration of Medical Images. *Image and Vision Computing*, v. 19, n. 1, p. 87–96, 2001.

FU, Y.; LEI, Y.; WANG, T.; CURRAN, W. J.; LIU, T.; YANG, X. Deep Learning in Medical Image Registration: A Review. *Physics in Medicine & Biology*, IOP Publishing, v. 65, n. 20, p. 20TR01, 2020.

FU, Y.; LEI, Y.; WANG, T.; HIGGINS, K.; BRADLEY, J. D.; CURRAN, W. J.; LIU, T.; YANG, X. LungRegNet: An Unsupervised Deformable Image Registration Method for 4D-CT Lung. *Medical Physics*, Wiley Online Library, v. 47, n. 4, p. 1763–1774, 2020.

GARDNER, M. W.; DORLING, S. Artificial Neural Networks (The Multilayer Perceptron)—A Review of Applications in the Atmospheric Sciences. *Atmospheric environment*, Elsevier, v. 32, n. 14-15, p. 2627–2636, 1998.

GERBRANDS, J. J. On the relationships between svd, klt and pca. *Pattern Recognition*, v. 14, n. 1, p. 375–381, 1981.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016.

GOODFELLOW, I.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, v. 27, 2014.

GOSHTASBY, A. A. *Theory and Applications of Image Registration*. [S.l.]: John Wiley & Sons, 2017.

HAFEMANN, L. G. *An Analysis of Deep Neural Networks for Texture Classification*. 89 p. Dissertação (Mestrado em Informática) — Universidade Federal do Paraná, Curitiba, PR, 2014.

HAO, Z. Deep Learning Review and Discussion of its Future Development. *MATEC Web of Conferences*, v. 277, p. 02035, 2019.

HASKINS, G.; KRUGER, U.; YAN, P. Deep Learning in Medical Image Registration: A Survey. *Machine Vision and Applications*, Springer, v. 31, n. 1, p. 1–18, 2020.

HAYKIN, S. *Redes Neurais: Princípios e Prática*. [S.l.]: Bookman Editora, 2007.

HERING, A.; HANSEN, L.; MOK, T. C. W.; CHUNG, A. C. S.; SIEBERT, H.; HÄGER, S.; LANGE, A.; KUCKERTZ, S.; HELDMANN, S.; SHAO, W.; VESAL, S.; RUSU, M.; SONN, G.; ESTIENNE, T.; VAKALOPOULOU, M.; HAN, L.; HUANG, Y.; YAP, P.-T.; BRUDFORS, M.; BALBASTRE, Y.; JOUTARD, S.; MODAT, M.; LIFSHITZ, G.; RAVIV, D.; LV, J.; LI, Q.; JAOUEN, V.; VISVIKIS, D.; FOURCADE, C.; RUBEAUX, M.; PAN, W.; XU, Z.; JIAN, B.; BENETTI, F. D.; WODZINSKI, M.; GUNNARSSON, N.; SJÖLUND, J.; GRZECH, D.; QIU, H.; LI, Z.; THORLEY, A.; DUAN, J.; GROßBRÖHMER, C.; HOOPES, A.; REINERTSEN, I.; XIAO, Y.; LANDMAN, B.; HUO, Y.; MURPHY, K.; LESSMANN, N.; GINNEKEN, B. van; DALCA, A. V.; HEINRICH, M. P. Learn2reg: Comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. *IEEE Transactions on Medical Imaging*, v. 42, n. 3, p. 697–712, 2023.

HOU, J.; GUERRERO, M.; CHEN, W.; D'SOUZA, W. D. Deformable Planning CT to Cone-Beam CT Image Registration in Head-And-Neck Cancer. *Medical Physics*, Wiley Online Library, v. 38, n. 4, p. 2088–2094, 2011.

HU, R.; WANG, H.; RISTANIEMI, T.; ZHU, W.; SUN, X. Lung CT Image Registration through Landmark-constrained Learning with Convolutional Neural Network. In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. [S.l.: s.n.], 2020. p. 1368–1371.

HU, Y.; MODAT, M.; GIBSON, E.; LI, W.; GHAVAMI, N.; BONMATI, E.; WANG, G.; BANDULA, S.; MOORE, C. M.; EMBERTON, M.; OURSELIN, S.; NOBLE, J. A.; BARRATT, D. C.; VERCAUTEREN, T. Weakly-Supervised Convolutional Neural Networks for Multimodal Image Registration. *Medical Image Analysis*, v. 49, p. 1–13, 2018.

JADERBERG, M.; SIMONYAN, K.; ZISSERMAN, A.; KAVUKCUOGLU, K. Spatial Transformer Networks. In: CORTES, C.; LAWRENCE, N.; LEE, D.; SUGIYAMA, M.; GARNETT, R. (Ed.). *Advances in Neural Information Processing Systems*. [S.l.]: Curran Associates, Inc., 2015. v. 28, p. 9.

JOSHI, R. R. *An Analysis of Rigid Image Alignment Computer Vision Algorithms*. 129 p. Dissertação (Mestrado em Ciência da Computação) — Georgia Southern University, Georgia, United States, 2012.

KER, J.; WANG, L.; RAO, J.; LIM, T. Deep Learning Applications in Medical Image Analysis. *IEEE Access*, IEEE, v. 6, p. 9375–9389, 2017.

- KLEIN, S.; STARING, M.; MURPHY, K.; VIERGEVER, M. A.; PLUIM, J. P. W. Elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, v. 29, n. 1, p. 196–205, 2010.
- KRISTIADI, A.; HEIN, M.; HENNIG, P. Being Bayesian, Even Just a Bit, Fixes Overconfidence in ReLU Networks. In: *Proceedings of the 37th International Conference on Machine Learning*. [S.l.: s.n.], 2020. p. 5436–5446.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In: PEREIRA, F.; BURGESS, C.; BOTTOU, L.; WEINBERGER, K. (Ed.). *Advances in Neural Information Processing Systems*. [S.l.]: Curran Associates, Inc., 2012. v. 25, p. 9.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. *Nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, v. 86, n. 11, p. 2278–2324, 1998.
- LECUN, Y.; KAVUKCUOGLU, K.; FARABET, C. Convolutional Networks and Applications in Vision. In: IEEE. *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. [S.l.], 2010. p. 253–256.
- LIN, G.; SHEN, W. Research on Convolutional Neural Network Based on Improved Relu Piecewise Activation Function. *Procedia Computer Science*, Elsevier, v. 131, p. 977–984, 2018.
- LITJENS, G.; KOOI, T.; BEJNORDI, B. E.; SETIO, A. A. A.; CIOMPI, F.; GHAFORIAN, M.; LAAK, J. A. V. D.; GINNEKEN, B. V.; SÁNCHEZ, C. I. A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, Elsevier, v. 42, p. 60–88, 2017.
- LIU, X.; GAO, K.; LIU, B.; PAN, C.; LIANG, K.; YAN, L.; MA, J.; HE, F.; ZHANG, S.; PAN, S. et al. Advances in Deep Learning-Based Medical Image Analysis. *Health Data Science*, Science Partner Journal, v. 2021, p. 8786793, Jun 2021.
- LONG, J.; SHELHAMER, E.; DARRELL, T. Fully Convolutional Networks for Semantic Segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2015. p. 3431–3440.
- LOYOLA-GONZALEZ, O. Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses from A Practical Point of View. *IEEE Access*, IEEE, v. 7, p. 154096–154113, 2019.
- MAASS, W. Networks of Spiking Neurons: The Third Generation of Neural Network Models. *Neural Networks*, v. 10, n. 9, p. 1659–1671, 1997.
- MAHAPATRA, D.; ANTONY, B.; SEDAI, S.; GARNAVI, R. Deformable Medical Image Registration Using Generative Adversarial Networks. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. [S.l.: s.n.], 2018. p. 1449–1453.
- MAMOSHINA, P.; VIEIRA, A.; PUTIN, E.; ZHAVORONKOV, A. Applications of Deep Learning in Biomedicine. *Molecular Pharmaceutics*, ACS Publications, v. 13, n. 5, p. 1445–1454, 2016.
- MANSILLA, L.; MILONE, D. H.; FERRANTE, E. Learning deformable registration of medical images with anatomical constraints. *Neural Networks*, v. 124, p. 269–279, 2020.

- MARCON, M. *Novel Deep Learning Methods Applied for Unconstrained Visual Object Recognition and 3D Modeling from Point Clouds*. 125 p. Tese (Doutorado em Informática) — Universidade Federal do Paraná, Curitiba, PR, 2020.
- MARCUS, D.; WANG, T.; PARKER, J.; CSERNANSKY, J.; MORRIS, J.; BUCKNER, R. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, v. 19, n. 9, p. 1498–1507, set. 2007. ISSN 0898-929X.
- MARSTAL, K.; BERENDSEN, F.; STARING, M.; KLEIN, S. SimpleElastix: A User-Friendly, Multi-Lingual Library for Medical Image Registration. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. [S.l.: s.n.], 2016. p. 574–582.
- MCCULLOCH, W. S.; PITTS, W. A logical Calculus of the Ideas Immanent in Nervous Activity . *The Bulletin of Mathematical Biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943.
- MOK, T. C.; CHUNG, A. C. Fast symmetric diffeomorphic image registration with convolutional neural networks. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2020. p. 4643–4652.
- MONTGOMERY, D. C.; RUNGER, G. C. *Applied Statistics and Probability for Engineers*. 3. ed. [S.l.]: John Wiley & Sons, 2002.
- NIELSEN, L. K. *Elastic Registration of Medical MR Images*. 111 p. Dissertação (Mestrado em Ciência da Computação) — University of Bergen, Bergen, Norway, 2003.
- OH, S.; KIM, S. Deformable Image Registration in Radiation Therapy. *Radiation oncology journal*, The Korean Society for Radiation Oncology, v. 35, n. 2, p. 101–111, Jun 2017.
- OISHI, A.; YAGAWA, G. Computational Mechanics Enhanced by Deep Learning. *Computer Methods in Applied Mechanics and Engineering*, Elsevier, v. 327, p. 327–351, 2017.
- PATIL, D. D.; DEORE, S. G. Medical Image Segmentation: A Review. *International Journal of Computer Science and Mobile Computing*, v. 2, n. 1, p. 22–27, 2013.
- PAUL, L. C.; SUMAN, A. A.; SULTAN, N. Methodological analysis of principal component analysis (pca) method. *International Journal of Computational Engineering & Management*, Citeseer, v. 16, n. 2, p. 32–38, 2013.
- PEARSON, K. Liiv. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, Taylor & Francis, v. 2, n. 11, p. 559–572, 1901.
- REDDY, G. P.; KUMAR, Y. V. P. Explainable AI (XAI): Explained. In: *2023 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)*. [S.l.: s.n.], 2023. p. 1–6.
- RODRIGUES, E. P. *Avaliação de Métricas para o Corregistro Não Rígido de Imagens Médicas*. 105 p. Tese (Doutorado em Física Aplicada à Medicina e Biologia) — Universidade de São Paulo, Ribeirão Preto, SP, 2010.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer International Publishing, 2015. p. 234–241. ISBN 978-3-319-24574-4.

ROSENBLATT, F. The perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, American Psychological Association, v. 65, n. 6, p. 386, 1958.

ROSENBLATT, F. Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms. *Cornell Aeronautical Lab Inc Buffalo NY*, 1961.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning Representations by Back-Propagating Errors. *Nature*, v. 323, n. 6088, p. 533–536, Oct 1986. ISSN 1476-4687.

SALVADEO, D. H.; MASCARENHAS, N. D.; MOREIRA, J.; LEVADA, A. L.; CORRÊA, D. C. Improving Face Recognition Performance Using RBPCA MaxLike and Information Fusion. *Computing in Science & Engineering*, v. 13, n. 3, p. 14–21, 2011.

SANCHES, M. K. *Aprendizado de Máquina Semi-Supervisionado: Proposta de um Algoritmo para Rotular Exemplos A Partir de Poucos Exemplos Rotulados*. 142 p. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) — Universidade de São Paulo, São Carlos, SP, 2003.

SANTOS, V. R. dos; PORSANI, J. L.; HIRATA, N. S. Automatic Classification of Metallic Targets Using Pattern Recognition of GPR Reflection: A Study in the IAG-USP Test Site, Sao Paulo (Brazil). In: IEEE. *Proceedings of the XIII Internacional Conference on Ground Penetrating Radar*. [S.l.], 2010. p. 1–4.

SILVA, I. N. D.; SPATTI, D. H.; FLAUZINO, R. A. Redes Neurais Artificiais para Engenharia e Ciências Aplicadas-Curso Prático. *São Paulo: Artliber*, 2010.

SKEIKA, E. L. *Utilização de Redes Neurais Completamente Convolucionais para Identificação e Medição de Crânios Fetais*. 101 p. Dissertação (Mestrado em Ciência da Computação) — Universidade Tecnológica Federal do Paraná, Ponta Grossa, PR, 2019.

SMITH, L. I. A Tutorial on Principal Components Analysis. *Department of Computer Science, University of Otago, New Zealand*, 2002.

SUZUKI, K. Overview of Deep Learning in Medical Imaging. *Radiological physics and technology*, Springer, v. 10, n. 3, p. 257–273, 2017.

TAVARES, J. A. *Deep-Learning Segmentation for Skin Lesion Images*. 58 p. Dissertação (Mestrado em Engenharia Elétrica) — Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, SP, 2018.

VOS, B. D. de; BERENDSEN, F. F.; VIERGEVER, M. A.; STARING, M.; IŞGUM, I. End-to-end Unsupervised Deformable Image Registration with A Convolutional Neural Network. In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. [S.l.]: Springer, 2017. p. 204–212.

VOS, B. D. de; BERENDSEN, F. F.; VIERGEVER, M. A.; SOKOOTI, H.; STARING, M.; IŞGUM, I. A Deep Learning Framework for Unsupervised Affine and Deformable Image Registration. *Medical Image Analysis*, v. 52, p. 128–143, 2019.

WRANGSJÖ, A.; PETTERSSON, J.; KNUTSSON, H. Non-rigid Registration Using Morphons. In: KALVIAINEN, H.; PARKKINEN, J.; KAARNA, A. (Ed.). *Image Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. p. 501–510. ISBN 978-3-540-31566-7.

YANI, M. et al. Application of Transfer Learning Using Convolutional Neural Network Method for Early Detection of Terry's Nail. In: IOP PUBLISHING. *Journal of Physics: Conference Series*. [S.l.], 2019. v. 1201, n. 1, p. 012052.

ZEILER, M. D.; FERGUS, R. Visualizing and Understanding Convolutional Networks. In: FLEET, D.; PAJDLA, T.; SCHIELE, B.; TUYTELAARS, T. (Ed.). *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014. p. 818–833. ISBN 978-3-319-10590-1.