POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Deep Learning-Based Multimodal Registration of Rodent Brain MRI and Histology after Traumatic Brain Injury

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Author: **Erica Bistacchia**

Student ID: 10867264
Advisor: Prof. Giacomo Boracchi
Co-advisors: Dr. Elisa R. Zanier, Dr. Federico Moro, Marcello De Salvo
Academic Year: 2024-25

Abstract

Rodent experiments on traumatic brain injury (TBI) induced by controlled cortical impact (CCI) enable detailed investigation of the mechanisms associated with disease progression via histological assessment at the microscopic level. At the same time, in vivo approaches such as magnetic resonance imaging (MRI) provide high translational value but lack the resolution of microscopy, and their underlying changes can be difficult to interpret. Aligning images of the same anatomical region across these modalities, facilitates the detection of structural damage and the understanding of microscopic alterations that contribute to macroscopic changes after injury. This task is challenging due to complex image deformations, substantial inter-modality differences and artifacts introduced during tissue preparation for histology, all of which limit the accuracy of traditional registration methods. Furthermore, most existing MRI-histology pipelines, originally developed for human prostate studies, are not directly transferable to rodent TBI studies due to anatomical differences. To address these challenges, this thesis presents an unsupervised MRI–histology registration pipeline that employs a CycleGAN for image-to-image translation, converting histology into MRI-like images and reducing the task to a monomodal problem, followed by a two-stage deep learning framework combining affine and thin-plate spline (TPS) transformations for spatial alignment. The proposed pipeline achieves an average Dice Similarity Coefficient (DSC) greater than 0.93 and a Mean Error Distance (MED) $\approx 0.3\ mm$. Qualitative inspection highlights that performance strongly depends on the GAN's ability to preserve anatomical structures while generating realistic MRI contrasts. These results demonstrate that the proposed approach outperforms state-of-the-art methods, offering more accurate and robust MRI-histology alignment in rodent TBI. This work has the potential to streamline experimental workflows, improve cross-modal interpretation and enhance the translational relevance of preclinical TBI studies.

Keywords: Image Registration, Deep learning, Histopathology, Magnetic Resonance, Traumatic Brain Injury.

Abstract in lingua italiana

Gli esperimenti sui roditori con trauma cranico indotto da impatto corticale controllato permettono un'analisi dettagliata dei meccanismi associati alla progressione della malattia tramite valutazioni istologiche a livello microscopico. Al contempo, approcci in vivo come la risonanza magnetica offrono alto valore traslazionale, pur con risoluzione inferiore, e i cambiamenti sottostanti possono risultare difficili da interpretare. L'allineamento di immagini della stessa regione anatomica, acquisite con modalità diverse, facilita l'identificazione di danni strutturali e la comprensione delle alterazioni microscopiche che contribuiscono ai cambiamenti macroscopici dopo il trauma. Il compito è complesso a causa di deformazioni complesse delle immagini, differenze sostanziali tra le modalità e artefatti introdotti durante la preparazione dei tessuti per l'istologia. Tali fattori riducono l'accuratezza dei metodi tradizionali, mentre gran parte delle pipeline esistenti per l'allineamento risonanza magnetica–istologia, sviluppate per studi sulla prostata umana, non sono direttamente trasferibili agli studi di trauma cranico su roditori a causa delle differenze anatomiche. Per affrontare queste sfide, questa tesi propone una pipeline non supervisionata per la registrazione risonanza magnetica–istologia, sfruttando una traduzione da immagine a immagine che trasforma il problema multimodale in uno monomodale. Una CycleGAN viene impiegata per tradurre le immagini istologiche in rappresentazioni risonanza magnetica. Successivamente, l'allineamento spaziale viene eseguito tramite un'architettura a due stadi. Il modello raggiunge un Dice score medio superiore a 0.93 e un errore medio di distanza $\approx 0.3\ mm$. Una valutazione qualitativa evidenzia come le prestazioni dipendano fortemente dalla capacità della GAN di preservare le strutture anatomiche generando al contempo contrasti realistici. I risultati dimostrano che questo approccio supera i metodi esistenti, offrendo un allineamento risonanza magnetica–istologia più accurato e robusto nel caso di roditori con trauma cranico. Questo lavoro ha il potenziale di ottimizzare i flussi sperimentali, migliorare l'interpretazione inter-modale e aumentare la rilevanza traslazionale degli studi preclinici sul trauma cranico.

Parole chiave: Registrazione di Immagini, Deep Learning, Istopatologia, Risonanza Magnetica, Trauma Cranico.

Contents

Introduction

Image registration is a fundamental technique in image processing, as it allows to integrate useful information from two or more images into a single, more informative representation. Its broad utility extends across various domains, including medical imaging, remote sensing and computer vision. In the medical field, there are multiple methods and modalities available to acquire images, each offering unique insights. As imaging technology progresses, the amount of information extracted from these images increases substantially. However, they are not necessarily redundant, since each imaging modality can provide different and complementary information. Thus, aligning two (or more) images depicting the same or corresponding anatomical structures, acquired at different time points, viewpoints, or with different imaging devices, becomes of key importance in this context.

Despite its critical role, medical image registration remains a challenging task due to several factors. First, images often undergo unknown geometric transformations, necessitating translation, rotation, scaling, or even complex non-linear warping to achieve proper alignment. Second, the histological process can introduce significant artifacts, including morphological changes, stain variations, tissue deformation such as shrinkage, stretching and tearing, and even tissue loss. Another major challenge arises when dealing with images from different modalities. In such cases, the appearance of corresponding features varies significantly, making registration particularly challenging. For example, Magnetic Resonance Imaging (MRI) highlights large-scale anatomical structures but lacks the cellular-level detail present in histopathological sections. This inherent appearance variability between modalities often precludes a general rule to establish robust structure correspondence.

To address these complexities, various approaches have been developed. Traditionally, methods can be broadly categorized into intensity-based and feature-based. Intensity-based methods directly relies on pixel intensity values to optimise a similarity metric that quantifies alignment accuracy between the images. While effective for monomodal registration and capable of high accuracy, they often struggle with multimodal registration due to disparate intensity profiles that may not reflect true anatomical correspondence and can be computationally intensive. Feature-based methods, conversely, consist in iden-

tifying and matching distinct features, such as point or contours, from both images based on which a geometric transformation is estimated. While robust to intensity variations, their performance is highly dependent on the reliable detection and accurate matching of salient and reproducible features, which is a challenging task in multimodal medical context.

In medical imaging, image registration is crucial for diagnosing and monitoring diseases by aligning multiple acquired images. For example, in oncology, registration between MRI and histopathology has been shown to improve cancer localization and classification in human prostate, enabling more accurate diagnosis and supporting treatment planning [1, 2]. Similarly, in the context of traumatic brain injury (TBI), where brain damage is a dynamic process evolving through cellular and molecular changes, registration facilitates the detection of structural damage, the assessment of changes over time, and the correlation of macroscopic abnormalities with microscopic alterations. For instance, the registration of high-resolution histopathological images with MRI scans is particularly valuable in TBI research. This allows for a detailed correlation between cellular-level damage, observable through histopathology (e.g., neuronal loss, axonal degeneration, and microglial activation), and larger-scale structural abnormalities detected by MRI (e.g., brain atrophy, haemorrhage, and white matter lesions) observed in vivo. By aligning these disparate modalities, researchers can gain a more comprehensive understanding of injury extent, validate imaging biomarkers, and ultimately improve diagnostic accuracy in TBI.

Despite the advancements in image registration techniques, significant challenges remain in complex scenarios such as TBI, where the brain lesion contribute to substantial artifacts in histopathological microscopy images. In this work, we propose a novel deep learning approach for the registration of MRI and histopathological sections of mouse brains with TBI. As a preliminary step, both histopathological and MRI images are segmented to isolate brain tissue and remove background regions, allowing the network to focus on the most relevant anatomical information. To address the challenges arising from the differences between imaging modalities, we employ a CycleGAN to transform histopathological images into MRI-like representations. Following domain translation, an unsupervised intensity-based network performs the registration in two stages: an affine transformation step followed by a deformable transformation step. The network is trained using the Mutual Information loss and estimates the geometric transformation parameters, which are subsequently applied to the original histopathological images.

The thesis is structured as explained hereafter. In Chapter 1 we present the mathematical formulation of the Image Registration task in a general setting. Then in Chapter 2, we

provide a fundamental and comprehensive overview of the medical aspects of TBI and the deep learning architectures employed in the proposed approach. Then, in Chapter 3 we present the latest advancements in state-of-the-art image registration techniques. Later we present the developed pipeline and the experimental evaluation in Chapter 4 and Chapter 5, respectively. Finally in Chapter 6, we comment the obtained results and discuss possible future improvements and developments of the conducted research.

1 | Problem Formulation

Image registration (IR), also known as image matching [3, 4], is the task of spatially aligning two (or more) images with the same or similar content, which may have been acquired at different time points, from different viewpoints, or using different imaging modalities, to enable accurate comparison, fusion, or joint analysis.

Let $I_M \in \mathbb{R}^{H \times W \times C}$ denote the *moving image* (also referred to as source or floating image) and $I_F \in \mathbb{R}^{H \times W \times C}$ the *fixed image* (also called target or reference image), respectively. Here, H and W are the spatial dimensions (height and width), and C is the number of channels (e.g., 1 for grayscale and 3 for RGB). We assume the moving and fixed images to have the same shape to simplify the formulation and to focus on the general case.

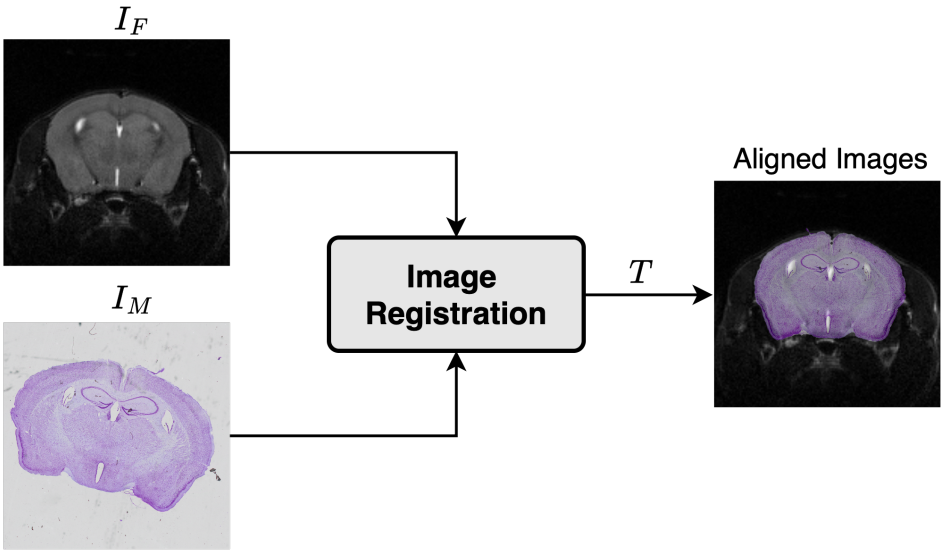


Figure 1.1: Image registration process on a sham mouse brain. The input consists of a histological section (I_M) and an MRI scan section (I_F). A spatial transformation T is estimated to align I_M to I_F, producing a fused image for joint analysis.

The objective of image registration is to estimate a geometric transformation T that maps each pixel location (i, j) to a new position (i', j'). Once the transformation is estimated,

the moving image is resampled according to T, using interpolation, yielding a registered version $I_M \circ T$, which ideally aligns with I_F. This process is illustrated in Figure 1.1.

Image registration is commonly formulated as a minimisation problem, seeking an optimal transformation T^*. This is expressed as:

$$T^* = \arg\min_{T \in \mathcal{T}} \mathcal{C}(T; I_M, I_F) \tag{1.1}$$

Here, $\mathcal{C}(T; I_M, I_F)$ represents a cost function that quantifies the degree of spatial disparity between the transformed moving image $I_M \circ T$ and the target image I_F, while set $\mathcal{T}$ denotes the space of all admissible transformations. The explicit definition of $\mathcal{C}$ is highly method-dependent (more detailed formulations of the IR problem are presented in Section 2.3.2).

In deep learning, *unsupervised* training is typically performed on a dataset of unaligned image pairs from different modalities, $\mathcal{D}_{unsup} = \{(I_M, I_F)_k\}_{k=1}^N$, where the goal is to estimate a transformation T minimising a cost function $\mathcal{C}$ without access to ground truth transformations. Conversely, *supervised* training relies on datasets of aligned image pairs $\mathcal{D}_{sup} = \{(I_M, I_F \circ T)_k\}_{k=1}^N$, with known ground truth transformations T.

2 | Theoretical Background

2.1. Medical Background

2.1.1. Traumatic Brain Injury

Traumatic brain injury is characterized by an alteration in brain function (such as a decreased level of consciousness, memory loss or neurological deficits), or by evidence of brain pathology (as identified through imaging techniques) resulting from an external mechanical force [5].

Among all common neurological disorders, TBI has the highest incidence and represents a major public health concern. Each year, it affects an estimated 50 to 60 million people worldwide, resulting in a global economic burden of approximately $400 billion. Critically, TBI is increasingly recognized not only as an acute event but as a chronic condition: patients often face long-term consequences, and the injury imposes enduring emotional, social, and caregiving challenges on both patients and their families [6].

Brain damage arising from TBI can be divided into primary injury, caused by the transmission of biomechanical energy to the brain, and secondary injury, which involves a series of cellular and molecular changes that are triggered immediately after the traumatic event and evolve over time in a continuum (Figure 2.1) [7].
The primary injury includes neuronal necrosis, blood vessel rupture, and diffuse axonal injury resulting from the rapid acceleration and deceleration forces involved. Secondary injury mechanisms occur in acute, subacute, and chronic phases and comprise a variety of cellular and molecular changes. These processes can progress over hours, days or even over a lifetime in some cases. A wide range of secondary neurotoxic mechanisms have been identified, including neuroinflammation, excitotoxicity, metabolic derangements, blood-brain barrier damage and cerebral edema. These events not only exacerbate the primary injury, but also contribute to the complexity and heterogeneity of TBI outcomes.
In addition to these pathological processes, the brain activates protective and regenerative responses such as angiogenesis, neurogenesis, gliogenesis, axonal sprouting, and synaptic

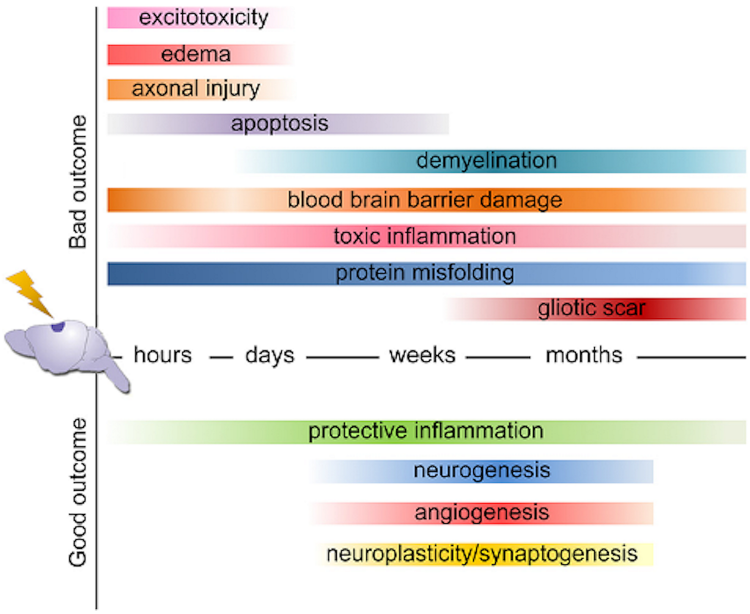


Figure 2.1: Toxic and protective events in TBI over time. Image source [7].

plasticity, which contribute to the recovery process (Figure 2.1).

Clinically, TBI is classified into mild, moderate and severe injury based on a combination of clinical criteria such as the Glasgow Coma Scale (GCS), duration of loss of consciousness (LOC), and neuroimaging findings [8]. However, patients with moderate-to-severe TBI represent a minority ($\sim$ 10%) of all cases of TBI presenting to hospital, while the remaining ($\sim$ 90%) are classified as mild TBI [6]. Despite being labelled as "mild", many of these injuries can still lead to significant brain damage, resulting in persistent symptoms and long-term cognitive or emotional deficits.

2.1.2. Controlled Cortical Impact injury model

Studying TBI directly in humans presents significant challenges due to the complexity, heterogeneity, and variability of injury mechanisms and patient's clinical history. Clinical studies are often limited by ethical constraints, and difficulties in accessing injured brain tissues. These factors hinder comprehensive research on pathological process progression and the efficacy of potential therapies.

Therefore, animal models play a crucial role in the study of TBI by providing controlled experimental settings to investigate the biomechanical, cellular, and molecular mechanisms involved. Despite their limitations, these models are essential to understand the underlying pathophysiology of TBI, particularly the progression of secondary injury mechanisms, and to develop potential therapeutic strategies that are otherwise difficult to assess di-

rectly in human studies [9]. Moreover, preclinical models offer the possibility to integrate longitudinal imaging and histopathological assessments, which is vital to bridge clinical observations with the biological mechanisms characteristic of TBI. Importantly, they also facilitate the discovery of new biomarkers, which is critical to patient stratification and to the improvement of clinical diagnosis and prognosis. Given the complexity and heterogeneity of TBI, various preclinical models have been developed to replicate specific pathological aspects of brain injury. While large animal models offer physiological similarities to humans, rodent (or murine) models are most used due to their lower cost, accessibility, and suitability for standardized experimental assessments.

One of the most widely adopted rodent models is the controlled cortical impact (CCI) model, which was used in this study. The CCI model consists of performing a craniotomy, a surgical procedure in which a small portion of the skull is removed to expose the brain surface. A piston, driven either by air or electromagnetic force, is then used to perform a controlled mechanical impact on the exposed cortical tissue, typically targeting the left parieto-temporal region. Injury parameters such as velocity, depth, and duration of impact are precisely controlled, allowing reproducible injuries with tuneable severity. After the impact, the skull opening is sealed and the scalp is sutured. Figure 2.2 illustrates the procedure.

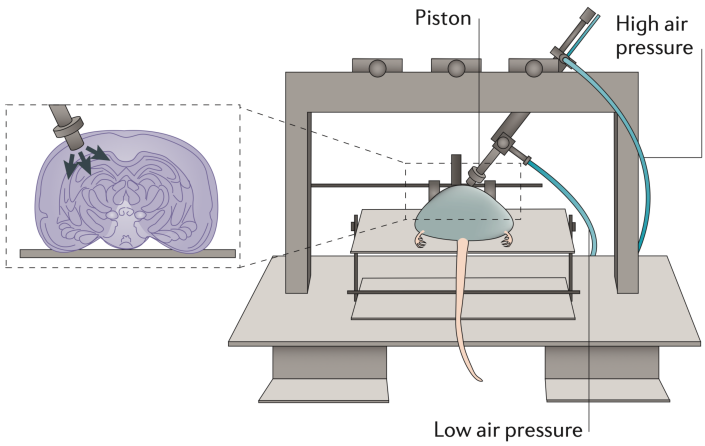


Figure 2.2: The CCI model uses an air- or electromagnetic-driven piston to deliver a mechanical impact to the brain at a known distance and velocity, inducing a reproducible and well-controlled injury. Image source [9].

2.1.3. Neuroimaging in preclinical TBI research

Neuroimaging plays a central role in both clinical and preclinical study of TBI.
In clinical practice, imaging techniques are indispensable for detecting and monitoring

structural abnormalities such as contusions, haemorrhages, cerebral edema, and brain atrophy. Early identification of such lesions is crucial for patient triage, prognosis, and treatment planning. The most used modalities are computed tomography (CT) and MRI. CT offers rapid acquisition and is highly sensitive to acute bleeding, making it the standard in emergency care. However, it has limited contrast resolution for soft tissues and often fails to detect subtle or diffuse injuries, particularly in cases of mild TBI. MRI, on the other hand, provides superior soft tissue contrast and a broader range of contrasts, allowing for the assessment of more subtle structural and physiological changes in the brain.

In the preclinical domain, neuroimaging techniques are equally critical, as in the animal models framework allows to monitor brain damage over time, observe the onset of pathological and recovery processes, and evaluate the efficacy of therapeutic interventions in a reproducible and translational context.

Within this framework, MRI and histological imaging are the most widely used and complementary techniques. MRI offers whole-brain, *in vivo*, non-invasive, non-ionising visualisation of anatomy and tissue alterations across time, while histology provides detailed *ex vivo* molecular and cellular information at a microscopic level. The integration of these two modalities is essential for a comprehensive understanding of TBI pathophysiology. However, MRI and histology operate at different spatial resolutions and physical principles, making direct comparison challenging. Histological sections can undergo significant deformation (such as shrinkage, folding, and tearing) during tissue processing, while MRI offers a more global but less detailed view. Aligning histological images with corresponding MRI scans allows researchers to correlate macroscopic anatomical findings with microscopic cellular changes, enabling more accurate localization and interpretation of pathological features. This multimodal integration enhances the biological understanding of disease progression and supports the development and validation of robust, translational imaging-based biomarkers with histopathological ground truth [10].

The following sections describe the principles and applications of MRI and histological imaging in preclinical TBI research.

Magnetic Resonance Imaging acquisition

MRI is a non-invasive imaging technique that exploits the physical principles of nuclear magnetic resonance (NMR) to produce detailed volumetric images of the brain (Figure 2.3).

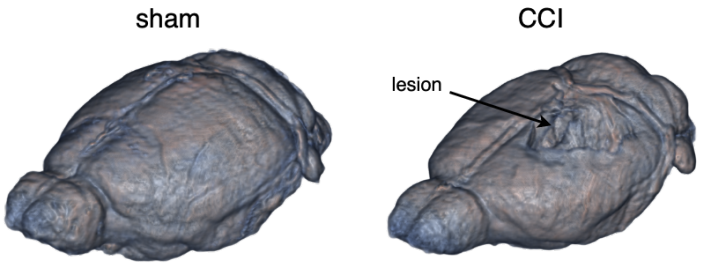


Figure 2.3: Brain volume reconstruction from MRI of sham and TBI mice.

Briefly, NMR exploits the magnetic properties of certain atomic nuclei, primarily hydrogen nuclei (protons), abundant in biological tissues such as water and fat. When placed in a strong external magnetic field, these nuclei align with the field. Applying radiofrequency (RF) pulses temporarily disturbs this alignment, and as the nuclei return to equilibrium (relaxation), they emit signals. These signals are detected and processed to create detailed images showing the location and properties of hydrogen-containing tissues. MRI contrast arises from differences in how hydrogen nuclei behave in various tissue environments. By tuning sequence parameters, specific tissue features can be emphasised. Principal MRI contrasts used in TBI studies include:

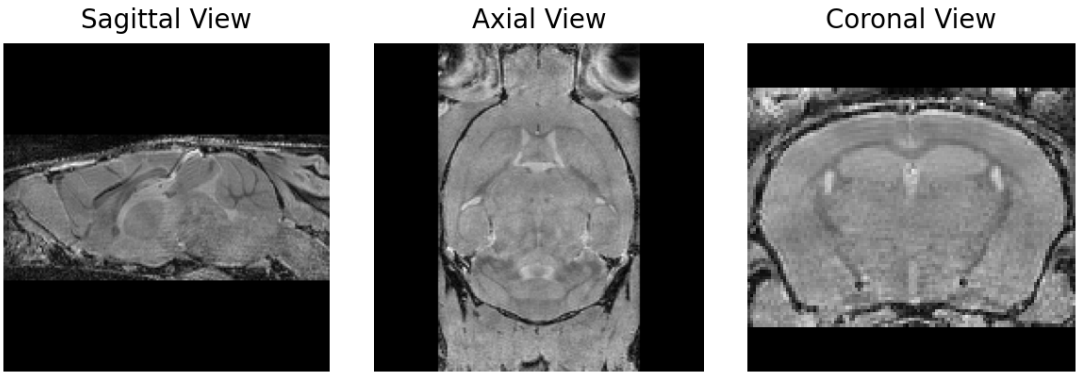
- *T1-weighted MRI*: generates images with contrast based on the relaxation properties of hydrogen atoms, where fat-rich tissues appear hyperintense and water-rich tissues relatively hypointense. T1-weighted images are critical for detecting gross structural changes such as tissue loss and haemorrhage.
- *T2-weighted MRI*: highlights differences in transverse relaxation times of hydrogen atoms, producing images in which regions with elevated water content, such as edema or inflammation, appear hyperintense. This contrast is particularly useful for detecting brain changes, such as swelling or tissue damage.

Figure 2.4 shows an example of T1-weighted MRI brain scans of sham and TBI mice.

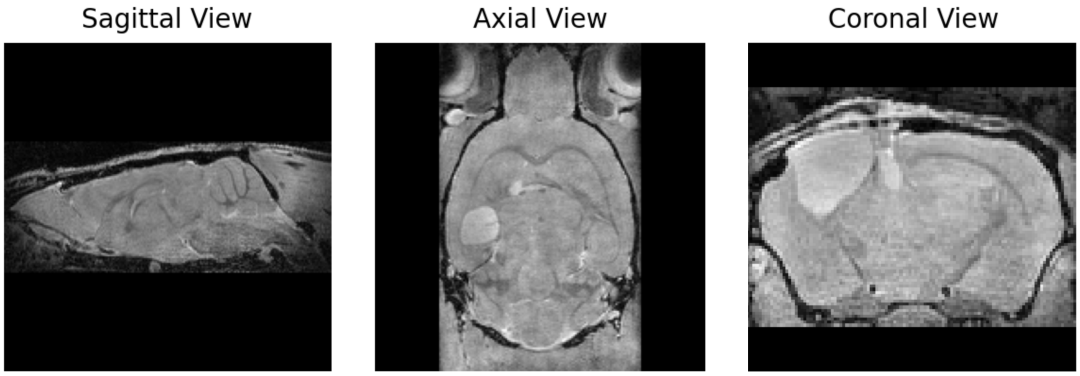
Histological Imaging

Histology is a microscopic imaging technique that provides insight into brain pathology by visualising tissue at the cellular level. In preclinical TBI research, histology is used to characterize pathological processes.

The process typically involves fixation of brain tissue, followed by sectioning into thin slices (e.g., 8–20 μm thick), staining, and high-resolution imaging under a microscope.



(a) Sham mice



(b) TBI mice

Figure 2.4: T1-weighted MRI of (a) sham and (b) TBI mouse brains. From left to right axial, sagittal, and coronal views are displayed.

Various stains and immunohistochemical markers are used to highlight distinct cellular components and pathological features (Figure 2.5):

- *Cresyl Violet* (a Nissl stain) highlights neuronal cell bodies and enables assessment of cell loss or structural damage. This stain helps reveal the extent of cortical and hippocampal neuronal damage often observed in TBI models.

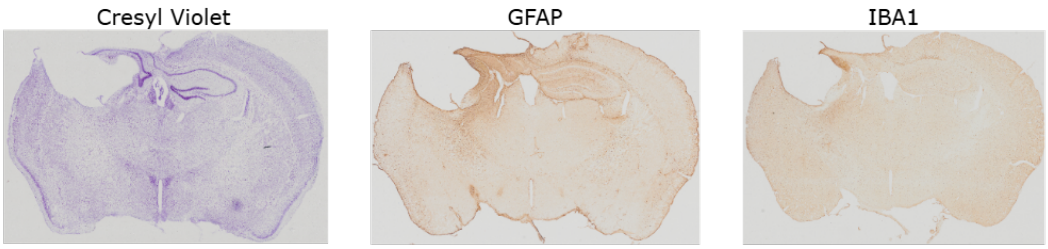


Figure 2.5: Histology with Cresyl Violet, GFAP and IBA1 staining in TBI mice brain.

- *GFAP* (glial fibrillary acidic protein) immunostaining detects reactive astrocytes that play a key role in the inflammatory and repair response following brain injury [11].
- *IBA1* immunostaining identifies microglia, the primary immune cells of the central nervous system which become activated in response to injury, contributing to neuroinflammation [11].

However, the histological preparation process inevitably introduces tissue alterations that do not reflect the true tissue morphology. Such changes are referred to as "artifacts", defined as artificial structures or tissue alterations on a microscopic slide resulting from an extraneous factor [12]. In brain histology, sectioning may cause folds, tears, or stretching, while fixation and dehydration can lead to shrinkage or uneven distortion. In some cases, partial tissue loss creates gaps unrelated to any anatomical structure. Staining procedures may introduce further variability in intensity and contrast, complicating subsequent analysis. These artifacts can directly impact downstream computational tasks, such as image registration, by introducing inconsistencies between histological and MRI data. Representative examples of artifacts commonly encountered in TBI mouse brain histology are shown in Figure 2.6.

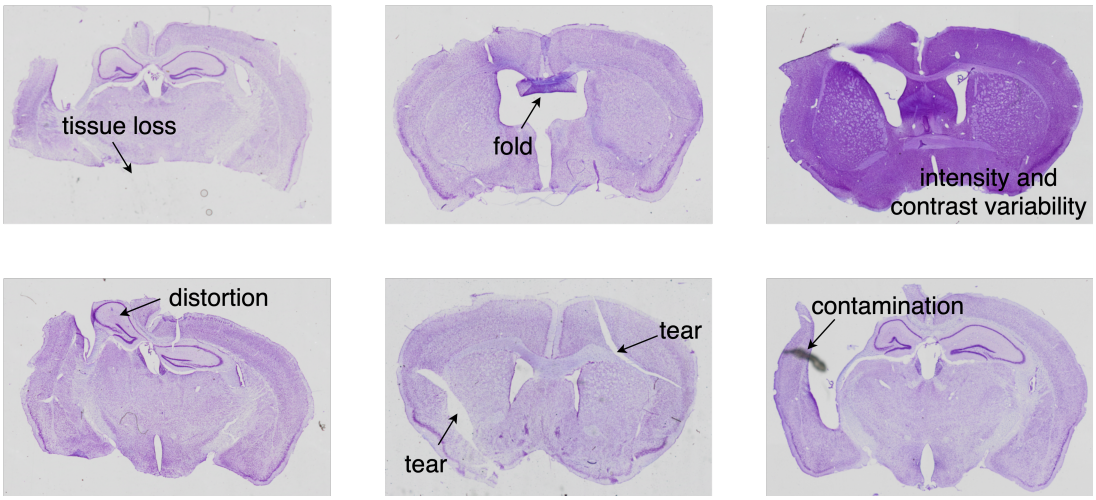


Figure 2.6: Representative examples of common artifacts in TBI mouse brain histology.

2.2. Technical Background

2.2.1. Image Preprocessing for Image Registration

Preprocessing is a crucial initial step for training a deep learning model. The discrepancy between the characteristics of MRIs and histological images affects the accuracy and robustness of automated registration methods. Therefore it is necessary to introduce an image preprocessing step to improve image quality, reduce the noise and standardize the data, allowing for more accurate and consistent alignment across modalities.

MRI preprocessing

N4 Bias Field Correction MRI often suffers from intensity inhomogeneities caused by imperfections in the scanner's magnetic field, resulting in smooth, low-frequency variations known as bias field. These artifacts degrade image quality and complicate downstream analysis. The N4 bias field correction algorithm [13, 14] is a widely used method for correcting these inhomogeneities using B-spline basis functions. The image formulation model assumed by the correction algorithm is:

$$I(x) = J(x)B(x) + n(x)$$

where $I(x)$ is the observed image intensity, $J(x)$ is the true intensity, $B(x)$ is the bias field and $n(x)$ is the noise [13]. By removing these intensity non-uniformities, it enhances the consistency of tissue intensities across the brain volume, thereby improving the reliability of the registration process.

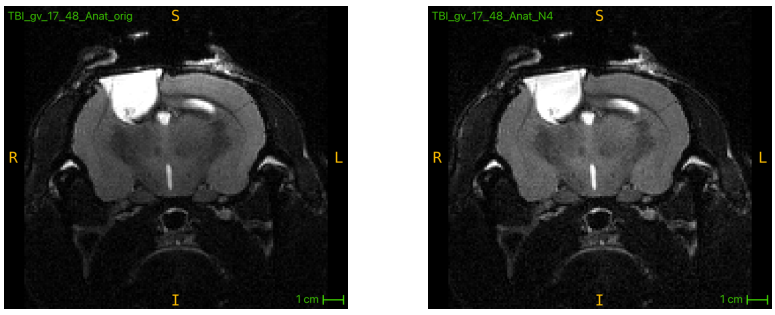


Figure 2.7: Original MRI coronal TBI mouse brain section exhibiting bias (left) and N4 bias field corrected brain section (right).

Skull Stripping Skull stripping is a critical preprocessing step of many neuroimaging analysis pipelines that isolates brain tissue by removing non-brain structures such as skull,

scalp, and eyes from MRI data. The presence of these extra-cranial tissues can affect intensity-based analyses and lead to inaccurate alignment during image registration. In our pipeline, we used the 3DR-NET [15] developed in our lab for mouse brain extraction, ensuring that subsequent registration was driven exclusively by brain tissue anatomy rather than confounding external structures.

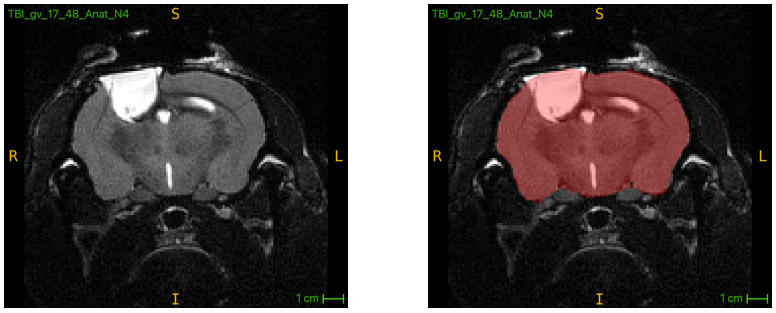


Figure 2.8: Original MRI coronal TBI mouse brain section (left) and brain mask tissue extraction (right).

Spatial Standardization and Resampling Images acquired from different scanners or protocols often differ in orientation, voxel size, and spatial resolution. Spatial standardization aligns images to a common coordinate system, facilitating direct comparison and registration. Resampling to an isotropic voxel size (equal resolution along all three axes) ensures uniform spatial sampling, which is critical for accurate geometric transformations and interpolation. Nearest-neighbour interpolation preserves discrete label values when resampling masks, although other interpolation methods may be used for intensity images to reduce artifacts.

Histological image preprocessing

Gaussian Blur Histological images have a much higher spatial resolution than MRI images, so it is generally necessary to downsample them before the registration process. However, directly downsampling a high-resolution image can introduce aliasing artifacts. To prevent this, a low-pass filtering step is applied before. The Gaussian blur is commonly used because it effectively smooths high-frequency details while preserving the overall image structure. It is achieved by convolving the image with a Gaussian kernel, typically defined as:

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

where σ is the standard deviation of the distribution and controls the smoothing effect and x, y represent the spatial locations. By applying this filter, high-frequency noise and fine

texture details are attenuated, reducing aliasing artifacts in the subsequent resampling step.

Resampling Processing histological images at their original high resolution often demands memory and computational resources beyond the capacity of typical systems. To address these practical limitations, the images are resampled to a lower resolution balancing detail preservation with resource constraints. In our pipeline, we applied bicubic interpolation during resampling, which uses the weighted average of the 16 nearest pixels to compute new pixel values. This method provides smoother and more visually accurate downsampling compared to simpler methods such as nearest-neighbour or bilinear interpolation. Bicubic interpolation ensures that sufficient anatomical detail is maintained for accurate cross-modal registration without overwhelming computational resources.

Macenko Stain Normalization Histological images exhibit variability in colour and staining intensity due to differences in tissue preparation, staining protocols, and imaging conditions. The Macenko method [16] addresses this by transforming RGB images into optical density (OD) space using the Beer–Lambert law:

$$\mathrm{OD} = -\ln(I + \varepsilon)$$

where I is the normalized RGB image and ε is a small constant to avoid numerical instability. In OD space, stain contributions become approximately linearly separable. To estimate the stain matrix, background pixels are excluded, and the OD values of the remaining foreground pixels are centered and subjected to SVD. The first two singular vectors represent the optical density directions of the primary stains.

To normalize the appearance of all histology images in the dataset, we first compute a target distribution of stain concentrations by applying this decomposition to a representative subset of the dataset. From this, we estimate the global mean and standard deviation of stain concentrations. For each image, the OD values are projected onto the stain vectors to obtain stain concentrations and normalized to match the global reference:

$$C' = \left(\frac{C - \mu_C}{\sigma_C}\right) \cdot \sigma_T + \mu_T$$

where C is the original concentration, μ and σ are the mean and standard deviation of the current stain concentration, and μ_T, σ_T are the target statistics computed on the dataset. The normalized concentrations are then used to reconstruct the OD image, which

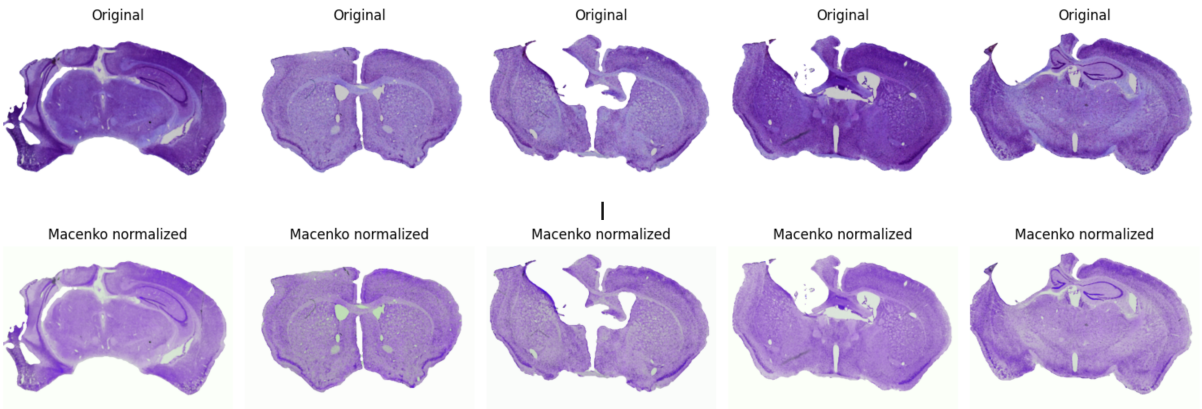


Figure 2.9: Stain normalized histological images with Macenko method.

is converted back to RGB space as:

$$I_{\text{norm}} = \exp(-\text{OD}_{\text{norm}})$$

This normalization reduces inter-sample variability and produces histological images with consistent colour profiles (Figure 2.9), enhancing the robustness of downstream steps.

2.2.2. Convolutional Neural Networks as Feature Extractors

Convolutional Neural Networks (CNNs) are a class of deep learning models designed to process and analyse visual data, such as images. Unlike fully connected neural networks (FCNNs), where each neuron is connected to every neuron in the adjacent layers, CNNs employ neurons that connect only to local regions of the input image. This sparse connectivity, combined with parameter sharing, allows CNNs to efficiently capture patterns and structures in the input image while enhancing computational efficiency. Moreover, convolutional layers enable the network to process images of arbitrary size, contrary to the classical fully connected layers that require a fixed input size. By learning hierarchical features, from simple edges and textures in early layers to complex and abstract structures in deeper layers, CNNs automatically extract meaningful representations directly from images. This is achieved through the sequential application of convolutional, activation, and pooling layers, which progressively transform input data into compact and discriminative feature representations.

This property has made CNNs the core of a wide range of image-related tasks in computer vision. CNNs, in particular, have become essential in medical imaging, where they are used to detect tumours, segment organs, and classify diseases. In addition to healthcare, CCNs are widely used in autonomous driving (e.g., object detection, lane recognition),

surveillance systems (e.g., face and activity recognition), remote sensing (e.g., land cover classification and disaster monitoring), and industrial inspection.
This section introduces the most relevant operations in CNNs, followed by an overview of the CNN architectures and deep learning techniques used in this thesis.

Convolution The convolution operation is the core mechanism that enables feature extraction in CNNs. At each convolutional layer, a set of learnable kernels (also called filters) K is applied to the input image or feature map I. Each kernel spans all input channels, and the convolution is computed as a sum over both the spatial and the channel dimensions. Assuming for simplicity a single-channel input, the convolution output at position (x, y) (Figure 2.10) is given by:

$$S(x,y) = (I * K)(x,y) = \sum_{i=0}^{h_k} \sum_{j=0}^{wk_k} I(x+i, y+j) \cdot K(i,j) \tag{2.1}$$

Sliding each kernel across the input produces a set of feature maps, each highlighting the presence of specific patterns learned during training. The region of the input that influences the activation of a particular neuron is called its receptive field. As the network depth increases, the receptive field grows, allowing deeper layers to integrate information from larger portions of the input image. Multiple filters per layer allow the network to capture a diverse set of features at different spatial locations and scales.

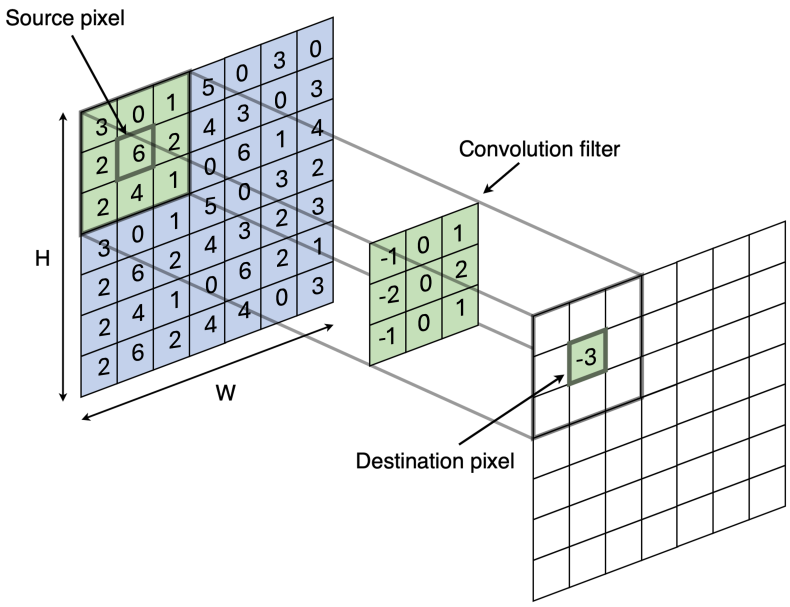


Figure 2.10: 2D Convolution filter of size 3×3 applied to single-channel feature map.

The spatial dimensions of the output feature map depend on the input size (H_{in}, W_{in}), kernel size (h_k, w_k), amount of padding (p), and stride (s).
The output shape is given by:

$$H_{out} = \left\lfloor \frac{H_{in} - h_k + 2p}{s} \right\rfloor + 1, \quad W_{out} = \left\lfloor \frac{W_{in} - w_k + 2p}{s} \right\rfloor + 1 \tag{2.2}$$

Activation After each convolution layer, an activation function introduces non-linearity, enabling the network to model complex relationships beyond linear transformations. The most common activation functions include the Rectified Linear Unit (ReLU), sigmoid, tanh, Leaky ReLU and softmax (Figure 2.11), each with distinct properties suited for different tasks.

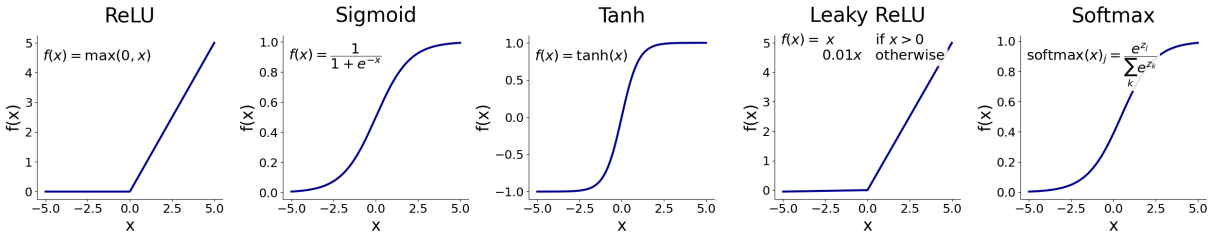


Figure 2.11: Common activation functions.

Pooling The pooling operation aggregates information contained in feature maps by summarising local neighbourhoods, typically using functions such as the maximum (max pooling) or the average (average pooling). For each window, max pooling outputs the maximum value (Figure 2.12), while average pooling computes the mean. This process reduces the spatial dimensions of the feature maps, leading to a more compact representations and high robustness to translations or distortions in the input image. Pooling is typically applied independently to each channel of the feature map and does not involve learnable parameters.

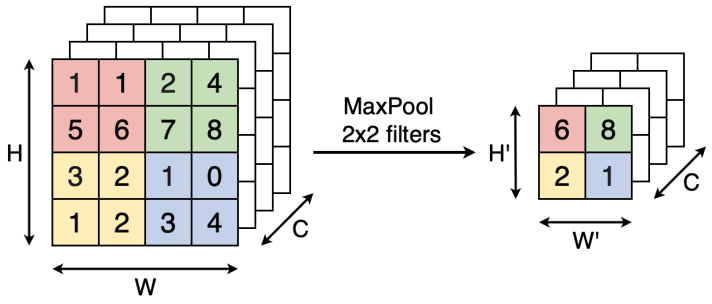


Figure 2.12: MaxPooling operation with filter of size 2×2 and stride 2.

Batch Normalization Batch normalization enhances training speed and stability by normalizing the activations within each layer. It addresses internal covariate shift by first normalizing the activations to have zero mean and unit variance across each mini-batch, and then rescaling and shifting them with learnable parameters to preserve the network's representational power. Formally, for the i-th activation in a mini-batch, the normalized value is computed as:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{2.3}$$

where x_i is the original activation before normalization, μ_B and σ_B denote the mean and standard deviation computed over the current mini-batch, and ε is a small constant added for numerical stability. The normalized activation $\hat{x}_i$ is then scaled and shifted by learnable parameters γ and β as:

$$y_i = \gamma \hat{x}_i + \beta \tag{2.4}$$

These parameters allow the network to restore the representational power if necessary. During training, batch statistics are used for normalization, while during inference, running averages accumulated during training are applied. In addition to accelerating convergence, batch normalization also acts as a regularizer, often reducing the need for other regularization techniques like dropout. However, batch normalization is sensitive to batch size: small batches may lead to noisy or unstable estimates of the mean and variance, degrading performance. In such cases, alternatives like Layer Normalization or Group Normalization may be more effective.

2.2.3. Data Augmentation

Data augmentation is a widely adopted technique in deep learning that aims to increase the diversity and size of the training dataset by applying different random transformations to the original data. Augmentation is especially critical when the available dataset is limited, as often occurs in medical imaging due to the high cost, time consumption, and expertise required for annotation, as well as privacy concerns that limit data sharing. In fact, for many tasks, obtaining sufficiently large and representative data remains a challenging problem. To address this issue, a variety of label-preserving operations are applied to the training samples and their corresponding labels. These typically involve geometric transformations (e.g., rotations, flips, scaling, elastic deformations) and intensity-based adjustments (e.g., brightness, contrast) that maintain the semantic content of the images. By artificially expanding the dataset, data augmentation helps reduce overfitting, improves model generalization, and enhances robustness against variations encountered in real-world scenarios. Careful selection of augmentation techniques is essential, as the

effectiveness of different transformations depends on the specific task and data characteristics. Choosing appropriate augmentations ensures that the generated samples remain realistic and relevant, thus maximizing the benefits of augmentation while avoiding potential negative impacts on model performance.

2.2.4. Transfer Learning and Fine Tuning

Training deep neural networks typically requires large amounts of data to avoid overfitting and achieve strong generalization. However, in many real-world applications, acquiring large and diverse datasets is challenging. Transfer learning offers a solution by leveraging models pretrained on large datasets (e.g., ImageNet dataset [17]), as a starting point for new tasks, rather than training a new model from scratch. These pretrained models act as feature extractors that capture general visual patterns, which are often transferable across different domains due to shared low-level image characteristics.
Fine-tuning complements transfer learning by adapting the pretrained model to the specific target domain. During fine-tuning, some or all layers of the model are further trained on the available target data, allowing the network to specialize its features for the new task. Typically, the earlier layers, which learn low-level features, are kept fixed or updated with a lower learning rate, while deeper layers, responsible for higher-level, task-specific features, are fine-tuned more extensively (Figure 2.13). This strategy not only accelerates convergence but also mitigates overfitting, particularly when training data is limited.

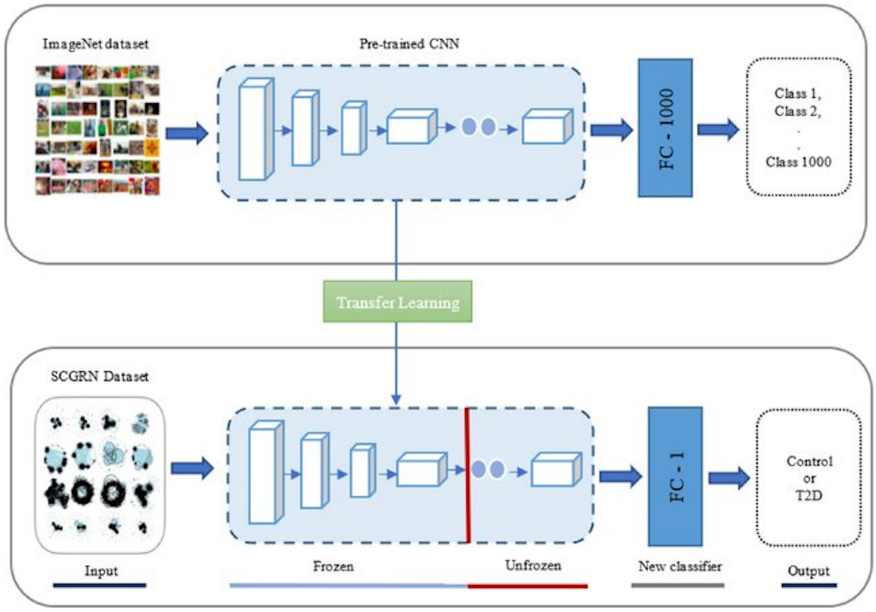


Figure 2.13: Transfer Learning applying fine tuning for a new task.

In medical imaging, particularly in multimodal registration tasks such as aligning MRIs and histopathological images, transfer learning and fine-tuning are valuable techniques for improving model performance and reducing training time. Although pretrained models are often trained on natural images, their learned features can sometimes be applied to medical image analysis, especially in cases where lower-level patterns such as edges and textures are important. However, their efficiency often decreases as the domain gap increases, particularly for tasks that need fine-grained, modality-specific understanding. As shown by Raghu et al. [18], transfer learning is particularly effective when the target dataset is small or the task is difficult, making it well-suited for medical imaging applications with limited annotated data.

VGG-16 Architecture Due to its simplicity and good performance, VGG-16 remains a widely adopted backbone in computer vision and medical imaging tasks, often used with pretrained weights to leverage transfer learning and reduce training time and resource requirements.

The VGG-16 architecture, proposed by Simonyan and Zisserman [19], is a deep CNN consisting of 13 convolutional layers and 3 fully connected layers, as shown in Figure 2.14. It uses small 3×3 convolutional filters with stride 1 and padding to preserve spatial resolution. These convolutional layers are organized into five blocks, each containing two or three convolutional layers with an increasing number of filters. Each convolutional layer is followed by a ReLU activation function, and max-pooling layers with 2×2 windows and stride 2 are applied after each block to progressively reduce spatial dimensions. The convolutional blocks extract hierarchical features, culminating in a flattened feature map that feeds into two fully connected layers, followed by a final classification layer.

With approximately 138 million parameters, the majority of which are in the fully connected layers, VGG-16 is relatively large and computationally demanding. Its considerable depth, resulting from stacking multiple small convolutional layers, enables the network to learn complex and increasingly abstract features. This depth, combined with the use of small 3×3 filters, allows for larger receptive fields through multiple successive convolutions while keeping the number of parameters manageable compared to using larger kernels.

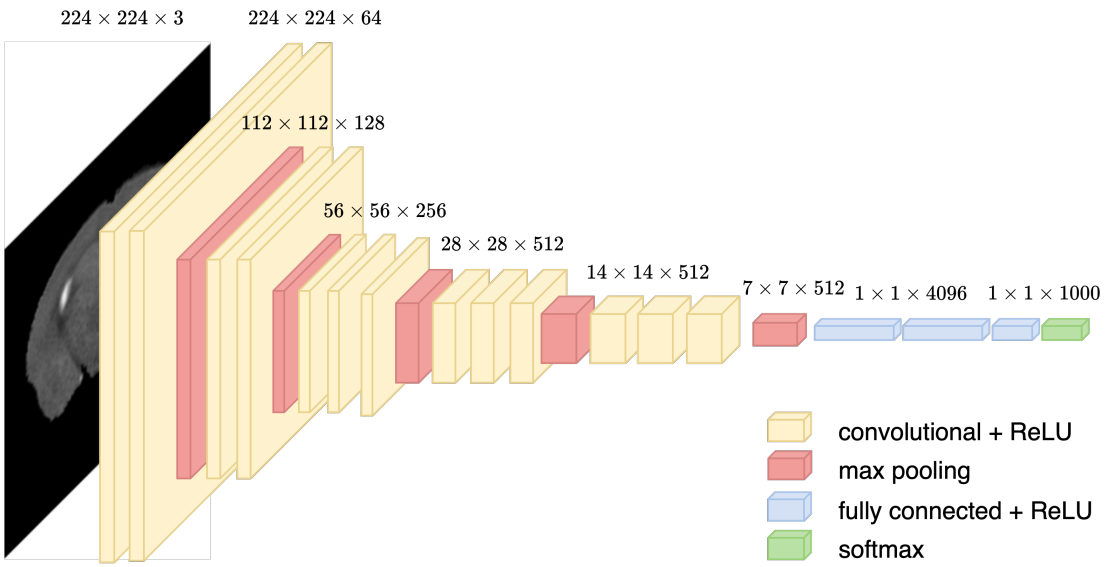


Figure 2.14: VGG-16 architecture.

2.2.5. Generative Adversarial Network for Image-to-Image Translation

Generative Adversarial Networks (GANs) are a class of neural networks widely used for image-to-image translation tasks, which aim to transform an image from a source domain into a corresponding image in a target domain. Formally, this task consists of learning a mapping $G : X \rightarrow Y$ between two image domains X (source) and Y (target), such that the output image $G(x)$ resembles images from the target domain Y while maintaining the structural and semantic properties of the input image $x \in X$.

GANs consist of two neural networks trained simultaneously in an adversarial setting: a *generator* G, that synthesizes images to resemble the target domain, and a *discriminator* D, that aims to correctly classify original images as real and generator-synthesized images as fake. Through this adversarial training, the generator learns to produce highly realistic images capable of fooling the discriminator.

Popular GAN architectures for image-to-image translation include Pix2Pix [20] and CycleGAN [21]. The main difference lies in the type of training data used. Pix2Pix consists of a single generator and single discriminator and requires aligned paired images between source and target domains and trains with a conditional adversarial loss alongside a reconstruction loss, enforcing pixel-wise correspondence. Instead, CycleGAN is designed for unpaired dataset, a common limitation in medical imaging where perfectly paired

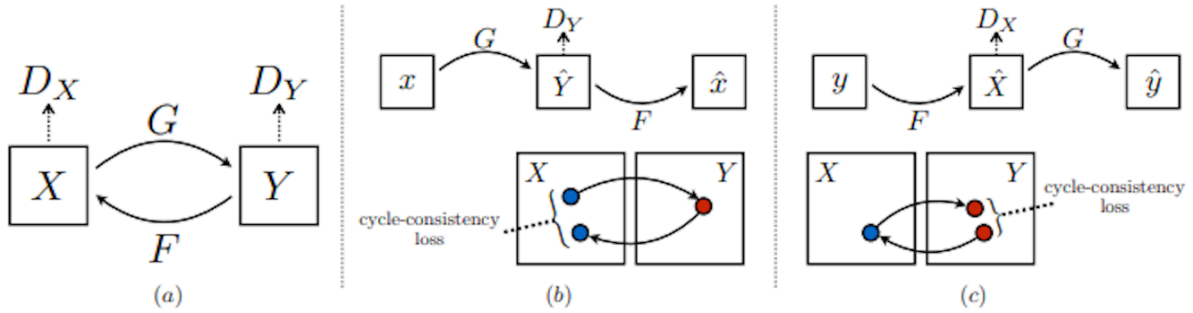


Figure 2.15: CycleGAN architecture for image-to-image translation task. Image source [21].

multimodal images are often unavailable or difficult to obtain. CycleGAN employs two generators and two discriminators to learn mappings in both directions simultaneously (Figure 2.15). Moreover, it leverages a cycle-consistency losses to ensure that translating an image to the target domain and back recovers the original image. This enables training when aligned pairs are unavailable.

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1] \tag{2.5}$$

$$\mathcal{L}_{id}(G, F) = \mathbb{E}_{y \sim p_{data}(y)}[\|G(y) - y\|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\|F(x) - x\|_1] \tag{2.6}$$

Both models employ convolutional generator networks (e.g, U-Net), and PatchGAN discriminators that classify image patches rather than entire images, focusing on the style of the images rather than the specific content. Additional losses such as identity and perceptual losses may be integrated to improve image quality and content preservation.

2.2.6. Geometric Transformation Models

All image registration pipelines rely on a geometric transformation model, which establishes how one image is spatially warped to align with another. These models vary in their degrees of freedom (DoF), offering increasing flexibility to capture geometric variations, as illustrated in Figure 2.16. They range from rigid transformations models with few parameters to complex deformable ones capable of representing local, non-linear distortions. Selecting the appropriate transformation model is critical. The choice should reflect the expected deformations between source and target images, the characteristics of the acquisition process, and the desired level of registration accuracy [22]. In medical imaging, this decision directly impacts registration performance: rigid models may suffice for brain scans exhibiting minimal deformation, whereas deformable models are essential when dealing with organs or tissues subject to significant anatomical variability [23] or trauma, as is the case in this research context.

Transformation models are typically categorized by increasing DoF as [22–24]:

- Rigid Transformations
- Affine Transformations
- Projective Transformations (Homographies)
- Deformable (Non-Rigid) Transformations

We will analyse each type of model separately.

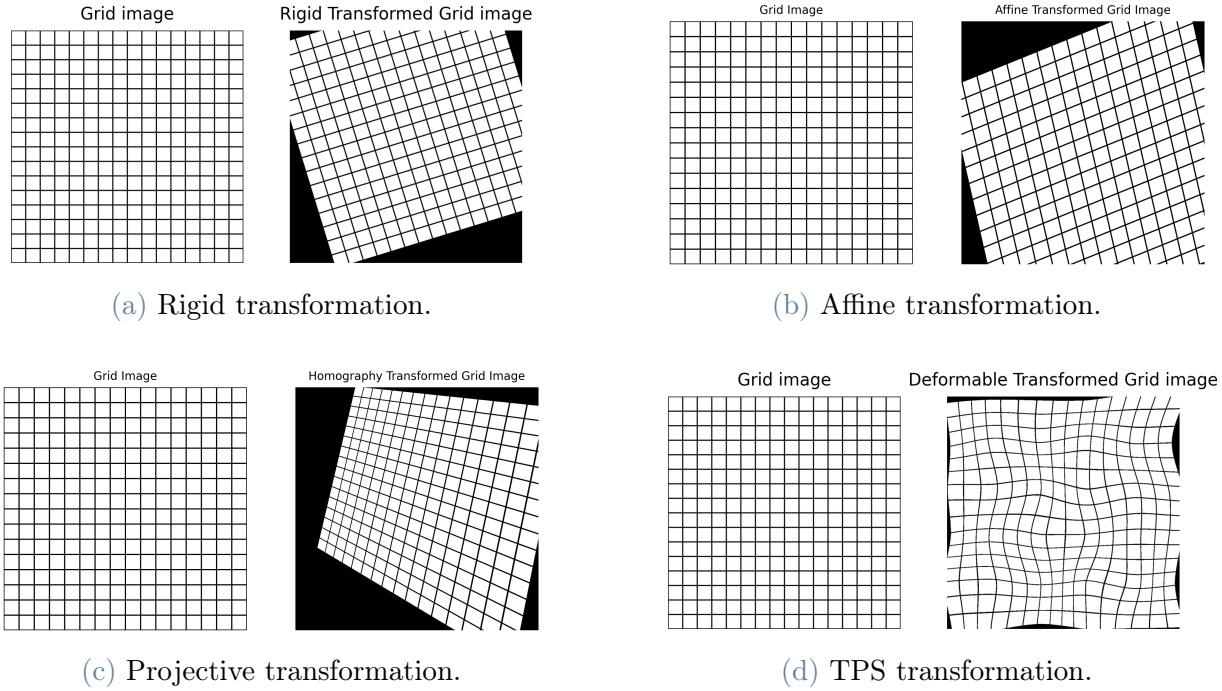
Rigid Transformation

A rigid transformation is an isometry, namely it preserves distances and angles between points, as shown in Figure 2.16a. It includes only rotation and translation and is suitable for cases where the object maintains its shape, such as bones or rigid body motion. In two dimensions, it is fully defined by three parameters: a rotation angle θ and a translation vector $[t_x, t_y]^T$. Mathematically, the transformation of a point $[x, y]^T$ to a new point $[x', y']$ is expressed as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{2.7}$$

Affine transformation

Affine transformations extend rigid transformations by incorporating scaling and shearing. They preserve collinearity (i.e, straight lines) and parallelism but not necessarily



(a) Rigid transformation.

(b) Affine transformation.

(c) Projective transformation.

(d) TPS transformation.

Figure 2.16: Comparison of different geometric transformation models.

distances or angles, as shown in Figure 2.16b. The general 2D affine transformation can be parametrized as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_4 & \theta_5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \theta_3 \\ \theta_6 \end{bmatrix} \tag{2.8}$$

where θ is the rotation angle, s_x and s_y account for the scaling along x and y axis, sh_x and sh_y are the shearing factors along x and y axis, and $[t_x, t_y]^T$ is the translation vector. These parameters allow for 6 degrees of freedom, making the affine transformations suitable for modelling more complex relationships, such as scanner-induced distortions.

Projective Transformation (Homography)

A projective transformation, also know as homography, allows mapping between two planes in projective space, as shown in Figure 2.16c. Like affine transformations, homographies preserve collinearity but do not preserve parallelism, distances, or angles. In two dimensions, it is defined by a 3×3 matrix H, applied in homogeneous coordinates:

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \Rightarrow \quad (x', y') = \left(\frac{x'}{w'}, \frac{y'}{w'} \right) \tag{2.9}$$

The matrix H has eight degrees of freedom (since it is defined up to scale), which control translation, rotation, scaling, shearing, and perspective distortion. This generalizes affine transformations by including the ability to model perspective effects (controlled by the parameters h_{31} and h_{32}), such as those caused by changes in viewpoint or camera angle.

Deformable (Non-Rigid) Transformation

Deformable transformations enable local and smooth variations in shape, allowing the modelling of non-linear deformations such as tissue elasticity or organ motion. These transformations are typically parametrized using control point grids (e.g., B-splines) or radial basis functions (e.g., Thin-Plate Splines). While deformable models require greater computational resources in terms of both time and memory, they are essential in medical imaging applications where soft tissue alignment is necessary, especially when tissue shape varies significantly across imaging modalities as in this case, or in longitudinal studies tracking anatomical changes over time in a patient.

Thin-Plate Splines (TPS) The TPS model interpolates a smooth deformation field between corresponding landmarks, as illustrated in Figure 2.16d. For a 2D point (x, y), the transformation function is defined as:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} f_x(x,y) \\ f_y(x,y) \end{bmatrix} = \mathbf{a_1} + \mathbf{a_x} \cdot x + \mathbf{a_y} \cdot y + \sum_{i=1}^{N} \mathbf{w_i} \cdot U\left(\|(x,y) - (x_i, y_i)\|\right) \tag{2.10}$$

Here, $U(r) = r^2 \log r$ is the radial basis function, and (x_i, y_i) are the control points in the input image arranged in an $N \times N$ grid. The affine coefficients $\mathbf{a_1}, \mathbf{a_x}, \mathbf{a_y} \in \mathbb{R}^2$ capture the global linear component of the transformation, while the weights $\mathbf{w_i} \in \mathbb{R}^2$ control the local non-linear deformations.

2.3. Image Registration

2.3.1. Classification of Image Registration Approaches by Imaging Modality

Image registration tasks can be broadly categorized according to the modalities of the image pairs involved. This classification is fundamental, as the nature of the image data significantly affects the design and performance of the registration method. Two primary scenarios are commonly considered: intra-modality and inter-modality registration.

Intra-Modality Registration Intra-modality registration, also known as monomodality registration, refers to the alignment of images acquired using the same imaging modality. For example, it can involve aligning multiple MRI brain scans from different time points in a longitudinal study, or registering CT scans of the thorax from different patients. Since these images share similar intensity distributions and structural characteristics, similarity metrics based on pixel-wise intensity comparisons (e.g., Sum of Squared Differences, Normalized Cross-Correlation) are usually effective. Typically, the main challenge lies in handling motion artifacts, noise and anatomical deformation.

Inter-Modality Registration Inter-modality registration, also known as multimodality registration, involves aligning images acquired from different imaging techniques, such as CT-MRI, MRI T2w-MRI T1w, or histology-MRI. The images usually exhibit distinct intensity patterns and structural representations, making direct pixel-wise comparison unreliable. To overcome this challenge, inter-modality registration typically relies on similarity metrics that are independent of modality, such as Mutual Information, along with robust feature extraction methods that capture structural correspondences beyond intensity values. In recent years, learning-based approaches have also become popular, using trained models to map features across modalities and improve registration accuracy.

2.3.2. Methodological Classification of Image Registration Approaches

An increasing number and variety of methods have been proposed to address the problem of image registration. As suggested by many researchers, a good taxonomy categorizes existing approaches into intensity-based and feature-based pipelines, in terms of their components and procedural details [22]. In recent years, learning-based methods have replaced traditional techniques in different steps of the registration process, showing strong

potential for automation, though several challenges still remain [4, 25].

Intensity-based approach

Intensity-based registration approach, also referred to as area-based, performs image alignment by comparing their intensity patterns across corresponding regions. Rather than relying on predefined features or landmarks, this approach uses the full image content to guide the alignment. It is applicable to both monomodal and multimodal settings and is especially effective when structural correspondence exists across the images.

These methods typically consist of three main components:

- **Similarity metric** $\mathcal{M}$: measures the accuracy of alignment between moving and fixed images based on voxel intensity values. Common choices include Sum of Squared Differences (SSD) and Normalized Cross-Correlation (NCC), typically used for images from the same modality, as well as Normalized Mutual Information (NMI), which is effective for aligning images from different modalities.
- **Geometric transformation model**: defines the geometric relationship between image pairs. Models can capture global (e.g., rigid, affine) or local transformations across images.
- **Optimisation method**: searches for the transformation parameters that best align the images by maximizing or minimising the chosen similarity metric. optimisation techniques include gradient-based methods, evolutionary strategies, or modern deep learning-based optimisers.

Figure 2.17 illustrates the general workflow of intensity-based image registration.

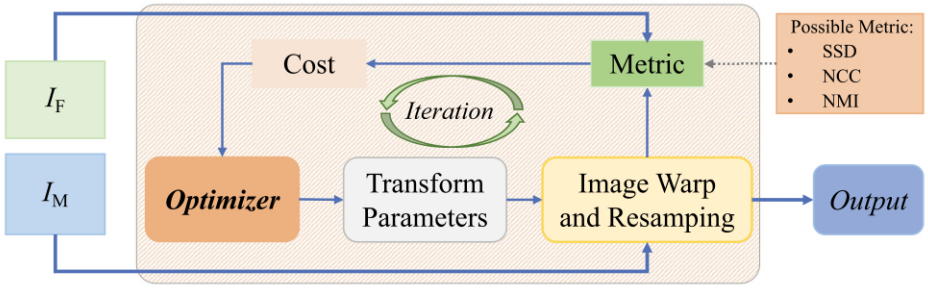


Figure 2.17: General outline of traditional intensity-based image registration approach. Image source [3].

Problem formulation Building on these components, the registration task can be expressed as an optimisation problem. The goal is to find the transformation T^* that

minimises a cost function combining a chosen similarity metric $\mathcal{M}$ and a regularization term $R(T)$, which promotes smoothness or physical plausibility of the transformation, namely:

$$T^* = \arg\min_{T \in \mathcal{T}} \mathcal{M}(I_F, I_M \circ T) + \lambda R(T) \tag{2.11}$$

Here, $\lambda \geq 0$ controls the trade-off between alignment accuracy and transformation regularity.

Feature-based approach

The feature-based registration approach aligns image pairs by extracting distinctive features from each image independently and establishing correspondences between them. Unlike intensity-based methods, this approach does not directly rely on voxel intensity values but instead focuses on salient local structures.

This approach typically consists of the following steps:

- **Feature detection**: detects keypoints that are distinctive, robust, and consistently identifiable across different images, such as corners, edges, or blobs. Common detectors include Harris, SIFT and ORB.
- **Feature description**: generates a vector representation of each keypoint local neighbourhood, encoding distinctive gradient or intensity patterns that are invariant to transformations like scale, rotation, and illumination changes.
- **Feature matching**: establishes correspondences by comparing descriptors from the fixed and moving images, typically by computing distances (e.g., Euclidean or Hamming), and selecting pairs with the smallest distance.

The matched feature pairs are then used to estimate the spatial transformation. Figure 2.18 illustrates the general workflow of feature-based image registration.

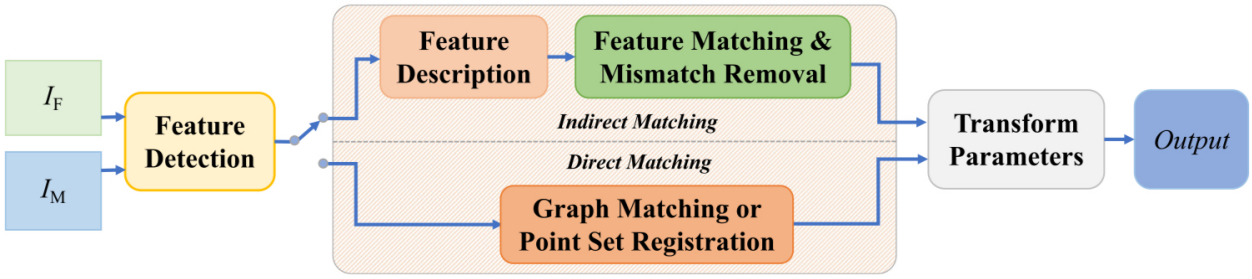


Figure 2.18: General outline of traditional feature-based image registration approach. Image source [3].

Problem formulation Let $\{p_i^F\}_{i=1}^N$ and $\{p_j^M\}_{j=1}^N$ denote the sets of detected keypoints in the moving and fixed images, respectively, with corresponding descriptors $\{d_i^F\}_{i=1}^N$, $\{d_j^M\}_{j=1}^N$. Feature correspondence pairs $\mathcal{C} = \{(i,j)\}_{i,j=1}^N$ are found by matching descriptor pairs that satisfy a similarity threshold τ, namely $(i,j) \in \mathcal{C} \iff \|d_i^F - d_j^M\| < \tau$. Using these matched pairs, the transformation T^* is estimated by minimising the geometric distance between corresponding points:

$$T^* = \arg\min_{T \in \mathcal{T}} \sum_{(i,j) \in \mathcal{C}} \|p_i^F - T(p_j^M)\|^2 \tag{2.12}$$

Robust estimation techniques such as RANSAC are often employed to exclude outliers, improving the accuracy of the transformation. Feature-based registration usually focuses on global transformation models, especially when keypoints are sparse.

SIFT: The Scale-Invariant Feature Transform (SIFT) is a feature extraction technique introduced by David G. Lowe in 2004 [26]. SIFT is designed to detect and describe local features that are invariant to scale, rotation, and partially invariant to affine distortion and illumination changes.

It consists of the following steps:

1. **Scale-space extrema detection**: Keypoints are identified as local extrema of the Difference-of-Gaussians (DoG) function in scale space.
2. **Keypoint localization**: Unstable points are discarded using a contrast threshold and edge response analysis.
3. **Orientation assignment**: Each keypoint is assigned one or more dominant orientations based on the gradient direction distribution around it.
4. **Keypoint descriptor**: A 128-dimensional vector is constructed using the gradient magnitudes and orientations in a local region around the keypoint.
5. **Keypoint matching**: Descriptors are matched across images using metrics like Euclidean distance.

SIFT is highly robust but relatively computationally expensive.

SURF: The Speeded-Up Robust Features (SURF) algorithm was developed by Bay et al. in 2006 [27]. It was proposed as a faster alternative to SIFT while retaining robustness to scale, rotation, and some affine transformations. SURF accelerates both detection and description by approximating operations used in SIFT.

It can be summarised into the following steps:

1. **Interest point detection**: Keypoints are detected as local maxima of the determinant of the Hessian matrix, approximated using box filters and integral images. Octaves are constructed by scaling the filter rather than the image as in SIFT, which reduces the computational cost.
2. **Keypoint localization**: Similarly to SIFT, non-maximal points and low-contrast points are eliminated.
3. **Orientation assignment**: Orientation is determined using Haar wavelet responses in the x and y directions within a circular region around the keypoint. A sliding window is used to select the dominant(/main) orientation.
4. **Keypoint descriptor**: A 64-dimensional descriptor is constructed by dividing the neighborhood into 4x4 subregions and summarising Haar wavelet responses.
5. **Keypoint matching**: Descriptors are matched using metrics such as Euclidean or Hamming distance.

SURF achieves improved speed and efficiency over SIFT, making it suitable for real-time applications.

LIFT: The Learned Invariant Feature Transform (LIFT) is a deep learning-based feature extraction pipeline proposed by Yi et al. in 2016 [28]. Unlike handcrafted methods as SIFT and SURF, LIFT learns the entire pipeline (detection, orientation estimation, and description) directly from data using a convolutional neural network (CNN). It is designed to be invariant to scale, rotation, and illumination, and is especially suited for challenging image conditions and cross-modality scenarios.

The LIFT pipeline consists of the following components:

1. **Keypoint detector**: A CNN-based module trained to predict the likelihood of a pixel being a good keypoint. The detector is trained using a patch-based triplet ranking loss to encourage repeatability.
2. **Orientation estimator**: After detecting keypoints, an orientation prediction network assigns a dominant orientation to ensure rotation invariance.
3. **Descriptor extractor**: A third CNN block generates a compact descriptor for each oriented patch, trained with a contrastive loss to maximize distinctiveness and robustness.

4. **Keypoint matching**: Descriptor vectors are matchedt across images using distance-based metrics, typically Euclidean distance.

By unifying the full pipeline into a trainable architecture, LIFT achieves improved robustness in complex image registration scenarios. However, the effectiveness of this approach is highly dependent on the quality and diversity of the training data, and domain-specific fine-tuning is frequently necessary to achieve optimal performance.

3 | Related Work

In recent years, attention has grown towards medical image registration, with increasing efforts to develop methods that allow accurate spatial alignment between different imaging modalities, giving rise to an increasing number and diversity of registration techniques. This procedure is crucial for an exhaustive biomedical analysis, particularly for understanding pathological changes in disease models by combining various and complementary information from different imaging sources [3]. In pathological contexts, multimodal imaging can be crucial for understanding the evolution of disease processes, monitoring pathology over time, characterizing the kinetics of biological changes, and identifying optimal intervention windows and novel targeted therapies [10].

Hereafter, we review the state-of-the-art methodologies used for medical image registration, starting with general MMIR methods (Section 3.1), followed by a focused excursus of histology-MRI registration approaches (Section 3.2) and finally the GAN-based image-to-image translation approaches used in the medical context for modality translation (Section 3.3).

3.1. Multi-modal Medical Image Registration

Image registration has long been recognized as a fundamental but challenging problem in medical imaging [29]. Its importance in pathological contexts has been reinforced by recent multimodal fusion studies. For instance, Guo et al (2024) [10] introduced SOmicsFusion, showing that only through registration of spatial metabolomics and biomedical imaging it is possible to capture pathological dynamics invisible in unimodal data. This illustrates why multimodal registration is essential in acute brain injury, where lesion progression and biological response kinetics require complementary information sources.

Despite decades of research and the development of a wide range of approaches, no universally applicable solution exists, particularly in complex multimodal and pathological contexts. Building upon the classification of IR approaches introduced in Section 2.3, here we analyse the core aspects and main limitations of both intensity-based and feature-based

approaches, used in the medical field. We then discuss how deep learning has been increasingly integrated into this framework, providing alternative strategies to address modality differences and improve registration.

Intensity-based approaches directly compare pixel intensities and are primarily valued in medical imaging for their effectiveness in multimodal registration. Information theory–based methods, such as mutual information (MI) [30], normalized mutual information (NMI) and other entropy-based metrics, have become the standard for multimodal registration. They quantify statistical dependence rather than direct intensity differences, making them robust to non-linear differences between modalities. Till now, research has primarily focused on mitigating MI high computational cost. Other similarity metrics as Sum of Squared Differences (SSD) and Normalized Cross-Correlation (NCC) are simpler to implement but less robust in the multimodal framework.

On the other hand, given the reduced texture of medical images, feature-based methods are rarely used for registration in medical research. Generally, the global structure and contours or edges commonly exist in both images of different modalities, thus providing matchable information [31, 32]. However, in histological brain images of rodents after TBI, these conditions are often not satisfied due to artifacts such as folds, tears and missing tissue (Section 2.1.3), making global structures and contours of the brain unreliable for registration. This limitation makes feature-based approaches insufficiently robust for registration in our setting.

Among deep learning-based approaches, GANs have gained attention for their capacity to perform image-to-image translation between modalities [21], effectively reducing the complexity of multimodal registration by transferring both images into a common modality domain. This enables a modality unification, simplifying the multimodal problem into a monomodal one. This approach mitigates the limitations related to the use of similarity metrics by reducing appearance differences, however the registration quality highly depends on the GAN's performance. The most relevant GAN architectures for modality translation are described in detail in Section Section 3.3.

3.2. MRI-Histology Image Registration

The majority of the existing literature on Neurological Image Registration has focused on radiological modalities such as CT and MRI while, to our knowledge, studies on MRI-histology registration have mostly been conducted on human prostate, with the aim of localizing and staging cancer. [1, 2, 33]. The intrinsic anatomical and pathological differences between human prostate tissue and rodent's brain with induced injury highlight

a significant research gap that remains to be addressed. In particular, the registration challenges in the context of TBI are amplified by complex, non-linear tissue deformations caused by edema, lesions, and morphological brain changes, features typically absent or less evident in prostate imaging scenarios.

Two significant contributions in prostate MRI-histology registration are ProsRegNet [1] and its successor RAPHIA [2] (by Shao et al.), both leveraging deep learning to enhance alignment accuracy but adopting different strategies.

ProsRegNet introduces spatial transformer networks to perform direct 2D registration between MRI and histological sections. It uses features extracted from both modalities in parallel and combined with a correlation layer, to estimate the transformation parameter through a regression network. It applies both affine and deformable geometric transformations sequentially to align the images and achieves improved accuracy and robustness compared to traditional or purely feature-based methods. However, it does not explicitly address the multimodal nature of the problem, as the network is trained on synthetic monomodal image pairs from both modalities. Shao et al. provide empirical evidence that, despite this limitation, the network generalizes well to multimodal MRI-histology data. Nevertheless, this remains an important limitation, since there is no guarantee the model is sufficiently robust to address the complex variability and deformation patterns typical of TBI.

RAPHIA builds upon ProsRegNet by automating several time-consuming preprocessing steps that were typically performed manually, such as prostate segmentation, orientation estimation of histological images (including rotation and flipping), and determining correspondences between MRI and histological sections. Crucially, it integrates a geometry-preserving image-to-image translation network that synthesizes MRI-like images from histology and vice versa, effectively bridging the multimodal domain gap that ProsRegNet does not directly address. These improvements allow RAPHIA to achieve expert-level registration accuracy even when used by novices, significantly enhancing its usability.

Despite such advances, applying prostate cancer-focused methods like ProsRegNet and RAPHIA to the rodent brain remains challenging. The brain in TBI models experiences more severe and heterogeneous tissue distortions than the prostate, complicating deformation modelling strategies. Additionally, differences in anatomical structure and scale limit the transferability of existing feature extraction and image translation networks. Nevertheless, RAPHIA's approach to multimodal training through geometry-preserving translation inspires our solution design, which aims to address the unique demands of mouse brain TBI MRI-histology registration.

The registration of histological images and MRI sections offers an important link between macroscopic *in vivo* imaging and microscopic *ex vivo* tissue pathology. Yet, to our knowledge, no prior studies have specifically addressed this task in rodent models of TBI. In this framework, unique challenges need to be addressed, including (i) substantial modality differences in resolution, contrast, and scale between MRI and histological sections, and (ii) complex non-linear tissue deformations caused by TBI-induced alterations (e.g., swelling, lesions, and partial tissue loss) as well as artifacts introduced during histological processing (Section 2.1.3). Overcoming these obstacles requires robust multimodal registration strategies capable of accounting for both inter-modality variability and tissue distortion.

3.3. Image-to-Image Translation for Histology-MRI

The substantial differences in appearance, contrast, resolution and dimensionality between histological images and MRIs pose a fundamental challenge in multimodal registration. Histology provides coloured, high-resolution, two-dimensional images that reveal cellular and molecular structures through the staining process, whereas MRI provides grayscale, lower-resolution, volumetric images highlighting tissue microstructure and water content. Several strategies have been proposed to bridge these modality differences, and recent advances increasingly rely on image-to-image translation techniques based on GANs [34].

Building on these ideas, recent works such as Leroy et al. (2021) [35], Wang et al. (2023) [36] and Heng et al. (2024) [37] propose GAN-based frameworks aimed at preserving geometry while translating between imaging modalities. Although these studies focus on image synthesis rather than registration, they demonstrate how modality translation can reduce appearance gaps and facilitate subsequent alignment, providing methodological inspiration for our work. Leroy et al. introduces a weakly paired MRI-to-histology synthesis method that balances structural fidelity and modality translation, highlighting the importance of maintaining anatomical accuracy during synthesis for reliable downstream analysis. Wang et al. propose a novel bidirectional generative model for synthesizing medical images between CT and MRI modalities using unpaired datasets. The model improves the original CycleGAN [21] by incorporating a dual contrast loss into the discriminators. Additionally, it integrates cross-entropy and structural similarity index (SSIM) into the standard loss functions to explicitly preserve important luminance and structural features during synthesis, enhancing the quality of the generated images. Heng et al. proposes enhancements such as the hinge loss and switchable normalization in CycleGAN architectures, improving medical image synthesis quality and stability, crucial for the registration

tasks. In this study they generate CT images from MRI data, which represent more close imaging domains compared to histology and MRI.

In TBI mouse brain applications, where tissue morphology undergoes complex alterations, image-to-image translation could be used to enable the registration process to focus on anatomical structure alignment rather than being confounded by modality-specific differences or artifacts introduced during histological image preparation. Integrating CycleGAN-based image synthesis as a preliminary step for the registration pipeline could facilitate network learning and improve registration accuracy and robustness, representing an advancement over classical multimodal registration methods based solely on similarity metric optimisation. Based on these ideas, the approach proposed in this thesis employs CycleGAN-based modality translation to enhance MRI-histology registration in rodent models of TBI.

4 | Proposed Solution

In this chapter we present the developed end-to-end pipeline for multimodal unsupervised registration of MRI and histological images of mouse brain with TBI. The proposed approach employs a CycleGAN to translate the histological brain image (moving image) into a synthetic MRI-like image. Subsequently, the synthetic image is passed, together with the original MRI section (fixed image) as input to a two-stage deep learning architecture that decomposes the registration task into global and local alignment stages, as illustrated in Figure 4.1. The affine and deformable transformation parameters estimated through these two steps are subsequently applied to the original histological image to obtain the final aligned image. This last step is performed only at inference, not during training.

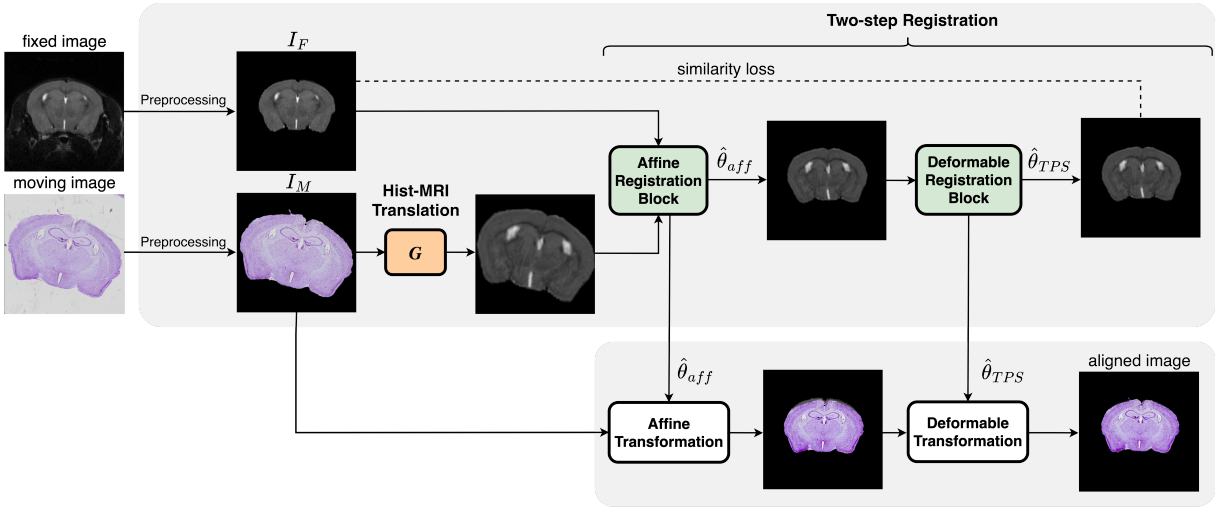


Figure 4.1: Overview of the pipeline for multimodal image registration.

Having introduced the overall pipeline, we now analyse its components in detail. We describe the preprocessing step, followed by the histology-to-MRI translation and the proper registration network architecture that estimates transformation parameters and outputs aligned images. Finally, we explain our training strategy, detailing the CycleGAN and Registration Network loss functions used for the various experimental settings.

4.1. Preprocessing step

Both MRI and histological sections are first automatically skull-stripped with a small deep-learning model developed within the institute and trained to perform binary semantic segmentation [15]. The resulting brain mask isolates the brain from surrounding tissues, removing dirt, bubbles, and other artifacts present in histological images that could influence the registration process.
Then, all sections are manually corrected for gross orientation errors (e.g., ± 90) and horizontal flipping, with identical transformations applied to their corresponding brain masks. Both images are cropped around the brain region, intensity-normalized to the range $[0, 1]$, and resampled to a fixed size of 256×256 pixels. Finally, to facilitate the registration process, the corresponding brain masks are applied to both images.

4.2. CycleGAN architecture for Histology-to-MRI Translation

Direct registration of histological images and MRIs is challenging due to intrinsic modality differences and artifacts from histology acquisition (Section 2.1.3). We address this challenge by using a CycleGAN to transform histological images into MRI-like representations.

The CycleGAN consists of two generators $G : X \to Y$, that translates histological images in the MRI domain, and $F : Y \to X$ that does the opposite domain translation; and two discriminators D_X and D_Y that determine the realism of the histological images and MRIs, respectively (Figure 4.2). The generators have a standard encoder–decoder architecture, where the encoder consists of 3 convolutional layers with 9 residual blocks and the decoder consists of 3 transposed convolution layers. The discriminators are PatchGAN classifiers that only penalizes structure at the scale of patches. Of the whole architecture, only the generator G is directly used in the pipeline to do domain unification (Figure 4.1). The discriminators D_X and D_Y and the reverse generator F are only used during CycleGAN training. The histological image is given as input to the generator G which transforms it into an MRI-like image, which is then used as input to the two-stage deep learning network.

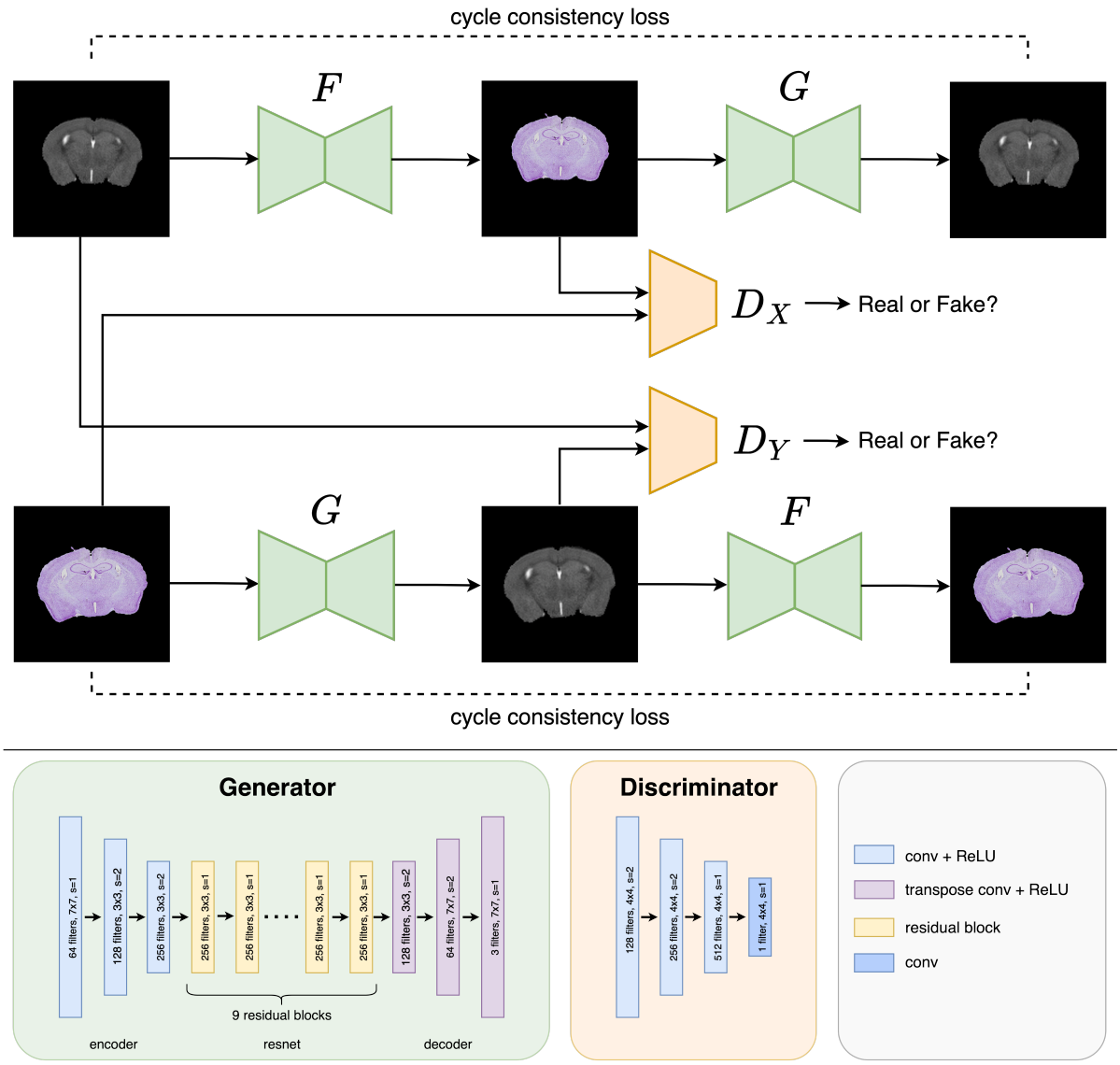


Figure 4.2: CycleGAN architecture for MRI-Histology translation.

4.3. Registration Network architecture

The registration network, based on ProsRegNet architecture [1], decomposes the alignment into global and local alignment stages, as illustrated in Figure 4.3.

- Step I: performs global alignment between moving and fixed images using an affine registration network, correcting for large-scale differences.
- Step II: refines the coarse registration result by applying a Thin-Plate Spline (TPS) transformation estimated through a TPS registration network, enabling the modelling of local non-linear deformations.

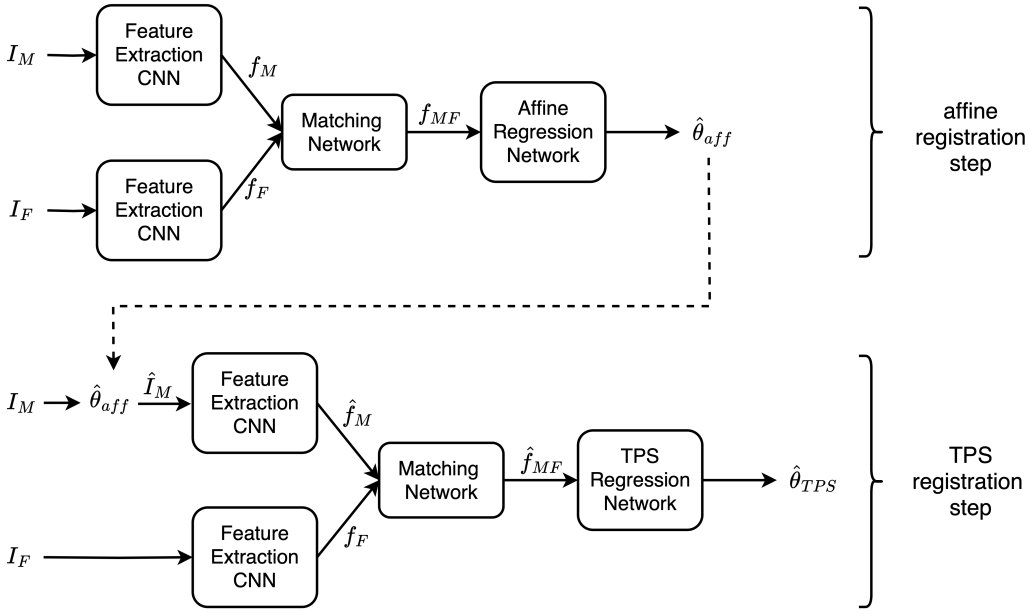


Figure 4.3: ProsRegNet: two-stage deep learning architecture.

Both steps share a common structure, illustrated in Figure 4.3. Moving I_M and fixed images I_F are fed into two feature extraction backbones with shared weights to generate feature maps f_M and f_F, respectively. These are then combined using a correlation layer. The resulting correlation map $f_M F$ is fed into a regression network that estimates geometric transformation parameters. The size of the parameter vector depends on the type of transformation parametrized. Subsequently, these parameters are used to warp the moving image with a differentiable spatial transformer layer.

Hereafter, we go through the architecture blocks in detail.

Feature Extraction Network The feature extraction backbone is the encoder of a CNN architecture (in our case, VGG-16). It processes the moving and fixed images in parallel to extract hierarchical, multi-scale features robust to modality differences. By encoding semantic and structural information, the extracted features serve as input for the subsequent matching and parameter regression stages.

Feature Matching with correlation layer To estimate spatial transformations, a feature correlation layer computes a matching feature map which captures spatial correspondences by comparing patches across the feature maps. After correlation, a normalization step (e.g., SoftMax or channel-wise standardization) regularizes the responses and emphasizes consistent matches across the spatial grid. The correlation layer produces a correlation map c_{MF} of size $H \times W$, containing pairwise correlation coefficients between

feature vectors, defined as:

$$c_{MF}(i,j,k) = \frac{cov[f_M(i,j) \cdot f_F(p,q)]}{std[f_M(i,j)]std[f_F(p,q)]}$$

where $k = p + H(q-1)$, and $f_M(i,j)$ and $f_F(p,q)$ are feature vectors at positions (i,j) and (p,q) in the moving and fixed feature maps, respectively.

Transformation Regression Networks The correlated feature volume is passed to a regression network that predicts geometric transformation parameters in two stages.

- Step I performs global alignment by regressing six parameters of an affine transformation modelling rotation, translation, scaling, and shearing. The 2D affine transformation is parametrized by a vector $\theta \in \mathbb{R}^6$ and defined as:

$$T(x,y) = \begin{bmatrix} 1+\alpha\theta_1 & \theta_2 \\ \theta_4 & 1+\alpha\theta_5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \theta_3 \\ \theta_6 \end{bmatrix} \tag{4.1}$$

  where (x,y) denotes a pixel location in the fixed image I_F and $\alpha = 0.001$ is a constant. This parametrization ensures that the estimated transformation remains close to the identity in the early training epochs, thereby stabilizing training.

- Step II refines the registration with a non-rigid transformation using a Thin-Plate Spline (TPS). The TPS is parametrized by a 6×6 control point grid, resulting in a parameter vector $\theta \in \mathbb{R}^7 2$. Similarly to the affine step, the predicted parameters are scaled by $\alpha = 0.001$ and added to an identity initialization, allowing gradual refinement from the identity mapping. This choice of grid size offers a good trade-off between deformation flexibility and model complexity.

Both regression networks are lightweight CNNs designed to preserve spatial context and are optimised using similarity losses (e.g., RMSD or MI) combined with regularization terms promoting smoothness and realism of the deformation field.

4.4. Training Loss

CycleGAN Loss

The loss used for the CycleGAN model training is a weighted sum of an adversarial loss along with a cycle consistency loss and a final identity loss.

Adversarial Loss The adversarial loss is employed to ensure that the generated images closely resemble real images, enhancing the generator to produce realistic synthesized images. For G and D_y, the adversarial loss can be expressed as:

$$L_{GAN}(G, D_Y) = \mathbb{E}_{y \sim data(y)}[(D_Y(y) - 1)^2] + \mathbb{E}_{x \sim data(x)}[(D_Y(G(x)))^2] \tag{4.2}$$

This corresponds to the Least Squares GAN (LSGAN) formulation, where the discriminator minimises an MSE objective instead of the binary cross-entropy (BCE) used in the original GAN.

Cycle Consistency Loss The cycle consistency loss ensures that the mappings between the domains X and Y are consistent and retain content. It enforces that translating from one domain to the other and back reconstructs the original image. It is defined as:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1] \tag{4.3}$$

Identity Loss The identity loss encourages the generator to preserve the image identity when given an input from the target domain, helping to maintain consistent colour mapping and improving the quality of generated images.

$$\mathcal{L}_{id}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[\|F(x) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\|G(y) - y\|_1] \tag{4.4}$$

Thus, the full CycleGAN objective is defined as:

$$\begin{aligned} \mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, X, Y) \\ + \lambda_{cyc} \cdot \mathcal{L}_{cyc}(G, F) \\ + \lambda_{id} \cdot \mathcal{L}_{cyc}(G, F) \end{aligned} \tag{4.5}$$

were λ_{cyc},λ_{id} control the relative importance of the different components. Setting cycle consistency loss weight $\lambda_{cyc} = 10.0$ is a common choice and effective to enforce structural consistency between domains. However it may constrain too much the generator, overpowering the adversarial loss and leading to unrealistic images in the target domain when the domains are highly distinct, as in this case. On the other hand, the identity loss weight λ_{id} controls the trade-off between enforcing appearance preservation and allowing flexibility in domain translation. Higher values encourage structural and colour fidelity, while lower values allow the generators to focus more on adversarial realism. It was set

empirically to $\lambda_{id} = 1.0$. This formulation balances the generation of realistic images in adversarial training and the preservation of the original content via cycle consistency. In preliminary experiments, we also tried replacing the MSE with the BCE in the adversarial loss, but this generated a lot of instability during training due to mode collapse and vanishing gradients, especially given the strong domain gap.

Registration Loss Functions

Mean Squared Error The Mean Squared Error (MSE), also known as the Sum of Squared Differences (SSD), when not normalized, measures the average squared intensity difference between the corresponding pixels in two images. It is defined as:

$$MSE = \frac{1}{|\Omega_I|} \sum_{p_i \in \Omega_I} (I_M(p_i) - I_F(p_i))^2 \tag{4.6}$$

where I_M and I_F denote the moving and fixed images respectively, $I_i(p), i = 1, 2$ is the pixel intensity at location p, Ω_I is the image domain, and $|\Omega_I|$ its cardinality.

MSE is widely used in image registration for its simplicity and computational efficiency. However, it assumes similar intensity distributions and is sensitive to intensity variations and noise, limiting its applicability in multimodal scenarios.

Root Mean Squared Difference The Root Mean Squared Difference (RMSD) is the square root of MSE:

$$RMSD = \sqrt{MSE} = \sqrt{\frac{1}{|\Omega_I|} \sum_{p_i \in \Omega_I} (I_M(p_i) - I_F(p_i))^2} \tag{4.7}$$

While mathematically equivalent to MSE in terms of optimisation, RMSD retains the original unit of pixel intensity, offering more intuitive interpretability. It is similarly employed as a loss function in deep image registration tasks when intensity consistency is assumed between modalities.

Mutual Information Mutual Information (MI) is a fundamental concept in information theory, originally introduced by C. Shannon [38] to quantify the amount of information shared between two random variables. It measures the reduction in uncertainty about one variable given knowledge of the other and captures all statistical dependencies between variables.

To understand MI, we first recall the concept of entropy. Also known as Shannon infor-

mation, entropy quantifies the average information contained in a random variable. For a discrete random variable X that takes values in a set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ and with probability distribution $\mathbb{P}(X)$ the entropy is defined as:

$$H(X) = -\sum_{x \in \mathcal{X}} \mathbb{P}(x)\ log\ \mathbb{P}(x)$$

Extending this to two variables X and Y, their joint entropy quantifies the uncertainty of the combined system:

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} \mathbb{P}(x,y)\ log\ \mathbb{P}(x,y)$$

while the conditional entropy of X given Y expresses the remaining uncertainty in X once Y is known:

$$H(X|Y) = -\sum_{x \in \mathcal{X}} \mathbb{P}(y) \sum_{y \in \mathcal{Y}} \mathbb{P}(x|y)\ log\ \mathbb{P}(x|y)$$

Based on these definitions, mutual information between X and Y measures how much knowing one variable reduces the uncertainty about the other. Formally, MI is defined as:

$$MI(X,Y) = H(X) + H(Y) - H(X,Y) \tag{4.8}$$

MI takes values in $[0, +\infty)$, and is zero if and only if the variables are statistically independent.

An alternative interpretation expresses MI as the Kullback–Leibler (KL) divergence between the joint distribution and the product of the marginals, namely:

$$MI(X,Y) = D_{\mathrm{KL}}\left(\mathbb{P}(X,Y) \,\middle\|\, \mathbb{P}(X)\mathbb{P}(Y)\right) = \sum_{x,y} \mathbb{P}(x,y) \log \frac{\mathbb{P}(x,y)}{\mathbb{P}(x)\mathbb{P}(y)} \tag{4.9}$$

This formulation highlights MI as a measure of how much the actual joint distribution deviates from the expectation under statistical independence.

Viola et al. [30] adapted the concept of mutual information for image registration by treating the intensity values of two images, I_M and I_F, as realizations of random variables. In this framework, the similarity between images is evaluated by analyzing their intensity distributions: $H(I_M)$ and $H(I_F)$ denote the entropies of the marginal intensity histograms, while $H(I_M, I_F)$ corresponds to the entropy of their joint histogram. The MI between

images is therefore expressed as:

$$MI(I_M, I_F) = H(I_M) + H(I_F) - H(I_M, I_F) \tag{4.10}$$

When the images are well aligned, their overlapping intensity distributions become more predictable, minimising joint entropy and thus maximizing MI. This property renders MI an especially powerful similarity measure for multimodal image registration, where simple metrics such as SSD fail due to complex intensity relationships. Equivalently, MI can be viewed as the reduction in uncertainty:

$$MI(I_M, I_F) = H(I_M) - H(I_M|I_F)$$
$$MI(I_M, I_F) = H(I_F) - H(I_F|I_M)$$

illustrating how knowledge of one image reduces the uncertainty about the other. In the context of training deep image registration networks, MI is commonly employed as a loss function by minimising its negative, enabling the network to optimise toward maximally informative alignments. However, leveraging MI in such settings introduces challenges not encountered in other metrics like the SSD. These include the need to compute differentiable, accurate estimates of joint and marginal distributions, increased computational complexity, and potential instabilities during optimisation. Various approaches have been proposed to mitigate these issues and ensure robust training with MI-based loss.

5 | Experiments and Results

This Chapter evaluates the effectiveness of the proposed pipeline in registering histological and MRI images of rodent brain and compares its performance with other state-of-the-art methods such as ProsRegNet [1].

5.1. Dataset

The imaging data we used to train and validate the proposed solution were collected at the *Istituto di Ricerche Farmacologiche Mario Negri*, Milan, Italy.
It includes in vivo T2-weighted MRIs and ex vivo histological brain images (Cresyl Violet stain) of sham and TBI rodents subjected to the CCI model, covering varying injury severity (Table 5.1). The MRI acquisitions consisted of coronal scans with a matrix size of $150 \times 150 \times 37$ voxels and a voxel spacing of $0.1 \times 0.1 \times 0.3$ mm^3. All rodents were sacrificed one month post-injury, and the brain tissue sections for histological analysis were cut at a thickness of 8 μm. The cohort consisted of male C57BL/6J mice, either 9 weeks (young) or 17 months (adult) at the time of injury.

The rodents belong to the following pathology groups:

- sham: animals that underwent the same surgical procedure as the TBI groups, but without induction of brain injury.
- single moderate TBI (smTBI): animals that sustained a CCI with a single injury of 0.5 mm depth.
- single severe TBI (ssTBI): animals that sustained a CCI with a single injury of 1.0 mm depth.

We preprocessed MRI volumes by applying N4 bias field correction to reduce intensity inhomogeneities, performing skull stripping to remove non-brain tissue, and applying spatial standardization. We then resampled the volumes to an isotropic resolution of $0.1 \times 0.1 \times 0.1$ mm^3 using nearest-neighbour interpolation and performed Z-score normalization to align intensity distributions across contrasts. Finally, we extracted coronal

brain sections from the preprocessed MRI volumes.
In parallel, we preprocessed histological images using the Macenko method for stain normalization, which standardizes colour appearance and reduces variability across samples.

Finally, since ground truth MRI-histology correspondences were not available, we established them through visual inspection based on overall anatomical similarity. This procedure may have introduced slight alignment uncertainty, potentially affecting the evaluation of registration accuracy, and should therefore be taken into account when interpreting experimental results. Table 5.1 provides a summary of the resulting collection of MRI and histological images across the three pathology groups.

Group	**n° rodents**	**n° T2-w MRI images**	**n° histology images**	**n° MRI-histology image pairs**
sham	12	252	163	110
smTBI	15	315	210	124
ssTBI	16	336	268	153

Table 5.1: Overview of the imaging data used in this study, reporting the number of rodents, T2-weighted MRI coronal sections, and histological images (CV stain) for each experimental group (sham, smTBI, ssTBI), as well as the number of MRI–CV histology image pairs.

Starting from imaging data summarised in Table 5.1, we defined two different datasets:

- D_{mono}: dataset composed of monomodal image pairs $(I_F, I_F \circ T)$ where both images belong to the same modality, either T2-w MRI or CV stained-histology. Each image I_F (from the T2-w MRI or Histology columns of Table 5.1) is paired with a randomly geometrically transformed version of itself $I_F \circ T$ to simulate acquisition and processing variability (Figure 5.1). Transformations include random rotations, translations, scaling, and thin-plate spline deformations to simulate acquisition and processing variability (see Section 2.1.3).
- D_{multi}: dataset composed of multimodal unregistered image pairs (I_F, I_M) (from the last column of Table 5.1), where I_F is a T2-w MRI coronal section and I_M the corresponding histological image section selected trough visual inspection by an expert operator within the institute.

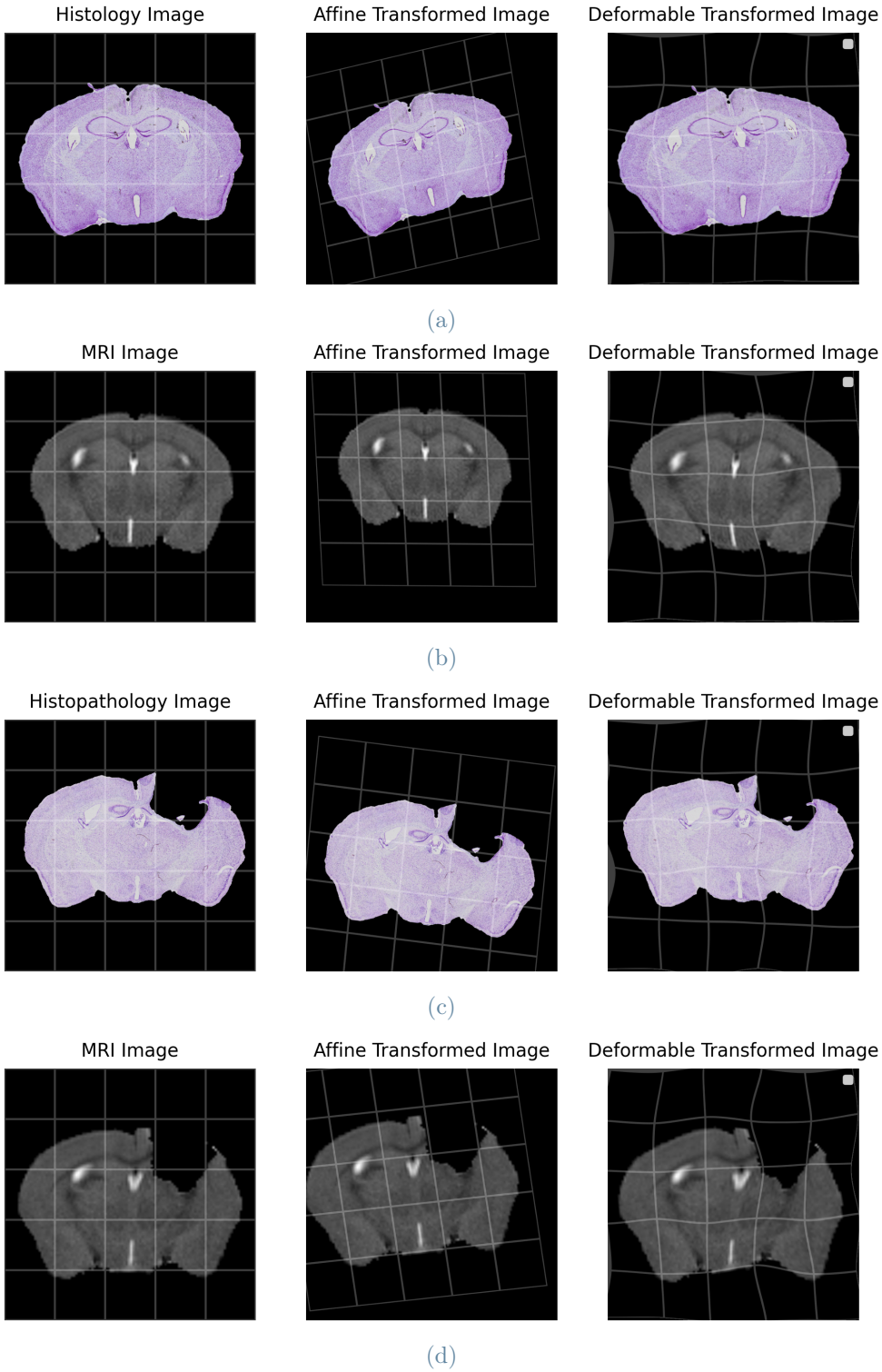


Figure 5.1: Examples of synthetic images generated with the affine and deformation model (overlayed with a grid) for sham and TBI rodents: (a, c) histological brain images and (b, d) MRI brain sections.

5.2. Models Configuration

We benchmarked the proposed pipeline against two reference implementations of ProsRegNet [1], trained under different settings. To ensure comparability, all registration networks were trained under a common optimisation protocol to ensure comparability, while the generative model (CycleGAN) followed its own dedicated configuration. Evaluation relied on 5-fold cross-validation described in detail in Section5.3. Below we outline the configurations for each model.

ProsRegNet To establish a baseline, we reproduced the ProsRegNet architecture by Shao et al. [1] and trained it on monomodal image pairs from D_{mono}. The model was optimised with the RMSD loss (Equation 4.7) to minimise pixel-wise differences. As shown by Shao et al., the model doesn't required paired MRI-histology slices to be trained and generalizes well even on multimodal image pairs.

Multimodal ProsRegNet To address the domain gap between monomodal training and multimodal evaluation, we trained the ProsRegNet architecture on multimodal image pairs from D_{multi} using the MI loss (Equation 4.9), which measures statistical dependence between modalities and is robust to intensity differences. Training hyperparameters were kept identical to the monomodal baseline so that performance differences reflect the change in input modality and loss.

Proposed pipeline The proposed pipeline employs a CycleGAN to translate histological images into MRI-like representations, thereby reducing the multimodal registration problem to a monomodal one.
We first trained the CycleGAN in an unpaired, fully unsupervised setting on the dataset D_{multi} using a modified cycle consistency loss (Equation 4.5, $\lambda_{cyc} = 10.0$, $\lambda_{id} = 1.0$) for 200 epochs with early stopping, using the Adam optimiser with an initial learning rate of 2×10^{-4} and a batch size of 1.
Subsequently, we trained the two-stage registration network on the generated MRI-like images, with CycleGAN weights frozen. We ran this stage optimising the RMSD loss between the warped MRI-like moving images and fixed MRI sections.

Unless otherwise specified, all registration networks were trained for 100 epochs with the Adam optimiser (initial learning rate 1×10^{-3}, batch size 16). A learning-rate scheduler reduced the learning rate by 5% when the validation loss did not improve for 5 consecutive epochs (minimum 1×10^{-6}). No early stopping was used for the registration networks.

5.3. Evaluation Method

5.3.1. K-fold Cross Validation

To evaluate the robustness of each model, we adopted a k-fold cross-validation strategy with k=5. This value is a common choice for limited datasets, as it provides a trade-off between obtaining a reliable variance estimate and retaining sufficient training data in each fold [39].
In this approach, the dataset is split into five folds; at each iteration, four folds are used for training and the remaining one for testing (Figure 5.2), and results are averaged across folds. Folds were generated under the following constraints:

- All images (both MRIs and histological images) from a given rodent were assigned to the same fold, ensuring that the model is never evaluated on data from a previously "seen" subject.
- Rodents with only histological images or only MRI were included exclusively in the training folds, never in the test fold, since evaluation requires paired multimodal images.

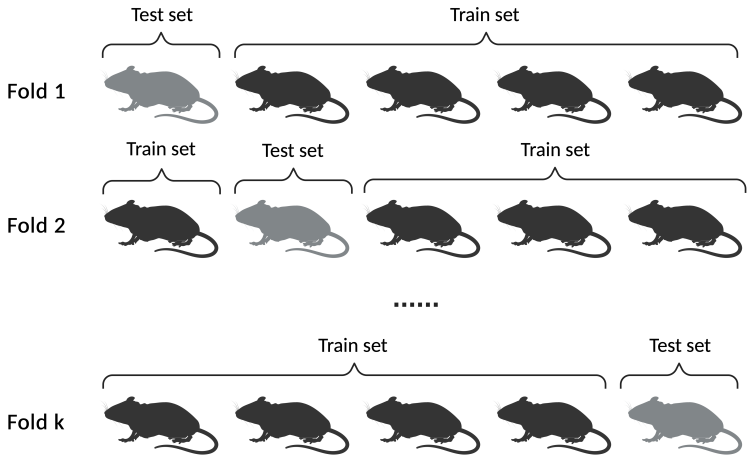


Figure 5.2: Experimental design of the k-fold cross-validation.

5.3.2. Evaluation Metrics

To comprehensively assess registration performance, we used the Dice similarity coefficient to evaluate the overlap of brain masks and the mean error distance to quantify point-to-point alignment accuracy between the transformed histological images and the corresponding MRI sections. These complementary metrics allow us to measure both

global and local registration accuracy.

Dice similarity coefficient The Dice similarity coefficient (DSC) measures the relative overlap of the brain masks of the two aligned images. It is defined as:

$$DSC = \frac{2|(M_M \circ T^*) \cap M_F|}{|(M_M \circ T^*)| + |M_F|} \tag{5.1}$$

where T^* is the estimated geometric transformation, $M_M \circ T^*$ denotes the brain mask of the moving image (histology) after applying the transformation, M_F the brain mask of the fixed image (MRI) and $|\cdot|$ is the cardinality of a set.

Mean error distance of landmark points The mean error distance (MED) measures the accuracy of point-to-point correspondences obtained after image registration.
Let T^* be the geometric transformation estimated during the registration process. Given N landmark points (LP) on the fixed MRI section and corresponding moving histological image, denoted by $(p_1, p'_1), ..., (p_N, p'_N)$, the MED for the transformation T^* is defined as:

$$MED = \frac{1}{N}\sum_{i=1}^{N} ED_i = \frac{1}{N}\sum_{i=1}^{N} \|p_i - T^*(p'_i)\|_2 \tag{5.2}$$

To evaluate this metric on the test set, we need to define relevant landmarks on the brain area. We identified these points according to the following criteria:

- relevant from an anatomical and pathological point of view,
- clearly visible and distinguishable across both imaging modalities,
- well distributed across the brain area,
- independently and consistently annotated in both image modalities.

Based on these, we defined 8 landmark points per brain. Table 5.2 summarises the anatomical definition of the selected landmarks for sham and TBI rodents, Figure 5.3 illustrates their locations.
For both MRI sections and histological images in our dataset we manually annotated these landmark points. In some cases, less than 8 points were available due to tissue distortion, folding, tearing, or other artifacts, which could affect the visibility of any of the defined landmarks. Finally, we computed the MED between corresponding available points after registration.

LP	Sham rodents	TBI rodents
1	Superior corner of left ventricle	Right extreme of corpus callosum on the lesions
2	Inferior corner of left ventricle	Left extreme of corpus callosum on the lesions
3	Superior corner of right ventricle	–
4	Inferior corner of right ventricle	–
5	Superior corner of third ventricle	–
6	Inferior corner of third ventricle	–
7	Left extreme of corpus callosum	–
8	Right extreme of corpus callosum	–

Table 5.2: Anatomical definition of landmark points (LP) in sham and TBI rodents. For TBI rodents, '–' indicates the same landmark definition as for sham.

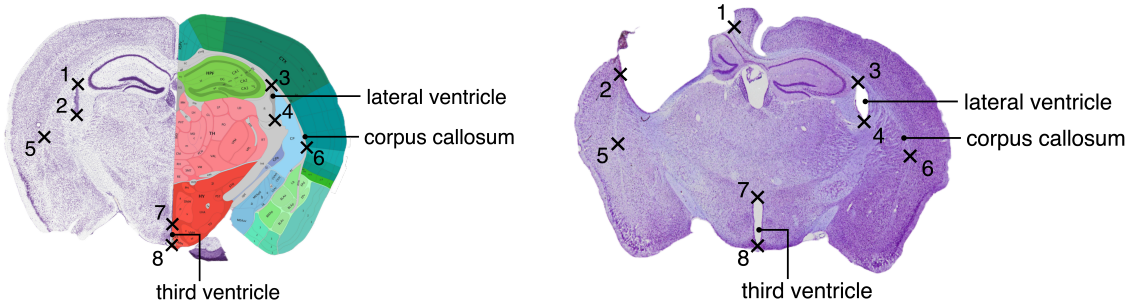


Figure 5.3: visualisation of the landmark points used for MED evaluation, shown on a sham mouse brain section. Image modified from Allen Brain ATLAS [40].

5.4. Experimental Results

We evaluated the models according to the 5-fold cross-validation strategy described in Section 5.3.1 and trained them with the configuration detailed in Section 5.2. Hereafter, we report DSC and MED (mm) as mean±SD within each pathology group, averaged across folds. A comparison across approaches is reported in Section 5.4.1.

ProsRegNet Figure 5.5 presents boxplots of these metrics, providing a visual overview of variability and distribution across folds and pathology groups. The results (Table 5.3) show that while the model captures coarse correspondences, performance is limited by the MRI–histology domain gap. Variability in MED across landmarks reflects the difficulty of multimodal alignment when contrast and intensity differences are not explicitly handled.

Rodent Group	DSC	MED LP1	MED LP2	MED LP3	MED LP4	MED LP5	MED LP6	MED LP7	MED LP8
sham	0.95±0.02	0.30±0.23	0.30±0.17	0.30±0.15	0.31±0.16	0.23±0.14	0.20±0.12	0.42±0.23	0.41±0.23
smTBI	0.93±0.03	0.33±0.21	0.37±0.21	0.41±0.31	0.42±0.23	0.33±0.22	0.28±0.18	0.41±0.26	0.51±0.31
ssTBI	0.93±0.02	0.27±0.13	0.35±0.17	0.33±0.18	0.37±0.23	0.26±0.14	0.26±0.14	0.38±0.21	0.44±0.21

Table 5.3: ProsRegNet registration pipeline performance across experimental groups (mean±SD for DSC and MED [mm]).

Multimodal ProsRegNet Table 5.4 summarises the metrics for the multimodal registration baseline trained with MI loss, across folds and pathology groups, and Figure 5.6 shows the corresponding distributions. This setting mitigates intensity and contrast differences and generally improves local alignment stability relative to the monomodal ProsRegNet baseline.

Rodent Group	DSC	MED LP1	MED LP2	MED LP3	MED LP4	MED LP5	MED LP6	MED LP7	MED LP8
sham	0.95±0.02	0.33±0.18	0.28±0.18	0.37±0.16	0.27±0.16	0.25±0.19	0.22±0.10	0.40±0.20	0.42±0.20
smTBI	0.93±0.02	0.31±0.17	0.39±0.16	0.43±0.23	0.36±0.23	0.26±0.15	0.25±0.15	0.38±0.20	0.42±0.19
ssTBI	0.93±0.02	0.29±0.14	0.28±0.16	0.40±0.22	0.42±0.22	0.24±0.15	0.25±0.16	0.36±0.18	0.46±0.19

Table 5.4: ProsRegNet registration pipeline trained in a multimodal setting using MI loss across experimental groups (mean±SD for DSC and MED [mm]).

Proposed pipeline Table 5.5 reports the quantitative results for the proposed CycleGAN-based pipeline, and Figure 5.7 illustrates per-fold distributions, across pathology groups. On average, the pipeline achieves high accuracy within each group, consistent with effective alignment after translation. At the same time, the results reveal variability across samples, suggesting that while the integration of the histology-to-MRI translation step benefits many cases, its effectiveness is not uniform across the dataset.

Rodent Group	DSC	MED LP1	MED LP2	MED LP3	MED LP4	MED LP5	MED LP6	MED LP7	MED LP8
sham	0.96±0.01	0.31±0.20	0.29±0.18	0.29±0.19	0.27±0.15	0.26±0.18	0.27±0.15	0.36±0.22	0.32±0.17
smTBI	0.94±0.02	0.25±0.16	0.28±0.18	0.36±0.23	0.39±0.27	0.27±0.19	0.22±0.14	0.35±0.21	0.43±0.27
ssTBI	0.94±0.01	0.27±0.14	0.25±0.15	0.29±0.20	0.39±0.33	0.22±0.15	0.21±0.15	0.35±0.23	0.49±0.35

Table 5.5: Proposed CycleGAN-based registration pipeline performance across experimental groups (mean±SD for DSC and MED [mm]).

Qualitative Evaluation of Histology-to-MRI Translation Step To better interpret the registration results, we qualitatively examined the CycleGAN translations used

as input to the registration network. As shown in Figure 5.4, the generated MRI-like images preserve anatomical structures in both sham and TBI rodents and capture the characteristic MRI contrast while reflecting the underlying histological information. This translation step reduces the domain gap between modalities and facilitates alignment. At the same time, variability in the quality of the generated images may account for the spread observed in the quantitative metrics across samples (Table 5.5).

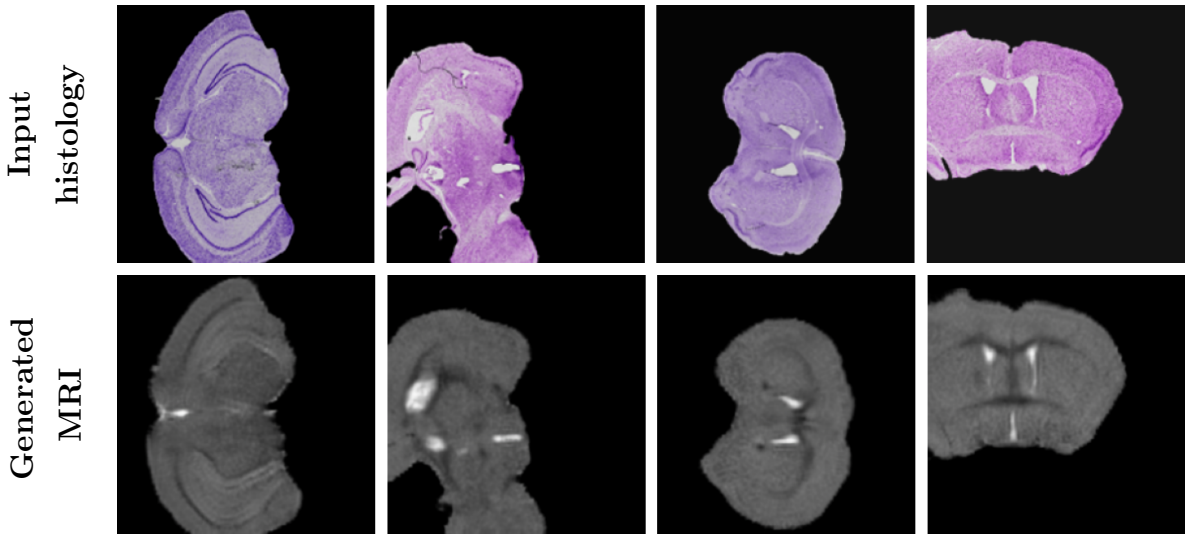
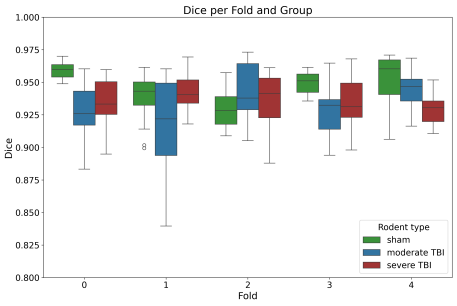
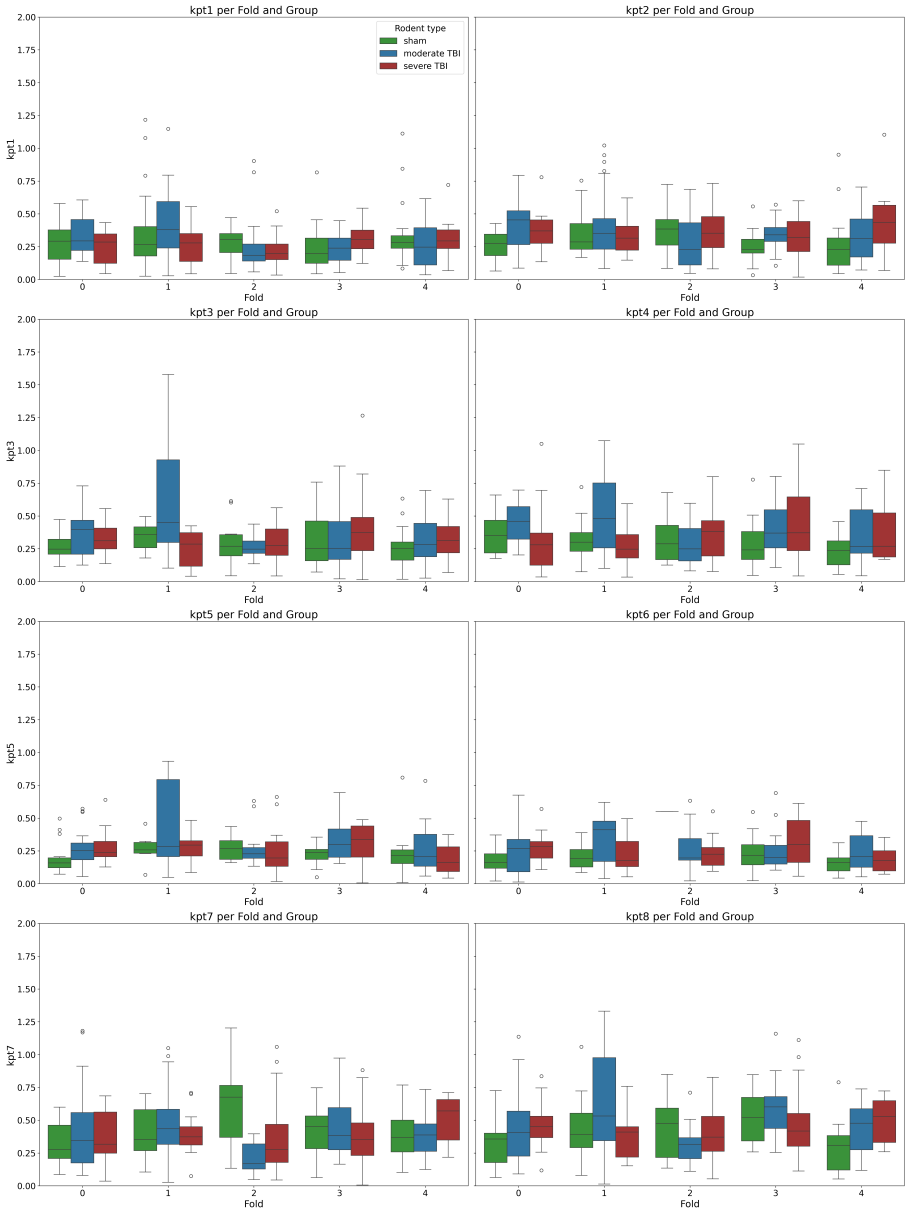


Figure 5.4: Histology-to-MRI translation with CycleGAN. From top to bottom: real histological images and the corresponding MRI-like generated images.

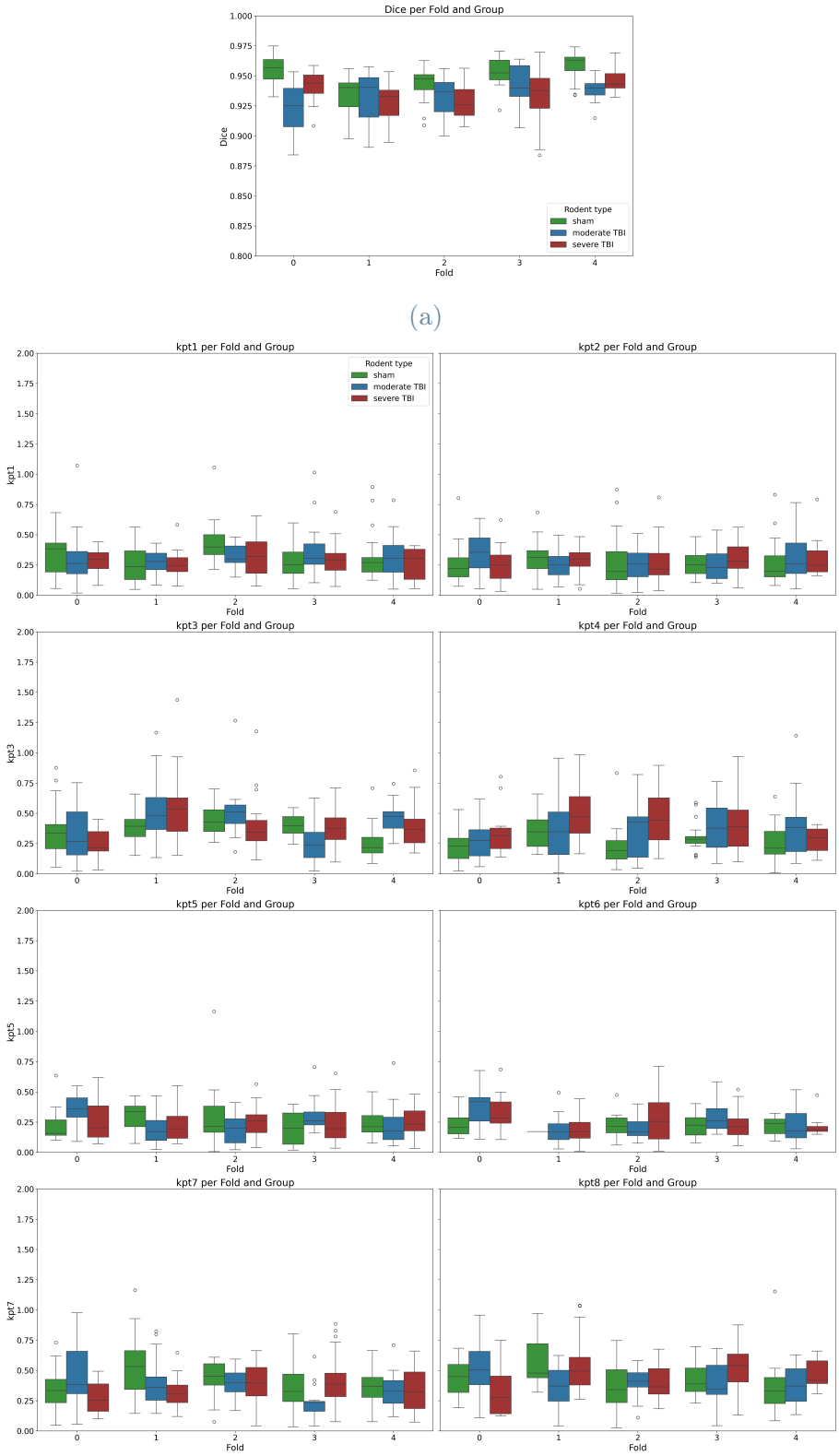


(a)



(b)

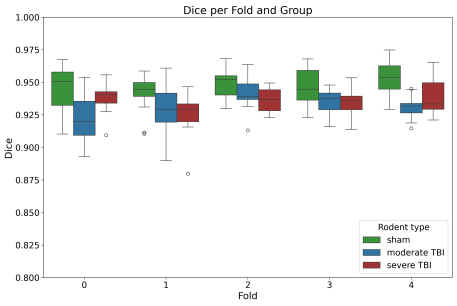
Figure 5.5: Boxplots of evaluation metrics grouped by fold and pathology for the ProsRegNet pipeline trained with RMSD loss: (a) DSC of brain masks and (b) MED of landmarks on the test set.



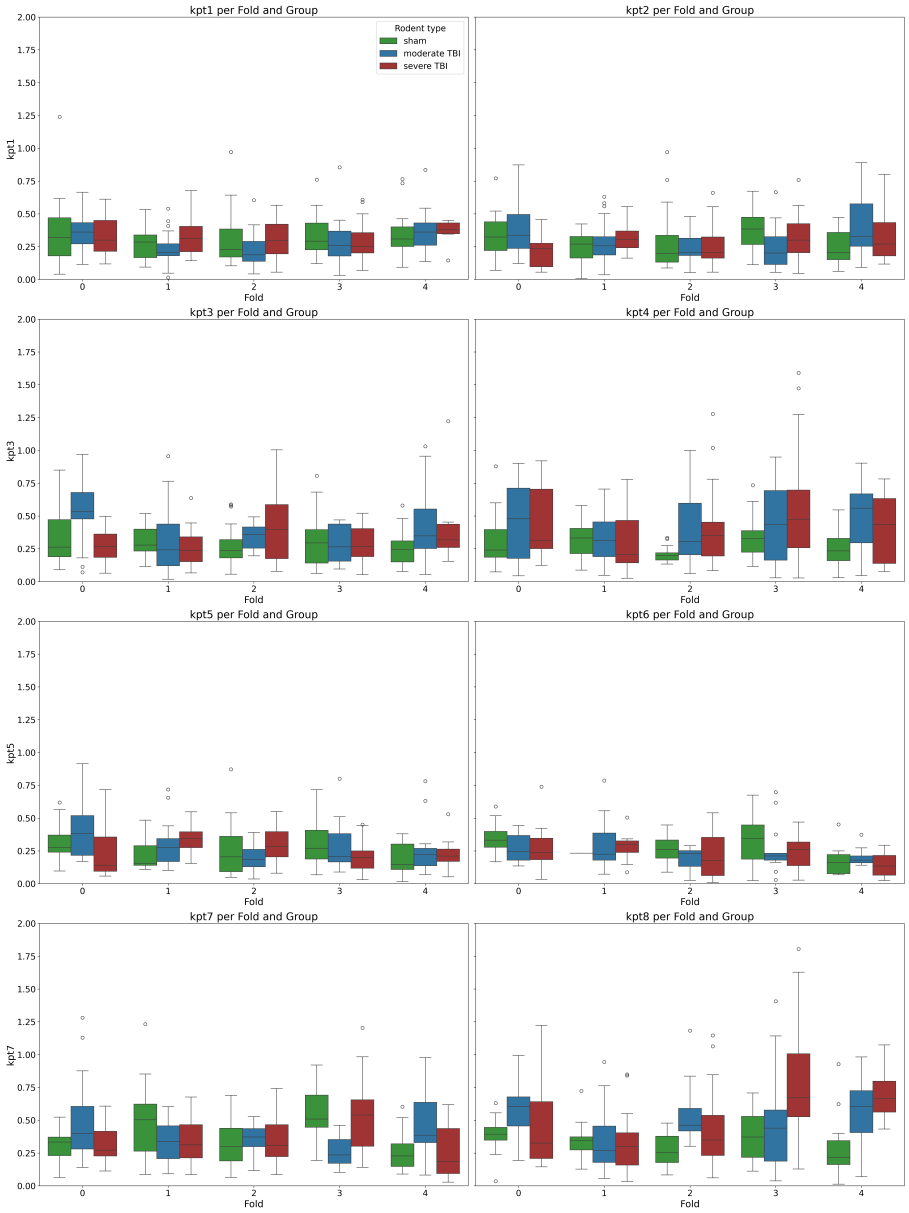
(a)

(b)

Figure 5.6: Boxplots of evaluation metrics grouped by fold and pathology for the Pros-RegNet pipeline trained on multimodal image pairs with MI loss: (a) DSC of brain masks and (b) MED of landmarks on the test set.



(a)



(b)

Figure 5.7: Boxplots of evaluation metrics grouped by fold and pathology for the proposed CycleGAN-based pipeline trained with RMSD loss: (a) DSC of brain masks and (b) MED of landmarks on the test set.

5.4.1. Model Comparison

It is important to note that the evaluation of registration performance in this study is influenced by the absence of ground-truth MRI–histology correspondences. Since image pairs were established through visual inspection based on anatomical similarity, some degree of alignment uncertainty is unavoidable. This limitation may contribute to variability in the reported metrics and should be considered when interpreting the results. Nevertheless, the consistent improvements observed across baselines and experimental groups indicate that the proposed pipeline provides a robust advantage despite this source of uncertainty.

Table 5.6 summarises the quantitative accuracies of the three methods in registering MRI and histological sections. The proposed CycleGAN-based registration pipeline achieves the lowest MED, while multimodal ProsRegNet attains the highest DSC.

Rodent Group	Registration Approach	Dice Coefficient	Mean Error Distance (mm)
sham	ProsRegNet	0.947	0.309
	Multimodal ProsRegNet	0.951	0.318
	Proposed Pipeline	**0.957**	**0.296**
smTBI	ProsRegNet	0.930	0.383
	Multimodal ProsRegNet	0.933	0.338
	Proposed Pipeline	**0.936**	**0.319**
ssTBI	ProsRegNet	0.936	0.333
	Multimodal ProsRegNet	0.934	0.350
	Proposed Pipeline	**0.936**	**0.309**

Table 5.6: Comparison of the registration approaches through the evaluation metrics grouped by rodent group.

Figure 5.8 shows representative registration results for one rodent from each experimental group. The images illustrate that the proposed CycleGAN-based model accurately predicts the geometric transformation, aligning the histological images to the corresponding MRI sections with high precision. Both quantitative and qualitative results highlight overall performance. Some misalignments occur in certain samples, likely due to inherent variability in the dataset, while visual inspection highlights both successful alignments and occasional errors, providing an intuitive understanding of registration performance.

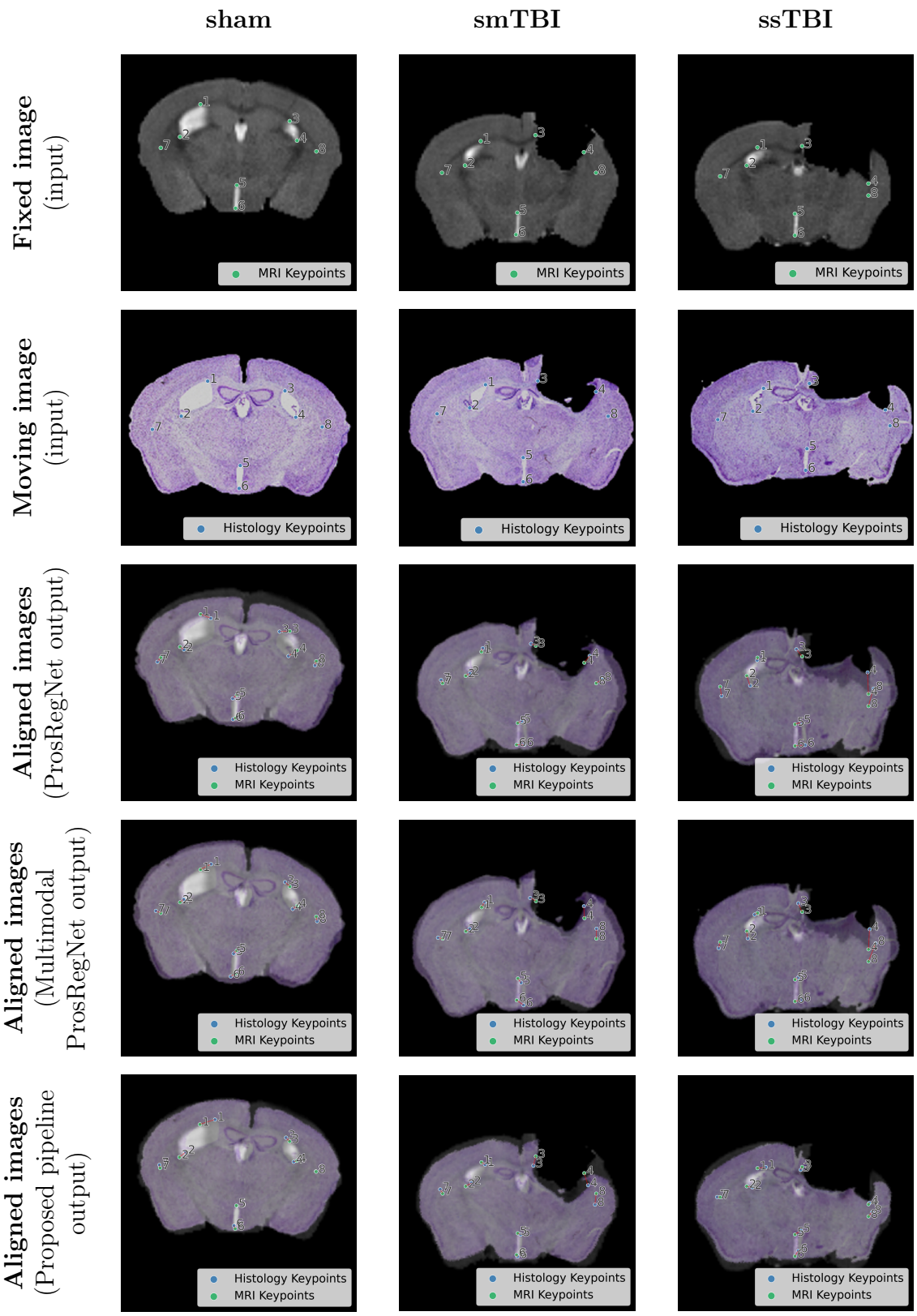


Figure 5.8: Registration results of three representative rodents, one for each pathology group. The green and blue dots represent landmarks on the MRI and histological sections, respectively. The red lines connect corresponding landmarks on aligned image pairs, illustrating registration error.

6 | Conclusions and Future Work

In this thesis, we developed a deep learning–based pipeline for the multimodal registration of *in vivo* T2-weighted MRI and *ex vivo* histological images of rodent brains following traumatic brain injury. The pipeline addresses the inherent challenges of aligning modalities with different resolutions, contrasts, and tissue deformations.

The proposed framework combines a CycleGAN-based cross-modality translation with a two-stage registration network. By translating histological images into MRI-like representations, the multimodal registration problem is reduced to a monomodal one. Quantitative evaluation using the Dice Similarity Coefficient and the Mean Error Distance showed consistent, sub-millimetre alignment across pathology groups (DSC $\approx$ 0.94–0.96, MED $\approx$ 0.30 mm). As detailed in Section 5.4.1, the proposed pipeline achieved the highest DSC together with the lowest MED, when compared with both the ProsRegNet baseline and the ProsRegNet variant trained with MI loss.
Qualitative inspection of the translated MRI-like images indicated good preservation of anatomical structures and generation of characteristic MRI contrast, which likely facilitated the downstream registration. Nonetheless, the quality of translations varied across samples, which is consistent with the observed spread in the quantitative metrics.

Despite these promising results, this study has some limitations. First, the lack of ground-truth MRI–histology correspondences introduces unavoidable alignment uncertainty. Second, the variability in the quality of CycleGAN-generated images can reduce the consistency of registration performance across samples. Third, despite including multiple injury severities and ages, the dataset employed contains artifacts (e.g., tears, folds, distortions) that may have affected model training and evaluation. Finally, manual landmark annotation, while necessary for evaluation, may introduce observer bias.

Looking forward, several directions could further improve and extend this work. Expanding the dataset with additional rodents, histological stains, imaging modalities (T1-W MRI or CT) or different neurological pathologies would increase model robustness and generalization. Refining the translation network using attention-based or transformer architectures could improve the fidelity and consistency of the MRI-like images, enhancing

registration accuracy. Integrating automatic landmark detection methods would reduce observer bias and streamline evaluation. Finally, developing more sensitive evaluation metrics capable of capturing local misalignments or structural preservation could provide deeper insights into registration quality.

Overall, the proposed pipeline represents a significant step toward accurate multimodal registration of MRI and histological images in rodent brain studies. By reducing the domain gap and enabling more precise alignment, this framework facilitates quantitative analysis of pathological changes following traumatic brain injury and lays the foundation for future research in advanced multimodal registration techniques.

Bibliography

[1] Wei Shao, L. Banh, C. A. Kunder, R. E. Fan, S. J. C. Soerensen, J. B. Wang, N. C. Teslovich, N. Madhuripan, A. Jawahar, P. Ghanouni, J. D. Brooks, G. A. Sonn, and M. Rusu. ProsRegNet: A deep learning framework for registration of MRI and histopathology images of the prostate. *Medical Image Analysis*, 68:101919, Feb 2021. doi: 10.1016/j.media.2020.101919.

[2] Wei Shao, Sulaiman Vesal, Simon J.C. Soerensen, Indrani Bhattacharya, Negar Golestani, Rikiya Yamashita, Christian A. Kunder, Richard E. Fan, Pejman Ghanouni, James D. Brooks, Geoffrey A. Sonn, and Mirabela Rusu. RAPHIA: A deep learning pipeline for the registration of MRI and whole-mount histopathology images of the prostate. *Computers in Biology and Medicine*, 173:108318, May 2024. doi: 10.1016/j.compbiomed.2024.108318.

[3] Xingyu Jiang, Jiayi Ma, Guobao Xiao, Zhenfeng Shao, and Xiaojie Guo. A review of multimodal image matching: Methods and applications. *Information Fusion*, 73: 22–71, Feb 2021. doi: 10.1016/j.inffus.2021.02.012.

[4] Yabo Fu, Yuxuan Lei, Tian Wang, Walter J. Curran, Tian Liu, and Xiaofeng Yang. Deep learning in medical image registration: A review. *Physics in Medicine & Biology*, 65(20):20TR01, Oct 2020. doi: 10.1088/1361-6560/ab843e.

[5] David K. Menon, Karen Schwab, Daniel W. Wright, Andrew I. Maas, Demographics, Clinical Assessment Working Group of the International, Interagency Initiative toward Common Data Elements for Research on Traumatic Brain Injury, and Psychological Health. Position statement: Definition of traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 91(11):1637–1640, Nov 2010. doi: 10.1016/j.apmr.2010.05.017.

[6] Andrew I. R. Maas, David K. Menon, Geoffrey T. Manley, Michael Abrams, Cecilia Åkerlund, Nada Andelic, Marcel Aries, Tom Bashford, Michael J. Bell, Yelena G. Bodien, and et al. Traumatic brain injury: Progress and challenges in prevention,

clinical care, and research. *The Lancet Neurology*, 21(11):1004–1060, Nov 2022. doi: 10.1016/S1474-4422(22)00309-X.

[7] Nino Stocchetti and Elisa R. Zanier. Chronic impact of traumatic brain injury on outcome and quality of life: A narrative review. *Critical Care*, 20(1):148, Jun 2016. doi: 10.1186/s13054-016-1318-1.

[8] Kim E. Saatman, Ann C. Duhaime, R. Bullock, Andrew I. R. Maas, Alex Valadka, Geoffrey T. Manley, and et al. Classification of traumatic brain injury for targeted therapies. *Journal of Neurotrauma*, 25(7):719–738, Jul 2008. doi: 10.1089/neu.2008.0586.

[9] Yan Xiong, Asim Mahmood, and Michael Chopp. Animal models of traumatic brain injury. *Nature Reviews Neuroscience*, 14(2):128–142, Feb 2013. doi: 10.1038/nrn3407.

[10] Ang Guo, Zhiyu Chen, Yinzhong Ma, Yueguang Lv, Huanhuan Yan, Fang Li, Yao Xing, Qian Luo, and Hairong Zheng. SOmicsFusion: Multimodal coregistration and fusion between spatial metabolomics and biomedical imaging. *Artificial Intelligence Chemistry*, 2(1):100058, 2024. ISSN 2949-7477. doi: 10.1016/j.aichem.2024.100058.

[11] Francesca Pischiutta, Elisa Micotti, James R. Hay, Isabella Marongiu, Elisa Sammali, Davide Tolomeo, Gianluigi Vegliante, Nino Stocchetti, and et al. Single severe traumatic brain injury produces progressive pathology with ongoing contralateral white matter damage one year after injury. *Experimental Neurology*, 300:167–178, Feb 2018. doi: 10.1016/j.expneurol.2017.11.003.

[12] Syed Ahmed Taqi, Syed Abdus Sami, Lateef Begum Sami, and Syed Ahmed Zaki. A review of artifacts in histopathology. *Journal of Oral and Maxillofacial Pathology*, 22(2):279, August 2018. doi: 10.4103/jomfp.JOMFP_125_15.

[13] Nicholas J. Tustison, Brian B. Avants, Paul A. Cook, Yuanjie Zheng, Alan Egan, Paul A. Yushkevich, and James C. Gee. N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, Jun 2010. doi: 10.1109/TMI.2010.2046908.

[14] Nicholas J. Tustison, Philip A. Cook, Andrew J. Holbrook, Hans J. Johnson, John Muschelli, Gabriel A. Devenyi, and et al. The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports*, 11:9068, Apr 2021. doi: 10.1038/s41598-021-87564-6.

[15] Marcello De Salvo. DS4HB-Workshop-CNNs. `https://github.com/MarcelloDeSalvo/DS4HB-Workshop-CNNs`, 2025.

[16] Marc Macenko, Marc Niethammer, J. S. Marron, David Borland, John T. Woosley, Xiaojun Guan, Charles Schmitt, and Nancy E. Thomas. A method for normalizing histology slides for quantitative analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1107–1110, Boston, MA, USA, Jun 2009. doi: 10.1109/ISBI.2009.5193250.

[17] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

[18] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding transfer learning for medical imaging. *arXiv preprint arXiv:1902.07208*, 2019. URL `https://arxiv.org/abs/1902.07208`.

[19] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2015. URL `https://arxiv.org/abs/1409.1556`.

[20] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*, 2018. URL `https://arxiv.org/abs/1611.07004`.

[21] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2020. URL `https://arxiv.org/abs/1703.10593`.

[22] Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21(11):977–1000, 2003. doi: 10.1016/S0262-8856(03)00137-9.

[23] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, Jul 2013. doi: 10.1109/TMI.2013.2265603.

[24] J.B.Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, Mar 1998. doi: 10.1016/s1361-8415(01)80026-8.

[25] Xuxin Chen, Ximin Wang, Ke Zhang, Kar-Ming Fung, Theresa C. Thai, Kathleen Moore, Robert S. Mannel, Hong Liu, Bin Zheng, and Yuchen Qiu. Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, 79:102444, 2022. ISSN 1361-8415. doi: 10.1016/j.media.2022.102444.

[26] David G. Lowe. Distinctive image features from scale-invariant keypoints. *In-*

ternational Journal of Computer Vision, 60:91–110, Nov 2004. doi: 10.1023/B:VISI.0000029664.99615.94.

[27] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF: Speeded up robust features. In Ales Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, volume 3951 of *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 2006. doi: 10.1007/11744023_32.

[28] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, and Pascal Fua. LIFT: Learned invariant feature transform. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, volume 9910 of *Lecture Notes in Computer Science*, pages 467–483. Springer, Cham, 2016. doi: 10.1007/978-3-319-46466-4_28.

[29] J. Ashburner and K. Friston. Multimodal image coregistration and partitioning—a unified framework. *NeuroImage*, 6(3):209–217, 1997. ISSN 1053-8119. doi: 10.1006/nimg.1997.0290.

[30] Paul Viola and William M. III Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24:137–154, Sep 1997. doi: 10.1023/A:1007958904918.

[31] Gang Wang, Zhicheng Wang, Yufei Chen, and Weidong Zhao. Robust point matching method for multimodal retinal image registration. *Biomedical Signal Processing and Control*, 19:68–76, 2015. doi: 10.1016/j.bspc.2015.03.004. URL `https://doi.org/10.1016/j.bspc.2015.03.004`.

[32] Z. Ghassabi, J. Shanbehzadeh, A. Sedaghat, and et al. An efficient approach for robust multimodal retinal image registration based on ur-sift features and piifd descriptors. *Journal of Image and Video Processing*, 2013:25, 2013. doi: 10.1186/1687-5281-2013-25. URL `https://doi.org/10.1186/1687-5281-2013-25`.

[33] Wei Shao, Indrani Bhattacharya, Simon J. C. Soerensen, Christian A. Kunder, Jeffrey B. Wang, Richard E. Fan, Pejman Ghanouni, James D. Brooks, Geoffrey A. Sonn, and Mirabela Rusu. Weakly supervised registration of prostate MRI and histopathology images. In M. de Bruijne and et al., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, volume 12904 of *Lecture Notes in Computer Science*, pages 86–96. Springer, Cham, 2021. doi: 10.1007/978-3-030-87202-1_10.

[34] Chen Qin, Bibo Shi, Rui Liao, Tommaso Mansi, Daniel Rueckert, and Ali Kamen. Unsupervised deformable registration for multi-modal images via disentangled representations, 2019. URL `https://arxiv.org/abs/1903.09331`.

[35] Amaury Leroy, Kumar Shreshtha, Marvin Lerousseau, Théophraste Henry, Théo Estienne, Marion Classe, Nikos Paragios, Vincent Grégoire, and Eric Deutsch. Magnetic resonance imaging virtual histopathology from weakly paired data. In *Proceedings of the MICCAI Workshop on Computational Pathology*, volume 156, pages 140–150, 2021. URL `https://proceedings.mlr.press/v156/leroy21a.html`.

[36] Jiayuan Wang, Q.M. Jonathan Wu, and Farhad Pourpanah. DC-cycleGAN: Bidirectional CT-to-MR synthesis from unpaired data. *Computerized Medical Imaging and Graphics*, 108:102249, Sep 2023. doi: 10.1016/j.compmedimag.2023.102249. URL `http://dx.doi.org/10.1016/j.compmedimag.2023.102249`.

[37] Yang Heng, Yinghua Ma, Fiaz Gul Khan, Ahmad Khan, and Hui Zeng. HLSNC-GAN: Medical image synthesis using hinge loss and switchable normalization in CycleGAN. *IEEE Access*, 12:55448–55464, 2024. doi: 10.1109/ACCESS.2024.3390245.

[38] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July 1948.

[39] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer, 2013. ISBN 978-1-4614-7137-0. doi: 10.1007/978-1-4614-7138-7.

[40] Allen Institute for Brain Science. Allen brain atlas, 2010. URL `https://portal.brain-map.org`. Accessed: 2025-08-21.

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude to Professor Boracchi for the extraordinary opportunity to work on my thesis in such an exciting field and for his expert guidance and patience throughout this journey.
My sincere thanks go to Dr. Elisa Zanier and Dr. Federico Moro for welcoming me into their research team. Their trust, invaluable advice, and constant support were essential during the most challenging months of this work. I am also grateful to Marcello, whose steady encouragement and practical insights were crucial to the success of this thesis.
Finally, I thank my parents for their unfailing support, my sister Alice for her inspiring determination, and all my friends and my boyfriend for sharing this journey with me. Completing this thesis has been a rewarding experience, and I truly appreciate everyone who made it possible.