POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

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

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Author: Erica Bistacchia

Advisor: Prof. Giacomo Boracchi

Co-advisor: Dr. Elisa R. Zanier, Dr. Federico Moro, Marcello De Salvo

Academic year: 2024-2025

1. Introduction

Traumatic brain injury (TBI) is a major global health concern, defined as an alteration in brain function or other evidence of neuropathology resulting from an external mechanical force [1]. It affects 50-60 million individuals annually and is associated with both acute and progressive secondary damage, including lesion expansion, edema and brain atrophy.

Preclinical models of TBI allow investigation of mechanisms associated with disease progression via histological assessments at the microscopic level. At the same time, translational *in vivo* approaches, such as magnetic resonance imaging (MRI), lack the resolution of microscopy but provide high translational value, even though the underlying changes can be difficult to interpret. For this reason, aligning two or more images of the same anatomical region acquired with diverse imaging modalities through multimodal image registration (MMIR) methods enables integration and comparison of complementary information across modalities. In this context, MMIR could facilitate the detection of structural damage, improve the evaluation of temporal changes, and enhance understanding of microscopic alterations that contribute to macroscopic changes after TBI [2].

Despite its importance, MMIR remains challenging due to several factors, mainly because different modalities, such as MRI and histology, may differ markedly in contrast and structure, often requiring complex nonlinear transformations for accurate alignment. Additionally, tissue preparation for histology often introduces artifacts, including morphological distortions, stain variations, shrinkage, stretching and tearing, or even tissue loss, making the registration process even more difficult. These challenges, together with the scarcity of publicly available datasets and the difficulty of acquiring large datasets due to time-consuming acquisition processes, make the registration task in the TBI context particularly complex to address using deep learning approaches.

Most state-of-the-art methods are designed for modalities exhibiting linear intensity relationships, such as T1- and T2-weighted MRI or MRI-CT scans, and thus cannot be directly applied to MRI-histology image registration. Furthermore, pipelines developed for MRI–histology alignment, primarily for the human prostate [3, 4], are not directly transferable to rodent TBI studies due to substantial

anatomical differences.
To address the lack of preclinical studies and the specific challenges of MMIR in rodent TBI, as identified by the *Traumatic Brain Injury and Neuroprotection Laboratory* at the *Istituto di Ricerche Farmacologiche Mario Negri*, Milan, Italy, this thesis introduces a fully unsupervised pipeline for multimodal registration of brain MRI and histology (Figure 2). The approach leverages a CycleGAN to translate histological images into MRI-like representations, converting the multimodal registration problem into a monomodal one. The resulting images are then aligned using a two-stage deep learning network, consisting of a global affine registration followed by a non-rigid thin-plate spline (TPS) transformation to refine local misalignments.

2. Problem Formulation

Image registration (IR), also known as image matching, is the task of spatially aligning two images $I_M, I_F \in \mathbb{R}^{H \times W \times C}$, referred to as the *moving* and *fixed images*, respectively. For simplicity, we assume both images have the same dimensions. The objective of IR is to estimate a geometric transformation T mapping each pixel location (i, j) in I_M to a new location (i', j') in I_F. Once the transformation is estimated, the moving image is resampled according to T, yielding a registered image $I_M \circ T$ aligned with I_F. This process is illustrated in Figure 1.

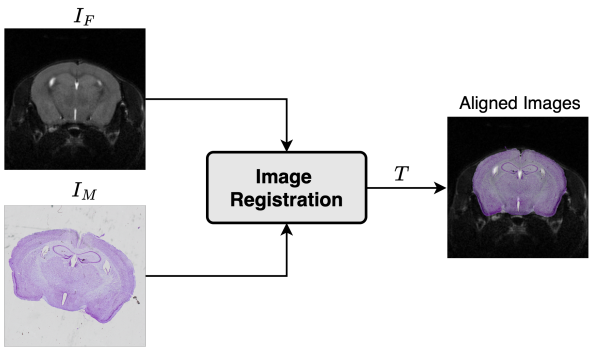

Figure 1: MRI-Histology registration of a mouse brain: given an MRI section (I_F) and a histological image (I_M), a transformation T is estimated to align I_M to I_F, enabling multimodal analysis.

Image registration is commonly formulated as an optimization problem, in which the goal is to find the transformation T^* that minimizes a cost function $\mathcal{C}$:

$$T^* = \arg\min_{T \in \mathcal{T}} \mathcal{C}(T; I_M, I_F) \tag{1}$$

Here, $\mathcal{T}$ denotes the space of allowable transformations, $\mathcal{C}(T; I_M, I_F)$ quantifies the spatial misalignment between $I_M \circ T$ and I_F. The choice of $\mathcal{C}$ depends on the method and typically reflects the modality, features, or intensity characteristics of the input images.
In deep learning, *unsupervised* training is typically performed on a dataset of unaligned image pairs from different modalities, $\mathcal{D}_{unsup} = \{(I_F, I_M)_k\}_{k=1}^N$, aiming to estimate a transformation T minimizing a cost function $\mathcal{C}$ without access to ground truth transformations. Conversely, *supervised* training relies on datasets of aligned image pairs $\mathcal{D}_{sup} = \{(I_F, I_M \circ T)_k\}_{k=1}^N$, with known ground truth transformations T.

3. Related Work

In recent years, interest in MMIR has increased, particularly in medical imaging, where accurate alignment between different modalities is essential. MMIR approaches can be broadly categorized as intensity-based, feature-based, or deep learning–based.
Intensity-based methods rely on direct comparison of pixel intensities. Metrics such as mutual information (MI) and normalized mutual information (NMI), which stem from information theory, are widely adopted for MMIR due to their robustness to non-linear intensity differences. In contrast, metrics like the sum of squared differences (SSD) common for unimodal registration, perform poorly in multimodal settings.
Feature-based methods detect edges, landmarks, or other structures, but in histology-MRI registration they are often unreliable due to tissue artifacts such as folds, tears, missing regions, and other morphological distortions that make global structures unsuitable for correspondence.
In histology-MRI registration, most prior work has focused on human prostate tissue. ProsRegNet [3] employs spatial transformer networks for direct 2D registration, improving accuracy compared to traditional pipelines. However, since it is trained on synthetic monomodal pairs, it may fail to generalize to datasets with significant cross-modality variability or

complex tissue deformations. RAPHIA [4] integrates a geometry-preserving image translation network, which facilitates registration across modalities. Nevertheless, both approaches are not directly transferable to rodent TBI studies due to anatomical differences and more severe histological artifacts.

Deep learning approaches have recently gained attention for MMIR, particularly GAN-based image-to-image translation, which reduces modality differences and converts a multimodal registration problem into a monomodal one, enabling subsequent alignment [5]. In particular, CycleGAN is widely used because its cycle-consistency mechanism preserves anatomical structures during translation, ensuring that the generated MRI-like images retain geometrical fidelity necessary for accurate registration.

Despite these advances, the combination of complex, TBI-induced deformations, edema, lesions, and severe histological artifacts in rodent brains remains challenging for MMIR. To address this, the proposed pipeline uses a CycleGAN to translate histology into MRI-like images, unifying the domains and enabling subsequent registration with a two-stage network.

4. Materials and Methods

4.1 Dataset

Imaging data were collected at the *Istituto di Ricerche Farmacologiche Mario Negri* and include *in vivo* T2-weighted MRIs and *ex vivo* histological brain images of sham and TBI rodents subjected to the controlled cortical impact (CCI) model, covering varying injury severity (Table 1). MRI acquisitions consist of coronal scans with a matrix size of $150 \times 150 \times 37$ voxels and a voxel spacing of $0.1 \times 0.1 \times 0.3 mm^3$. Rodents were sacrificed one month post-injury, and brain tissues were sectioned at $8\mu m$ thickness for histology. Rodents were grouped as follows:

- *sham*: rodents not subjected to any injury.
- *single moderate TBI* (smTBI): rodents subjected to a single CCI with 0.5 mm depth.
- *single severe TBI* (ssTBI): rodents subjected to a single CCI of 1.0 mm depth.

Starting from imaging data summarized in Table 1, we defined two datasets:

- D_{mono}: monomodal image pairs $(I_F, I_F \circ T)$ where both images belong to the same modality, either T2-w MRI or histology. Each image I_F (from the MRI or Histology columns of Table 1) is paired with a randomly geometrically transformed version of itself $I_F \circ T$ to simulate acquisition and processing variability.
- D_{multi}: multimodal unregistered image pairs (I_F, I_M), where I_F is an MRI section and I_M the corresponding histological image selected through visual inspection (last column of Table 1).

Group	MRI	Histology	Paired MRI-Histology
sham	252	163	110
smTBI	315	210	124
ssTBI	336	268	153

Table 1: Overview of imaging data.

4.2 Proposed Pipeline

Here we propose an unsupervised pipeline for multimodal registration of rodent brain MRI and histological section following TBI. Our approach leverages a CycleGAN to translate the histological image (I_M) into a MRI-like representation, which is then aligned to the original MRI (I_F) via a two-stage deep learning network performing global and local registration. The resulting affine and deformable transformations are applied to the original histological image I_M for final registration. Figure 2 shows the pipeline overview (top), including details of the CycleGAN architecture (bottom left) and the registration block (bottom right). We assume that each MRI-histology image pair corresponds to the same coronal brain section, providing direct spatial correspondence for registration.

Preprocessing MRI and histological sections are skull-stripped using a deep-learning model to isolate the brain from surrounding tissues. They are then manually corrected for orientation errors ($\pm 90°$) and flips, applying the same corrections to the brain masks. Finally, images are cropped around the brain, intensity-normalized to [0,1] and resampled to 256×256 pixels.

Histology-to-MRI Translation The CycleGAN architecture used for histology-to-MRI translation is shown in Figure 2. It consists of two generators, $G : X \rightarrow Y$ (histology to MRI)

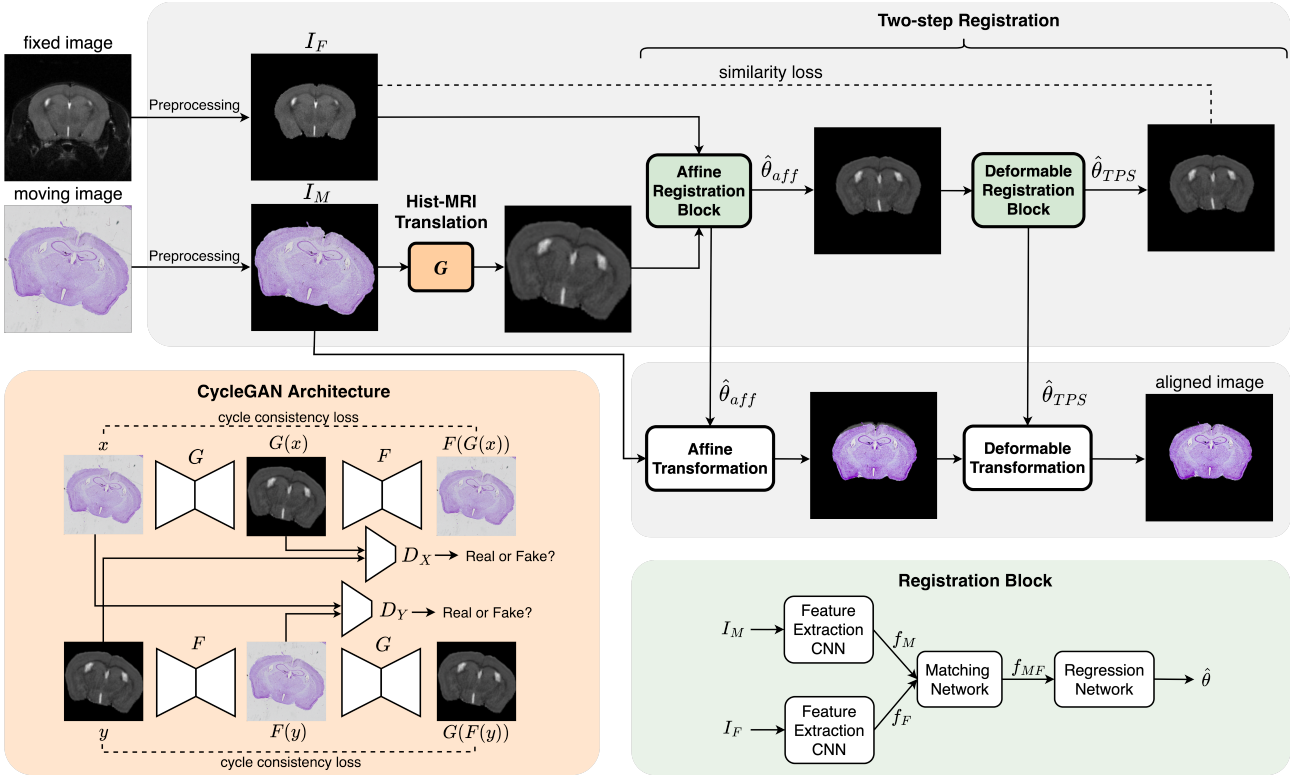

Figure 2: Overview of the pipeline for multimodal image registration.

and $F : Y \to X$ (MRI to histology), and two discriminators, D_X and D_Y. The generator uses an encoder–decoder architecture with three convolutional layers, nine residual blocks, and three transposed convolution layers, while the discriminator adopts a PatchGAN design that penalizes local structural inconsistencies. In the proposed pipeline, only G is used to produce MRI-like representations from histological inputs. However, D_X and D_Y ensure the realism of generated MRIs and histological images.

Two-step Registration The registration network (Figure 2) employs ProsRegNet architecture: first performs global alignment using an affine transformation, then refined with a thin-plate spline (TPS) transformation. Both stages share the same structure, referred to as the Registration Block in Figure 2. The images I_M and I_F are passed through feature extraction VGG16 backbones with shared weights, producing feature maps f_M and f_F. These are combined using a correlation layer, and the resulting correlation map f_{MF} is fed into a regression network that estimates geometric transformation parameters $\hat{\theta}$, whose dimensionality depends on the transformation type. Finally, $\hat{\theta}$ is applied to warp I_M via a differentiable spatial transformer layer.

The affine transformation is parametrized by $\theta \in \mathbb{R}^6$ and defined as:

$$T(x, y) = \begin{bmatrix} 1 + \alpha\theta_1 & \theta_2 \\ \theta_4 & 1 + \alpha\theta_5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \theta_3 \\ \theta_6 \end{bmatrix} \quad (2)$$

where (x, y) denotes a pixel location in I_F and $\alpha = 0.001$ is a constant that keeps the transformation close to the identity in the early training epochs, stabilizing learning. For local refinement, the TPS deformation is parametrized by a 6×6 control-point grid, yielding $\theta \in \mathbb{R}^{72}$. This grid resolution provides a trade-off between flexibility and model complexity. As in the affine stage, predicted parameters are scaled by $\alpha = 0.001$ and added to an identity initialization, ensuring gradual and stable refinement.

4.3 Training Method

We first trained the CycleGAN in an unsupervised setting on D_{multi} with a modified cycle consistency loss (Eq. 3), $\lambda_{cyc} = 10.0$ and $\lambda_{id} = 1.0$, for 200 epochs with early stopping. We optimized the network with Adam optimizer (initial learning rate 2×10^{-4}, batch size 1).

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) = &\ \mathcal{L}_{GAN}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{GAN}(F, D_X, Y, X) \\ &+ \lambda_{cyc} * \mathcal{L}_{cyc}(G, F) \\ &+ \lambda_{id} * \mathcal{L}_{id}(G, F)\end{aligned} \tag{3}$$

Subsequently, we trained the two-stage registration network on the generated MRI-like images with frozen CycleGAN weights for 100 epochs. We applied Adam optimizer (initial learning rate of 1×10^{-3}, batch size 16) and minimized the root mean squared difference (RMSD) between warped MRI-like images and fixed MRIs:

$$RMSD = \sqrt{\frac{1}{|\Omega_I|} \sum_{p_i \in \Omega_I} (I_M(p_i) - I_F(p_i))^2} \tag{4}$$

4.4 Reference Models

In this work, we consider two reference models based on ProsRegNet architecture [3].

ProsRegNet To establish a baseline, we reproduced ProsRegNet architecture [3] and trained it on monomodal image pairs from D_{mono}, minimizing the RMSD loss (Eq. 3). As shown by Shao et al., the model doesn't required paired slices to be trained and generalizes well even on multimodal image pairs.

Multimodal ProsRegNet To evaluate if monomodal training limits ProsRegNet, we trained a variant of the same network in a true-multimodal setting, by using image pairs from D_{multi} with the MI loss. MI is defined as the Kullback–Leibler divergence between the joint distribution and the product of the marginals:

$$M(I_M, I_F) = \sum_{x,y} \mathbb{P}(x, y) \log \frac{\mathbb{P}(x, y)}{\mathbb{P}(x)\mathbb{P}(y)} \tag{5}$$

The MI loss handles intensity distribution differences across modalities by maximizing their shared information, enhancing MMIR.

4.5 Evaluation Method

We evaluated the proposed pipeline and the reference models with 5-fold cross-validation on multimodal image pairs. Networks were trained for 100 epochs without early stopping, using Adam optimizer (initial learning rate 1×10^{-3}, batch size 16). A scheduler reduced the learning rate by 5% after 5 epochs without validation loss improvement (minimum learning rate 1×10^{-6}). To assess registration performance, we used the Dice similarity coefficient (DSC) between transformed moving and fixed brain masks, and the mean error distance (MED), to measure global and local accuracy. The MED measures the point-to-point correspondence accuracy as:

$$MED = \frac{1}{N} \sum_{i=1}^{N} \|p_i - T^*(p_i')\|_2 \tag{6}$$

where T^* is the estimated transformation, and $(p_1, p_1'), ..., (p_N, p_N')$ are landmark points in the fixed and moving images (Figure 3).

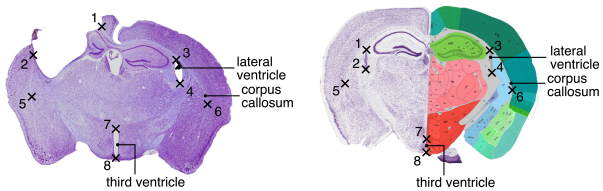

Figure 3: Landmark points labelled by our experts on (left) a sham and (right) a TBI mouse.

5. Experimental Results

We assessed the proposed pipeline by comparing it with two reference models. Table 2 shows the proposed pipeline consistently improves in DSC and MED metrics across pathology groups, with more evident gains in sham than in TBI rodents.

Status	Registration Approach	DSC	MED
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 2: Comparison of registration performance (DSC, MED) across approaches.

To better understand these results, we qualitatively inspected CycleGAN-generated histology-to-MRI translations (Figure 4), exhibiting realistic contrast synthesis and minor geometric deviations. These deviations are more pronounced in TBI cases, likely due to the more complex anatomical changes and artifacts, which can

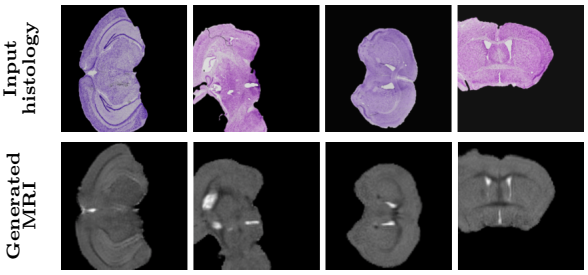

Figure 4: Histology-to-MRI translation.

limit registration accuracy. Landmark overlays in Figure 5 confirm robust alignment across sham and TBI rodents. These findings highlight the critical role of the CycleGAN in preserving anatomical structures while generating realistic contrasts for effective alignment.

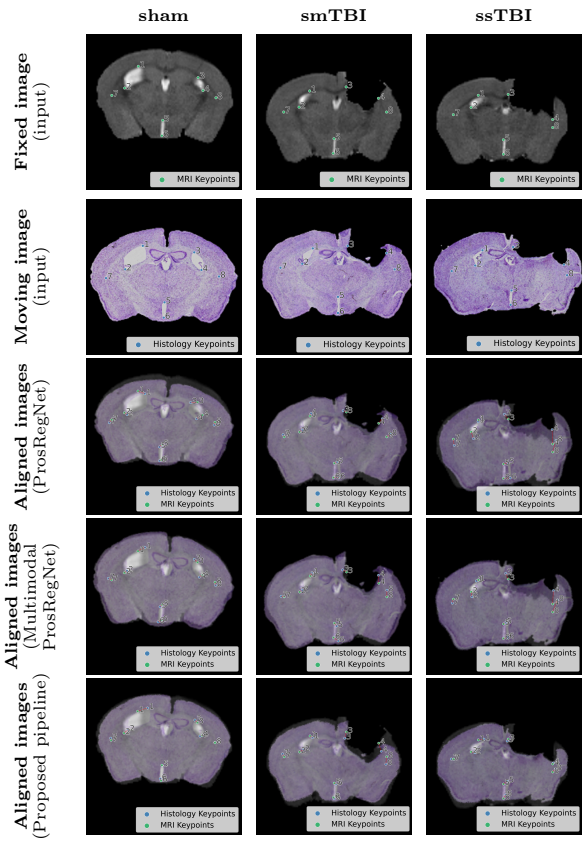

Figure 5: Registration results across groups. Green and blue dots show MRI and histology landmarks, red lines indicate misalignment.

6. Conclusions

This thesis addressed the alignment of *in vivo* MRI and *ex vivo* histological images of rodent brains after TBI. We proposed a fully unsupervised multimodal registration pipeline combining CycleGAN-based modality translation with a two-stage registration network performing affine and TPS alignment. By converting the multimodal problem into a monomodal one, this approach enables more accurate alignment. Evaluation showed consistent improvements over monomodal and multimodal baselines in DSC and MED metrics. Although gains are modest in some TBI cases, the method demonstrates robustness and provides a reliable cross-modality registration framework.

Future work will extend the pipeline to different histological stains and neurological pathologies to evaluate its adaptability. Additionally, we aim to explore alternative image-to-image translation architectures that may better capture subtle TBI-related changes. These efforts aim to improve registration quality and generalizability, for more comprehensive and reliable TBI neuroanatomical analysis.

References

[1] David K. Menon, Karen Schwab, Daniel W. Wright, and et al. Position statement: Definition of traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 91 (11):1637–1640, Nov 2010.

[2] Ang Guo, Zhiyu Chen, Yinzhong Ma, and et al. SOmicsFusion: Multimodal coregistration and fusion between spatial metabolomics and biomedical imaging. *Artificial Intelligence in Chemistry*, 2(1):100058, 2024.

[3] Wei Shao, L. Banh, C. A. Kunder, and et al. ProsRegNet: A deep learning framework for registration of MRI and histopathology images of the prostate. *Medical Image Analysis*, 68:101919, Feb 2021.

[4] Wei Shao, Sulaiman Vesal, Simon J.C. Soerensen, and et al. 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.

[5] Chen Qin, Bibo Shi, Rui Liao, and et al. Unsupervised deformable registration for multimodal images via disentangled representations. *arXiv preprint*, 2019.