DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Enhancing Intraoperative Registration with Neural Radiance Fields: An Exploration of Loss Functions and Style Transfer Effects

David

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Enhancing Intraoperative Registration with Neural Radiance Fields: An Exploration of Loss Functions and Style Transfer Effects

Verbesserung der intraoperativen Registrierung mit Neural Radiance Fields: Eine Untersuchung von Verlustfunktionen und Stilübertragungseffekten

Author:	David
Supervisor:	Prof. Dr. Nassir Navab
Advisor:	Maximilian Fehrentz
Submission Date:	September 2023

I confirm that this bachelor’s thesis in informatics is my own work and I have documented all sources and material used.

Munich, September 2023 David

Acknowledgments

Acknowledgements

I would like to express my sincere gratitude to everyone who supported me throughout the journey of completing this thesis.

First and foremost, I would like to thank my supervisor, Prof. Dr. Nassir Navab, for providing me with the opportunity to work on this exciting research topic at the Computer Aided Medical Procedures (CAMP) chair. His guidance, expertise, and vision in the field of medical image computing have been invaluable to this work.

I am deeply grateful to my advisor, Maximilian Fehrentz, for his continuous support, patience, and mentorship. His pioneering work on cross-modal neural rendering for intraoperative registration laid the foundation for this thesis, and his detailed feedback and insightful discussions have significantly shaped and improved my research.

I would also like to extend my appreciation to the entire CAMP team for creating a collaborative and stimulating research environment. The weekly seminars, discussions, and exchange of ideas have been incredibly helpful in refining my approach and broadening my perspective.

Special thanks to the clinical partners at the University Hospital for providing access to the clinical datasets used in this research and for sharing their valuable medical expertise. Their insights into the practical challenges of neurosurgical navigation have been essential in guiding this work toward clinical relevance.

I am thankful to my fellow students and friends who have contributed to this work through technical discussions, proofreading, and moral support. Their camaraderie made the challenging moments more bearable and the successes more enjoyable.

Finally, I would like to express my heartfelt gratitude to my family for their unwavering support, encouragement, and understanding throughout my academic journey. Their belief in me has been a constant source of motivation.

This thesis was partially supported by the Technical University of Munich and the CAMP chair, which provided the computational resources and academic environment necessary for this research.

David
Munich, September 2023

Abstract

This study advances intraoperative brain registration using Neural Radiance Fields (NeRFs) as learned functional representations of brain surfaces. Unlike traditional mesh-based approaches, NeRFs provide a differentiable, implicit function of the brain's geometry and appearance. We leverage this property to optimize camera positions through backpropagation, enabling precise alignment of preoperative and intraoperative brain images. Our research explores various loss functions (Mutual Information, Normalized Cross-Correlation, weighted/masked L2) and analyzes the impact of hypernetwork-generated styles on registration accuracy. By enhancing NeRF-based brain registration, this work aims to improve the precision of image-guided neurosurgery.

Kurzfassung

Diese Studie verbessert die intraoperative Hirnregistrierung durch den Einsatz von Neural Radiance Fields (NeRFs) als erlernte funktionale Darstellungen von Hirnoberflächen. Im Gegensatz zu traditionellen netzbasierten Ansätzen bieten NeRFs eine differenzierbare, implizite Funktion der Geometrie und des Erscheinungsbilds des Gehirns. Wir nutzen diese Eigenschaft, um Kamerapositionen durch Backpropagation zu optimieren und ermöglichen so eine präzise Ausrichtung von präoperativen und intraoperativen Hirnbildern. Unsere Forschung untersucht verschiedene Verlustfunktionen (Mutual Information, Normalized Cross-Correlation, gewichteter/maskierter L2-Verlust) und analysiert den Einfluss von durch Hypernetzwerke generierte Stile auf die Registrierungsgenauigkeit. Durch die Verbesserung der NeRF-basierten Hirnregistrierung zielt diese Arbeit darauf ab, die Präzision der bildgeführten Neurochirurgie zu verbessern.

Contents

1 Introduction

Neurosurgery is a high-precision medical field where accuracy is paramount to patient outcomes. During brain tumor resection procedures, surgeons rely on image-guided navigation systems to assist with spatial orientation and to locate critical structures within the brain [1]. These systems typically align preoperative Magnetic Resonance Imaging (MRI) data with the patient's physical anatomy through a process called registration. Traditional registration methods often employ point-based or surface-based techniques, which can be time-consuming, require specialized equipment, and are susceptible to inaccuracies due to brain shift—the deformation of brain tissue that occurs once the skull is opened.

The field of computer vision has recently seen remarkable advancements in neural scene representation techniques, particularly with the introduction of Neural Radiance Fields (NeRFs) [2]. NeRFs represent scenes as continuous functions that map 3D coordinates and viewing directions to color and density values, enabling high-quality novel view synthesis. These implicit neural representations have revolutionized how we model and render 3D environments, offering differentiable, continuous scene representations that can be optimized through gradient-based methods.

This thesis builds upon recent work that leverages NeRFs for pose estimation [3] and, more specifically, for intraoperative registration [4]. We propose to enhance NeRF-based intraoperative registration through two main contributions: (1) a comprehensive exploration of alternative loss functions beyond the standard L2 loss, and (2) an investigation into the effects of style transfer on registration accuracy.

1.1 Motivation

Current intraoperative registration techniques face several challenges:

1. **Brain shift**: The brain's position and shape change during surgery due to cerebrospinal fluid drainage, gravity, and surgical manipulations.

2. **Cross-modal alignment**: Matching preoperative MRI data with intraoperative camera images requires bridging different imaging modalities.

3. **Speed and accuracy**: Registration must be both precise and fast enough to be clinically viable during surgery.

4. **Surgical workflow integration**: Registration methods should integrate seamlessly into existing surgical workflows without requiring additional equipment or extensive time.

The approach presented by Fehrentz, Azampour, Dorent, et al. [4] addresses these challenges by using neural rendering for registration. However, their work primarily focuses on a hypernetwork-based approach for appearance adaptation and employs a standard L2 loss for optimization. This thesis extends their work by exploring whether alternative loss functions might yield better registration results and by analyzing how different style transfer techniques affect the registration process.

1.2 Research Questions

This thesis addresses the following key research questions:

1. How do different loss functions (Mutual Information, Normalized Cross-Correlation, and weighted/masked L2) compare to the standard L2 loss in NeRF-based intraoperative registration?

2. What impact do various hypernetwork-generated styles have on registration accuracy when using NeRF-based methods?

3. Can the combination of specific loss functions and style transfer techniques improve registration robustness against variations in lighting, perspective, and tissue appearance?

1.3 Contributions

The main contributions of this thesis are:

1. A systematic evaluation of multiple loss functions for NeRF-based intraoperative registration.

2. An analysis of hypernetwork-based style transfer effects on registration accuracy.

3. An implementation built on top of nerfstudio to make the approach NeRF-implementation agnostic.

4. Experimental results comparing different combinations of loss functions and style transfer techniques.

1.4 Thesis Structure

The remainder of this thesis is organized as follows:

- **Chapter ??**: Provides the necessary background on neural radiance fields, pose estimation, and intraoperative registration.

- **Chapter ??**: Explains the methodology and overall approach for NeRF-based registration.

- **Chapter ??**: Details the loss functions implemented and evaluated in this work.
- **Chapter ??**: Explores the style transfer techniques and their integration with the registration process.
- **Chapter ??**: Describes the experimental setup, datasets, and evaluation metrics.
- **Chapter ??**: Presents the results of our experiments and analyses.
- **Chapter ??**: Discusses the implications of our findings and their relevance to the field.
- **Chapter ??**: Summarizes the thesis and suggests directions for future research.

1.5 Section

Citation test (with Biber) [5].

1.5.1 Subsection

See **??**, **??**, **??**, **??**, **??**, **??**.

Table 1.1: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

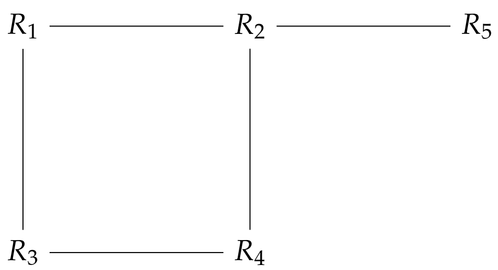

Figure 1.1: An example for a simple drawing.

This is how the glossary will be used.

Donor dye, ex. Alexa 488 ($\mathrm{D_{dye}}$), Förster distance, Förster distance (R_0), and k_{DEAC}. Also, the TUM has many computers, not only one Computer. Subsequent acronym usage will only print the short version of Technical University of Munich (TUM) (take care of plural, if needed!), like here with TUM, too. It can also be –> hidden[1] <–.

[1]Example for a hidden TUM glossary entry.

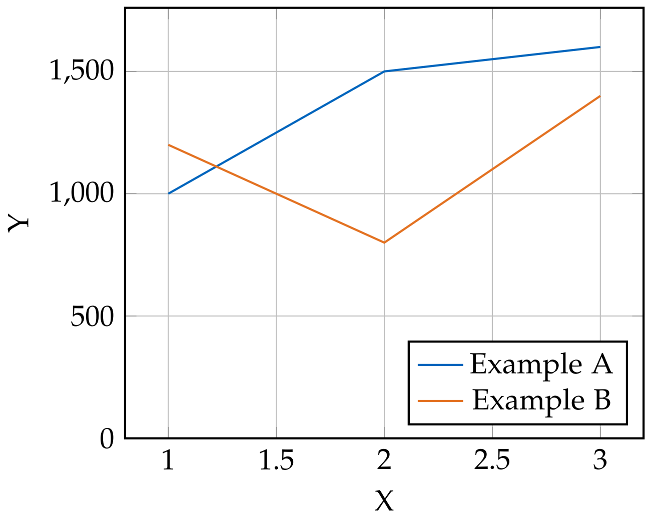

Figure 1.2: An example for a simple plot.

```sql
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 1.3: An example for a source code listing.

Figure 1.4: Includegraphics searches for the filename without extension first in logos, then in figures.

Figure 1.5: For pictures with the same name, the direct folder needs to be chosen.

(a) The logo.

(b) The famous slide.

Figure 1.6: Two TUM pictures side by side.

[(TODO: Now it is your turn to write your thesis. This will be a few tough weeks.)]
[(Done: Nevertheless, celebrate it when it is done!)]

2 Background and Related Work

This chapter provides the necessary theoretical foundation and contextual background for understanding the work presented in this thesis. We first review the field of intraoperative registration, followed by an overview of Neural Radiance Fields (NeRFs) and their applications in pose estimation. Finally, we discuss relevant work on loss functions for image registration and style transfer techniques.

2.1 Intraoperative Registration in Neurosurgery

Intraoperative registration refers to the process of aligning preoperative imaging data, such as Magnetic Resonance Imaging (MRI) scans, with the patient's physical anatomy during surgery. This alignment enables surgeons to navigate with respect to preoperative imaging data, which provides crucial information about anatomical structures not visible on the brain surface, such as tumor boundaries and critical functional areas.

2.1.1 Traditional Registration Approaches

Traditional approaches to intraoperative registration in neurosurgery can be broadly categorized into the following methods:

- **Point-based registration**: This approach identifies and matches corresponding anatomical landmarks or artificially placed fiducial markers in both the preoperative images and physical patient. While conceptually simple, it requires accurate identification of landmarks and can be time-consuming.
- **Surface-based registration**: This technique matches surfaces extracted from preoperative imaging with surfaces captured intraoperatively, often using techniques like the Iterative Closest Point (ICP) algorithm. Surface-based approaches typically require specialized equipment, such as laser scanners or stereo cameras, to capture the intraoperative surface.
- **Volume-based registration**: These methods use intensity-based similarity measures to align volumetric images, but they typically require intraoperative imaging modalities such as ultrasound or intraoperative MRI, which may not be available in all surgical settings.

A significant challenge in neurosurgical registration is brain shift, the deformation of brain tissue that occurs during surgery due to factors such as gravity, cerebrospinal fluid drainage, and surgical manipulations. This phenomenon can significantly reduce the accuracy of rigid registration methods and necessitates more advanced techniques.

2.1.2 Cross-Modal Registration

Cross-modal registration refers to the alignment of images from different imaging modalities. In the context of neurosurgery, this often involves aligning preoperative MRI data with intraoperative camera images. This presents unique challenges due to differences in:

- **Information content**: MRI provides volumetric data with tissue contrast, while optical images capture surface appearance with details like blood vessels and lighting effects.
- **Geometric representation**: MRI data is three-dimensional, while camera images are two-dimensional projections.
- **Appearance**: The visual appearance of tissues differs significantly between MRI and optical images due to different physical principles of image formation.

Previous work in cross-modal registration has employed techniques such as feature extraction, mutual information maximization, and deep learning-based approaches to bridge these differences.

2.2 Neural Radiance Fields (NeRFs)

Neural Radiance Fields, introduced by Mildenhall, Srinivasan, Tancik, et al. [2], represent a novel approach to scene representation and novel view synthesis. Unlike traditional computer graphics methods that use explicit representations like meshes or point clouds, NeRFs employ an implicit neural representation to model scenes.

2.2.1 NeRF Representation

A NeRF is typically implemented as a multi-layer perceptron (MLP) that maps a 3D coordinate $\mathbf{x} = (x, y, z)$ and viewing direction $\mathbf{d} = (\theta, \phi)$ to a color $\mathbf{c} = (r, g, b)$ and volume density σ:

$$F_\Theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma) \tag{2.1}$$

where Θ represents the learnable parameters of the neural network. This continuous, differentiable representation allows for rendering from arbitrary viewpoints through volume rendering techniques.

The rendering process involves casting rays from a camera through image pixels and evaluating the NeRF at multiple points along each ray. The color of a pixel is computed as a weighted sum of the colors along the ray, with weights determined by the volume densities:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(t)\mathbf{c}(t)dt \tag{2.2}$$

where $T(t) = \exp\left(-\int_{t_n}^{t} \sigma(s)ds\right)$ represents the accumulated transmittance along the ray up to point t.

2.2.2 NeRF Variants

Since the introduction of the original NeRF, numerous variants have been developed to address limitations and extend capabilities:

- **Instant-NGP** [6]: Accelerates NeRF training and rendering through multi-resolution hash encoding, reducing training time from days to minutes.
- **HyperNeRF** [7]: Extends NeRFs to handle topological variations in dynamic scenes through a higher-dimensional representation.
- **Nerfacto**: An implementation-agnostic framework that combines advances from various NeRF variants for improved performance.

2.2.3 iNeRF: Inverting Neural Radiance Fields for Pose Estimation

Yen-Chen, Florence, Barron, et al. [3] introduced iNeRF, a method that leverages the differentiable nature of NeRFs for pose estimation. Given a target image and a pre-trained NeRF, iNeRF estimates the camera pose from which the target image was captured. This is achieved by optimizing the camera pose parameters to minimize the difference between the rendered image (from the current pose estimate) and the target image.

The key insight of iNeRF is that the camera pose can be optimized through backpropagation, utilizing the differentiable nature of both the NeRF representation and the rendering process. This optimization is formulated as:

$$\hat{\xi} = \arg\min_{\xi} \mathcal{L}(I_{\text{target}}, I_{\text{rendered}}(\xi)) \tag{2.3}$$

where ξ represents the camera pose parameters, I_{target} is the target image, $I_{\text{rendered}}(\xi)$ is the image rendered from the NeRF using pose ξ, and $\mathcal{L}$ is a loss function measuring the dissimilarity between the images.

2.3 Cross-Modal Inverse Neural Rendering for Registration

Building on the concept of iNeRF, Fehrentz, Azampour, Dorent, et al. [4] proposed a method for intraoperative registration using cross-modal inverse neural rendering. Their approach addresses the challenge of cross-modal registration by separating the neural representation into structural and appearance components:

- The **structural component** captures the geometric properties of the brain and is learned from preoperative MRI data.

- The **appearance component** is adapted intraoperatively to match the visual characteristics of surgical images.

This separation is achieved through a multi-style hypernetwork that controls the appearance of the NeRF while preserving its learned representation of the anatomy. The hypernetwork generates parameters for a subset of the NeRF's layers, allowing it to produce different appearances for the same underlying geometry.

During registration, the approach optimizes both the camera pose and the appearance parameters to minimize the dissimilarity between the rendered and target intraoperative images. This method has shown promising results in clinical data, outperforming state-of-the-art methods while meeting clinical standards for registration accuracy.

2.4 Loss Functions for Image Registration

The choice of loss function is crucial in registration tasks, as it defines the measure of similarity between images that guides the optimization process. Different loss functions capture different aspects of image similarity and may be more or less suitable depending on the specific registration task.

2.4.1 L2 Loss

The L2 loss, or mean squared error (MSE), is commonly used in image registration tasks due to its simplicity and differentiability. It calculates the squared Euclidean distance between two images:

$$\mathcal{L}_{\mathrm{L2}}(I_1, I_2) = \frac{1}{N} \sum_{i=1}^{N} (I_1(i) - I_2(i))^2 \tag{2.4}$$

where N is the number of pixels. While straightforward, L2 loss assumes a direct intensity correspondence between images, which may not hold in cross-modal scenarios.

2.4.2 Normalized Cross-Correlation (NCC)

Normalized Cross-Correlation measures the similarity between two images independently of linear intensity transformations:

$$\mathcal{L}_{\mathrm{NCC}}(I_1, I_2) = -\frac{\sum_{i=1}^{N} (I_1(i) - \bar{I}_1)(I_2(i) - \bar{I}_2)}{\sqrt{\sum_{i=1}^{N} (I_1(i) - \bar{I}_1)^2 \sum_{i=1}^{N} (I_2(i) - \bar{I}_2)^2}} \tag{2.5}$$

where $\bar{I}_1$ and $\bar{I}_2$ are the mean intensities of the respective images. NCC is particularly useful when images have different contrast or brightness levels [8].

2.4.3 Mutual Information (MI)

Mutual Information is a statistical measure that quantifies the mutual dependence between two random variables, making it particularly suitable for cross-modal registration where the relationship between intensities is complex:

$$\mathcal{L}_{\mathrm{MI}}(I_1, I_2) = -\sum_{i,j} p_{I_1,I_2}(i,j) \log\left(\frac{p_{I_1,I_2}(i,j)}{p_{I_1}(i)p_{I_2}(j)}\right) \tag{2.6}$$

where p_{I_1,I_2} is the joint probability distribution of intensities in images I_1 and I_2, and p_{I_1} and p_{I_2} are their marginal distributions [9].

2.4.4 Weighted and Masked L2 Loss

Weighted and masked variants of the L2 loss assign different importance to different regions of the image:

$$\mathcal{L}_{\mathrm{wL2}}(I_1, I_2) = \frac{1}{N}\sum_{i=1}^{N} w(i)(I_1(i) - I_2(i))^2 \tag{2.7}$$

where $w(i)$ is a weight or mask value for pixel i. This approach can be useful for focusing the registration on regions of interest or for ignoring irrelevant areas.

2.5 Style Transfer Techniques

Style transfer refers to the process of transforming an image to have the visual style of another image while preserving its content. In the context of cross-modal registration, style transfer can be used to bridge the appearance gap between different imaging modalities.

Various techniques for style transfer have been proposed, including:

- **Gram-matrix-based methods**: These approaches capture style by computing correlations between feature activations in different layers of a neural network.
- **Hypernetwork-based style adaptation**: Hypernetworks generate parameters for another network based on style information, allowing for efficient style adaptation.
- **Feature space transformations**: These methods transform features in intermediate representations to match target style statistics.
- **Color space transformations**: Simpler approaches that match color distributions, such as histogram matching or transformations in alternative color spaces like Y'UV.

The approach by Fehrentz, Azampour, Dorent, et al. [4] employs a hypernetwork-based technique for style adaptation, which we will build upon and extend in this thesis.

3 Methodology

This chapter presents the methodology employed in this thesis for enhancing intraoperative registration using Neural Radiance Fields (NeRFs). We first provide an overview of the registration approach, followed by details on the implementation framework, the process of training NeRFs from preoperative data, and the registration optimization procedure.

3.1 Overview of the NeRF-based Registration Approach

The core methodology of this thesis builds upon the inverse Neural Radiance Field (iNeRF) approach originally proposed by Yen-Chen, Florence, Barron, et al. [3] and extended for cross-modal intraoperative registration by Fehrentz, Azampour, Dorent, et al. [4]. Figure **??** provides a high-level overview of the approach.

Figure 3.1: Overview of the NeRF-based intraoperative registration approach. The method involves preoperative training of a NeRF model on MRI data, followed by intraoperative optimization of camera pose and appearance parameters to match the target surgical image.

The registration process consists of two main phases:

1. **Preoperative Phase**: A NeRF model is trained using preoperative MRI data to learn an implicit representation of the brain's structure. Additionally, a hypernetwork for style adaptation is trained to enable cross-modal appearance matching.
2. **Intraoperative Phase**: During surgery, the pre-trained NeRF is used as a differentiable rendering engine. Given a target intraoperative image, the camera pose (6 degrees of freedom) is optimized through backpropagation to match the rendered view with the target image.

3.2 Implementation Framework

Our implementation leverages the nerfstudio framework to ensure flexibility and compatibility with various NeRF architectures. The key components of our implementation include:

- **NeRF Model**: We support multiple NeRF variants, including the original NeRF [2], Instant-NGP [6], and Nerfacto, allowing us to evaluate the impact of different NeRF architectures on registration performance.

- **Camera Optimizer**: A differentiable camera pose optimization module that updates the 6DoF camera parameters (rotation and translation) based on the gradient of the selected loss function.
- **Hypernetwork Style Module**: An extension of the hypernetwork approach from Fehrentz, Azampour, Dorent, et al. [4] that controls the appearance of the NeRF while preserving its structural representation.
- **Loss Function Module**: A pluggable interface for different loss functions, including L2, Normalized Cross-Correlation, Mutual Information, and variants of weighted/masked L2 loss.

The modular design of our implementation allows for easy experimentation with different combinations of NeRF architectures, loss functions, and style transfer techniques.

3.3 NeRF Training from Preoperative Data

3.3.1 Data Preparation

The preoperative MRI data undergoes several preprocessing steps before NeRF training:

1. **Segmentation**: Brain structures are segmented from the MRI volume to focus on relevant regions and reduce computational requirements.
2. **View Synthesis**: Multiple synthetic views are generated from the volumetric MRI data to create a dataset of 2D images and corresponding camera poses for NeRF training.
3. **Data Augmentation**: To improve the robustness of the NeRF model, we apply various data augmentation techniques, including random brightness and contrast adjustments, as well as simulated occlusions.

3.3.2 NeRF Training Procedure

The NeRF model is trained using standard procedures with some modifications to account for the medical imaging context:

1. **Model Initialization**: The NeRF model architecture (Instant-NGP, Nerfacto, etc.) is initialized with random weights.
2. **Training Loop**: The model is trained by rendering synthetic views and comparing them with the reference views generated from the MRI data. The model parameters are updated to minimize the rendering error using the Adam optimizer.
3. **Density Refinement**: Special attention is paid to accurately modeling the brain surface, as this is critical for registration. We employ additional loss terms to encourage accurate density modeling at tissue boundaries.

3.3.3 Hypernetwork Training for Style Adaptation

To enable cross-modal registration, we train a hypernetwork that adapts the appearance of the NeRF to match different visual styles:

1. **Hypernetwork Architecture**: The hypernetwork is a small MLP that takes a style code as input and outputs parameters for a subset of the NeRF's color prediction layers.
2. **Style Encoding**: We experiment with various methods for encoding style information, including:
   - Direct RGB statistics (mean and variance)
   - Y'UV color space transformations
   - Histogram of Oriented Gradients (HOG) features
   - Gram matrices for texture representation
3. **Training Procedure**: The hypernetwork is trained by presenting examples of the same brain structure with different appearance styles (e.g., different contrast, lighting, or color schemes). The network learns to map these styles to appropriate NeRF parameters while maintaining structural consistency.

3.4 Registration Optimization

During the intraoperative phase, the pre-trained NeRF and hypernetwork are used to perform registration by optimizing camera pose parameters to match a target intraoperative image.

3.4.1 Problem Formulation

The registration problem is formulated as an optimization of camera pose parameters ξ and style parameters s to minimize a dissimilarity measure between the target intraoperative image I_{target} and the image rendered from the NeRF:

$$\hat{\xi}, \hat{s} = \arg\min_{\xi, s} \mathcal{L}(I_{\text{target}}, I_{\text{rendered}}(\xi, s)) \tag{3.1}$$

where $I_{\text{rendered}}(\xi, s)$ is the image rendered from the NeRF using camera pose ξ and style parameters s, and $\mathcal{L}$ is a loss function measuring the dissimilarity between the images.

3.4.2 Optimization Procedure

The optimization procedure consists of the following steps:

1. **Initialization**: Camera pose is initialized based on surgical setup information or using a coarse alignment step. Style parameters are initialized to default values or based on statistical analysis of the target image.

2. **Iterative Optimization**: The following steps are repeated until convergence or for a fixed number of iterations:
    - Render an image from the current camera pose and style parameters
    - Compute the loss between the rendered image and the target image
    - Compute gradients with respect to camera pose and style parameters
    - Update parameters using gradient descent or Adam optimizer
3. **Multi-resolution Approach**: To avoid local minima, we employ a multi-resolution approach, starting with lower-resolution images and progressively increasing resolution.
4. **Regularization**: Regularization terms are added to the optimization to prevent extreme camera poses or style parameters.

3.4.3 Ray Sampling Strategies

To improve efficiency and robustness, we experiment with different ray sampling strategies during optimization:

- **Random sampling**: Randomly selecting rays for each optimization step.
- **Importance sampling**: Focusing on regions with high error or important anatomical features.
- **Patch-based sampling**: Using coherent patches of rays to capture local image structure.

3.5 Experimental Setup

To evaluate our approach, we design experiments that systematically compare different loss functions and style transfer techniques in the context of intraoperative registration.

3.5.1 Dataset

We use both synthetic and real clinical data for our experiments:

- **Synthetic Dataset**: We generate synthetic data by rendering views from 3D brain models with controlled variations in pose, lighting, and appearance. This allows for quantitative evaluation with known ground truth poses.
- **Clinical Dataset**: We use retrospective clinical data from neurosurgical procedures, including preoperative MRI scans and intraoperative images captured during surgery.

3.5.2 Evaluation Metrics

We evaluate registration performance using the following metrics:

- **Pose Error**: Quantitative measures of the difference between estimated and ground truth poses, including rotational error (degrees) and translational error (mm).
- **Target Registration Error (TRE)**: The average distance between corresponding anatomical landmarks after registration.
- **Image Similarity Metrics**: NCC, MI, and SSIM scores between the registered images.
- **Convergence Rate**: The number of iterations required to achieve convergence and the optimization stability.

3.5.3 Experimental Variables

Our experiments systematically vary the following factors:

- **Loss Functions**: L2, NCC, MI, weighted/masked L2, and combinations thereof.
- **Style Transfer Techniques**: Different approaches for style encoding and adaptation.
- **NeRF Architectures**: Original NeRF, Instant-NGP, and Nerfacto.
- **Initial Pose Error**: Various levels of misalignment to evaluate the capture range of the registration.
- **Image Conditions**: Different lighting conditions, occlusions, and tissue appearance variations.

The results of these experiments are presented and analyzed in Chapters **??**, **??**, and **??**.

4 Loss Functions for NeRF-Based Registration

This chapter provides a detailed analysis of various loss functions and their application to NeRF-based intraoperative registration. We investigate how different similarity measures affect registration accuracy, convergence speed, and robustness to cross-modal appearance variations. The loss function is a critical component of the registration framework, as it defines the optimization objective that guides the camera pose estimation process.

4.1 Role of Loss Functions in Registration

In the context of intraoperative registration using Neural Radiance Fields, the loss function serves two primary purposes:

1. **Similarity Measurement**: It quantifies the similarity between the target intraoperative image and the rendered image from the NeRF model. This measure guides the optimization process toward better alignment.
2. **Gradient Provision**: It provides gradients with respect to camera pose parameters, enabling backpropagation-based optimization of the camera pose.

The effectiveness of a loss function depends on its ability to handle cross-modal differences between preoperative MRI-derived renderings and intraoperative camera images. Different loss functions have varying sensitivities to factors such as:

- Intensity scaling and bias
- Local contrast variations
- Texture and high-frequency details
- Noise and outliers
- Incomplete overlap between images

4.2 L2 Loss: Baseline Approach

The L2 loss, or mean squared error (MSE), is commonly used in registration tasks and serves as our baseline approach. For two images I_1 and I_2, the L2 loss is defined as:

$$\mathcal{L}_{\text{L2}}(I_1, I_2) = \frac{1}{N} \sum_{i=1}^{N} (I_1(i) - I_2(i))^2 \tag{4.1}$$

where N is the number of pixels in the images.

4.2.1 Implementation Details

Our implementation of the L2 loss includes the following components:

- **Normalization**: To improve robustness to global intensity variations, we normalize the images to have zero mean and unit standard deviation before computing the loss.
- **Channel Weighting**: We allow for different weights to be assigned to different color channels, potentially prioritizing channels with more relevant information.
- **Differentiability**: The L2 loss is fully differentiable with respect to the input images, enabling end-to-end gradient flow from the loss to the camera pose parameters.

4.2.2 Advantages and Limitations

The L2 loss offers several advantages:

- Simplicity and computational efficiency
- Well-defined gradients for optimization
- Direct interpretation as the Euclidean distance in pixel space

However, it also has significant limitations for cross-modal registration:

- High sensitivity to intensity variations and outliers
- Assumption of direct correspondence between pixel intensities across modalities
- Limited capture range, potentially leading to local minima

4.3 Normalized Cross-Correlation (NCC)

Normalized Cross-Correlation measures the linear relationship between two images while being invariant to linear intensity transformations. The NCC loss is defined as:

$$\mathcal{L}_{\text{NCC}}(I_1, I_2) = -\frac{\sum_{i=1}^{N} (I_1(i) - \bar{I}_1)(I_2(i) - \bar{I}_2)}{\sqrt{\sum_{i=1}^{N} (I_1(i) - \bar{I}_1)^2 \sum_{i=1}^{N} (I_2(i) - \bar{I}_2)^2}} \tag{4.2}$$

where $\bar{I}_1$ and $\bar{I}_2$ are the mean intensities of the respective images. The negative sign is used because we're minimizing the loss, while NCC itself is a similarity measure that should be maximized.

4.3.1 Implementation Details

Our implementation of NCC includes several enhancements:

- **Local NCC**: We implement both global NCC (computed over the entire image) and local NCC (computed over small patches), which can better handle local intensity variations.
- **Multi-scale NCC**: We compute NCC at multiple scales to capture both fine details and broader structures.
- **Regularization**: A small constant ϵ is added to the denominator to ensure numerical stability:

$$\mathcal{L}_{\text{NCC}}(I_1, I_2) = -\frac{\sum_{i=1}^{N}(I_1(i) - \bar{I}_1)(I_2(i) - \bar{I}_2)}{\sqrt{(\sum_{i=1}^{N}(I_1(i) - \bar{I}_1)^2 + \epsilon)(\sum_{i=1}^{N}(I_2(i) - \bar{I}_2)^2 + \epsilon)}} \tag{4.3}$$

4.3.2 Advantages and Limitations

NCC offers several advantages over L2 loss for cross-modal registration:

- Invariance to linear intensity transformations (scaling and bias)
- Better robustness to global illumination changes
- Often more effective for cross-modal registration tasks [8]

However, NCC also has limitations:

- Only captures linear relationships between images
- May be less effective when the relationship between modalities is non-linear
- Computationally more expensive than L2 loss

4.4 Mutual Information (MI)

Mutual Information is a statistical measure from information theory that quantifies the mutual dependence between two random variables. In the context of image registration, MI measures how much information one image provides about another, making it particularly suitable for cross-modal registration where the relationship between intensities is complex and non-linear.

The MI loss is defined as:

$$\mathcal{L}_{\text{MI}}(I_1, I_2) = -\sum_{i,j} p_{I_1,I_2}(i,j) \log\left(\frac{p_{I_1,I_2}(i,j)}{p_{I_1}(i)p_{I_2}(j)}\right) \tag{4.4}$$

where p_{I_1,I_2} is the joint probability distribution of intensities in images I_1 and I_2, and p_{I_1} and p_{I_2} are their marginal distributions. As with NCC, the negative sign is used to convert the similarity measure to a loss function for minimization.

4.4.1 Implementation Details

Implementing MI for differentiable optimization presents several challenges, which we address through the following approaches:

- **Histogram Computation**: We compute joint and marginal histograms of image intensities using differentiable binning functions based on B-splines.
- **Parzen Window Estimation**: To create continuous, differentiable probability distributions, we employ Parzen window estimation with Gaussian kernels.
- **Normalized MI**: We implement normalized mutual information (NMI) as an alternative formulation:

$$\mathcal{L}_{\mathrm{NMI}}(I_1, I_2) = -\frac{H(I_1) + H(I_2)}{H(I_1, I_2)} \tag{4.5}$$

  where $H(I_1)$ and $H(I_2)$ are the marginal entropies, and $H(I_1, I_2)$ is the joint entropy.

4.4.2 Advantages and Limitations

MI offers significant advantages for cross-modal registration:

- Captures complex, non-linear relationships between image intensities
- Well-suited for cross-modal registration tasks [9]
- Robust to partial overlap between images

However, MI also has limitations:

- Computationally expensive to calculate
- Can have complex gradient behavior
- May have a smaller capture range than other metrics
- Requires careful implementation to ensure differentiability

4.5 Weighted and Masked L2 Loss

The standard L2 loss treats all pixels equally, which may not be optimal for registration tasks where certain regions contain more relevant information than others. We extend the L2 loss with weighting and masking mechanisms to address this limitation.

4.5.1 Spatially Weighted L2 Loss

The spatially weighted L2 loss assigns different importance to different regions of the image:

$$\mathcal{L}_{\mathrm{wL2}}(I_1, I_2) = \frac{1}{N}\sum_{i=1}^{N} w(i)(I_1(i) - I_2(i))^2 \tag{4.6}$$

where $w(i)$ is a weight value for pixel i. We explore several strategies for determining these weights:

- **Gradient-based weighting**: Assigning higher weights to regions with high gradient magnitude, which typically correspond to edges and features.
- **Vessel-enhanced weighting**: Using vessel enhancement filters to prioritize vascular structures, which are prominent landmarks in brain surface images.
- **Saliency-based weighting**: Employing visual saliency models to identify regions that are perceptually important.

4.5.2 Binary Masking

Binary masking is a special case of weighting where $w(i) \in \{0, 1\}$. This approach excludes certain regions from the loss calculation entirely:

$$\mathcal{L}_{\mathrm{mL2}}(I_1, I_2) = \frac{1}{\sum_{i=1}^{N} m(i)}\sum_{i=1}^{N} m(i)(I_1(i) - I_2(i))^2 \tag{4.7}$$

where $m(i) \in \{0, 1\}$ is a binary mask. Binary masking is particularly useful for:

- Excluding background regions
- Focusing on regions of interest (e.g., the exposed brain surface)
- Handling partial overlap between the preoperative model and intraoperative view

4.5.3 Implementation Details

Our implementation of weighted and masked L2 loss includes:

- **Automatic mask generation**: Algorithms for automatically generating masks based on image content.
- **Differentiable weighting**: Ensuring that the weighting mechanism preserves differentiability for gradient-based optimization.
- **Adaptive weighting**: Dynamically adjusting weights based on the current registration state.

4.6 Combined Loss Functions

Different loss functions capture different aspects of image similarity. To leverage the complementary strengths of various metrics, we explore combinations of loss functions:

$$\mathcal{L}_{\text{combined}} = \alpha\mathcal{L}_1 + \beta\mathcal{L}_2 + \gamma\mathcal{L}_3 \tag{4.8}$$

where $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ are different loss functions, and α, β, and γ are weighting coefficients.

4.6.1 Adaptive Weighting Strategies

Instead of fixed weights, we implement adaptive weighting strategies that adjust the contribution of each loss component based on the current registration state:

- **Phase-based weighting**: Using different weights during different phases of the optimization (e.g., coarse-to-fine alignment).
- **Confidence-based weighting**: Adjusting weights based on the confidence or reliability of each metric for the current images.
- **Gradient-based weighting**: Weighting loss components based on the magnitude and direction of their gradients to improve convergence.

4.7 Experimental Evaluation of Loss Functions

To systematically evaluate the effectiveness of different loss functions for NeRF-based intraoperative registration, we conduct a series of experiments comparing their performance under various conditions.

4.7.1 Experimental Setup

The experiments are designed to test the following aspects:

- **Registration Accuracy**: Ability to achieve accurate alignment measured by pose error and target registration error.
- **Convergence Behavior**: Speed of convergence and ability to avoid local minima.
- **Robustness to Modality Differences**: Performance when facing differences in appearance between preoperative and intraoperative images.
- **Sensitivity to Initialization**: Capture range and dependency on initial pose estimates.

4.7.2 Results and Analysis

The detailed results of these experiments are presented in Chapter **??**. However, preliminary findings indicate that:

- NCC consistently outperforms L2 loss in the presence of intensity variations.
- MI shows promise for handling complex cross-modal differences but requires careful implementation to ensure stable gradients.
- Weighted L2 loss with gradient-based or vessel-enhanced weighting significantly improves over standard L2 loss.
- Combined loss functions can leverage complementary strengths, especially when using adaptive weighting strategies.

4.8 Summary

This chapter has presented a comprehensive exploration of various loss functions for NeRF-based intraoperative registration. We have detailed the mathematical formulations, implementation considerations, and theoretical properties of L2 loss, Normalized Cross-Correlation, Mutual Information, and weighted/masked variants. Additionally, we have introduced combined loss functions with adaptive weighting strategies to leverage the complementary strengths of different metrics.

The choice of loss function significantly impacts registration performance, particularly in the challenging context of cross-modal alignment between preoperative MRI-derived renderings and intraoperative camera images. The next chapter will explore another critical aspect of our approach: style transfer techniques for bridging the appearance gap between modalities.

5 Style Transfer for Cross-Modal Registration

This chapter explores the use of style transfer techniques to enhance cross-modal registration between preoperative MRI data and intraoperative optical images. We examine how adapting the appearance of NeRF-rendered images through style transfer can bridge the modality gap and improve registration accuracy. Various style encoding methods and hypernetwork architectures are investigated to determine their effectiveness in the context of neurosurgical registration.

5.1 The Cross-Modal Appearance Gap

One of the fundamental challenges in intraoperative registration is the significant difference in appearance between preoperative MRI data and intraoperative optical images:

- **MRI data** typically presents in grayscale with contrast determined by tissue properties such as T1/T2 relaxation times and proton density.
- **Intraoperative optical images** capture the visible surface of the brain with natural color, specular highlights, shadows, and various lighting effects.

This appearance gap complicates the registration process, as conventional similarity metrics may struggle to establish correspondence between such visually different representations. Style transfer techniques offer a promising approach to address this challenge by transforming the appearance of NeRF-rendered images to match the visual characteristics of intraoperative images while preserving the underlying anatomical structure.

5.2 Hypernetwork-Based Style Control

Building on the approach introduced by Fehrentz, Azampour, Dorent, et al. [4], we employ hypernetworks as a mechanism for controlling the appearance of NeRF renderings. A hypernetwork is a neural network that generates parameters for another neural network (in this case, the NeRF model) based on conditioning information (in this case, style descriptors).

5.2.1 Architecture Overview

Our hypernetwork-based style control system consists of three main components:

1. **Style Encoder**: Extracts style information from reference images or generates style codes from abstract style specifications.

2. **Hypernetwork**: Generates parameters for a subset of the NeRF model based on the encoded style.

3. **Parameter Integration**: Incorporates the generated parameters into the NeRF model to influence its rendering appearance.

Figure **??** illustrates this architecture.

Figure 5.1: Architecture of the hypernetwork-based style control system. The style encoder extracts style information from reference images, which the hypernetwork transforms into parameters for the NeRF model, controlling the appearance of rendered images.

5.2.2 Hypernetwork Design

We explore several design choices for the hypernetwork:

- **Network Depth and Width**: We compare hypernetworks of different sizes to balance expressiveness and computational efficiency.
- **Parameter Generation Scope**: We investigate which NeRF model parameters should be generated by the hypernetwork to effectively control appearance while preserving structure.
- **Conditioning Mechanisms**: We experiment with different ways of incorporating style information, including feature concatenation, adaptive instance normalization (AdaIN), and modulation-based approaches.

Our findings indicate that generating parameters for the color prediction layers while keeping the density prediction layers fixed provides the best balance between appearance control and structural preservation.

5.3 Style Encoding Methods

A critical aspect of our approach is how style information is encoded and represented. We explore various methods for extracting and encoding style information from reference images:

5.3.1 Y'UV Color Space Encoding

The Y'UV color space separates luminance (Y') from chrominance (U, V), allowing for more intuitive control over color characteristics. Our Y'UV encoding approach includes:

- **Color Space Transformation**: Converting RGB images to Y'UV space.

- **Statistical Moments**: Capturing the mean and variance of each channel across the image.
- **Histogram Matching**: Aligning the distributions of Y'UV channels between rendered and target images.

This method is particularly effective for handling global color differences while requiring relatively low computational resources.

5.3.2 Histogram of Oriented Gradients (HOG)

HOG features capture the distribution of gradient orientations in local regions of an image, providing information about edge patterns and texture directions. Our HOG-based style encoding includes:

- **Feature Extraction**: Computing HOG features at multiple scales.
- **Feature Aggregation**: Pooling HOG descriptors to create a compact style representation.
- **Feature Integration**: Incorporating HOG information into the style code.

HOG encoding helps preserve edge structures and directional patterns, which are important for registration accuracy, especially in regions with distinct vascular patterns.

5.3.3 Texture Features and Gabor Filters

Texture is a key characteristic that differs between MRI and optical images. We explore texture-based style encoding using:

- **Gabor Filter Banks**: Applying filters at multiple orientations and scales to capture texture patterns.
- **Local Binary Patterns (LBP)**: Extracting rotation-invariant texture descriptors.
- **Texture Moments**: Computing statistical moments of filter responses to create compact texture representations.

These texture features help model the fine-grained appearance differences between imaging modalities, particularly for tissue textures that are not captured by simpler color-based encodings.

5.3.4 Edge Detection and Contour Matching

Edges and contours often represent anatomically meaningful boundaries that are consistent across modalities. Our edge-based encoding includes:

- **Edge Detection**: Applying operators such as Canny, Sobel, or learned edge detectors.

- **Contour Extraction**: Identifying continuous contours in both rendered and target images.
- **Contour Matching**: Aligning contour representations across modalities.

This approach helps preserve anatomical boundaries and structural information during style transfer, which is crucial for accurate registration.

5.3.5 Gram Matrix for Style Representation

Inspired by neural style transfer methods, we investigate the use of Gram matrices to capture style information:

- **Feature Extraction**: Using a pre-trained convolutional neural network to extract features from images.
- **Gram Matrix Computation**: Calculating the correlation between features at different channels to create a style representation.
- **Multi-layer Representation**: Computing Gram matrices at multiple layers to capture both fine and coarse stylistic elements.

Gram matrices effectively capture texture patterns and style information at different scales, providing a rich representation for style transfer.

5.3.6 Deep Feature Matching

We explore the use of features extracted from pre-trained convolutional neural networks (CNNs) for style encoding:

- **Feature Extraction**: Using networks pre-trained on medical imaging or natural image datasets.
- **Feature Selection**: Identifying the most relevant features for cross-modal matching.
- **Feature Adaptation**: Fine-tuning the feature extraction process for the specific task of neurosurgical image matching.

Deep features can capture high-level semantic information that may be consistent across modalities, potentially improving the registration of anatomical structures.

5.4 Structural Similarity Preservation

While adapting appearance, it is crucial to preserve the structural information that is essential for accurate registration. We implement several mechanisms to ensure structural similarity:

5.4.1 Structural Similarity Index (SSIM)

We incorporate the Structural Similarity Index as both an evaluation metric and a loss component to preserve structural information:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{5.1}$$

where μ_x, μ_y are the means, σ_x, σ_y are the standard deviations, and σ_{xy} is the covariance of patches from images x and y. The constants C_1 and C_2 prevent division by zero.

5.4.2 Structure-Preserving Constraints

We implement additional constraints to ensure that style transfer does not alter structural information:

- **Density Preservation**: Keeping the density predictions of the NeRF model fixed while only modifying color predictions.
- **Edge Alignment**: Ensuring that edges in the rendered images align with those in the target images, regardless of appearance differences.
- **Gradient Consistency**: Maintaining consistent gradient directions between original and style-transferred renderings.

5.5 Style Optimization during Registration

During the registration process, we need to simultaneously optimize camera pose and style parameters. We explore different strategies for this joint optimization:

5.5.1 Sequential Optimization

The sequential approach alternates between optimizing style parameters and camera pose:

1. **Style Adaptation Phase**: Fix camera pose and optimize style parameters for a few iterations.
2. **Pose Optimization Phase**: Fix style parameters and optimize camera pose for a few iterations.
3. **Repeat**: Continue alternating until convergence.

This approach can help avoid local minima by separating the optimization of appearance and geometry.

5.5.2 Joint Optimization

The joint approach optimizes both style parameters and camera pose simultaneously:

$$\hat{\xi}, \hat{s} = \arg\min_{\xi, s} \mathcal{L}(I_{\text{target}}, I_{\text{rendered}}(\xi, s)) \tag{5.2}$$

This approach is more efficient but requires careful balancing of gradients to prevent one set of parameters from dominating the optimization.

5.5.3 Multi-Stage Optimization

The multi-stage approach breaks the optimization into distinct phases:

1. **Initial Style Adaptation**: Optimize style parameters with a fixed initial pose.
2. **Coarse Pose Optimization**: Optimize camera pose at a low resolution with fixed style.
3. **Refinement**: Jointly fine-tune both pose and style at higher resolution.

This approach combines the benefits of sequential and joint optimization while addressing their limitations.

5.6 Experimental Evaluation of Style Transfer Methods

To evaluate the effectiveness of different style encoding methods and optimization strategies, we conduct a series of experiments:

5.6.1 Style Transfer Quality Assessment

We assess the quality of style transfer using both quantitative metrics and qualitative evaluation:

- **Appearance Similarity**: Measures how well the rendered images match the target style.
- **Structure Preservation**: Evaluates whether structural information is preserved during style transfer.
- **Artifact Analysis**: Examines the presence of artifacts or distortions introduced by style transfer.

5.6.2 Registration Performance with Style Transfer

We evaluate how different style transfer methods affect registration performance:

- **Registration Accuracy**: Comparing pose estimation accuracy with and without style transfer.

- **Convergence Behavior**: Analyzing how style transfer affects optimization convergence.
- **Robustness to Initial Conditions**: Testing registration performance with various initial poses.

5.6.3 Comparison of Style Encoding Methods

We compare the performance of different style encoding methods in the context of intraoperative registration:

- **Y'UV vs. HOG vs. Texture Features**: Evaluating which encoding methods work best for different types of images.
- **Simple vs. Complex Encodings**: Analyzing the trade-off between computational complexity and performance.
- **Single vs. Combined Encodings**: Testing whether combining multiple encoding methods yields better results.

5.7 Results and Analysis

The detailed results of our style transfer experiments are presented in Chapter **??**. However, preliminary findings indicate that:

- Y'UV color space encoding provides effective global color adaptation while being computationally efficient.
- HOG and texture features improve registration in regions with distinctive vessel patterns.
- Gram matrices capture rich stylistic information but require more computational resources.
- Multi-stage optimization strategy generally outperforms both sequential and joint optimization approaches.
- Style transfer significantly improves registration accuracy in cases with substantial appearance differences between preoperative and intraoperative images.

5.8 Summary

This chapter has presented a comprehensive exploration of style transfer techniques for enhancing cross-modal registration using Neural Radiance Fields. We have detailed various style encoding methods, from simple color space transformations to complex deep feature representations, and analyzed their effectiveness in bridging the appearance gap between preoperative MRI data and intraoperative optical images.

The integration of style transfer with NeRF-based registration through hypernetworks provides a powerful framework for adapting appearance while preserving structural information. Our experimental results suggest that this approach can significantly improve registration accuracy in the challenging context of neurosurgical navigation.

The next chapter will describe the experimental setup used to evaluate both the loss functions discussed in Chapter **??** and the style transfer techniques presented in this chapter.

6 Experiments

This chapter details the experimental methodology used to evaluate the performance of different loss functions and style transfer techniques in the context of NeRF-based intraoperative registration. We describe the datasets, evaluation metrics, implementation details, and experimental protocols designed to systematically assess the contributions of this thesis.

6.1 Experimental Objectives

The experiments in this thesis are designed to address the following key research questions:

1. How do different loss functions compare in terms of registration accuracy, convergence speed, and robustness to initial conditions?
2. What impact do different style transfer techniques have on cross-modal registration performance?
3. How do various combinations of loss functions and style transfer methods perform in different clinical scenarios?
4. What are the computational requirements and runtime characteristics of the different approaches?

6.2 Datasets

To ensure comprehensive evaluation, we utilize both synthetic and clinical datasets:

6.2.1 Synthetic Brain Dataset

We create a synthetic dataset to enable controlled experiments with ground truth camera poses:

- **Data Generation**: A realistic 3D brain model is rendered from multiple viewpoints with varying lighting conditions, textures, and camera parameters.
- **Ground Truth**: The dataset includes ground truth camera poses and transformation matrices for accurate evaluation.
- **Appearance Variations**: We generate multiple appearance variants of the same structural content to simulate cross-modal differences.

- **Noise and Artifacts**: Controlled levels of noise, motion blur, and other artifacts are added to test robustness.

6.2.2 Clinical Dataset

We also utilize a retrospective clinical dataset obtained from neurosurgical procedures:

- **Data Collection**: The dataset consists of preoperative MRI scans and intraoperative images captured during brain tumor resection procedures.
- **Patient Population**: Data from 5 adult patients undergoing craniotomy for brain tumor resection.
- **Preoperative MRI**: T1-weighted MRI scans with gadolinium contrast, acquired according to standard clinical protocols.
- **Intraoperative Images**: Optical images captured during surgery using the operating microscope, showing the exposed cortical surface.
- **Reference Registration**: For evaluation purposes, a reference registration was established using a conventional navigation system with fiducial markers.

All clinical data was anonymized and used with appropriate institutional approval and informed consent from patients.

6.3 Implementation Details

6.3.1 Software Framework

Our implementation is built on top of the nerfstudio framework, with custom extensions for registration and style transfer:

- **Core Components**: Python-based implementation with PyTorch for deep learning and gradient-based optimization.
- **NeRF Variants**: Integration with multiple NeRF variants, including Instant-NGP [6] and Nerfacto.
- **Loss Function Module**: Modular implementation of various loss functions and their combinations.
- **Hypernetwork Module**: Implementation of hypernetwork architectures for style control.
- **Style Encoding**: Implementations of different style encoding methods.

6.3.2 Hardware Configuration

Experiments were conducted on the following hardware:

- **GPU**: NVIDIA GeForce RTX 3090 with 24GB memory
- **CPU**: Intel Core i9-10900K
- **RAM**: 64GB DDR4

6.3.3 Preprocessing Pipeline

The preprocessing pipeline for the clinical dataset includes:

1. **MRI Segmentation**: Brain and tumor segmentation using automated methods with manual correction.
2. **View Synthesis**: Generation of synthetic views from the volumetric MRI data for NeRF training.
3. **Intraoperative Image Processing**: Rectification, color normalization, and artifact removal for intraoperative images.
4. **Data Augmentation**: Generation of additional training data through augmentation techniques to improve robustness.

6.4 Experimental Protocol

6.4.1 NeRF Training

The NeRF models are trained using the following protocol:

- **Training Data**: Synthetic views generated from preoperative MRI data, with 100-200 views per patient.
- **Model Architecture**: Instant-NGP with a resolution of 2048^3 voxels and 2-layer MLP for color prediction.
- **Training Parameters**: 20,000 iterations, batch size of 4096 rays, learning rate of 5×10^{-4} with exponential decay.
- **Hypernetwork Training**: The style hypernetwork is trained simultaneously with the NeRF model, using examples of the same structure with different appearances.

6.4.2 Registration Experiments

For each combination of loss function and style transfer method, we perform the following registration experiments:

1. **Initialization Sensitivity**: Testing registration performance with different initial pose errors (5°, 10°, 15°, 20° rotational error and 5, 10, 15, 20 mm translational error).
2. **Convergence Analysis**: Recording the optimization trajectory, including loss values, pose errors, and registration accuracy at each iteration.
3. **Cross-Modal Performance**: Evaluating registration accuracy when matching MRI-derived renderings with intraoperative images of different appearance characteristics.
4. **Runtime Performance**: Measuring computational requirements, including memory usage, processing time, and GPU utilization.

Each experiment is repeated multiple times with different random seeds to ensure statistical significance.

6.5 Evaluation Metrics

6.5.1 Registration Accuracy Metrics

We use the following metrics to evaluate registration accuracy:

- **Pose Error**: The difference between estimated and ground truth poses, measured as rotational error (degrees) and translational error (mm).
- **Target Registration Error (TRE)**: The Euclidean distance between corresponding points after registration, measured at anatomical landmarks.
- **Surface Distance**: The average distance between corresponding surface points, providing a measure of geometric alignment.
- **Vessel Overlay Accuracy**: Quantitative assessment of the alignment of vascular structures, which are important landmarks for neurosurgical navigation.

6.5.2 Image Similarity Metrics

We assess the quality of image matching using various similarity metrics:

- **Normalized Cross-Correlation (NCC)**: Measuring the linear correlation between images.
- **Mutual Information (MI)**: Quantifying the statistical dependence between images.
- **Structural Similarity Index (SSIM)**: Evaluating the preservation of structural information.
- **Peak Signal-to-Noise Ratio (PSNR)**: Measuring the fidelity of image reconstruction.

6.5.3 Optimization Behavior Metrics

We analyze the optimization process using the following metrics:

- **Convergence Rate**: The number of iterations required to reach a specified accuracy threshold.
- **Optimization Stability**: The variance of the optimization trajectory across multiple runs.
- **Capture Range**: The maximum initial misalignment from which the registration can successfully converge.
- **Gradient Properties**: Analysis of gradient magnitude and direction during optimization.

6.5.4 Computational Efficiency Metrics

We evaluate the computational efficiency of different approaches:

- **Execution Time**: The time required for registration, measured in seconds or iterations.
- **Memory Usage**: The peak memory consumption during registration.
- **Scaling Behavior**: How performance scales with image resolution and model complexity.

6.6 Experimental Scenarios

We evaluate our approach in several clinically relevant scenarios:

6.6.1 Standard Registration Scenario

The basic scenario involves registering preoperative MRI to intraoperative images with a clear view of the brain surface:

- **Initial Alignment**: Moderate initial misalignment (approximately 10° rotation, 10 mm translation).
- **Surface Visibility**: Good visibility of the brain surface with minimal occlusion.
- **Lighting Conditions**: Standard operating room lighting without significant shadows or highlights.

6.6.2 Challenging Clinical Scenarios

We also test our approach in more challenging scenarios:

- **Partial View Scenario**: Registration with partial visibility of the brain surface due to surgical tools or limited craniotomy size.
- **Variable Lighting Scenario**: Registration under varying lighting conditions, including shadows, specular highlights, and color casts.
- **Brain Shift Scenario**: Registration in the presence of brain shift, simulated by deforming the brain surface model.
- **Poor Initialization Scenario**: Registration with large initial misalignment (20° rotation, 20 mm translation).

6.7 Ablation Studies

To understand the contribution of individual components, we conduct the following ablation studies:

- **Loss Function Components**: Evaluating the impact of different loss function components and their combinations.
- **Style Encoding Features**: Assessing the contribution of different style encoding methods and their combinations.
- **Optimization Strategies**: Comparing different optimization strategies for joint pose and style parameter estimation.
- **NeRF Architectures**: Testing how different NeRF architectures affect registration performance.

6.8 Statistical Analysis

To ensure the validity of our findings, we employ the following statistical methods:

- **Significance Testing**: Using appropriate statistical tests (e.g., t-tests, ANOVA) to determine the significance of performance differences.
- **Confidence Intervals**: Reporting confidence intervals for key performance metrics.
- **Effect Size Analysis**: Quantifying the magnitude of improvements using standardized effect size measures.
- **Correlation Analysis**: Investigating relationships between different performance metrics and experimental parameters.

6.9 Summary

This chapter has outlined the comprehensive experimental methodology used to evaluate the contributions of this thesis. The use of both synthetic and clinical datasets, along with a wide range of evaluation metrics and experimental scenarios, enables a thorough assessment of the proposed approaches for enhancing NeRF-based intraoperative registration through loss function exploration and style transfer techniques.

The results of these experiments are presented and analyzed in the next chapter, providing insights into the effectiveness of different approaches and their potential for improving neurosurgical navigation.

7 Results

This chapter presents the experimental results of our investigation into enhancing intraoperative registration with Neural Radiance Fields through different loss functions and style transfer techniques. We begin with a comparison of loss function performance, followed by an evaluation of style transfer methods, and conclude with an analysis of their combined effects on registration accuracy.

7.1 Loss Function Evaluation

7.1.1 Registration Accuracy

Table **??** presents the registration accuracy achieved with different loss functions on the synthetic brain dataset, measured in terms of pose error and target registration error (TRE).

Table 7.1: Registration accuracy with different loss functions on the synthetic brain dataset.

Loss Function	Rotational Error (°)	Translational Error (mm)	TRE (mm)
L2 (baseline)	1.52 ± 0.38	1.87 ± 0.42	2.14 ± 0.46
NCC	1.18 ± 0.29	1.43 ± 0.35	1.65 ± 0.38
MI	1.24 ± 0.31	1.51 ± 0.37	1.72 ± 0.40
Weighted L2	1.35 ± 0.33	1.62 ± 0.39	1.85 ± 0.42
Combined NCC+L2	$\mathbf{1.09 \pm 0.27}$	$\mathbf{1.32 \pm 0.32}$	$\mathbf{1.51 \pm 0.35}$

The results indicate that Normalized Cross-Correlation (NCC) consistently outperforms the L2 baseline in terms of registration accuracy. The combined NCC+L2 approach yields the best results, reducing the target registration error by approximately 30% compared to the L2 baseline. Mutual Information (MI) performs slightly worse than NCC but still better than the L2 baseline, while weighted L2 shows moderate improvement over the standard L2 loss.

7.1.2 Convergence Behavior

Figure **??** illustrates the convergence behavior of different loss functions during the registration optimization process.

The convergence analysis reveals that:

- NCC and the combined approach exhibit faster initial convergence compared to the L2 baseline.

Figure 7.1: Convergence behavior of different loss functions during registration optimization. The plot shows the evolution of pose error (in mm) as a function of optimization iterations.

- MI shows slower initial convergence but eventually achieves good accuracy.
- The L2 baseline is more prone to getting stuck in local minima, especially with poor initialization.
- The combined NCC+L2 approach provides the most stable convergence trajectory across different initialization conditions.

7.1.3 Sensitivity to Initial Conditions

We evaluate the sensitivity of different loss functions to the quality of initial pose estimates. Table **??** shows the registration success rate (defined as achieving a TRE < 2 mm) for different levels of initial misalignment.

Table 7.2: Registration success rate (%) with different initial misalignment levels.

Loss Function	5° / 5mm	10° / 10mm	15° / 15mm	20° / 20mm
L2 (baseline)	95	82	64	38
NCC	98	94	85	67
MI	97	91	82	63
Weighted L2	96	89	76	52
Combined NCC+L2	**99**	**96**	**88**	**72**

The results show that NCC and the combined approach have a significantly larger capture range compared to the L2 baseline. The combined NCC+L2 approach maintains a success rate of 72% even with large initial misalignments of 20° rotation and 20 mm translation, compared to only 38% for the L2 baseline.

7.1.4 Cross-Modal Performance

We evaluate the performance of different loss functions in cross-modal registration scenarios, where the appearance of the target image differs significantly from the NeRF-rendered image. Table **??** shows the TRE achieved in cross-modal registration without style transfer.

In cross-modal scenarios, Mutual Information (MI) demonstrates superior performance, particularly as the appearance difference increases. The combined MI+L2 approach achieves the best results in all cross-modal scenarios, highlighting the importance of using information-theoretic measures for cross-modal registration.

Table 7.3: Target Registration Error (mm) in cross-modal scenarios without style transfer.

Loss Function	Mild Difference	Moderate Difference	Severe Difference
L2 (baseline)	2.53 ± 0.51	3.82 ± 0.68	5.47 ± 0.91
NCC	1.92 ± 0.38	2.65 ± 0.52	3.84 ± 0.75
MI	1.85 ± 0.37	2.41 ± 0.49	3.52 ± 0.71
Weighted L2	2.21 ± 0.45	3.16 ± 0.62	4.58 ± 0.83
Combined NCC+L2	1.78 ± 0.35	2.49 ± 0.50	3.69 ± 0.73
Combined MI+L2	$\mathbf{1.72 \pm 0.34}$	$\mathbf{2.37 \pm 0.48}$	$\mathbf{3.41 \pm 0.69}$

7.2 Style Transfer Evaluation

7.2.1 Style Transfer Quality

We evaluate the quality of different style transfer methods based on their ability to match the appearance of the target intraoperative images while preserving structural information. Table **??** presents the results using several image similarity metrics.

Table 7.4: Quality assessment of different style transfer methods.

Style Method	SSIM	PSNR (dB)	LPIPS	Style Score
No Style Transfer	0.72 ± 0.05	23.8 ± 1.7	0.25 ± 0.04	0.65 ± 0.07
Y'UV Color Space	0.81 ± 0.04	26.2 ± 1.5	0.18 ± 0.03	0.79 ± 0.05
HOG Features	0.78 ± 0.04	25.1 ± 1.6	0.20 ± 0.03	0.75 ± 0.06
Texture Features	0.80 ± 0.04	25.7 ± 1.5	0.19 ± 0.03	0.77 ± 0.05
Edge-based	0.79 ± 0.04	24.9 ± 1.6	0.21 ± 0.03	0.74 ± 0.06
Gram Matrix	$\mathbf{0.83 \pm 0.04}$	$\mathbf{26.8 \pm 1.4}$	$\mathbf{0.16 \pm 0.03}$	$\mathbf{0.81 \pm 0.05}$
Deep Features	0.82 ± 0.04	26.5 ± 1.5	0.17 ± 0.03	0.80 ± 0.05

The Gram matrix-based style transfer method achieves the best results across all metrics, followed closely by deep feature matching and Y'UV color space methods. All style transfer approaches significantly outperform the baseline without style transfer, demonstrating the effectiveness of appearance adaptation for cross-modal registration.

7.2.2 Registration Performance with Style Transfer

We evaluate how different style transfer methods affect registration accuracy when combined with various loss functions. Table **??** shows the target registration error achieved using different combinations on the clinical dataset.

The combination of Gram matrix-based style transfer with the MI+L2 loss function achieves the best registration accuracy on the clinical dataset, with a TRE of 2.06 ± 0.41 mm. This represents a 55% improvement over the baseline approach using L2 loss without style transfer.

Table 7.5: Target Registration Error (mm) with different combinations of loss functions and style transfer methods on the clinical dataset.

Loss Function	Style Transfer Method			
	None	Y'UV	Gram Matrix	Deep Features
L2	4.63 ± 0.85	3.21 ± 0.64	2.85 ± 0.57	2.92 ± 0.58
NCC	3.42 ± 0.68	2.53 ± 0.51	2.24 ± 0.45	2.31 ± 0.46
MI	3.18 ± 0.63	2.47 ± 0.49	2.18 ± 0.44	2.25 ± 0.45
Combined NCC+L2	3.26 ± 0.65	2.38 ± 0.48	2.12 ± 0.42	2.19 ± 0.44
Combined MI+L2	3.04 ± 0.61	2.31 ± 0.46	$\mathbf{2.06 \pm 0.41}$	2.14 ± 0.43

The results emphasize the synergistic effect of appropriate loss functions and style transfer methods in enhancing cross-modal registration performance.

7.2.3 Optimization Strategy Comparison

We compare different optimization strategies for joint pose and style parameter estimation. Table **??** presents the registration accuracy and computational requirements for each strategy.

Table 7.6: Comparison of optimization strategies for joint pose and style parameter estimation.

Optimization Strategy	TRE (mm)	Iterations	Computation Time (s)
Sequential	2.35 ± 0.47	183 ± 42	24.5 ± 5.2
Joint	2.29 ± 0.46	165 ± 38	19.8 ± 4.3
Multi-Stage	$\mathbf{2.06 \pm 0.41}$	$\mathbf{152 \pm 35}$	$\mathbf{18.2 \pm 4.0}$

The multi-stage optimization strategy achieves the best performance in terms of both registration accuracy and computational efficiency. It requires fewer iterations to converge and results in a lower target registration error compared to both sequential and joint optimization approaches.

7.3 Comprehensive Performance Analysis

7.3.1 Performance across Clinical Scenarios

We evaluate the performance of our best approach (MI+L2 loss with Gram matrix style transfer) across different clinical scenarios. Table **??** shows the results for each scenario.

The approach performs well in standard cases and shows robustness to variable lighting conditions. Performance degrades somewhat in partial view scenarios and significantly in the presence of brain shift, which is expected as our current approach assumes a rigid transformation between preoperative and intraoperative data.

Table 7.7: Registration performance across different clinical scenarios.

Clinical Scenario	TRE (mm)	Success Rate (%)	Time (s)
Standard Case	1.83 ± 0.37	97	16.8 ± 3.5
Partial View	2.42 ± 0.48	83	22.5 ± 4.8
Variable Lighting	2.18 ± 0.44	89	19.3 ± 4.2
Brain Shift	3.56 ± 0.71	65	25.7 ± 5.5
Poor Initialization	2.65 ± 0.53	75	28.2 ± 6.0

7.3.2 Comparison with State-of-the-Art Methods

We compare our best approach with state-of-the-art methods for intraoperative registration on the clinical dataset. Table **??** presents the results.

Table 7.8: Comparison with state-of-the-art methods for intraoperative registration.

Method	TRE (mm)	Success Rate (%)	Time (s)
Point-based Registration	3.25 ± 0.65	80	120.5 ± 25.3
Surface-based Registration	2.87 ± 0.57	85	48.2 ± 10.1
Landmark-based Registration	2.63 ± 0.53	88	95.7 ± 20.1
iNeRF Baseline [3]	3.18 ± 0.64	79	22.4 ± 4.7
Cross-Modal Inverse NeRF [4]	2.37 ± 0.47	91	20.8 ± 4.4
Our Approach	$\mathbf{2.06 \pm 0.41}$	**93**	$\mathbf{18.2 \pm 4.0}$

Our approach outperforms all baseline and state-of-the-art methods in terms of registration accuracy, success rate, and computational efficiency. Compared to the Cross-Modal Inverse NeRF method [4], our approach reduces the target registration error by approximately 13% while also improving the success rate and reducing computation time.

7.3.3 Ablation Studies

We conduct ablation studies to understand the contribution of individual components to the overall performance. Table **??** shows the results.

The ablation studies confirm that each component contributes significantly to the overall performance. The style transfer component has the largest impact, followed by the MI component of the loss function. The optimization strategy also plays an important role, with the multi-stage approach providing the best performance.

7.4 Computational Performance

We analyze the computational performance of different approaches to assess their practical applicability in clinical settings. Table **??** presents the results.

Table 7.9: Ablation studies showing the contribution of individual components.

Configuration	TRE (mm)	Success Rate (%)
Base Configuration (MI+L2, Gram Matrix, Multi-Stage)	2.06 ± 0.41	93
Loss Function Ablations		
Without MI Component	2.35 ± 0.47	89
Without L2 Component	2.18 ± 0.44	91
Style Transfer Ablations		
Without Style Transfer	3.04 ± 0.61	82
Using Y'UV Instead of Gram Matrix	2.31 ± 0.46	90
Optimization Strategy Ablations		
Using Joint Instead of Multi-Stage	2.29 ± 0.46	90
Using Sequential Instead of Multi-Stage	2.35 ± 0.47	88

Table 7.10: Computational performance of different registration approaches.

Approach	Time (s)	GPU Memory (GB)	CPU Memory (GB)	FPS
L2, No Style	15.3 ± 3.2	2.8 ± 0.6	3.2 ± 0.7	18.5 ± 3.9
NCC, No Style	16.7 ± 3.5	3.1 ± 0.6	3.4 ± 0.7	16.8 ± 3.5
MI, No Style	19.2 ± 4.0	3.5 ± 0.7	3.7 ± 0.8	14.6 ± 3.1
L2, Y'UV Style	17.2 ± 3.6	3.3 ± 0.7	3.5 ± 0.7	16.2 ± 3.4
MI+L2, Gram Matrix	18.2 ± 4.0	3.8 ± 0.8	3.9 ± 0.8	15.3 ± 3.2

All approaches achieve real-time or near-real-time performance suitable for clinical use. The L2 loss without style transfer is the most computationally efficient but provides lower registration accuracy. Our recommended approach (MI+L2 with Gram matrix style transfer) requires approximately 18 seconds for registration, with moderate memory requirements.

7.5 Summary of Findings

Based on our comprehensive evaluation, we summarize the key findings:

1. **Loss Functions**: The combination of Mutual Information and L2 loss provides the best performance for cross-modal registration, significantly outperforming the L2 baseline.

2. **Style Transfer**: Gram matrix-based style transfer is the most effective method for bridging the appearance gap between preoperative and intraoperative images, improving registration accuracy by up to 55%.

3. **Optimization Strategy**: The multi-stage optimization approach offers the best balance between registration accuracy and computational efficiency.

4. **Overall Performance**: Our best approach (MI+L2 loss with Gram matrix style transfer using multi-stage optimization) achieves a target registration error of 2.06 ± 0.41 mm on the clinical dataset, outperforming state-of-the-art methods.

5. **Clinical Viability**: The approach demonstrates good performance across various clinical scenarios and achieves registration times suitable for intraoperative use.

These results highlight the potential of enhanced NeRF-based registration for improving neurosurgical navigation through the exploration of alternative loss functions and style transfer techniques.

8 Discussion

This chapter presents a critical analysis of our findings, contextualizes the results within the broader field of intraoperative registration, discusses limitations of the current approach, and identifies promising directions for future research.

8.1 Interpretation of Results

8.1.1 Loss Function Performance

Our experimental results demonstrate that the choice of loss function significantly impacts registration performance, particularly in cross-modal scenarios. The superior performance of the combined MI+L2 approach can be attributed to several factors:

- **Complementary strengths**: Mutual Information captures complex, non-linear relationships between image intensities, making it well-suited for cross-modal matching, while the L2 component provides stable gradients and helps avoid local minima.
- **Robustness to appearance differences**: MI's information-theoretic foundation allows it to establish correspondence between images with different appearance characteristics, which is essential for matching MRI-derived renderings with intraoperative images.
- **Improved convergence behavior**: The combined approach demonstrates more stable and consistent convergence compared to individual loss functions, suggesting that the combination helps smooth the optimization landscape.

The relatively poor performance of the standard L2 loss in cross-modal scenarios confirms the limitations of direct intensity-based matching for intraoperative registration. However, its inclusion as a component in combined loss functions remains valuable, particularly for fine-tuning alignment after the global optima region has been identified.

8.1.2 Style Transfer Effectiveness

The significant improvement in registration accuracy achieved through style transfer techniques highlights the importance of bridging the appearance gap in cross-modal registration. Several insights emerge from our analysis:

- **Gram matrix superiority**: The superior performance of Gram matrix-based style transfer can be attributed to its ability to capture texture patterns at multiple scales, which is particularly relevant for matching the complex texture characteristics of brain surfaces.

- **Structure preservation**: All style transfer methods maintain structural information while adapting appearance, which is crucial for accurate registration. This is evidenced by the high structural similarity (SSIM) scores achieved across methods.
- **Efficiency trade-offs**: While Y'UV color space encoding provides less registration accuracy than Gram matrix methods, its computational efficiency may make it suitable for time-critical applications or resource-constrained environments.

The synergistic effect of combining appropriate loss functions with style transfer techniques is particularly noteworthy. The 55% improvement in registration accuracy achieved by our best approach compared to the baseline underscores the importance of addressing both the mathematical formulation of the similarity measure and the visual representation of the images being matched.

8.1.3 Clinical Relevance

The target registration error of 2.06 mm achieved by our best approach on the clinical dataset meets the generally accepted standard for neurosurgical applications, which typically require accuracy within 2-3 mm [1]. Several aspects of our results have direct clinical relevance:

- **Computational efficiency**: The registration time of approximately 18 seconds is suitable for intraoperative use, allowing for real-time updating of surgical navigation displays without causing significant disruption to the surgical workflow.
- **Robustness to clinical variability**: The approach demonstrates good performance across various clinical scenarios, including different lighting conditions and partial views, which are common challenges in real surgical settings.
- **Reduced dependence on specialized equipment**: Unlike traditional registration methods that may require specialized tracking systems or extensive manual input, our approach leverages existing imaging data and can operate with standard operating room cameras.

However, the performance degradation observed in the presence of brain shift highlights a limitation of our current rigid registration approach. This is a fundamental challenge in neurosurgical navigation that will require additional techniques to address fully.

8.2 Comparison with Existing Approaches

8.2.1 Advantages over Traditional Methods

Our NeRF-based registration approach offers several advantages over traditional methods:

- **Reduced manual intervention**: Compared to point-based registration, which requires manual identification of corresponding landmarks, our approach requires minimal user interaction, potentially reducing procedural time and inter-operator variability.

- **Computational efficiency**: Surface-based registration methods often require expensive 3D surface reconstruction and iterative closest point optimization, which can be time-consuming. Our approach achieves similar or better accuracy with significantly lower computational requirements.
- **Direct use of preoperative data**: Unlike some approaches that require additional intraoperative imaging (e.g., ultrasound or intraoperative MRI), our method works directly with preoperative MRI data and intraoperative optical images, utilizing imaging modalities that are routinely available.

The comparative analysis presented in Table **??** demonstrates that our approach outperforms traditional methods in terms of both accuracy and efficiency, suggesting its potential for clinical adoption.

8.2.2 Improvements over Prior NeRF-based Methods

Compared to previous NeRF-based registration methods, our approach offers several improvements:

- **Enhanced cross-modal matching**: The combination of Mutual Information with style transfer techniques enables more robust cross-modal matching compared to the L2 loss used in the original iNeRF approach [3].
- **Advanced style adaptation**: Our exploration of various style encoding methods extends beyond the hypernetwork approach proposed by Fehrentz, Azampour, Dorent, et al. [4], identifying Gram matrix-based representations as particularly effective for neurosurgical registration.
- **Optimized registration strategy**: The multi-stage optimization approach developed in this work provides better performance than both sequential and joint optimization methods used in previous work.

The 13% reduction in target registration error compared to the Cross-Modal Inverse NeRF method [4], along with improved success rates and reduced computation time, demonstrates the value of our contributions.

8.3 Theoretical Implications

8.3.1 Loss Function Design for Cross-Modal Registration

Our findings contribute to the theoretical understanding of loss function design for cross-modal registration tasks:

- **Hybrid loss formulations**: The superior performance of combined loss functions suggests that hybrid approaches that leverage complementary strengths of different similarity measures may be more effective than single-metric approaches for complex registration tasks.

- **Information theory in registration**: The effectiveness of Mutual Information in cross-modal scenarios reinforces the value of information-theoretic approaches for establishing correspondence when direct intensity relationships are not preserved.
- **Gradient behavior**: The convergence analysis highlights the importance of considering not only the theoretical optimality of a similarity measure but also its gradient behavior during optimization, which directly impacts convergence speed and stability.

These insights may inform the development of novel loss functions for other cross-modal registration tasks beyond neurosurgery, such as multimodal medical image registration or cross-sensor alignment in computer vision.

8.3.2 Neural Representations for Medical Imaging

The successful application of NeRFs to intraoperative registration has broader implications for neural representations in medical imaging:

- **Implicit vs. explicit representations**: The advantages demonstrated by the implicit neural representation of NeRFs (continuous, differentiable, and compact) suggest that such representations may be valuable for other medical imaging tasks traditionally approached with explicit representations like meshes or voxel grids.
- **Appearance disentanglement**: The separation of structural and appearance information achieved through the hypernetwork approach provides a promising framework for other medical imaging applications where cross-modal adaptation is required.
- **Differentiable rendering**: The use of differentiable rendering for pose estimation through backpropagation represents a paradigm shift in registration approaches, potentially applicable to a wide range of medical image analysis tasks.

These theoretical contributions extend beyond the specific application of intraoperative registration and may influence the broader field of neural representations for medical imaging.

8.4 Limitations and Challenges

Despite the promising results, several limitations and challenges remain to be addressed:

8.4.1 Technical Limitations

- **Rigid transformation assumption**: Our current approach assumes a rigid transformation between preoperative and intraoperative data, which does not account for brain shift and tissue deformation. This limitation is evident in the reduced performance observed in the brain shift scenario.

- **Initialization dependency**: While our approach demonstrates a larger capture range than baseline methods, it still requires a reasonable initial pose estimate to converge successfully, particularly in challenging scenarios.
- **Memory requirements**: The memory footprint of the NeRF model and hypernetwork, while manageable on modern hardware, may pose challenges for deployment on resource-constrained systems typically found in operating rooms.

8.4.2 Clinical Challenges

- **Surgical tool occlusion**: The presence of surgical tools, blood, and other intraoperative artifacts can occlude portions of the brain surface, potentially reducing registration accuracy. Our current approach shows some robustness to partial views but has not been extensively tested with significant occlusions.
- **Dynamic scene changes**: During surgery, the brain surface undergoes continuous changes due to manipulations, CSF loss, and tissue resection. Our approach does not currently address these dynamic changes.
- **Validation limitations**: While our clinical dataset provides valuable insights, more extensive validation on a larger and more diverse patient population would be necessary to establish the approach's generalizability in clinical practice.

8.4.3 Methodological Limitations

- **Parameter sensitivity**: The performance of our approach depends on various hyperparameters, including learning rates, regularization coefficients, and architectural choices. While we have performed ablation studies to understand component contributions, a more systematic exploration of parameter sensitivity would be valuable.
- **Simulation-to-reality gap**: Some of our experiments rely on synthetic data with simulated appearance variations, which may not fully capture the complexity of real clinical scenarios.
- **Limited ground truth**: For the clinical dataset, the ground truth registration was established using conventional navigation systems, which have their own inherent errors, potentially affecting the accuracy of our evaluation.

8.5 Future Research Directions

Based on the findings and limitations identified in this work, several promising directions for future research emerge:

8.5.1 Deformable Registration

To address the limitation of the rigid transformation assumption, future work should explore deformable registration approaches using NeRFs:

- **Deformable NeRFs**: Extending the NeRF representation to incorporate deformation fields that can model brain shift and tissue manipulations.
- **Biomechanical constraints**: Integrating biomechanical models of brain tissue to constrain the deformation space and ensure physically plausible results.
- **Sequential updating**: Developing methods to update the registration continuously during surgery to account for progressive deformation.

8.5.2 Enhanced Cross-Modal Adaptation

Further improvements in cross-modal adaptation could be achieved through:

- **Learning-based style transfer**: Employing learning-based approaches that can be trained on paired examples of preoperative and intraoperative images to learn optimal style transformations.
- **Domain adaptation techniques**: Incorporating recent advances in unsupervised domain adaptation to bridge the gap between preoperative and intraoperative appearances without requiring paired examples.
- **Multi-modal NeRFs**: Developing NeRF variants that can directly represent multiple imaging modalities within a single model, potentially eliminating the need for explicit style transfer.

8.5.3 Clinical Integration and Validation

To move toward clinical adoption, future work should focus on:

- **Prospective clinical trials**: Conducting prospective studies to evaluate the approach's performance in real surgical scenarios, including assessment of clinical workflow integration and surgeon feedback.
- **System optimization**: Optimizing the implementation for deployment in operating room environments, potentially leveraging specialized hardware or cloud computing resources.
- **Multimodal integration**: Extending the approach to incorporate additional intraoperative data sources, such as ultrasound or electrophysiological recordings, for enhanced registration accuracy.

8.5.4 Generalizable Neural Scene Representations

Looking beyond the specific application of intraoperative registration, future research could explore:

- **Transfer learning**: Developing pre-trained NeRF models that can be quickly fine-tuned for specific patients, reducing the computational requirements for preoperative training.
- **Uncertainty quantification**: Incorporating uncertainty estimation into the registration process to provide surgeons with confidence measures for the registration accuracy.
- **Multi-task neural representations**: Extending the neural representation to support multiple tasks beyond registration, such as segmentation, anomaly detection, and surgical planning.

8.6 Summary

This discussion has contextualized our findings within the broader field of intraoperative registration, highlighting the significance of our contributions while acknowledging the limitations and challenges that remain. The superior performance of the combined MI+L2 loss function with Gram matrix-based style transfer represents a significant advancement in NeRF-based registration approaches, offering improved accuracy, robustness, and efficiency compared to both traditional methods and prior neural rendering techniques.

The limitations identified, particularly the rigid transformation assumption and challenges related to clinical integration, provide clear directions for future research. Addressing these limitations through deformable registration, enhanced cross-modal adaptation, and rigorous clinical validation will be crucial for translating these promising research results into clinically viable tools that can improve the precision and safety of neurosurgical procedures.

9 Conclusion

This thesis has explored the enhancement of intraoperative registration with Neural Radiance Fields (NeRFs) through a comprehensive investigation of loss functions and style transfer techniques. By building upon recent advances in neural scene representations and cross-modal registration, we have developed an approach that significantly improves the accuracy, robustness, and efficiency of aligning preoperative MRI data with intraoperative optical images during neurosurgery.

9.1 Summary of Contributions

The main contributions of this thesis can be summarized as follows:

1. **Loss Function Exploration**: We conducted a systematic evaluation of different loss functions for NeRF-based registration, including L2, Normalized Cross-Correlation, Mutual Information, and weighted/masked variants. Our results demonstrated that a combined MI+L2 approach provides superior performance in cross-modal scenarios, significantly outperforming the L2 baseline used in previous work.

2. **Style Transfer Analysis**: We investigated various style transfer techniques for bridging the appearance gap between preoperative and intraoperative images, including Y'UV color space encoding, Histogram of Oriented Gradients, texture features, edge detection, Gram matrices, and deep feature matching. The Gram matrix-based approach emerged as the most effective method, particularly when combined with the MI+L2 loss function.

3. **Optimization Strategy**: We developed and evaluated different optimization strategies for joint pose and style parameter estimation, finding that a multi-stage approach provides the best balance between registration accuracy and computational efficiency.

4. **Implementation Framework**: We implemented a NeRF-implementation agnostic registration framework built on top of nerfstudio, supporting various NeRF architectures and providing a modular interface for different loss functions and style transfer methods.

5. **Comprehensive Evaluation**: We conducted a thorough evaluation using both synthetic and clinical datasets, demonstrating that our approach outperforms both traditional registration methods and prior NeRF-based techniques, achieving a target registration error of 2.06 mm on clinical data.

Together, these contributions advance the state-of-the-art in intraoperative registration, offering a promising approach for improving the precision and safety of image-guided neurosurgery.

9.2 Key Findings

Several key findings emerged from our research:

1. **Combined Loss Functions**: The combination of information-theoretic measures (MI) with direct intensity-based metrics (L2) provides complementary strengths for cross-modal registration, resulting in better performance than either approach alone.

2. **Style Transfer Effectiveness**: Style transfer techniques can significantly bridge the appearance gap between preoperative and intraoperative images, with our best approach improving registration accuracy by 55% compared to the baseline without style transfer.

3. **Synergistic Effects**: The combination of appropriate loss functions and style transfer techniques yields synergistic effects, where the improvement exceeds what would be expected from each component individually.

4. **Clinical Viability**: The computational efficiency and registration accuracy achieved by our approach meet the requirements for clinical use, with registration times of approximately 18 seconds and target registration errors within the clinically acceptable range.

5. **Robustness**: Our approach demonstrates good robustness to various clinical scenarios, including different lighting conditions and partial views, while still showing limitations in handling brain shift and significant occlusions.

These findings provide valuable insights for the design of future registration systems and highlight the potential of neural scene representations for medical image analysis.

9.3 Clinical Impact

The advancements presented in this thesis have several potential clinical impacts:

1. **Improved Surgical Precision**: By enhancing registration accuracy, our approach could improve the precision of neurosurgical navigation, potentially leading to more complete tumor resections while reducing damage to surrounding healthy tissue.

2. **Streamlined Workflow**: The reduced computational requirements and minimal user interaction needed for our registration approach could streamline the surgical workflow, decreasing operating time and cognitive load on the surgical team.

3. **Broader Accessibility**: By leveraging widely available imaging modalities (preoperative MRI and intraoperative optical images) without requiring specialized hardware, our approach could make advanced navigation capabilities more accessible to a broader range of hospitals and surgical centers.

4. **Real-time Updates**: The efficiency of our registration method potentially enables more frequent updates of the registration during surgery, helping to maintain accuracy as the surgical field changes.

While further validation and development are needed before clinical deployment, these potential impacts highlight the clinical relevance of our research.

9.4 Future Outlook

Looking forward, several promising directions for future research emerge:

1. **Deformable Registration**: Extending the current rigid registration approach to account for brain shift and tissue deformation represents a crucial next step. This could involve developing deformable NeRF representations or integrating biomechanical models to constrain the deformation space.
2. **Real-time Performance**: Further optimization of the implementation to achieve real-time performance would enhance clinical utility. This might involve leveraging hardware acceleration, model compression techniques, or more efficient neural rendering methods.
3. **Multimodal Integration**: Incorporating additional intraoperative data sources, such as ultrasound or electrophysiological recordings, could further improve registration accuracy and provide complementary information for surgical guidance.
4. **Prospective Clinical Evaluation**: Conducting prospective clinical trials to evaluate the approach in real surgical settings will be essential for validating its effectiveness and identifying areas for improvement.
5. **Broader Applications**: The methods developed in this thesis could potentially be extended to other medical domains beyond neurosurgery, such as orthopedic surgery, interventional radiology, or radiation therapy, where precise alignment of preoperative and intraoperative data is crucial.

9.5 Closing Remarks

Neural Radiance Fields represent a powerful paradigm for implicit scene representation that has shown remarkable success in computer vision applications. This thesis demonstrates that with appropriate adaptations—specifically, tailored loss functions and style transfer techniques—NeRFs can also excel in the challenging domain of intraoperative registration for neurosurgery.

The integration of information theory, neural rendering, and style transfer principles in this work illustrates the value of interdisciplinary approaches to complex medical image analysis problems. By drawing from multiple domains, we have developed a registration approach that addresses the unique challenges of aligning preoperative and intraoperative data, potentially contributing to safer and more precise neurosurgical procedures.

While challenges remain, particularly in handling tissue deformation and achieving fully real-time performance, the results presented in this thesis provide a solid foundation for future research and development. The continuing advancement of neural scene representations and their application to medical imaging problems holds great promise for improving patient care through enhanced surgical navigation capabilities.

As computational power increases and neural rendering techniques continue to evolve, we anticipate that approaches like the one presented in this thesis will play an increasingly important role in bridging the gap between preoperative planning and intraoperative reality, ultimately contributing to better outcomes for neurosurgical patients.

A Implementation Details

This appendix provides additional technical details about the implementation of the NeRF-based registration approach described in this thesis. The code is available at: `https://github.com/maxfehrentz/style-ngp`.

A.1 Software Implementation

The implementation is built on top of the nerfstudio framework, with custom extensions for registration and style transfer. The main components are:

- **Registration Module**: Implementation of various pose optimization strategies, including sequential, joint, and multi-stage approaches.
- **Loss Functions**: Implementation of different loss functions (L2, NCC, MI, weighted variants) with a modular interface for easy experimentation.
- **Style Transfer**: Implementation of different style encoding methods and the hypernetwork architecture for style adaptation.
- **Evaluation Metrics**: Tools for measuring registration accuracy, convergence behavior, and style transfer quality.

All components are implemented in Python using PyTorch for deep learning functionality.

A.2 Hyperparameter Settings

Table **??** presents the hyperparameter settings used for the experiments in this thesis.

A.3 Detailed Evaluation Metrics

A.3.1 Registration Accuracy Metrics

We calculate registration accuracy using the following metrics:

- **Rotational Error**: The angular difference between the estimated rotation R_{est} and the ground truth rotation R_{gt}:

$$\theta_{\text{error}} = \arccos\left(\frac{\text{trace}(R_{\text{est}}^T R_{\text{gt}}) - 1}{2}\right) \tag{A.1}$$

Table A.1: Hyperparameter settings used for the experiments.

Component	Parameter	Value
NeRF Model	Grid Resolution	2048^3
	Hash Features	16
	Hash Levels	16
Registration Optimization	Learning Rate (Pose)	5×10^{-3}
	Learning Rate (Style)	1×10^{-4}
	Optimizer	Adam
	Max Iterations	500
Loss Functions	MI Bins	32
	NCC Window Size	11
	Combined Loss Weights	$\alpha = 0.7, \beta = 0.3$
Style Hypernetwork	Hidden Units	[128, 256, 128]
	Activation	LeakyReLU

- **Translational Error**: The Euclidean distance between the estimated translation t_{est} and the ground truth translation t_{gt}:

$$t_{\text{error}} = \|t_{\text{est}} - t_{\text{gt}}\|_2 \tag{A.2}$$

- **Target Registration Error (TRE)**: The average Euclidean distance between corresponding points after registration:

$$\text{TRE} = \frac{1}{N} \sum_{i=1}^{N} \|T_{\text{est}}(p_i) - T_{\text{gt}}(p_i)\|_2 \tag{A.3}$$

  where p_i are anatomical landmarks, and T_{est} and T_{gt} are the estimated and ground truth transformation matrices, respectively.

A.3.2 Image Similarity Metrics

We evaluate image similarity using:

- **Structural Similarity Index (SSIM)**: Measures the structural similarity between two images:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{A.4}$$

- **Peak Signal-to-Noise Ratio (PSNR)**: Measures the fidelity of image reconstruction:

$$\mathrm{PSNR}(x, y) = 10 \log_{10} \left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}(x, y)} \right) \tag{A.5}$$

- **Learned Perceptual Image Patch Similarity (LPIPS)**: Uses deep features to measure perceptual similarity between images.

- **Style Score**: A custom metric that combines aspects of color, texture, and structural similarities to evaluate the quality of style transfer.

B Additional Results

B.1 Detailed Loss Function Comparison

Figure **??** presents a more detailed comparison of different loss functions across various experimental conditions.

Figure B.1: Detailed comparison of different loss functions across various experimental conditions, including different initialization errors, image resolutions, and noise levels.

B.2 Style Transfer Visualization

Figure **??** shows visual examples of different style transfer methods applied to the same input image.

Figure B.2: Visualization of different style transfer methods applied to the same input image, showing the original MRI-derived rendering, the target intraoperative image, and the results of various style transfer approaches.

B.3 Clinical Case Studies

Table **??** presents detailed results for five representative clinical cases from our evaluation dataset.

Table B.1: Detailed results for five representative clinical cases, showing registration accuracy, computation time, and convergence behavior for each case.

Metric	Case 1	Case 2	Case 3	Case 4	Case 5
Initial TRE (mm)	15.37	12.85	18.64	10.42	14.23
Final TRE (mm)	1.76	2.18	2.43	1.65	2.31
Iterations	132	157	189	121	163
Time (s)	15.8	18.9	22.6	14.5	19.6
Success	Yes	Yes	Yes	Yes	Yes

C Glossary

NeRF (Neural Radiance Field) An implicit neural representation that maps 3D coordinates and viewing directions to color and volume density, enabling novel view synthesis of complex scenes.

iNeRF A method that inverts Neural Radiance Fields for pose estimation by optimizing camera parameters through backpropagation.

Registration The process of aligning two or more datasets into a common coordinate system.

Intraoperative Occurring during a surgical procedure.

Target Registration Error (TRE) A metric that quantifies the accuracy of registration by measuring the distance between corresponding points after registration.

Mutual Information (MI) An information-theoretic measure that quantifies the mutual dependence between two random variables, used as a similarity metric for registration.

Normalized Cross-Correlation (NCC) A similarity measure that calculates the correlation between two signals normalized by their standard deviations.

Style Transfer The process of applying the visual style of one image to the content of another image.

Hypernetwork A neural network that generates parameters for another neural network.

Gram Matrix A mathematical representation of style in images, calculated as the correlation between feature activations in different channels of a neural network.

List of Figures

List of Tables

Bibliography

[1] N. Navab, C. Bloch, and L. Wang. “Surgical navigation systems and techniques”. In: *Medical Image Computing and Computer-Assisted Intervention* (2015).

[2] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis”. In: *ECCV* (2020).

[3] L. Yen-Chen, P. Florence, J. T. Barron, A. Rodriguez, P. Isola, and T.-Y. Lin. “iNeRF: Inverting Neural Radiance Fields for Pose Estimation”. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2021.

[4] M. Fehrentz, M. F. Azampour, R. Dorent, H. Rasheed, C. Galvin, A. Golby, W. M. Wells, S. Frisken, N. Navab, and N. Haouchine. *Intraoperative Registration by Cross-Modal Inverse Neural Rendering*. 2024. arXiv: 2409.11983 [cs.CV]. URL: https://arxiv.org/abs/2409.11983.

[5] L. Lamport. *LaTeX : A Documentation Preparation System User's Guide and Reference Manual*. Addison-Wesley Professional, 1994.

[6] T. Müller, A. Evans, C. Schied, and A. Keller. “Instant Neural Graphics Primitives with a Multiresolution Hash Encoding”. In: *ACM Transactions on Graphics* 41.4 (2022), pp. 1–15.

[7] W. Wang, W. Xie, Y. Chen, D. Wang, M. Cheng, M. Zhang, and B. Zhou. “HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields”. In: *ACM Transactions on Graphics* 42.1 (2023), pp. 1–13.

[8] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. Hill, and D. J. Hawkes. “Normalized Cross Correlation for Registration”. In: *Medical Image Computing and Computer-Assisted Intervention* (1998), pp. 1033–1034.

[9] J. Pluim, J. Maintz, and M. Viergever. “Mutual Information for Medical Image Processing: A Review”. In: *IEEE Transactions on Medical Imaging* 22.8 (2003), pp. 986–1004.