**MSc Project - Reflective Essay**

|  |  |
| --- | --- |
| **Project Title:** | Automated Layer Selection for Efficient Fine-Tuning of Medical Image Segmentation Models |
| **Student Name:** | Denis Bolonin |
| **Student Number:** | 240029096 |
| **Supervisor Name:** | Kit Mills Bransby |
| **Programme of Study:** | MSc Computer Science |

This project was conceived to address a critical and growing challenge in the field of medical artificial intelligence: the immense computational cost associated with adapting large, pre-trained models for specialized tasks. While foundational models like CSWin-UNet have demonstrated state-of-the-art performance in medical image segmentation, their full fine-tuning demands significant resources, creating a barrier for many clinical and research institutions. The core objective of my dissertation was to design and evaluate an automated framework for parameter-efficient fine-tuning. The proposed solution sought to move beyond manual or heuristic-based layer freezing by intelligently identifying which parts of a model to update.

**The Strengths and Weaknesses of the Project**

Reflecting on this project, I believe its primary strength lies in addressing a genuinely practical problem with clinical relevance. However, the path to the final experimental design was more iterative than initially anticipated, and this evolutionary process itself became a valuable learning experience.

My initial approach was methodical and conservative. I began with CIFAR-10-C experiments using ResNet18 to validate my understanding of Surgical Fine-Tuning and TPGM in a controlled, well-understood environment. This proved essential for building confidence in the implementation before tackling the complexity of medical imaging. Once satisfied with the basic functionality, I transitioned to CSWin-UNet, initially following a similar paradigm by creating a blurred version of Synapse dataset to simulate domain shift - essentially applying the same distribution shift concept but within medical imaging.

The project's scope evolved as I gained deeper insights into the problem domain. After establishing that the methods worked on blurred Synapse data, I sought additional tasks and incorporated KiTS23 for kidney tumor segmentation. It was only later in the project that I realized the fascinating challenge of sequential learning across three distinct but related tasks, leading to the addition of LiTS17 for liver segmentation. This progression from single-task domain adaptation to complex multi-task sequential learning represents one of the project's key strengths - the willingness to adapt and pursue more ambitious goals as understanding deepened.

The systematic evaluation of three distinct approaches (Surgical Fine-Tuning, TPGM, and their hybrid combination) provided valuable insights into the trade-offs between knowledge preservation and task adaptation. I was particularly pleased with how effectively both TPGM and Surgical Fine-Tuning prevented catastrophic forgetting, completely recovering performance for anatomical structures that suffered total knowledge loss under standard fine-tuning. The computational efficiency gains - requiring only 1 hour per fine-tuning step compared to 8 hours for individual pre-training - demonstrate clear practical value.

However, the project also revealed several limitations that I must acknowledge. The incremental development approach, while methodologically sound, meant that some early design decisions may have influenced later experimental choices in suboptimal ways. More significantly, the hybrid approach that I hypothesized would combine the best of both methods failed to consistently outperform either individual technique. This was a reminder that in machine learning, combining successful methods doesn't always yield additive benefits - sometimes competing regularization strategies can constrain model flexibility.

The evaluation was also limited to sequential rather than truly continual learning scenarios, and I lacked access to real clinical environments for validation. While the progression from simple experiments to complex multi-task scenarios demonstrates good experimental methodology, the scope remained constrained to abdominal CT imaging, limiting generalizability.

**Future Work**

Looking forward, several research directions emerge from this work. The most immediate extension would be continuous learning scenarios where new data arrives incrementally, more closely mimicking clinical deployment. The evolution of my experimental design - from single domain shifts to multi-task sequential learning - suggests that even more complex scenarios involving mixed anatomical regions or imaging modalities could provide valuable insights.

I would also like to explore adaptive method selection - automatically choosing between different fine-tuning approaches based on task characteristics and available computational resources. The distinct performance patterns I observed between TPGM and Surgical Fine-Tuning suggest that different clinical scenarios might benefit from different approaches.

Cross-modal adaptation represents another compelling direction. Extending the framework beyond CT imaging to MRI, ultrasound, and X-ray would significantly broaden its clinical applicability. Additionally, incorporating uncertainty-guided adaptation, where confidence measures inform parameter update decisions, could improve both efficiency and reliability.

Perhaps most importantly, this work needs clinical validation studies to assess real-world performance and quantify practical computational savings. Collaborating with medical institutions to deploy and evaluate these methods in clinical workflows would provide crucial insights into their practical viability.

**Critical Analysis of the Relationship Between Theory and Practical Work**

This project highlighted both the power and limitations of translating theoretical advances into practical applications. The initial CIFAR-10-C experiments served as a crucial bridge between theory and medical applications, allowing me to understand the fundamental behaviors of Surgical Fine-Tuning and TPGM in a controlled environment before tackling domain-specific challenges.

The theoretical foundations - Surgical Fine-Tuning's feature space decomposition and TPGM's learnable regularization - provided solid starting points. However, their adaptation to medical imaging required careful consideration of domain-specific challenges like severe class imbalance and high inter-patient variability. The progression from simple domain shift (blurred images) to complex sequential learning scenarios revealed how theoretical concepts must be adapted and scaled for real-world complexity.

The theoretical promise of combining these approaches into a hybrid method proved more complex in practice. While both methods individually succeeded in their design goals, their integration introduced competing constraints that sometimes hindered performance. This experience reinforced the importance of empirical validation and the recognition that theoretical elegance doesn't always translate to practical superiority.

The evolution of my experimental design - from validating basic concepts to exploring complex multi-task scenarios - demonstrated how practical work can inform and extend theoretical understanding. The discovery that three-task sequential learning presented unique challenges led to insights that wouldn't have emerged from simpler experimental setups.

**Awareness of Legal, Social, Ethical Issues, and Sustainability**

This project operates within a complex landscape of ethical and regulatory considerations. Patient data privacy was paramount throughout the research, requiring careful attention to data anonymization and secure processing protocols. The use of publicly available datasets mitigated some privacy concerns but highlighted the ongoing challenge of accessing diverse, representative medical data for AI research.

The potential for algorithmic bias represents a significant concern. Medical AI systems must perform equitably across different patient populations, yet training data often exhibits demographic imbalances. While this project focused on technical efficiency rather than fairness explicitly, future work must address these bias considerations systematically.

From a sustainability perspective, the project's core motivation - reducing computational requirements - aligns with growing concerns about the environmental impact of large-scale AI training. By demonstrating that effective model adaptation requires significantly fewer computational resources than training from scratch, this work contributes to more sustainable AI practices in healthcare.

The broader social implications concern healthcare accessibility. If efficient fine-tuning methods can make advanced medical AI accessible to institutions with limited computational resources, they could help democratize high-quality medical imaging analysis. However, this potential benefit must be balanced against ensuring that cost-effective solutions don't compromise diagnostic accuracy or patient safety.

**Conclusion**

This project provided valuable insights into the practical challenges of deploying efficient fine-tuning methods for medical image segmentation. The iterative development process - from simple CIFAR-10-C validation to complex multi-task medical scenarios - proved as educational as the final results themselves.

While the technical results demonstrate clear advantages over standard approaches, the experience reinforced the complexity of translating research advances into clinical practice. The most significant learning was the importance of understanding trade-offs rather than seeking universal solutions. TPGM's balanced performance across tasks may be more valuable than methods that excel in specific scenarios but fail elsewhere.

The project's evolution from initial concept to final implementation highlighted the value of iterative development and the willingness to pursue more ambitious goals as understanding deepens. While the hybrid approach didn't achieve the hoped-for improvements, the systematic exploration of different combinations provided valuable insights into the limitations of current methods.

Moving forward, I believe the most impactful research will emerge from sustained engagement with real clinical workflows, where the messy complexity of medical practice can inform and validate technical advances. This project represents a step in that direction, demonstrating how careful experimental progression from simple to complex scenarios can reveal both opportunities and limitations in automated fine-tuning approaches.