Main Independent Study Specifications:

Proptosis Medical Screening Computer Vision

Enrollment Request: CS 596 (3 credits for Aryan Sajith)

Semester: Spring 2025

Advisor: Professor Erik Learned-Miller

Descriptive Title:

Leveraging Self-Supervised Learning for Early Medical Screening: A Pilot Study in Proptosis Detection

Statement of Objectives:

Ethical Considerations & Data Usage:

At the outset of this study, we emphasize that no real human data will be used in the initial phase. The first stage will rely solely on synthetic, publicly available, or simulated datasets that do not involve human subjects. Only in the second phase, when applying the methodology to proptosis detection, will real human data be considered. At that stage, we will ensure all necessary ethical approvals, privacy safeguards, and informed consent procedures are in place before proceeding with any data collection or analysis.

Primary Objectives:

This study aims to explore the viability of self-supervised learning techniques as a foundation for enhancing medical facial screening, with a specific focus on the early detection of proptosis—a potential indicator of various systemic diseases like Thyroid Eye Disease, Tumors, and so on. The project will proceed in two main phases:

1. **Proof-of-Concept via a Toy Problem:**

Initially, we will design and execute experiments on a controlled, non-human toy dataset (e.g. distinguishing between various animals, reptiles, fish, etc.) to validate the self-supervised approach. By training on a large, diverse set of unlabeled data and subsequently fine-tuning on smaller, specialized subsets, we intend to:

- Assess the effectiveness of self-supervised representations in scenarios where annotated data is limited.
- Quantify improvements in classification performance after domain-specific fine-tuning.
- Establish robust evaluation metrics that can be transferred to more complex, real-world problems.

2. **Application to Proptosis Screening:**

Building upon the insights gained from the toy problem, the second phase will apply these methodologies to the medical domain by targeting proptosis detection. Objectives include:

- Developing an algorithm capable of screening for subtle facial changes indicative of proptosis.
- Investigating the transferability of learned features from the toy dataset to clinical imagery.
- Exploring how early detection through automated facial analysis can serve as a non-invasive, preliminary screening tool to prompt further medical evaluation.

Through this study, we aim not only to advance the state of self-supervised learning in low-data regimes but also to contribute a novel approach to medical screening that bridges modern computer vision techniques with practical clinical applications. Ultimately, the findings are hoped to inform future innovations in diagnostic tool development and early disease intervention strategies.

Planned Activities:

Throughout the Semester

- **Collaborative Infrastructure:**
  - Set up a shared GitHub repository for code, data, and documentation.
  - Schedule bi-weekly team meetings (via Zoom or in person) to discuss progress, challenges, and next steps.
  - Maintain shared documentation (using tools like Google Docs or Notion) to track decisions, literature reviews, and experiment logs.
- **Role Assignments:**
  - Distribute responsibilities among team members (e.g., literature review, data management, model implementation, experiment coordination, and documentation) to ensure balanced contributions and cross-checking.

Weeks 1–2: Project Kickoff & Literature Review

- **Kickoff Meeting & Planning:**
  - Introduce team members and establish clear roles and responsibilities.
  - Define project objectives, deliverables, and communication protocols.
- **Literature & Background Review:**
  - Study key works on self-supervised learning techniques, with emphasis on methods such as contrastive learning and representation learning.
  - Review foundational material on face recognition and clinical implications of proptosis.
  - Compile summaries and annotations of relevant papers (e.g., recent self-supervised learning advances and clinical studies on proptosis).
- **Infrastructure Setup:**
  - Create and organize the GitHub repository with initial documentation outlining project scope and timeline.
  - Set up project management tools (e.g., Trello, Asana) to track tasks and milestones.

Weeks 3–4: Phase 1 – Toy Problem: Data Preparation and Baseline Model Implementation

- **Data Selection & Preprocessing:**
  - Identify a suitable toy dataset (e.g., images categorizing different animal species).
  - Assign team members to collect, clean, and perform exploratory analysis of the dataset.
- **Baseline Implementation:**
  - Develop a self-supervised learning model (e.g., using contrastive learning or a simple autoencoder) on the toy dataset.

    - Implement a supervised baseline model for direct comparison (e.g., a simple CNN or MLP).
- **Team Collaboration:**
    - Organize pair programming sessions or code reviews to ensure all team members understand the implementations.
    - Document initial design choices and share code updates via GitHub.

Weeks 5–6: Phase 1 – Experimentation and Cross-Comparative Testing

- **Model Training & Evaluation:**
    - Train both the self-supervised model and the supervised baseline on the toy dataset.
    - Perform cross-comparative tests to evaluate:
    - Feature representation quality
    - Performance metrics (accuracy, precision, recall)
    - Robustness in low-data regimes.
- **Analysis Meetings:**
    - Convene team discussions to interpret experimental results and refine experimental protocols as needed.
    - Record observations and lessons learned in the project log.
- **Documentation:**
    - Update the GitHub repository with experiment scripts, logs, and preliminary analysis reports.

Weeks 7–8: Phase 1 Synthesis & Transition Planning

- **Results Synthesis:**
    - Consolidate findings from the toy problem experiments.
    - Prepare an interim report that summarizes the effectiveness of self-supervised learning in a controlled setting.
- **Phase 2 Planning:**
    - Develop a strategy for applying the refined self-supervised approach to proptosis detection.
    - Outline requirements for the clinical data (e.g., ethical clearances, data anonymization) and establish communication with clinical partners.
- **Task Redistribution:**
    - Reassign or adjust roles for Phase 2 tasks, ensuring smooth transition and continuity.
- **Interim Presentation:**
    - Hold an internal review meeting to present findings from Phase 1 and the proposed plan for Phase 2.

Weeks 9–11: Phase 2 – Proptosis Screening: Data Acquisition, Fine-Tuning, and Implementation

- **Clinical Data Acquisition & Preprocessing:**
  - Submit Institutional Review Board proposal for appropriate permissions(could submit several weeks earlier for permissions to arrive by this time)
  - Work with clinical collaborators to obtain an anonymized dataset of facial images suitable for proptosis screening.
  - Preprocess the images (e.g., normalization, alignment, and augmentation) and perform exploratory analysis to understand data characteristics.
- **Model Fine-Tuning:**
  - Leverage the self-supervised model from Phase 1 and fine-tune it on the proptosis dataset.
  - Experiment with transfer learning techniques to adapt representations learned from the toy problem to clinical data.
- **Experimental Evaluation:**
  - Set up experiments comparing the fine-tuned model to existing screening benchmarks (if available) or a baseline supervised model.
  - Evaluate performance metrics such as sensitivity (to capture early signs), specificity, and overall accuracy.
- **Team Coordination:**
  - Ensure regular check-ins to share progress, address challenges (e.g., low-data issues or overfitting), and update the shared repository with new findings.

Week 12: Phase 2 – Comprehensive Comparative Analysis & Validation

- **Deep Dive Analysis:**
  - Compare the performance of the fine-tuned self-supervised model against traditional approaches and the baseline model from Phase 1.
  - Analyze trade-offs such as computational efficiency, data requirements, and diagnostic accuracy.
- **Discussion & Troubleshooting:**
  - Hold a dedicated session to discuss any discrepancies, challenges, or unexpected results.
  - Adjust model parameters or data processing methods based on the analysis.
- **Preparation for Final Synthesis:**
  - Begin drafting sections of the final research report that integrate both phases' findings.

Week 13: Final Synthesis, Report Compilation, & Presentation

- **Final Report & Documentation:**
  - Collaboratively write a comprehensive research paper that:
  - Reviews the methodology and experiments from both phases
  - Synthesizes insights from the toy problem and proptosis screening applications
  - Discusses limitations, trade-offs, and future directions in applying self-supervised learning for medical screening.
- **Team Presentation:**

    - Prepare a final presentation summarizing key findings, methodologies, and practical implications of the research.
    - Conduct a rehearsal presentation session to ensure clarity and cohesion.
- **Final Repository & Deliverables:**
    - Finalize the GitHub repository with complete code, documentation, and experiment logs.
    - Submit the final report and prepare materials for any subsequent publication or funding proposals.

Evaluation Criteria:

In-Depth Paper:

- A well-structured final report documenting both phases of the study (toy problem and proptosis screening) with comprehensive methodology, experimental details, and comparative analyses.
- Inclusion of challenges encountered, key insights gained, and a discussion of potential future directions for self-supervised learning in medical screening.

GitHub Repository:

- A well-organized repository containing all code, datasets (or links), and experiment logs.
- A detailed README outlining project objectives, setup instructions, implementation details for both phases, and step-by-step guidelines for reproducing the experiments.

Meetings:

- Bi-weekly meetings to discuss progress, share updates, and address challenges, with documented minutes or meeting logs demonstrating effective collaboration among the team members.