SUNY Potsdam
Capstone Senior Project Proposal
Author: Brendon Vineyard
Date: January 24, 2025
Course: CIS 405 - Senior Project

Introduction

Proposed Project: Multi-Modal Sentiment and Emotion Classification Using Computer Vision and Natural Language Processing

What Will Be Produced?

My goal with this project is to build a machine learning model that can look at both images and text together to classify sentiment and emotions. This kind of analysis could be useful for understanding how people feel in posts where they express themselves in a mix of photos and captions, like on social media. Most existing models only analyze text or images separately, which often misses out on the full picture when people combine both forms to express their emotions. By working with both image and text data, I hope to develop a model that gives a more complete, accurate picture of sentiment and emotion.

To do this, I'll train a Convolutional Neural Network (CNN) to pick up on emotional cues from images and a Natural Language Processing (NLP) model (probably something like BERT) to analyze the sentiment in text. The two outputs will then be combined through a "fusion" layer, which is just a way to join their outputs. I'm planning to experiment with methods like simple concatenation or a cross-attention layer for this fusion. The end result will be a model that could be applied in areas like content analysis on social media, mental health tracking, or even market research where understanding people's emotions in a well-rounded way is valuable.

Preparation

This project builds off a lot of what I've learned so far in my major, especially from courses and projects focused on machine learning, neural networks, and deep learning:

- **Relevant Coursework**: In CIS 421 (Artificial Intelligence), I've had hands-on experience with a bunch of foundational AI concepts and models. I've also had projects on genetic algorithms, feedforward networks, and CNNs, which has taught me about everything from model training and tuning to evaluating performance. This background is solid prep for the CNN part of this project.

- **Project Experience**: Recently, I did a project involving a multi-phase CNN model for breast cancer detection. This was my first serious experience with CNNs, and it taught me about dealing with complex datasets, choosing network architectures, and improving model accuracy. Working on that project made me realize the importance of data prep and parameter tuning, which will be super useful when working with image data for emotion analysis in this proposal.

- **Tools and Frameworks:** I've used TensorFlow, Keras, and Google Colab in various projects, and I plan to use these again for this model. I'm already comfortable setting up and training models in these platforms, and Colab's GPU support has been helpful for more compute-heavy training. TensorFlow's flexibility will be useful since I'll be experimenting with different architectures and fusion techniques.

- **Learning Fusion Techniques**: While I haven't directly worked with fusion techniques for multi-modal models before, I've done some research on them. I understand the basics of how CNN and NLP outputs can be combined, and I'm ready to explore more methods like cross-attention, which lets the model learn how the text and image features interact. This project will be my first go at actually implementing these ideas.

Practice

This project is a chance for me to bring together a lot of what I've been learning in AI, as well as push into some newer territory:

1. **Multi-Modal Data Processing**: This project will be my first time working with multi-modal data, which just means combining both image and text analysis in a single system. Multi-modal AI is becoming a big deal as more applications try to interpret different types of data together, and getting this experience now will give me valuable insight into the challenges and benefits of combining data sources.
2. **Experimenting with Fusion Techniques**: I'll get to explore a few different ways to combine the CNN and NLP outputs, from straightforward concatenation to using cross-attention layers. These are new areas for me, so it'll be a good chance to experiment and see which approach is more effective in getting accurate sentiment predictions.
3. **Open-Ended Project Design**: Unlike a typical class assignment, this project is much more open-ended, meaning I'll have to make a lot of decisions as I go. This will build my project management and problem-solving skills, as I'll be balancing experimentation with making progress and deciding on next steps based on ongoing evaluations.
4. **Real-World Applications**: What excites me most about this project is that it could actually be used in the real world. From social media analysis to mental health tools,

there are a ton of areas where a well-rounded emotional analysis model like this could add value. This could also be a nice portfolio piece that shows I can handle a mix of AI techniques and apply them in creative ways.

Deliverables

Here's what I plan to deliver by the end of this project:

- **Trained Multi-Modal Model**: The final deliverable will be a fully trained AI model that can analyze pairs of images and text to classify sentiment and emotions. I'll test it across a few different datasets to ensure it can handle different types of data and still produce accurate classifications. The fusion technique I use to combine CNN and NLP outputs should ideally improve the model's accuracy over separate single-mode models.

- **Project Documentation**: I'll write up a detailed documentation file that explains the model architecture, data processing steps, training configurations, and any key decisions made during development. This will be useful for anyone who wants to understand or replicate my work.

- **Performance Analysis Report**: Finally, I'll create a report on how well the model performed, including metrics like accuracy, precision, and confusion matrices. This will help illustrate any strengths or weaknesses in the model and show how effective the fusion technique was. I might include some examples of test cases to make the analysis clearer.

Administration
Expected Credit Hours: 3
Expected Enrollment Semester: Spring 2025

Appendices

Appendix A: Grading Rubric

Grade	Description
4.0	Fully functioning model with strong performance; comprehensive documentation; evaluation report comparing fusion techniques, and a final presentation with example outputs.
3.5	Model works with moderate accuracy; includes documentation, a detailed evaluation report, and

	a presentation.
3.0	Model functions but may lack accuracy; includes basic documentation and performance report.
2.5	Model with limited functionality or incomplete documentation. Presentation has limited results.
2.0	Minimal functionality and documentation; incomplete presentation with limited or no performance analysis.

Appendix B: Tentative Schedule

Week	Task
1	Define project goals and gather datasets; sketch out the model structure.
2	Prepare and preprocess the image and text data.
3-4	Develop and train the CNN model for image-based emotion analysis.
5-6	Develop and train the NLP model for text-based sentiment analysis.
7	Choose fusion techniques and begin combining CNN and NLP outputs.
8-9	Train and fine-tune the multi-modal model; log performance metrics.
10	Conduct evaluations, comparing fusion methods.
11	Complete documentation and final testing.
12	Final presentation prep, including demonstration.