CRCV REU Week #2

David Rivard

06/06/2025

Overview

- Introduction
- Assignments
- Project Overview
- Research Papers
- Next Steps
- References



Introduction

University: Lake Superior State University

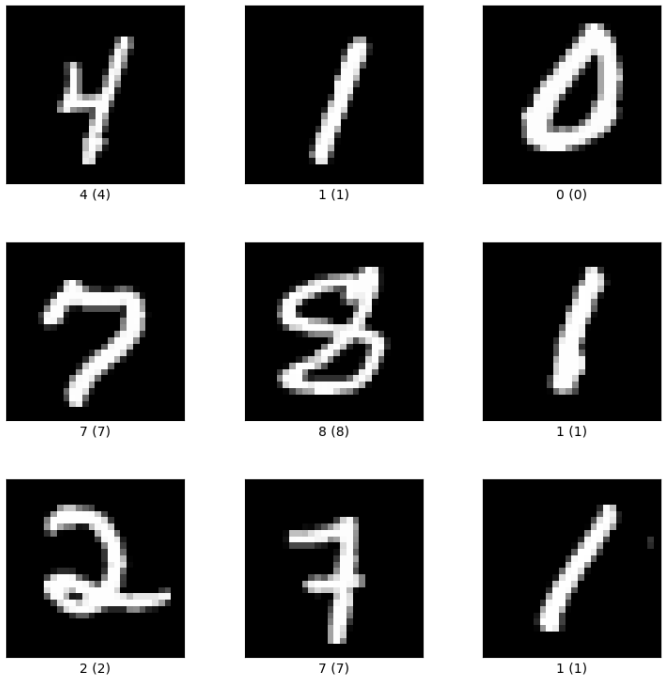
Major: Computer Science

Grade: Senior

Hometown: Escanaba, Michigan

Courses: Data Structures in Java, Python, Programming I-III, Advanced Data Structures and Algorithms, Discrete Programming and Data Structures, Discrete Proofs, Senior Design

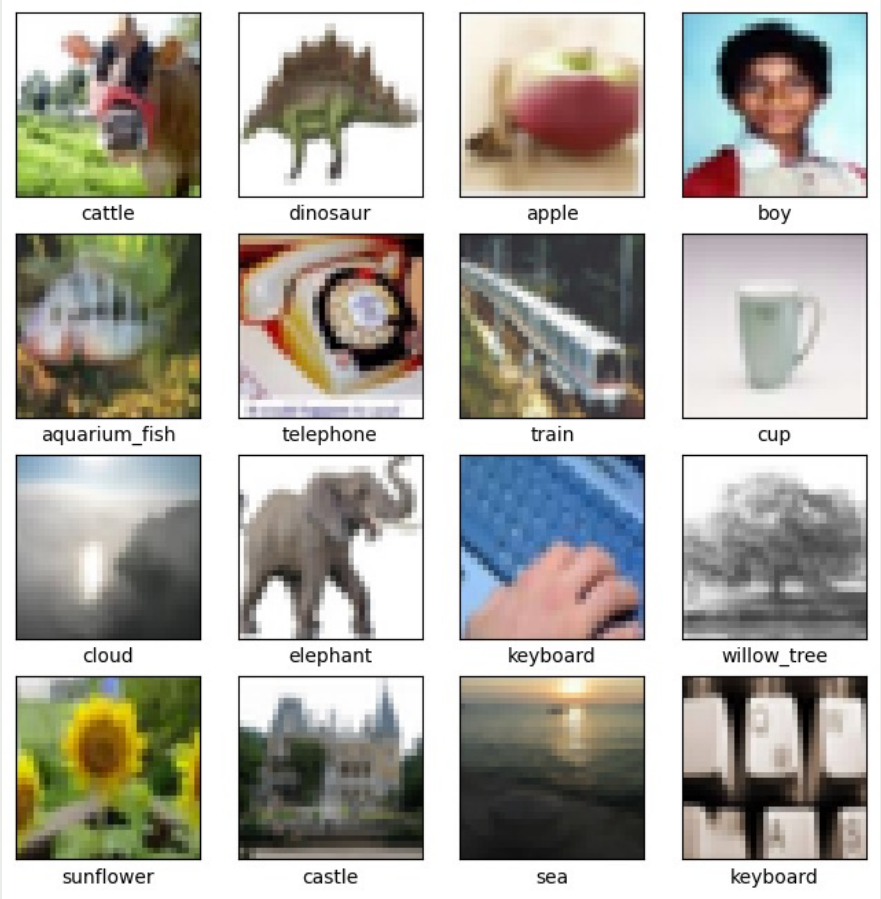
Assignment 0



- **Dataset**: MNIST
  - **10** classes
  - Handwritten digits
  - Grayscale 28x28 images
- **Tool**: Keras
- **Model**:
  - Simple CNN with two convolutional layers, dropout, and softmax.
- **Results**:
  - 99% test accuracy
  - 98% train accuracy

- **Dataset**: CIFAR-100
  - **100** classes
  - Small images (**32x32**)
  - Low resolution and high noise
  - Similar classes with varying poses
- **Model**: ResNet-based CNN
- **Experiments**: 45 configs with different:
  - **Optimizers**: Adam, SGD, RMSProp
  - **Loss**: CrossEntropy and LabelSmoothing
  - **Learning Rates**: 0.01, 0.001, 0.0005
- **Best Accuracy**: ~67.9%
  - (with Adam + Label Smoothing)

Assignment 1



Project Overview: GEO-HiPO

- **Goal**: Predict place names from images using a multi-modal LLM.
- **Model**: Qwen2.5
- **Key Ideas**:
  - Preference-based training using great-circle distance.
  - Coarse-to-fine curriculum (continent down to city).
  - Natural language output (like “Milan, Italy”).
  - Visual cues only, no labels.
  - **Mentors**: Parth Parag Kulkarni, Rohit Gupta, and Dr. Shah

Project Direction

- Met with Parth Kulkarni and Rohit Gupta.
- **Discussed**:
  - Expectations
  - More research papers.
  - Model we'll be working with.
  - Datasets that we'll be using.
  - Preference optimization vs. classification.

Project Understanding

- Uses Vision-Language Models (VLMs) with Preference Optimization.
- **Architecture**: VLM fine-tuned on ranked image pairs.
- **Loss**: Learn to prefer images that have more geographical information.
- **Supervision**: Learn based on geographic distance not class labels.
- **Focus**: Generalize across the world with different lighting and seasons.

Project Progress

- Read the provided research papers over.
- Parsed IM2GPS and IM2GPS3k datasets.
  - Sample sets are ready.
- Cloned and setup fine-tuning repo.
- Started training locally with my **RTX 3080**.
- Ran into a plethora of issues.

Research Papers

- **“GeoReasoner: Geo-Localization with Reasoning in Street Views.”**
  - Uses visual cures and spatial reasoning for street-level geolocation.
- **“GaGA: Towards an Interactive Global Geolocation Assistant.”**
  - Improves geolocation assistant that refines predictions with multi-step reasoning and user input.
- **“Pigeon: Predicting image geolocations.”**
  - Combines geocell classes, contrastive pretraining, and cluster refinement to perform global image geolocation. Better than top GeoGuessr player.

Research Papers Continued

- **“Direct Preference Optimization.”**
  - Introduces loss function that directly optimizes model outputs based on ranked preferences, removing the need of reward models.
- **“SimPO: Simple Reference-Free Preference Optimization.”**
  - An approach to preference learning that doesn’t need reference scores and instead uses binary preferences to train models.
- **“Self-Paced Curriculum Learning by Dynamic Diffi-culty.”**
  - Improves learning by dynamically adjusting training difficulty, gradually introducing harder samples, modeling human learning to help generalization.

Next Steps

- Get environment set up on cluster.
- Confirm small-scale fine-tuning job on cluster.
- Start analyzing performance and error.
- **Start experimenting**:
  - With data augmentation.
  - With different image captions.
  - With deeper models: Qwen2.5-32B, Qwen2.5-72B.

References

- **Mentors**: Parth Parag Kulkarni, Rohit Gupta, and Dr. Shah
- PyTorch Classification Tutorial: https://docs.pytorch.org/tutorials/beginner/blitz/cifar10_tutorial.html
- ResNets: https://www.digitalocean.com/community/tutorials/writing-resnet-from-scratch-in-pytorch
- Data Augmentation: https://docs.pytorch.org/vision/stable/transforms.html
- Research on Residual Learning: https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf

References Continued

- R. Rafailov, J. Liu, T. Li, et al. “Direct Preference Optimization.” NeurIPS 2023.
- J. Li, M. T. Nguyen, Z. Chen, et al. “GeoReasoner: Geo-Localization with Reasoning in Street Views.” ICML, 2024.
- A. Coker, A. Geli, X. Li, et al. “GaGA: Towards an Interactive Global Geolocation Assistant.” arXiv:2412.08907, 2024.
- D. Simianer and F. F´evotte. “SimPO: Simple Reference-Free Preference Optimization.” 2024.
- J. Turrin, T. Uva, V. Strisciuglio, et al. “Self-Paced Curriculum Learning by Dynamic Difficulty.” ICML, 2019.
- Haas, L., Skreta, M., Alberti, S., & Finn, C. (2024). Pigeon: Predicting image geolocations - CVF open access.

THANK YOU!

Any Questions?

David Rivard

06/06/2025

Lake Superior State University