CS589: Project Details

Overview

- The course project is an opportunity for you to apply what you have learned in class to a problem of your interest.
- Your are encouraged to select a topic and work on your own project. Potential projects usually fall into these two tracks:
  - **Applications.** If you're coming to the class with a specific background and interests (e.g. biology, engineering, physics), we'd love to see you apply deep neural networks to problems related to your particular domain of interest. Pick a real-world problem and apply deep neural networks to solve it.
  - **Models.** You can build a new model (algorithm) with deep neural networks, or a new variant of existing models, and apply it to tackle vision tasks. This track might be more challenging, and sometimes leads to a piece of publishable work.

Timeline

1. **Parts**
    a. **Proposal: 7/26/2025**
    b. **Milestone: 8/9/2025**
    c. **Final write-up: 8/15/2015**
2. **Can’t use extension “late days” for these deadlines!!!**

Projects as a mini-conference

1. You will write a paper with your team.
    a. A suggested format will make sure you cover the right kinds of topics.
2. TA and I will grade all the final write-ups at the same time as the reviews. I will not use the review scores directly

Project Proposal

The project proposal should be concise (200-400 words). Your proposal should contain:

- Who are the (1~2) group members? What will each person do? (This needs to be a separate detailed paragraph)
- If the project is shared with another class, which portion of the work will be counted for each class? (This needs to be a separate detailed paragraph)
- What is the problem that you will be investigating? Why is it interesting?
- What data will you use? If you are collecting new datasets, how do you plan to collect them?
- What method or algorithm are you proposing? If there are existing implementations, will you use them and how? How do you plan to improve or modify such implementations?
- What reading will you examine to provide context and background?
- How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g. plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics or statistical tests)?

Submission: Please upload a PDF file named `<your ID>_proposal.pdf` to Gradescope. One submission for each group is sufficient.

Project Milestone

Your project milestone report should be between 2 - 3 pages using the provided template. The following is a suggested structure for your report:

- Title, Author(s)
- Introduction: this section introduces your problem, and the overall plan for approaching your problem
- Problem statement: Describe your problem precisely specifying the dataset to be used, expected results and evaluation
- Technical Approach: Describe the methods you intend to apply to solve the given problem
- Intermediate/Preliminary Results: State and evaluate your results upto the milestone

Submission: Please upload a PDF file to Gradescope. Please coordinate with your teammate and **submit only under ONE of your accounts**, and add your teammate on Gradescope.

Final Submission

Your final write-up should be between **6 - 8** pages using the provided template. After the class, we will post all the final reports online so that you can read about each others' work. If you do not want your writeup to be posted online, then please let us know at least a week in advance of the final writeup submission deadline.

Final Presentation

All project teams will present their work at the end of the semester. We will have one or two poster sessions. Each team should prepare a poster (similar in style to a conference poster) and present it during the allocated time. If applicable, live demonstrations of your software are highly encouraged.

Project Ideas

Friend or Foe? Analysis of training on out-of-distribution data on fixed test set

Motivation: Given a fixed training and test set (Example: CUB dataset), how can you select new training samples such that model performance on test set improves?

Relevant Areas: classification, Active learning, Self-Supervised Learning.

Good?

Bad?

Is charizard a dragon? zero-shot categorization/attribute discovery of art work

Motivation: Given an artwork, can we understand what's in the artwork without training any model on the artwork itself? Does the discovered attribute align with human perception?

Relevant Areas: Vision-Language models, Attribute discovery, Human evaluation.

- Dragon
- Lizard
- Orange
- Fire

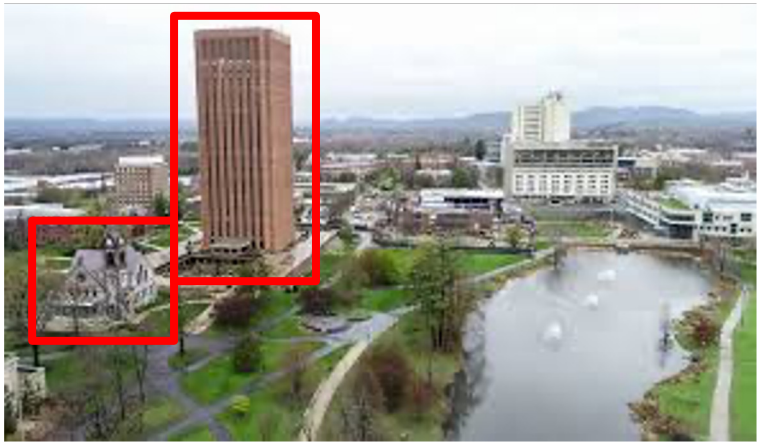
- Old man
- Heart
- Bull
- Eyes

What makes Amherst look like Amherst?

Motivation: Given multiple images of different places of Amherst, can we find what are the characteristic visual elements that make Amherst unique?

Relevant Areas: Clustering, Coreset selection.

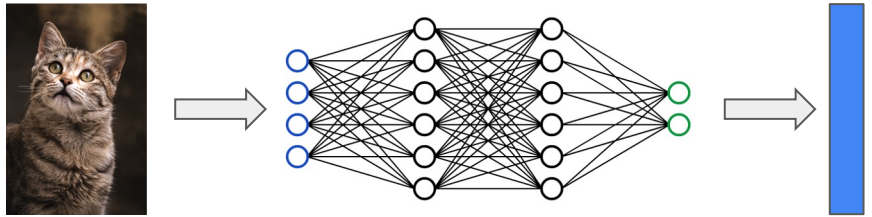
Reference: http://graphics.cs.cmu.edu/projects/whatMakesParis/paris_sigg_reduced.pdf

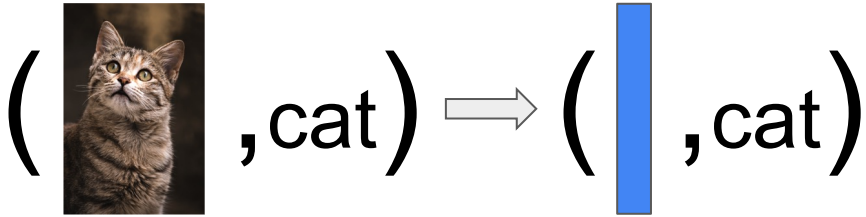


Representation learning

Train a model that takes an image and produces a representation vector that is useful for downstream tasks.

- Use self-supervised learning to learn useful representations of data.
- How useful are your representations for classification?
- Can you do well with other downstream tasks?





https://openreview.net/forum?id=xm6YD62D1Ub

Audio classification with CNNs

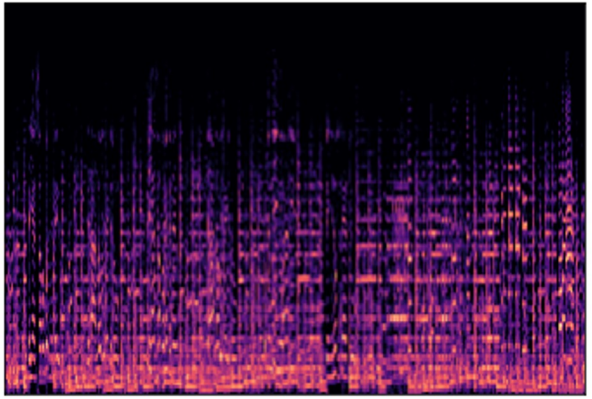
What's the best way to classify music genre?

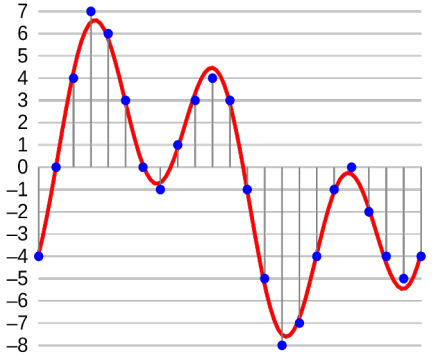
A very common approach to machine learning with sound converts the audio sample into a spectrogram image which is then passed to a CNN.

Can you come up with a better approach? Are there modifications to the kernel size, conv layer, activation function, etc. that would be better for a spectrogram image compared to a natural image? Or maybe you can think of an alternative architecture to CNNs that could be used with the raw waveform?

The GTZAN dataset provides song samples and spectrogram images with corresponding genre labels. You can use this data to classify which music genre a corresponding audio sample represents. Alternatively, find your own audio dataset!

Label: “Blues”





Mini-Spotify: neural network based music retrieval

Motivation: Can you build a simple recommender system for suggesting similar songs given a query song? What information would you require to find similar songs?

Relevant Areas: Metric learning, Information retrieval.

Datasets: https://gist.github.com/alexanderlerch/e3516bffc08ea77b429c419051ab793a

Object localization based on textual description

Applying Visual Grounding in Real-world Scenarios

Visual grounding is the process of linking textual descriptions to specific regions in an image.

For example, find if a certain item is present in a picture of a grocery store aisle.

Image Input

Chat question answering

Q. Is there soda here?
Model Ans. No

Q. Where is the orange juice?
Model Ans.

Emotion-based Style Transfer

Adjust the artistic style transfer based on detected emotions in the content image.

Use a melancholic painting style for sad expressions and lively, vibrant styles for joyful ones.

In the examples to the right, the sad face painting uses dull colours whereas the happy one uses bright colors.

Text Augmentation using LLMs

Leverage LLMs to boost performance in text classification by generating augmented text.

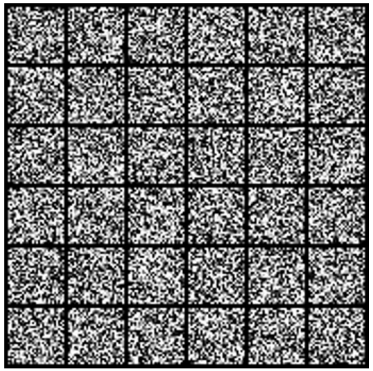
Primarily target under-represented categories to create a balanced dataset.

Generate augmented texts for the following sentences:
"I absolutely loved this movie! Best I've seen in a while."
"The product broke after just one week of use."
"It's okay, but I've had better meals at this price point."

1. "I absolutely loved this movie! Best I've seen in a while."
   - "This movie was fantastic! Haven't seen anything this good in ages."
   - "Truly adored this film! It's the best I've watched in recent times."
   - "I was totally impressed by this movie! Tops anything I've viewed in a long time."
2. "The product broke after just one week of use."
   - "The item failed in only a week of usage."
   - "After using it for a week, the product malfunctioned."
   - "Within a week of use, the product stopped working."
3. "It's okay, but I've had better meals at this price point."
   - "It's decent, but I've tasted better for the same price."
   - "It's fine, but I've experienced better dishes at this cost."
   - "The meal was alright, but I've had more impressive ones for the price."

Mini diffusion model for MNIST

- Build a baseline diffusion model for MNIST. Then try to reduce the model's size to save computation while maintaining reasonable performance.
- Various lightweight model design techniques can be considered including depthwise conv, quantization and pruning.



TA: Junyan

The evolution of zero-shot Cats vs Dogs Classification

- Design and evaluate three types of **zero-shot** cats-vs-dogs classification models: kNN, CNN, and CLIP (multimodal LLM such as BLIP-2 is a bonus)
- Zero-shot means that the model is not trained using cats-vs-dogs data. For CNN classification models, since most of them are trained on ImageNet and ImageNet contains dogs and cats, you are not allowed to use the pre-trained classification head. Instead, use the feature from the pre-trained backbone to do zero-shot classification.

Reference:
https://www.kaggle.com/competitions/dogs-vs-cats/data