

MAIS 202 - Project Deliverable 1

1. Choice of dataset

The chosen dataset is the "Tagged Anime Illustrations" dataset from Kaggle. This dataset contains a rich collection of tagged and labeled artwork depicting characters from Japanese anime, sourced from two image boards: danbooru and moeimouto.

Reasons for choosing this dataset:

1. Rich metadata: The extensive tagging system provides a wealth of information about each image, allowing for various machine learning approaches.
2. Visual data: Working with image data offers opportunities to explore computer vision techniques and models.
3. Cultural relevance: Anime is a significant part of popular culture, making this dataset interesting and potentially applicable to real-world scenarios.
4. Diverse applications: The dataset's structure allows for multiple project directions, from classification to generative modeling.
5. Challenging aspects: Dealing with potential mislabeled images and ensuring appropriate content filtering presents real-world data cleaning challenges.

2. Methodology

a. Data Preprocessing

The dataset requires some preprocessing due to its nature and potential challenges:

1. Content filtering: Implement a secondary screening process to ensure all images are indeed SFW, possibly using a pre-trained NSFW image classifier.

2. Tag preprocessing:

- Normalize tags (lowercase, remove special characters)
- Handle multi-word tags
- Perform frequency analysis to identify the most common tags

3. Image preprocessing:

- Resize images to a consistent dimension
- Normalize pixel values
- Augment data if necessary (rotations, flips, color jittering)

4. Split data into training, validation, and test sets, ensuring balanced representation of tags across sets.

b. Machine Learning Model

Proposed model: Multi-label image classification using a Convolutional Neural Network (CNN)

The goal is to predict multiple tags for each anime illustration. A CNN is well-suited for this task because:

1. CNNs are designed to work with image data, automatically learning relevant features.
2. They can be adapted for multi-label classification by using a sigmoid activation in the output layer.

Alternative models considered:

1. Transfer Learning with pre-trained models (e.g., ResNet, VGG):

- Pros: Potentially better performance, especially with limited data
 - Cons: May require more computational resources
2. Object Detection model (e.g., YOLO, Faster R-CNN):
- Pros: Could identify and localize specific anime character features
 - Cons: Requires bounding box annotations, which this dataset may not have

c. Evaluation Metrics

For this multi-label classification task, the following metrics can be used:

1. Mean Average Precision (mAP): This metric summarizes the precision-recall curve for each tag and provides an overall performance measure.
2. F1 Score: The harmonic mean of precision and recall, calculated per tag and then averaged.
3. Hamming Loss: Measures the fraction of incorrectly predicted labels to the total number of labels.
4. Exact Match Ratio: The proportion of samples where all labels are correctly predicted.
5. Label Ranking Average Precision (LRAP): Evaluates the fraction of higher-ranked labels that are in the correct labels.

Baseline to beat: achieve a mean Average Precision of at least 0.7 and an F1 Score of 0.65 across all tags. These baselines will be refined after initial experiments with the data.

3. Application

I suggest developing a web application where users can upload anime-style illustrations and receive predicted tags. The application will have the following features:

User Input:

- An upload button allowing users to submit their own anime-style illustrations
- Option to use a webcam to capture hand-drawn sketches

Processing:

- The uploaded image will be preprocessed similarly to the training data
- The trained model will generate tag predictions

User Output:

- Display the uploaded image alongside a list of predicted tags
- Show confidence scores for each predicted tag
- Provide a brief explanation of what each tag represents in the context of anime illustrations

Additional Features:

- Allow users to correct or add tags, potentially for future model improvement
- Implement a gallery of recently tagged images (with user permission)
- Provide tag-based search functionality for the gallery
- Integrate a language model to generate Stable Diffusion prompts using the extracted tags
 - This feature would take the predicted tags and use them to create textual prompts suitable for image generation models like Stable Diffusion

This application would be useful for anime artists looking to tag their work automatically, for fans trying to understand the elements present in an anime illustration, or for generating new anime-style images based on existing artwork tags.