**Disney Animated Character Recognition**

First Progress Report

October 10, 2024

1. **Group Information**

Group 9: Kaitlyn Le, Anh Tran, and Benjamin Tutka

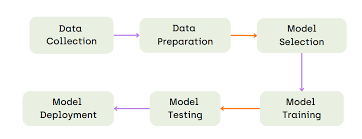
1. **Problem Statement**

Over the past century, Walt Disney Studios has produced dozens of movies featuring many beloved characters. From classic icons like Mickey Mouse to more recent icons like Elsa, there’s no doubt that these characters have created a lasting impact across generations. As times change and technology advances, there comes a need for new consumer experiences and innovative tools for the creative minds behind these characters [1]. Through this project, our goal is to build a model that recognizes Disney animated characters from various movie scenes. With art styles evolving from hand-drawn to 3D and CGI over the years, we are interested to see if these characters remain recognizable across different animation styles, settings, or outfits.

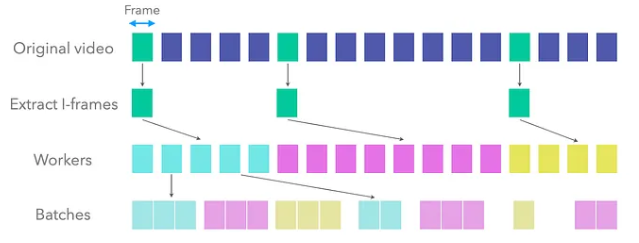
Hypothetically, our model could be used to identify a specific character across all the films or shows in which they appear, which could then be used to create a more enhanced recommendation system (for Disney’s streaming platform Disney+ for instance). If a user likes a particular art style, perhaps the model could help suggest other movies or TV shows that are of a similar aesthetic. Additionally, it could also recommend users content based on the types of characters they like. This could be particularly interesting for characters who make cameo appearances in other films. Spotting a fan-favorite character in a movie they weren’t expecting to see them in could bring joy to the viewer, enhancing customer satisfaction.

Animated character recognition is actually something the Walt Disney Company has been working on in recent years. Using traditional machine learning methods like HOG+SVM, engineers from Disney’s Direct-to-Consumer & International Organization (DTCI) developed the first automated tagging pipeline, which provided the framework for a facial detection and recognition model. The models worked particularly well for live-action characters, but it struggled to recognize animated characters that didn’t have normal human features and proportions (e.g., Lightning McQueen from *Cars*). In response, they decided to implement transfer learning to fine tune a Faster-R CNN model with PyTorch to improve detection of animated characters. The model was trained on a dataset including animated characters with both human-like and non-human designs.

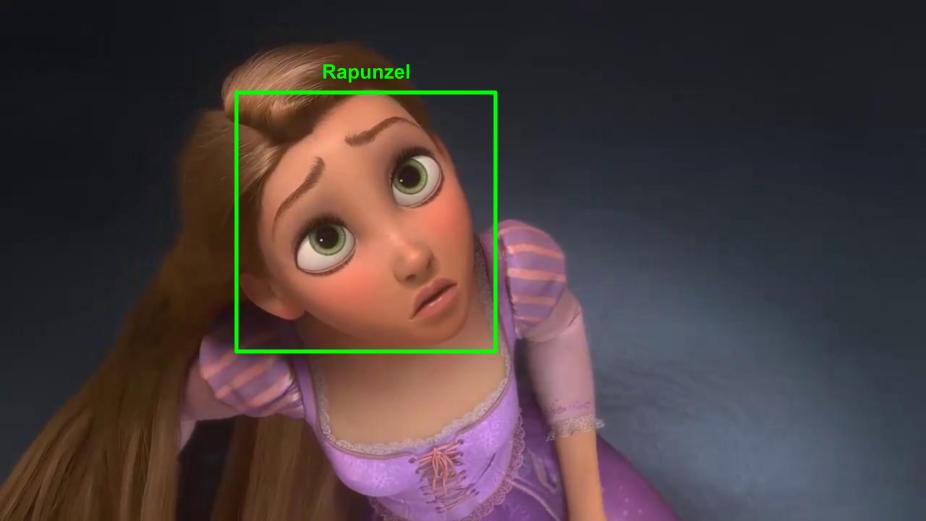
1. **Data**

For our project, we have created a dataset consisting of images of selected Disney animated characters. Each group member has chosen four distinct, non-overlapping Disney characters, resulting in a total of 12 characters across the group. Our dataset consists of about 300 images in total (25 for each character).

To construct our dataset, images were derived from video clips and/or gathered from various movie scenes, ensuring there were multiple scenes and backgrounds to provide a diverse range of visual contexts. We extracted images from these videos, selecting every few frames to represent different poses and expressions of the characters. For characters where videos weren’t as feasible, we decided to use still snapshots from the movie instead. The dataset comprises of dynamic images that were extracted from movie clips, facilitating a comprehensive approach to character recognition.



Each frame or image will be annotated with character labels indicating the name of each character. This labeling is essential for training our model, which will classify images based solely on character identity. Before training, we will preprocess the dataset to standardize all images, including resizing them to a uniform dimension. The standardized image will consist of cropping the images around just the faces specifically. Furthermore, we want to convert them to a consistent file format (e.g., JPEG or PNG), and organize the data in a structured storage format to give easy access during model training.

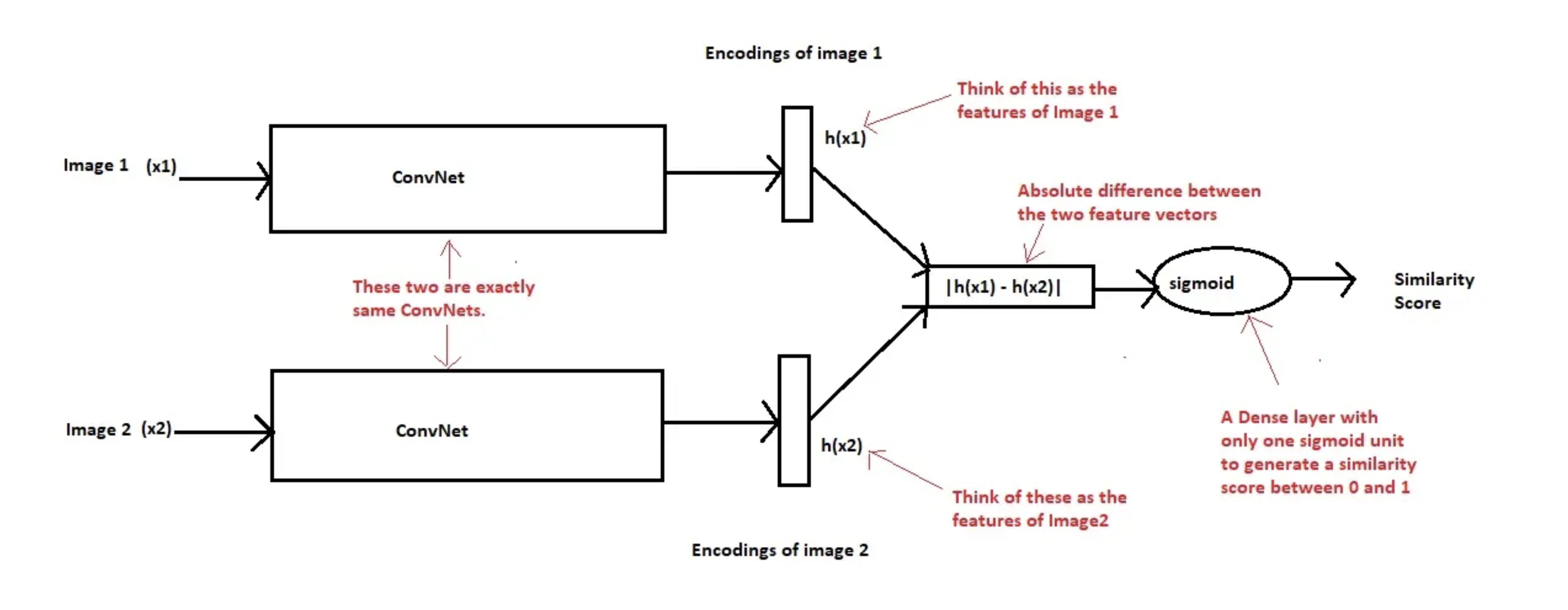


1. **Machine Learning Model**

To perform facial recognition of Disney characters, we will utilize a Siamese Convolutional Neural Network (CNN) model. While traditional CNNs are typically employed for image classification tasks—where they learn and extract features to predict specific class labels—we aim to incorporate one-shot learning into our approach to address the challenges of facial recognition.

In standard image classification, models require large datasets and extensive training for each class, which may not be ideal for facial recognition tasks, particularly when limited examples are available for some characters. One-shot learning, on the other hand, allows the model to identify and distinguish faces after seeing only one or a few examples of each character.

To achieve this, we will leverage similarity learning instead of the conventional classification. Rather than predicting a class label directly, the Siamese CNN will learn to measure the similarity between pairs of images. The model will output a similarity score, indicating whether two images belong to the same character or not. This approach is particularly suited for facial recognition, as it enables the model to generalize well even with limited data, making it an ideal solution for recognizing a wide variety of Disney characters based on their facial features.



**Evaluation**

In this project, we will evaluate the performance of our Disney character facial recognition model using accuracy, precision, and recall. Since our dataset does not exhibit class imbalance, it is appropriate to use these standard performance metrics to measure the effectiveness of the model.

**Accuracy**

Accuracy measures the proportion of correct predictions (both true positives and true negatives) out of all predictions made by the model. It gives an overall assessment of how well the model performs across all classes.

**Precision**

Precision focuses on the positive class and measures how many of the positive predictions made by the model were actually correct. It is useful when the cost of false positives is high.

**Recall**

Recall (or Sensitivity) measures the ability of the model to correctly identify all positive instances in the dataset. It is crucial when the cost of missing positive instances (false negatives) is high.

1. **Contribution**

For our project, each team member frequently communicated and collaborated with one another. We all worked together on the report, but one person took the lead on a specific section. Kaitlyn helped propose the initial project idea and was primarily responsible for developing the Problem Statement section, describing why this project is important and what has already been done for it in the past. Ben was primarily responsible for the Data section, describing the data collection process and how we plan to use the data. Anh played a big role in researching previously used models that helped the team gain a better understanding of the project. She was primarily responsible for the Machine Learning Model section, explaining the model we are implementing and the different evaluation metrics we will use. All members contributed to the data collection process, with each person gathering 25 different frames for four different Disney characters, resulting in a dataset of about 300 images.

**References**

1. <https://medium.com/pytorch/how-disney-uses-pytorch-for-animated-character-recognition-a1722a182627>
2. <https://ieeexplore.ieee.org/abstract/document/8302003>
3. <https://arxiv.org/pdf/2208.11012>
4. <https://www.kaggle.com/code/lxdnz254/art-image-classification>
5. <https://ieeexplore.ieee.org/document/9133143>