TOM AND JERRY IMAGE CLASSIFICATION PROJECT

1. Introduction

1.1 Project Overview

This project aims to develop a deep learning model that can detect whether Tom, Jerry, or both characters are present in a given frame from the *Tom and Jerry* show. The model leverages CNN (Convolutional Neural Networks) to recognize patterns and features unique to each character.

1.2 Objective

The objective is to create a model capable of identifying which characters (Tom, Jerry, or both) are present in a specific frame of the animation. This can be applied to automated analysis of episodes, video summarization, or scene extraction.

2. Materials and Experimental Evaluation

2.1 Dataset

1. Dataset Source : https://www.kaggle.com/datasets/balabaskar/tom-and-jerry-image-classification/data

Sources: Tom & Jerry cartoon show videos from Internet

Collection Methodology : Frames were extracted from various episodes of the Tom and Jerry show. These frames were manually labeled based on the presence of characters (Tom and Jerry).

2. Data Format: Labeling for these images is done manually (by going through images one by one to tag them as 1 of the 2 outcomes), so the accuracy of ground_truth is 80%

Labeled images are separated into 2 different folders as given.

Folder – DATASET_tom_and_jerry

SubFolder tom_jerry_1 - contains images with both 'tom' and 'jerry'

SubFolder tom_jerry_0 - contains images without both the characters

2.2 Preprocessing:

Frame Resizing: All images were resized to 1280 x 720 pixels to standardize input sizes for the CNN.

Normalization: The pixel values were normalized to the range [0, 1] to make training easier.

Data Augmentation: Data augmentation techniques like horizontal flipping, zooming, rotation, and cropping were applied to increase the diversity of the dataset and avoid overfitting.

Number of Classes: There are 2 different classes in the data.

Class Distribution : tom_jerry_1 class makes up 66.4 percent of data and tom_jerry_0 makes up 33.6 percent of data.

3.Methodology

We built a CNN to classify the presence of characters in the frames. CNN is well-suited for image classification tasks as it efficiently captures spatial and hierarchical patterns.

3.1 CNN Layers : The architecture of the CNN model includes:

Input Layer: Accepts images of size 1280 x 720

Convolutional Layers: Extract features from images by detecting edges, textures, and other patterns specific to Tom and Jerry.

Pooling Layers: Reduce the spatial dimensions, keeping the important features and reducing computational costs.

Fully Connected Layers: Combines all features extracted by previous layers to make a final prediction.

Output Layer: 2 output neurons representing the two classes (with Tom and Jerry,without tom and Jerry).

3.2 Transfer Learning

A pre-trained CNN model (such as ResNet or VGG16) was used with fine-tuning, transferring the knowledge learned from large-scale image classification tasks to this project. This helped speed up the training process and improve model performance.

3.3 Hyperparameters

Learning Rate: 0.0001 (tuned through experimentation)
Optimizer: Adam optimizer
Loss Function: Categorical Cross-Entropy Loss (since this is a multi-class classification problem)
Batch Size: 32
Epochs: 20

4. Model Training and Evaluation:

4.1 Training Process

The model was trained using the training dataset with data augmentation applied to improve the generalization of the model. Early stopping was used to halt training when the validation accuracy stopped improving to avoid overfitting.

4.2 Evaluation Metrics

- **Accuracy**: 80% of correctly classified frames.
- **Precision, Recall, and F1-Score**: Measured for each class (with Tom and Jerry, without Tom and jerry).

Classification Report :

	precision	recall	f1-score	support
0	0.87	0.82	0.84	307
1	0.68	0.76	0.72	156
accuracy			0.80	463
macro avg	0.78	0.79	0.78	463
weighted avg	0.81	0.80	0.80	463

Figure 1 : Classification Report

- **Confusion Matrix**: A confusion matrix was used to visualize how well the model distinguished between the Two categories.

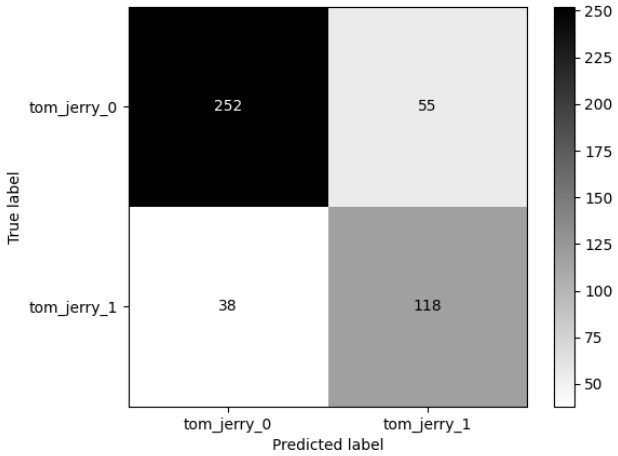

Figure 2: Confusion Matrix

5 Results

5.1 Training and Validation Accuracy

The model achieved an accuracy of around **82%** on the validation set after 20 epochs.

5.2 Test Set Performance

On the test set, the model achieved an accuracy of **83%**, showing that it generalizes well to new frames from the show. The confusion matrix showed that the model performed better at detecting both characters together, while distinguishing between only Tom or only Jerry was slightly harder.

6. Model Interpretation

6.1 Feature Maps and Grad-CAM

To interpret the model’s decision-making process, Grad-CAM (Gradient-weighted Class Activation Mapping) was used to visualize the regions of the image that were most influential in the model's prediction. This showed that the model focused on specific areas like Tom’s facial features and Jerry's smaller size.

7. Conclusion

This project successfully created a CNN model capable of detecting the screen presence of Tom and Jerry with good accuracy. The results show that the model is able to recognize both characters in various scenes, though performance could be improved with more complex networks or larger datasets.

8. Future Work

Improving the Dataset: Adding more frames with varied backgrounds and lighting conditions could improve model robustness.

Model Optimization: Fine-tuning hyperparameters or using deeper networks could improve accuracy further.

Deploying the Model: This model could be deployed as a real-time frame analyzer for video streaming platforms or used to automate video analysis tasks.

Extending to Other Characters: The model could be expanded to recognize additional characters from the show, such as Spike or Tuffy.Challenges:

There are images that can be challenged during training in image classification, as these images are distorted in the original size or shape and color of the characters.

9.Reference

1. S. C. Agrawal and R. K. Tripathi, "Cartoon Face Detection and Recognition with Emotion Recognition," 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), Chennai, India, 2023, pp. 1-4, doi: 10.1109/ICDSAAI59313.2023.10452501. keywords: {Industries;Emotion

recognition;Face recognition;Artificial neural networks;Media;Face detection;Task analysis;cartoon;deep neural network;face recognition},
2. https://www.kaggle.com/datasets/balabaskar/tom-and-jerry-image-classification/data
3. T. Guo, J. Dong, H. Li and Y. Gao, "Simple convolutional neural network on image classification," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 2017, pp. 721-724, doi: 10.1109/ICBDA.2017.8078730. keywords: {Image classification;Biological neural networks;Feature extraction;Convolutional codes;Machine learning;Training;Convolutional neural network;Deep learning;Image classification;learning rate;parametric solution},
4. Hussain, M., Bird, J.J., Faria, D.R. (2019). A Study on CNN Transfer Learning for Image Classification. In: Lotfi, A., Bouchachia, H., Gegov, A., Langensiepen, C., McGinnity, M. (eds) Advances in Computational Intelligence Systems. UKCI 2018. Advances in Intelligent Systems and Computing, vol 840. Springer, Cham. https://doi.org/10.1007/978-3-319-97982-3_16