**Movie Poster Classification - Final Report**

John Durso & Jacob Stephens

**1 Introduction**

Our objective was to classify the genre a movie poster belongs to. Our goal was to achieve the highest performance by experimenting with and comparing several different models. Our proposed approach was to use and compare the following models' performance: baseline Multi-layer Perceptron (MLP), VGG-16 (Visual Geometry Group), ResNet-50, and EfficientNet-b0. We would evaluate these models on several evaluation metrics such as accuracy, recall, precision, and F1 score. The classification of movie posters has many applications. Primarily, our model could play a role in any movie recommendation system and could recommend to a user genres similar to the ones they like, based on poster images. Additionally, our model could be used to identify the most popular genres or discover any trends within or between genres. For instance, how similar are the posters between romance and comedy and what does that say about the genres? There are many applications and discoveries to be made using this model as a guide.

**2 Dataset**

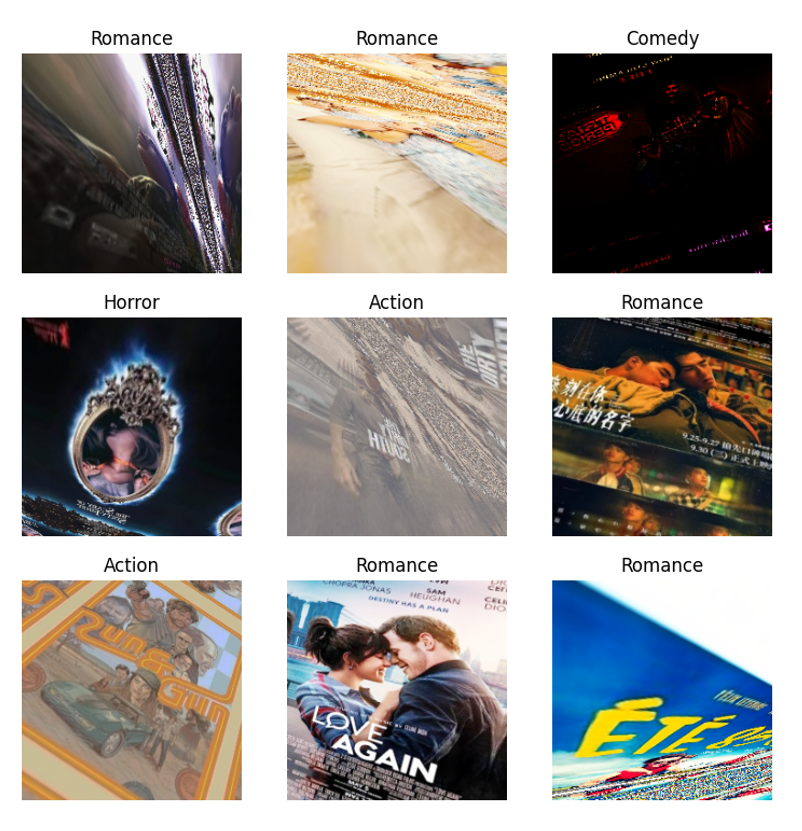
We used a singular dataset we found on Kaggle: “Four-Genre Movie Poster Images” created by Zulkarnian Saurav. The dataset contains 1,325 poster images with roughly 300 posters for each of the four genres. Below is the data distribution of each genre as well as several statistical measures of the dataset:

| **Genre** | **Poster count** |
| --- | --- |
| Action | 337 |
| Comedy | 321 |
| Horror | 398 |
| Romance | 269 |
| Total | 1,325 |

Clearly shown in the dataset, the “horror” genre has the highest poster count in our dataset and the “romance” genre has the fewest. The mean posters for a genre is about 331 posters. The median poster amount is 321 posters. The standard deviation is 46.01.

Here are some randomly-chosen samples from the dataset from each genre:

However, posters formatted like those above were not inputs to our model. We employed several pre-processing techniques on our data to increase performance: resizing (to 224x224), data augmentation (rotation, zooming, and flipping), pixel normalization, and separate data image loaders for train/validation set and test set. Below are some sample images that have been processed with the aforementioned techniques:



**3 Methods**

We experimented with transfer learning using pre-trained CNN architectures and fine-tuned them on our dataset. Transfer learning allows us to leverage knowledge learned from large-scale datasets and adapt it to our specific task with relatively small amounts of data. The candidate models are:

VGG16 - a convolutional neural network architecture consisting of 16 layers, including 13 convolutional layers and 3 fully connected layers. It is known for its simplicity and uniform architecture, with 3x3 convolutional filters and max-pooling layers,

ResNet50 - a variant of the ResNet (Residual Network) architecture, comprising 50 layers, including residual blocks with skip connections. These skip connections enable training of very deep networks by mitigating the vanishing gradient problem. ResNet50 has shown excellent performance on various image classification tasks.

EfficientNet\_B0 - represents a family of convolutional neural networks that achieve state-of-the-art performance through efficient scaling of network depth, width, and resolution. The base model, EfficientNet-B0, consists of 7 layers, including multiple stages of depthwise-separable convolutions and squeeze-and-excitation blocks.

These models are well-established in the computer vision community and Each candidate model has its unique architecture and computational characteristics. To establish a baseline for comparison, we also implemented a simple baseline model based on a multi-layer perceptron (MLP). This baseline model serves as a reference point for evaluating the performance of more complex CNN architectures. The MLP consists of two fully connected layers and is trained on flattened image inputs.

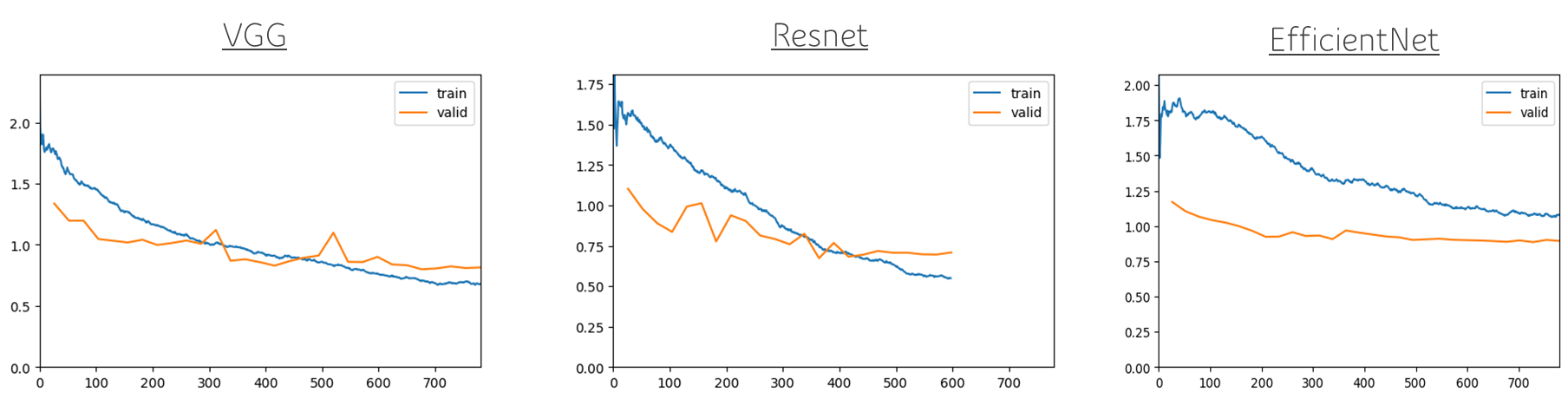
**4 Experiment Setup**

We defined hyperparameter configurations and conducted experiments with varying batch transformations, learning rates, and data augmentation settings to optimize model performance. These experiments enabled us to identify the most effective configurations for each model architecture and refine our approach accordingly.

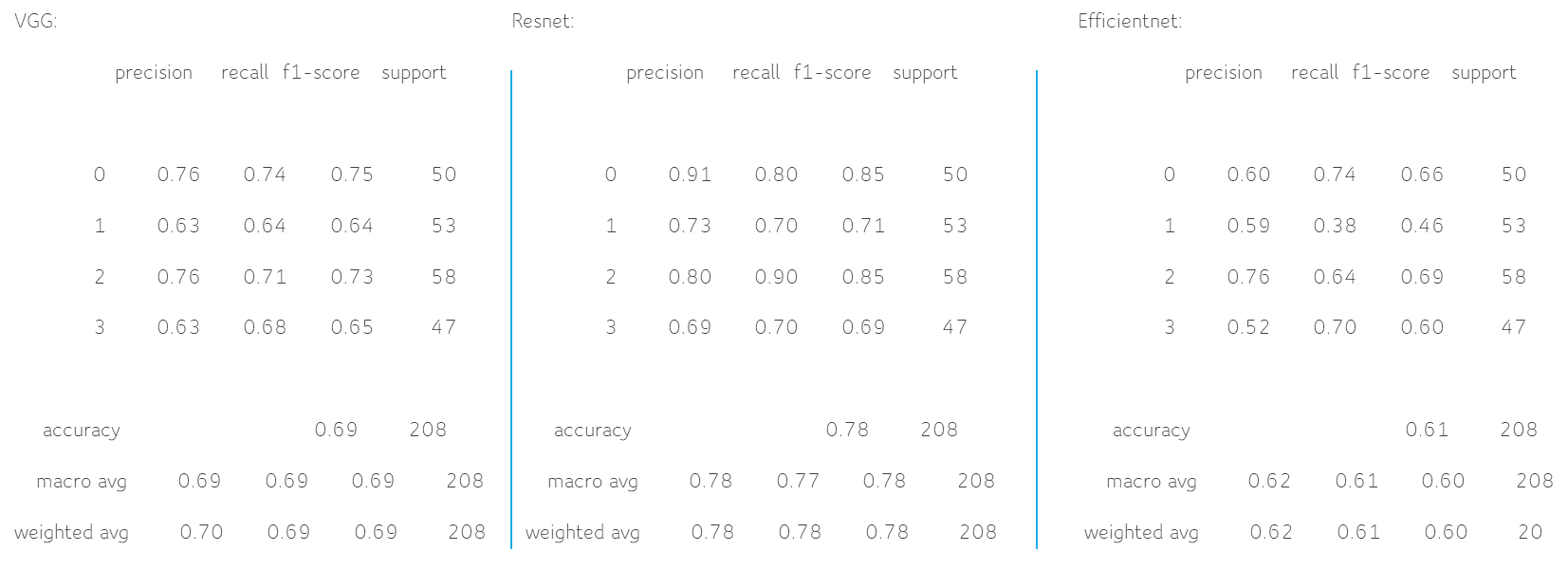
**5 Results**

ResNet50 achieved the highest accuracy on the test set, followed by VGG16 and EfficientNet-B0. Confusion matrices and evaluation metrics provided insights into model performance for each genre, highlighting areas of strength and areas for improvement. Below are some figures.

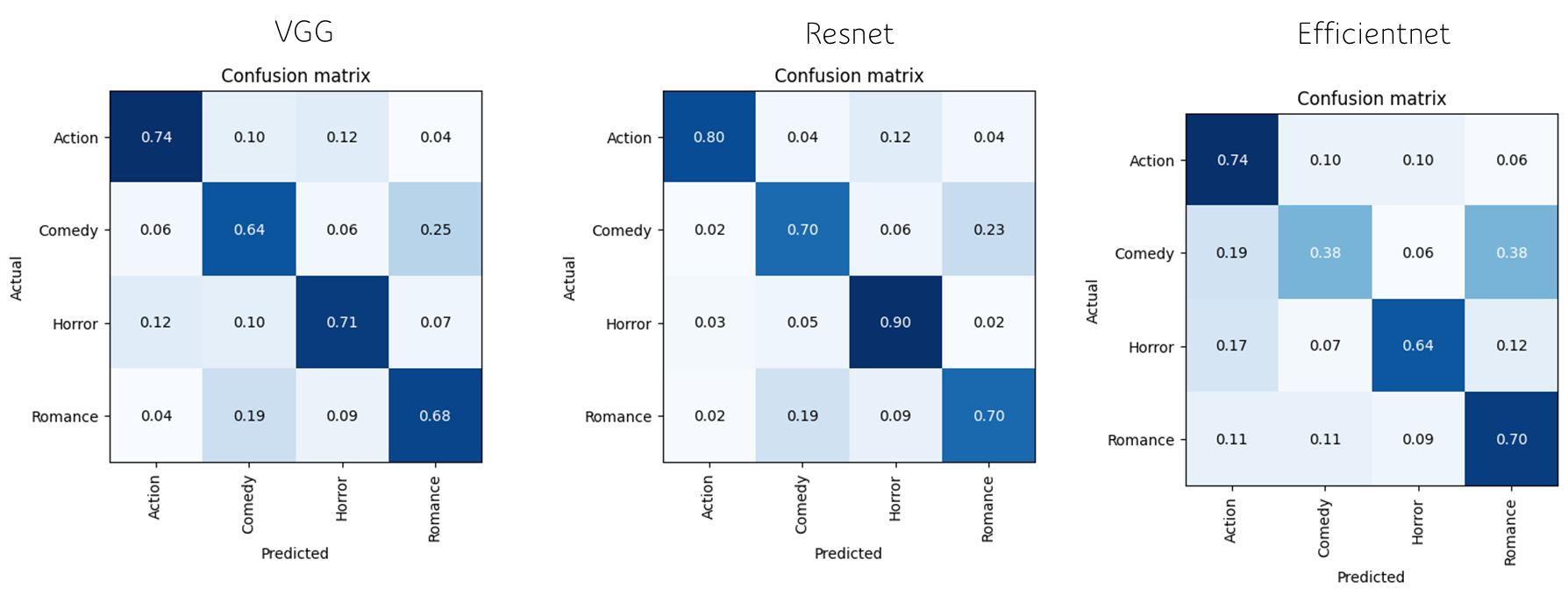
Training Loss vs Validation Loss

****

Test Set Classification Report

****

Test Set Confusion Matrices

****

**6 Discussion**

Although for some genres we achieved high accuracy, like horror with a 0.90 accuracy, the genres romance and comedy were commonly mistaken for each other. This confusion makes sense, given the similarities and interconnectedness of the two genres. A challenge we faced were the models, especially VGG-16 and ResNet-50 tended to severely overfit, so the utilization of regularization was essential to maintaining high accuracy in spite of the overfitting. Another challenge was the relatively low data amounts. An average of only 300 posters per genre was a limiting factor; more posters would have increased our performance and relieved us from our heavy reliance on data augmentation. Further improvements for our model would be to, of course, increase the amount of data we had to work with. Additionally, if we had more time, we could have run more experiments and tuned our parameters more to cater our model more strongly to our dataset. Lastly, an improvement would be the addition of cross-fold validation in order to have some statistical stability. It would also improve performance to include a system which recognizes facial expressions in order to distinguish emotion present on the cover.

**7 Conclusion**

In essence, we achieved our initial problem statement by developing a high-performing model that classifies movie posters to their correct genre. In the process, we discovered insights about the similarity of romance and comedy as well as the distinctiveness of horror. Although there were some challenges, we achieved a model which has many applications surrounding the data and trends of movie posters and cinema at large.