

Case Study: Movie Poster Classification

Research Question: How accurately can an image classification model identify the genre specific components of movie posters?

Motivation:

Every year more than 500 films are released in the US and Canada, and thousands more are released worldwide. In an industry with so much competition, a movie's marketing campaign is crucial to determining a movie's success at the box office. Importantly, a movie's poster, which the public views in the form of advertisements in the theater, on billboards, in online advertisements, etc., must be engaging while also conveying to a target audience member that this movie is for them. This results in movies of different genres, which have distinct audiences, setting themselves apart from each other through their stylistic choices. For example, a horror movie might look darker and more solemn than a comedy movie poster with bright colors and humorous expressions on people's faces. Hence, it is important for marketing executives and designers to understand the conventions of the movie genre that they are producing a poster for.

Your Goal:

As a data scientist, you are equipped with the skills to take publicly available image data and turn it into insight about the differences between different film genre's posters. You will do this by exploring the brightness and RGB differences (and any other visual features that stand out to you). Then, you will create (or adapt an existing) convergent neural network and train it on hundreds of movie posters and then test its efficacy in classifying movies into genres based on its poster alone. You can also add to this project by training the model to identify specific objects (e.g. faces) or by including movie titles in the analysis (since these are featured heavily in most movie posters). At the end of the project, you should be able to quantify the efficacy of your model and how this ties into the specific genres that you have chosen. Your deliverable will be a GitHub repository containing your movie poster dataset after scraping posters off the web, your exploratory analysis and the code for your image classification model (including the accuracy of the model).

Case Study Repository with further instructions: <https://github.com/Nehach73/CS3-Case-Study>