



# The Movie Poster Predictor

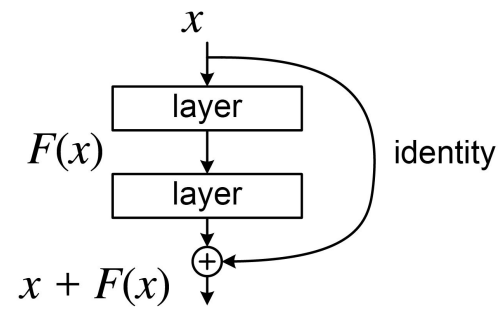
By: Cole Balacek, Xingjian Du, Skylar Bi

# Our Project Goals

- Our goal with this project is to make an AI that will be able to take in the .jpg file of a movie poster and predict what genre is and what its score is
- This will help movie producers test their posters to make sure that they both correctly communicate the genre of the film and are ones the audience will likely associate with high quality movies



# AI Structure and Dataset



- We are using a type of structure called a Residual Neural Network or Resnet
  - A Resnet is a type of Convolution Neural Network or CNN
    - Was made to overcome the vanishing gradient problem, where the accuracy gets poorer after multiple convolutions
    - The main feature of this network is that it incorporates the original X value back into its final predictions by adding in residual connections.
- The dataset we are using was taken from kaggle.com and was posted by NEHA
  - The set included a large number of features for each data point, however what we will only be taking into consideration the poster image, the score, and the genre



# Potential Problem

- Since this project only takes into consideration the movie's poster, genre, and score, it does miss other details that would affect the design of a movie poster such as
  - When the poster came out and the general trends of the time
  - Who the director is
  - Who the producer is
  - Who is starring in the film
  - What other IP it might be connected to
  - What team designed the poster
- Any number of these factors could affect a posters design but we sadly do not have access to a dataset that includes them

I don't understand  
this trilogy at all





# Dataset setup

```
# Define the custom dataset class
class MoviePosterDataset(Dataset):
    def __init__(self, csv_file, poster_dir, transform=None):
        self.data = pd.read_csv(csv_file, encoding='latin1')
        self.poster_dir = poster_dir
        self.transform = transform

        # Check if the poster image file exists for each IMDB ID
        self.data["file_exists"] = self.data["imdbId"].apply(lambda x: os.path.exists(os.path.join(self.poster_dir, f"{x}.jpg")))
        self.data = self.data[self.data["file_exists"]]
        # Discretize the IMDB scores into intervals
        self.data["score_label"] = pd.cut(self.data["IMDB Score"], bins=score_intervals, labels=labels, include_lowest=True)
        # Filter out unusual labels
        self.data = self.data[self.data["score_label"].isin(labels)]
        self.data = self.data.reset_index(drop=True)
```

- We are setting up our dataset by creating a dataset subclass called MoviePosterDataset
  - It reads the input csv and returns the proper values
- We have compressed the score values from 0-10 to 0-5 to help increase the accuracy of our guesses
- Our posters are transformed to a size of 224x224 pixels, placed in a tensor, and have their color rated based on how far it differentiates between the average rgb value for pictures (we got the average and deviation from ImageNet)

```
# Define the transforms
transform = transforms.Compose([
    transforms.Resize((224, 224)),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
])
```

# Training

- We used a standard torchvision resnet50 neural network in and trained the model in 10 epochs
  - A resnet 50 is the same as a standard resnet but with 50 layers of convolution
- A progress bar was also included so the user can know how far along it is in its training progress.
- We split our dataset into training, evaluation, and test sections





# End Result



- Our accuracy for the score was 66% while our genre crashed before it could finish
- Our low accuracy suggests a low level of correlation between our the poster image and the movie's score
- It might have also been do to a flaw in the dataset, such as how it uses movies from a variety of time periods
  - For example: Good posters might generally be paired with good movies but a good poster 30 years ago might not be the same as a good poster today, so the old high quality movie would be associated with the new low quality movies







Thank you for your Time

Q&A