Inferring Movie Genres From Their Poster

Mostafa Dewidar

mdewidar@Stanford.edu

Computer Science, Stanford University

Stanford Computer Science

Background

Movie posters are one way to encode the most information about a movie in one picture. If broken down successfully, such a dense representation can provide a lot of information about a movie that can be helpful to viewers, producers, and distributors. In this paper, I attempt to apply many different CNN architectures, namely resnet101, VGG19, and AlexNet to the problem of Movie-poster-to-genre classification to see if they can solve it and if so, which are better adapted to the problem to suggest methods to improve on performance in the future. Pobar et al. used ML-kNN, RAKEL and Naïve Bayes on a dataset of around 6739 posters and achieved a top accuracy with Naïve Bayes of 38% . Wi et al. used a Gram layer in a CNN to extract style features from the poster before feeding it into the network and were able to achieve an accuracy of 45%. Barney and Kaya used a custom resnet34 implementation, a custom CNN architecture as well as ML-KNN and were able to achieve a top accuracy of 38.26% . Kundalia et al. Balanced the poster data collected from IMDb to make sure all genres had an equal number of posters associated with them, used inceptionV3 and were able to report a remarkable 84% accuracy.

Problem Statement

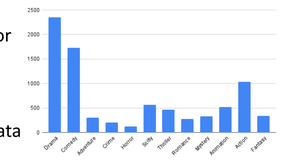
Problem:	Classify Movies to genres using posters.
INPUT:	Movie Poster Image.
Networks:	Alexnet, VGG19, Resnet101.
Output:	Movie Genre Classification.
Metrics:	Accuracy, Precision, Recall.

Dataset

8252 poster images from IMDb dataset. All images were resized to be (300,180,3) and the data was split into 80:10:10 training:validation:test split. each poster had an associated set of genre labels that ranged from 1 to 3 associated genres. I pre-processed the data to only in- clude one label for each poster. The label was selected according to alphabetical order of genre names appearing for the poster, a random selection scheme that should maintain the distribu- tion of the data. The distribution of the labels for the data can be viewed in Table 1.

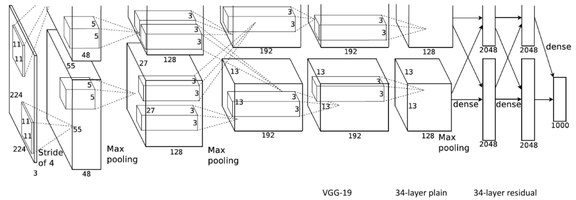
Genre	Examples
Drama	2351
Comedy	1731
Adventure	306
Crime	204
Horror	121
Sci-Fi	567
Thriller	467
Romance	276
Mystery	334
Animation	522
Action	1035
Fantasy	338

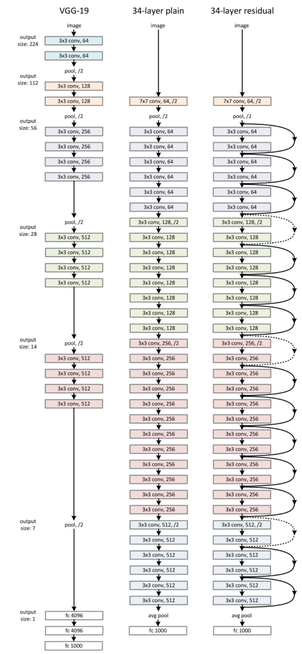
Table 1. Data Distribution by Genre



Methods

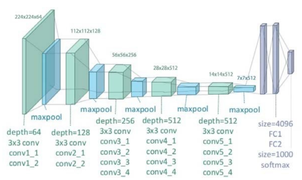
AlexNet is a large computer vision model with more than 60 million parameters. AlexNet has 650,000 neurons, five convolutional layers followed by 3 max pooling layers and three fully connected layers. AlexNet popularize the regularization technique of dropout whereby neurons are dropped (activations set to zero) during training with random probability so that they have no effect on the outcome for the training batch and then used normally during inference to make sure that the network doesn't overweight some neurons while ignoring others thereby nudging it to learn complex features of the data.





The Resnet (Residual Network) architecture was invented to solve the problem of increasing inefficiency of training and optimization in very deep neural networks, since it had been shown that depth can be very helpful to the accuracy of CNNs. Resnets are built on the corollary of the hypothesis that states that if one can asymptotically approximate a complicated function with multiple stacked non-linear layers, then the same can be said for residual functions.

VGG16 and VGG19 architecture were an attempt to see the effect of increasing depth on effectiveness of CNNs. The net contains eight layers with weights; the first five are convolutional and the remaining three are fully-connected. The output of the last fully-connected layer is fed to a 1000-way softmax which produces a distribution over the 1000 class labels for the ImageNet challenge.



Experiments & Results

Training seemed to converge on a result after 4 epochs as we can see that both training a validation loss converge to the same value and training loss starts to go lower than validation loss indicating that we might be over fitting if we train for more epochs.

Epoch	Accuracy	Precision	Recall
1	0.3539	0.1510	0.1546
2	0.3787	0.1546	0.1531
3	0.4115	0.1747	0.1848
4	0.4175	0.1811	0.1837

Table 4. Training Results for VGG-19

Confusion matrices shows that the network is skewed to predicting the most popular genres when it's confused or prompted by posters that are towards the edges of the distribution. This may signal that additional pre-processing of the data could have yielded better results.

Epoch	Accuracy	Precision	Recall
1	0.3472	0.1693	0.1412
2	0.3648	0.1704	0.1421
3	0.3854	0.1744	0.1515
4	0.3896	0.1756	0.1588

Table 3. Training Results for Alexnet

A peak accuracy of 42% was achieved. Networks were able to achieve similar results which suggests that architecture is less important for this problem although deeper networks VGG19 and ResNet101 got marginally better results which suggests depth may be useful

Epoch	Accuracy	Precision	Recall
1	0.3727	0.2136	0.1947
2	0.3848	0.1922	0.1712
3	0.4218	0.2034	0.1905
4	0.4212	0.2273	0.1872

Table 2. Training Results for ResNet 101

It seems that the network is fitting everything unfamiliar into buckets of what its most likely to find in the dataset. We can see that a good number of the Drama movies are being classified as Comedy and vice versa. When the networks are confused about one label, they just tend to predict one of those two. We can also see that Action is another label that gets thrown around because it's over-represented in the data. All of this suggests that the networks are learning the distribution of the data to a good degree but haven't seen enough examples from all the Genres to be able to make predictions with high precision.

Conclusion and Future Work

The best model accuracy was 42% which is substantially better than random guessing and is similar to State-of-the-Art when no pre-processing of the dataset is involved. In the future, creating a better dataset should be strongly considered. Creating a bigger dataset might involve creating more than one poster for existing movies or creating fictional poster movies for each genre to augment the data. Another thing that needs to be done is to balance the dataset so that each Genre has equal representation. One more thing to consider is feature engineering. Using segmentation and object recognition, or text-recognition on the posters and feeding the outputs of these networks as inputs to the genre-classification network will add valuable information.

Acknowledgements

Thanks to the CS 231n team, Pranav Hari, my Family and Friends

References

[1] Gabriel Barney and Kris Kaya. Predicting genre from movie posters. *Stanford CS 229: Machine Learning*, 2019. 1

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 2

[3] Yin-Fu Huang and Shih-Hao Wang. Movie genre classifica- tion using svm with audio and video features. In *Interna- tional Conference on Active Media Technology*, pages 1–10. Springer, 2012. 2

[4] Marina Ivasic-Kos, Miran Pobar, and Luka Mikec. Movie posters classification into genres based on low-level features. In *2014 37th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, pages 1198–1203. IEEE, 2014. 1

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural net- works. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Wein- berger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. 2

[6] Kaushil Kundalia, Yash Patel, and Manan Shah. Multi-label movie genre detection from a movie poster using knowledge transfer learning. *Augmented Human Research*, 5(1):1–9, 2020. 1

[7] Eric Makita and Artem Lenskiy. A multinomial proba- bilistic model for movie genre predictions. *arXiv preprint arXiv:1603.07849*, 2016. 1

[8] Mel-Leo Rosal. U.s. film industry statistics [2022]: Facts about the u.s. film industry, May 2022.

[9] Karen Simonyan and Andrew Zisserman. Very deep convo- lutional networks for large-scale image recognition, 2014.

[10] Jeong A Wi, Soojin Jang, and Youngbin Kim. Poster-based multiple movie genre classification using inter-channel fea- tures. *IEEE Access*, 8:66615–66624, 2020.