Classification of Book Genres By Cover and Title

Holly Chiang[1] Yifan Ge[2] and Connie Wu[3]

***Abstract*— This paper discusses the classification of books purely based on cover image and title, without prior knowledge or context of author and origin. Several methods were implemented to assess the ability to distinguish books based on only these two characteristics. First we used a color-based distribution approach. Then we implemented transfer learning with convolutional neural networks on the cover image along with natural language processing on the title text. We found that image and text modalities yielded similar accuracy which indicate that we have reached a certain threshold in distinguishing between the genres that we have defined. This was confirmed by the accuracy being quite close to the human oracle accuracy.**

I. INTRODUCTION

There is a saying that you shouldn't judge a book by its cover, but for most readers the cover forms their first impression of a book. It has been shown that the cover design has a significant impact on the sales of a book, with book sales often shooting up after a change in design. Our goal is to create a model that can determine how representative a cover is of its genre, as a method to later evaluate if the more a book cover resembles others in its genre, the higher the book sales.

We fed our algorithm the cover image and the cover's title, which often are what book consumers look at first whether it is in store or online. Without any other information such as prior customer purchases or similar book associations, we want to guess the genre purely based on these two characteristics. Surprisingly it is very difficult for even a human to distinguish between books of different categories, especially if the title is vague or if the book cover does not allow for much inference.

The motivation for solving this problem is for designing covers of new books that want to come onto market with a relatively unknown author. This study would show what types of features concerning covers and titles are most important for determining a book's genre, and subsequently how a consumer perceives such a book.

We classify books into five genres: Business, Fantasy, History, Science-Fiction, and Romance such as in Figure 1 by using a multi-class SVM. These categories are large enough to allow us to easily extract large training and test sets and have covers distinct enough that our methods should have some success.

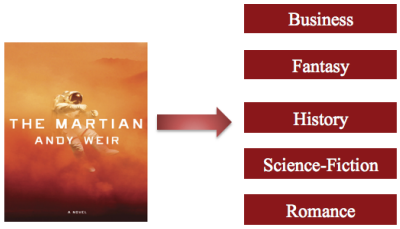


Fig. 1: We classify book covers into five genres: Business, Fantasy, History, Science-Fiction, and Romance

II. RELATED WORK

There were no previous attempts that we could find aimed at tackling our particular problem of classifying novels by genres. However one work that was tangentially similar was Classifying Paintings by Artistic Genre [1]. Zujovic et al. tackled the problem by extracting features from gray scale and color images. Texture and edge features were extracted from gray scale images using Steerable Filter Decomposition and the Canny edge detector. Color image features were extracted using an 8-bin histogram for each of the HSV values. They then tried multiple different classifiers including Naive Bayes, k-nearest neighbors with 1 and 10 nearest neighbors, SVM, a neural network with 4 hidden layers, and a J48 decision tree. The strong point of this approach is they were able to try a wide variety of classifiers, with accuracy ranging from 47.6 to 68.3. The weakness of the approach lies in the fact that the features extracted using Steerable Filter Decomposition included edge information, constraining their results. Similar to this approach, we tried binning the HSV histogram and finding the image complexity in terms of number of edges.

To extract features from book covers, we also looked into previous studies on measuring image complexity. We think it would be a good indicator of genres. Image

complexity is best represented by compression level. However, compression level is costly to compute. On the other hand, spatial information displays a strong correlation with compression level [2]. Yu discussed how the mean, RMS, and standard deviation of spatial information relate to the image complexity.

In conjunction, recent studies have shown that generic image descriptors extracted from convolutional neural networks are powerful when used in combination with SVM or softmax classifiers in visual recognition tasks [3]. Earlier layers have "general" features such as color blobs and edge detectors whereas later layer features are more specific to the dataset such as distinguishing unique characteristics of objects in the same category[4]. This is why we chose to use a neural network trained on ImageNet, since it covers a variety of images from a large database.

Finally, we looked into correlating a novel's title to its genre. Originally we wanted to find the conditional probability a book belonged to each genre given a particular word, but quickly realized that titles tended to be unique, with many words appearing only once across the entire data set. To address this, we instead used word2vec [5] pretrained using the skip gram model on the Google News dataset. word2vec was able to return metrics even for most of the unique words in the dataset. Its main drawback was that the metrics it returned were not specifically trained to tell how indicative a word was of a genre.

III. Dataset and Features

To obtain the data set, we looked into several options: the Google Books API, Google image search, and the OpenLibrary API. Each of these options had its advantages and drawbacks. The Google Books website has higher quality images than the others. However, it requires Google account login information and its API is more geared towards accessing the user's own book collection. When we attempted to collect book cover images from Google image search, we found the results returned were quite inconsistent, especially the quality and size. This had the potential to cause issues later on. Finally, we decided to use OpenLibrary's API. Its API is well documented and maintained. Compatibility with RESTful API makes OpenLibrary's data very accessible with Python urllib function calls. It also provides an interface to obtain a list of books by genre. However, since OpenLibrary book cover images are user uploaded, the quality of the images are very inconsistent. Also, all the images are relatively small with 350×500 pixels. Since the features we extract don't depend on the details, small images mostly won't hurt our performance but will improve our processing speed. For images with blank or no useful information, we developed some pre-processing algorithms to improve the quality of our dataset, which will be discussed in detail in that subsection.

Our dataset obtained from OpenLibrary.org consists of a total of 6000 images from the five genres: Business, Fantasy, History, Science-Fiction, and Romance. The assumption we made here was that these genres have small overlaps and are relatively easy to differentiate from their covers. Among these 6,000 book cover images, we have 1,000 training images and 200 test images for each genre.

A. Pre-Processing

Among the images we obtained from OpenLibrary, there were some images with single color blocks or were extremely low-resolution with minimal text. Since our project is focused on using cover images to extract information about the genre, pure text covers do not provide the requisite information. To improve the dataset quality, we decided to research and implement some pre-processing algorithms to filter the original dataset.

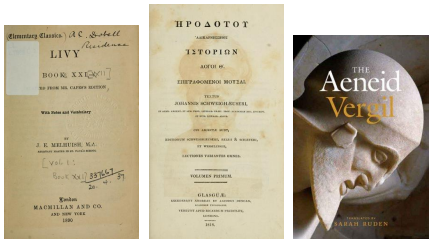
The image processing steps we used are listed in order below:

1) Convert the image to gray scale
2) Use median filter to blue the image
3) Adaptive threshold to isolate features of an image
4) Compute the ratio between the number of thresholded pixels over the total pixel count

Finally, we discarded images with thresholded pixel counts less than 20% of the total image pixels. After tuning the parameter in the filter and adaptive thresholding algorithm, we were able to successfully distinguish low quality images from the dataset. An example of processing three book covers are shown in Fig. 2 and Fig. 3. In the example, we picked three images where we want to discard 2a and keep 2b and 2c. Notice that differentiating 2b from 2a is harder. We accomplished this by aggressively using large median filter blocks and picking the best threshold value.

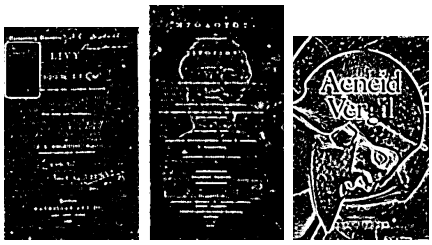
B. ImageNet

We have extracted features from a fully-connected fc7 (second to last) layer of a convolutional neural network model, pre-trained on ImageNet from Krizhevsky et al. as shown in Figure 4 [6]. The neural network



(a) (b) (c)

Fig. 2: Original book cover images, a) is low quality cover need to discard, b) and c) are good images need to keep.



(a) (b) (c)

Fig. 3: Processed book cover images, the threshold pixel ratio for a), b) and c) are 17.49%, 23.37% and 71.97%, respectively

was trained on 1.3 million high-resolution images in the LSVRC-2010 ImageNet training set. Each vector has length of 4096 features.

Taking each image in the test set, we fed it as input to the CNN and using the extracted fc7 layer, we associated each of these vectors with the respective genre label. After, we normalized the features by dividing by the norm of the vector. Then we stored the feature vectors and their labels in a dictionary to be combined with the extracted NLP features.

C. Stanford NLP

We extracted features from book covers using two different NLP classifiers as shown in Figure 5. The first classifier we used was the Stanford NLP classifier, which is a probabilistic softmax classifier. The features it used were character n-grams of size 1 to 8, the prefixes and suffixes, and the string length bucketed into 10, 20, or 30 characters.

The second classifier used was word2vec, which produces word vectors from a text corpus using skip-gram. We fed it the Google News dataset and used the extracted vector to find the distance of each word in the title to a word representing each genre. We used

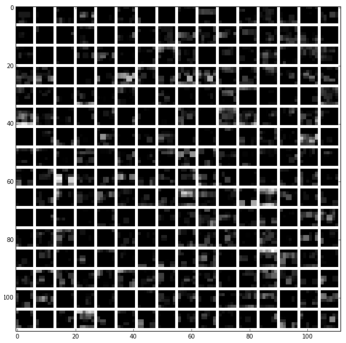


Fig. 4: A fully-connected layer of the ImageNet CNN

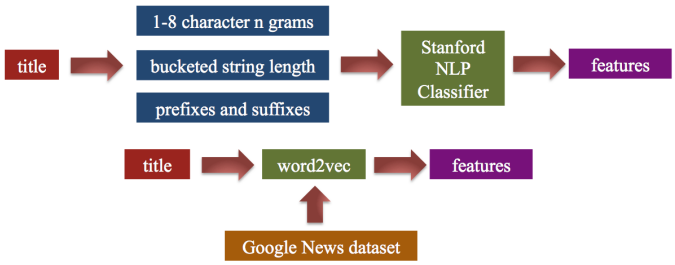


Fig. 5: NLP feature extraction

softmax to find the probability of the title belonging to any genre category.

From each of the two classifiers, we made feature vectors consisting of the probabilities of the title belonging to each of the five categories, and one hot encoding for the overall predicted category.

D. Image Processing

Intuitively, different genres are more likely to have different color schemes. For example, Science-Fiction may have brighter colors, such as yellow and red, whereas history tends to have a bland colors such as beige and brown. Thus, the first set of image features we constructed were based on the color distribution in HSV space. We converted images to HSV space and binned the pixels into 16 color schemes. These color schemes are defined by shattering of the image spectrum shown in Fig. 6a. Then we used a 2% threshold value to convert the histogram to a 16 bit one-hot feature vector. The average histogram is visually represented in Fig. 6. As can be seen, each histogram has a different color pattern. The average histogram of history genre is starkly different whereas the other histograms are subtly distinct.

Besides using the color histogram, we also looked into the image complexity. Image complexity can be

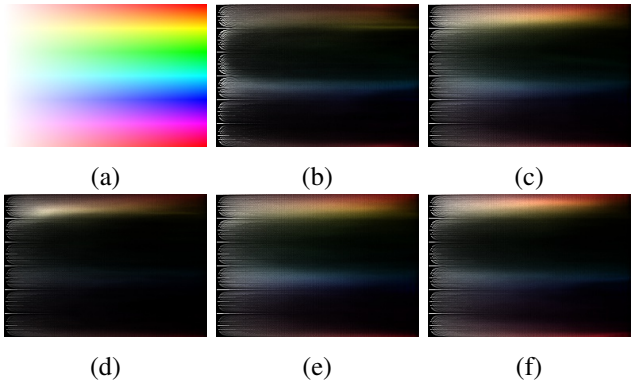


Fig. 6: HSV histogram. a) HSV color spectrum map; b) Business, c) Fantasy, d) History, e) Science-Fiction and f) Romance

best represented by the compression level of an image. However, getting the compression level can be costly. That's why we also looked into spatial information, which has decent correlation with complexity[2] and is much more efficient to calculate. Hence, we added spatial information vectors to the feature set.

IV. Methods

A. Multi-class SVM

After combining the features, we used a multi-class support vector machine to classify our training set. It is similar to the binary classification, where we fit classifiers to each class and look for the largest distance from a hyper-plane. However, now we use a one-versus-rest approach to determine which of the five categories an image most belongs to, where the category that outputs the highest distance will determine the category of the image [7]. The dual optimization problem for the classifier is stated below [8]:

$$\min_{a} \frac{1}{2}\alpha^{\top} y_i y_j K(x_i, x_j)\alpha + e^{\top}\alpha$$

$$\text{subject to } y^{\top}\alpha = 0$$

$$0 \leq \alpha \leq C, i = 1, ..., n$$

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2)$$

Since science-fiction and fantasy books are not necessarily separable, soft margins are incorporated. Similarly, we used the radial basis function kernel since it allows for higher dimensions in the data space allowing for more flexibility in separating the data.

V. Results

Since we used supervised learning, we were able to quantify the performance of our algorithm. It was interesting to observe that the results were similar for transfer learning on the cover image versus NLP on the title text hovering around 60%. This is quite close to the average human trial accuracy we found which was around 73% done on over 10 human subjects.

We originally tried three different approaches: image processing, transfer learning, and NLP. However, we found that when we tried to train with both the image processing features and the transfer learning features the overall accuracy went down. This was because all the image processing features we came up with were already incorporated in the neural net, and so adding in the image processing features ended up decreasing the power of the pre-trained convolutional neural network. Therefore our best model combined the CNN and NLP features since they had the most divergent features and therefore best represent the span of features per image. The results of the CNN and NLP combinations are below.

TABLE I: Transfer Learning

Genres	TP	FN	FP	Precision	Recall
Business	157	43	63	71%	79%
Fantasy	111	89	104	52%	56%
History	121	78	61	66%	61%
Science-Fiction	92	108	89	51%	46%
Romance	131	69	70	65%	66%
Total	612	387	387	38.7%	61.3%

TABLE II: Natural Language Processing

Genres	TP	FN	FP	Precision	Recall
Business	152	45	37	80%	77%
Fantasy	88	108	107	45%	45%
History	117	79	67	64%	60%
Science-Fiction	87	109	116	43%	44%
Romance	125	72	86	59%	63%
Total	569	413	413	42.1%	57.9%

TABLE III: Transfer Learning and NLP

Genres	TP	FN	FP	Precision	Recall
Business	159	38	60	73%	81%
Fantasy	109	87	97	53%	56%
History	120	76	59	67%	61%
Science-Fiction	93	103	84	53%	47%
Romance	132	65	69	66%	67%
Total	613	369	369	37.6%	62.4%

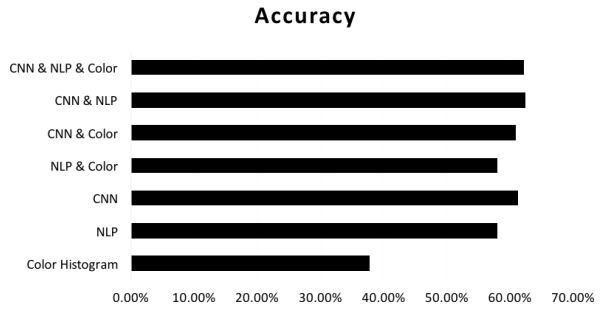


Fig. 7: Total Accuracy of Various Methods

Over all the genres, Science Fiction and Fantasy consistently had the worst accuracy, as shown in the tables I-III above. This can be explained by the similarity of the covers and titles of books in the two genres, so that many science fiction books are classified as fantasy and vice versa. There is some difficulty in that many books can be classified into more than one genre, and also were genres that we did not address in our classifier.

Along with SVM, we also tried to use softmax to train the model. After the training, the prediction from softmax had 57.3% accuracy with transfer learning features. This is lower than 62.4% result we got from SVM. Thus, we decided to use SVM in the final model.

VI. CONCLUSIONS

The convolutional neural network features performed the best out of the methods that were attempted due to the fact that the original ImageNet model was trained on many more images than any of our other methods. We were constrained from doing the same for the NLP and Image Processing techniques from of resources since training on such a set would take multiple GPUS over several weeks to train. Therefore given access to those resources we would like to train over the entire image database with our specific task to see if performance can be improved.

The OpenLibrary API JSON output also has related keywords for each book in the database. Given more time, it would be interesting to investigate whether the NLP portion of the project can be used on these words to provide more context for book predictions.

Ultimately we would like to see if the ability to identify of the book is directly correlated with its sales. That way this can help publishers and authors recognize which book cover type most attracts new readers.

REFERENCES

[1] Zujovic, Jana, et al. "Classifying paintings by artistic genre: An analysis of features & classifiers." Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on. IEEE, 2009.

[2] Yu, Honghai, and Stefan Winkler. "Image complexity and spatial information." Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on. IEEE, 2013.

[3] Razavian, Ali S., et al. "CNN features off-the-shelf: an astounding baseline for recognition." Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on. IEEE, 2014.

[4] Yosinski, Jason, et al. "How transferable are features in deep neural networks?." Advances in Neural Information Processing Systems. 2014.

[5] Goldberg, Yoav, and Omer Levy. "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method." arXiv preprint arXiv:1402.3722 (2014).

[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[7] Fan, Rong-En, et al. "LIBLINEAR: A library for large linear classification." The Journal of Machine Learning Research 9 (2008): 1871-1874.

[8] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." The Journal of Machine Learning Research 12 (2011): 2825-2830.

[9] Jia, Yangqing, et al. "Caffe: Convolutional architecture for fast feature embedding." Proceedings of the ACM International Conference on Multimedia. ACM, 2014.