Multi-Label Book Classification for Automatic Tag Generation

Isha Jain Kritika Bakshi Tania

Ashu Jain, Assistant Professor (IT Department)

IT Department, ADGITM, Shastri Park, New Delhi 110053

Abstract. Book covers communicate information to potential readers, but can that same information be learned by computers? We propose a novel approach to classify books given their cover images using an ensemble of different multi-label classification techniques, and hence generate the most appropriate tags for each of them. The purpose of this research is to investigate whether relationships between books and their covers can be learned. However, determining the genre of a book is a difficult task because covers can be ambiguous and genres can be overarching. Despite this, we show that a CNN can extract features and learn underlying design rules set by the designer to define a genre. Using deep learning, we can bring the large amount of resources available to the book cover design process. In addition, we present a new challenging dataset that can be used for many pattern recognition tasks

1. Introduction

"Don't judge a book by its cover" is a common English idiom meaning not to judge something by its outward appear- ance. Although, it still happens when a reader encounters a book. The cover of a book is often the first interaction and it creates an impression on the reader. It starts a conversation with a potential reader and begins to draw a story revealing the contents within. But, what does the book cover say? What are the clues that the book cover reveals? While the visual clues can communicate information to humans, we explore the possibility of using computers to learn about a book by its cover. Machine learning provides the ability to use a large amount of resources to the world of design. By bridging the gap between design and machine learning, we hope to use a large dataset to understand the secrets of visual design.

We propose a method deriving a relationship between book covers and their genre automatically. The goal is to determine if genre information can be learned based on the visual aspects of a cover created by the designer. This research can aid the design process by revealing underlying information, help promotion and sales processes by providing automatic genre suggestion, and be used in computer vision fields.

The difficulty of this task is that books come with a wide variety of book covers and styles, including nondescript and misleading covers. Unlike other object detection and classification tasks, genres are not concretely defined. Another problem is that there is a massive amount of books exist and it is not suitable for exhaustive search methods.

To tackle this task, we present the use of an artificial neural network. The concept of neural networks and neural coding is to use interconnected nodes to work together to capture information. Early neural network-like models such as multilayer perception learning were invented in the 1970s but fell out of

favor . More recently, artificial neural networks have been a focus of state-of-the-art research because of their successes in pattern recognition and machine learning. Their successes are in part due to the increase in data availability, increase in processing power, and introduction of GPUs .

Convolutional Neural Networks (CNN) , in particular, are multilayer neural networks that utilize learned convolutional kernels, or filters, as a method of feature extraction. The general idea is to use learned features rather than pre-designed features as the feature representation for image recognition. Recent deep CNNs combine multiple convolutional layers along with fully-connected layers. By increasing the depth of the network, higher level features can be learned and discriminative parts of the images are exaggerated . These deep CNNs have had successes in many fields including digit-recognition , and large-scale image recognition.

The contribution of this paper is to demonstrate that connections between book genres can be learned using only the cover images. To solve this task, we used the concept of transfer learning and developed a CNN based system for book cover genre classification. AlexNet pre-trained on ImageNet is adapted for the task of genre recognition. We also reveal the relationships automatically learned between genres and book covers.

Secondly, we created a large dataset containing 137,788 books in 32 classes made of book cover images, title text, author text, and category membership. This dataset is very challenging and can be used for a variety of tasks some of which include text recognition, font analysis, and genre predic- tion. Furthermore, although AlexNet pre-trained on ImageNet has already achieved state-of-the-art results on document clas- sification, we had a limited accuracy which indicates the high level of difficulty of the proposed dataset.

2. Literature Review

Brian Kenji Iwana , Syed Tahseen Raza Rizvi, Sheraz Ahmed , Andreas Dengel, “Multi-Label Book Classification using Automatic Tag Generation”. In this paper, the author has explained that Visual design is intentional and serves a purpose. It has a rich history and exploring the purposes of design has been extensively analyzed by designers but is a relatively new field in machine learning.
Techniques have been used to identify artistic styles and qualities of paintings and photographs ,etc used deep CNNs to learn and copy the artistic style of paintings. Similarly, the goal of this trial is to learn the stylistic qualities of the work, but we go beyond to learn the underlying meaning behind the style. However, most of these methods use designed features or features specific to the task. In a more general sense, document classification tackles a similar problem in that it classes documents into architectural categories. In particular, deep CNNs have been successful in document classification.

3. Convolutional Neural Network

Modern CNNs are made up of three components: con-
volutional layers, pooling layers, and fully-connected layers. The convolutional layers consist of feature maps produced by repeatedly applying filters across the input. The filters represent shared weights and are trained using backpropaga- tion. The feature maps resulting from the applied filters are down-sampled by a max pooling layer to reduce redundancy improving the computational time for future layers. Finally, the last few layers of a CNN are made up of fully-connected layers. These layers are given a vector representation of the images from a preceding pooling layer and continue like standard feedforward neural

networks.

A. *AlexNet*

The network used for our book cover classification is inspired from the work of Krizh. We used a pre-trained network on ImageNet . By pre-training AlexNet on a very large dataset such as ImageNet, its possible to take advantage of the learned features and transfer it to other applications. Initializing a network with transferred features has shown to improve generalization . To accomplish this, we remove the original softmax output layer for the 1,000- class classification of ImageNet and replace it with a 30- class softmax for the experiment. Subsequently, the training is continued using the pre-trained parameters as an initialization. The network architecture is as follows. The network consists of a total of eight layers, where the first five are convolutional layers followed by three fully-connected layers.

4. Experimental Results

A. DataSet Preparation :

The dataset was collected from the book cover images and genres listed by Amazon.com. The full dataset contains 137,788 unique book cover images in 32 classes as well as the title, author, and subcategories for each respective book. Each book's class is defined as the top categories under "Books" in the Amazon.com marketplace. However, for the experiment we refined the dataset into 30 classes of 1,900 books in each class. To equalize the number of books in each class, books were chosen at random to be included in the experiment. The two categories, "Gay & Lesbian" and "Education & Teaching," were not used for the experiment because they only contain 1,341 and 1,664 books respectively, thus not having enough representation in the dataset. Also, when the dataset was collected, each book was assigned to only a single category. If the book belonged to multiple categories, one was chosen at random. We randomized and split the dataset into 90% training set and 10% test set. No pruning of cover images and no class membership corrections were done. In addition, we resized all of the images to fit 227px by 227px by 3 color channels for the input of the AlexNet and 56px by 56px by 3 color channels for LeNet.

B. Evaluation

The pre-trained AlexNet with transfer learning resulted in a test set Top 1 classification accuracy of 24.7% , 33.1% for Top 2, and 40.3% for Top 3 which are 7.4, 5.0, and 4.0 times better than random chance respectively. As comparison, using the modified LeNet, we had a Top 1 accuracy of 13.5%, Top 2 accuracy of 21.4%, and Top 3 accuracy of 27.8%. The AlexNet performed much better on this dataset than the LeNet. Considering that CNN solutions are state-of-the-art for image and document recognition, the results show that classification of book cover designs is possible, although a very difficult task. Table I shows the individual Top 1 accuracies for each genre. In every class except "Christian Books & Bibles," the AlexNet performed better. For most cases, AlexNet had more than twice as good Top 1 accuracy compared to LeNet.

C. Analysis

In general, most cover images have either a strong activation toward a single class or are ambiguous and could be part of many classes at once. Figure 1 shows examples of books classified in the "Cookbooks, Food & Wine" category. When the cover contained an image of food, the CNN predicted the correct class and with a high probability. But, the covers with more ambiguous images resulted in a low confidence. The misclassified examples in Fig. 1 (b) failed for understandable reasons; the first two are ambiguous and can reasonably be classified as "Self-Help" and "Science & Math" respectively

The final example had a strong probability of being in "Comics & Graphic Novels" and "Children's

Books" because the cover image features an illustration of a vehicle. Many books contain misleading covers like these examples and correct classification would be difficult even for a human without reading the text. Figure 2 reveals another example of misleading cover images, but for the "Biographies & Memoirs" category. The difficulty of this category comes from a high rate of sharing qualities with other categories causing substantial ambiguity of the genre itself. A high number of misclassifications from the "Biographies & Memoirs" category went into "History." We also observed a similar relationship between "Comics & Graphic Novels" and "Children's Books" and between "Medical Books" and "Science & Math." This shows that the AlexNet network was able to automatically learn relationships between categories based solely on the cover images. The figure clearly shows the large central cluster of difficult covers as well as the confident correctly classified covers near each axis. For classes such as "Politics & Social Sciences" and "Christian Books & Bibles," the strong softmax responses are sparse and it is reflected in their very low recognition accuracy

5. Proposed Approach

Analysis of the results reveals that AlexNet was able to learn certain high-level features of each category. Some of these correlated features may be objects such as portraits for "Biographies & Memoirs" or food for "Cookbooks, Food & Wine." Other times it is colors, layout, or text. In this section, we explore the design principles that the CNN was able to automatically learn.

A. Color Matters

In the absence of distinguishable features, the CNN has to rely on color alone to classify covers. Because of this, many classes get associated to certain colors for books with limited features. Shown in Fig. 4, the AlexNet relates white to "Self- Help," yellow to "Religion & Spirituality," green to "Science & Math," blue to "Computers & Technology," red to "Medical Books," and black to "Biographies & Memoirs." Although, classifying simple book covers by color alone causes many misclassifications to occur.

The color association does not only restrict itself to simple book covers. Despite having active book covers, the tone of book covers were also important for classification. For exam- ple, "Cookbooks, Food & Wine" often features food and are commonly by shades of beige and tan. Likewise, there is a high representation of gardening books in the "Crafts, Hobbies & Home" class, therefore, green books are commonly classified in that genre. Also, the tone of the book can define the mood, so "Children's Books" commonly have designs with yellow or bright backgrounds and "Science Fiction & Fantasy" books usually have black or dark backgrounds. The AlexNet was able to successfully capture the mood of book genres by grouping books of certain moods to respective genres.

B. Objects Matter

The image on book covers is usually the thing that first attracts potential readers to a book. It should be no surprise that the object featured on the cover has an effect on how it gets classified. What is surprising about the results of our experiment is how the network is able to distinguish different genres but with common objects. For instance, featuring people on the cover is common among many genres, but the type of person or how the person is dressed determines how the book gets classified. Figure 6 shows four genres that centrally display humans, but have discriminating features that make the classes separable.

The structure and layout of the book cover also makes a difference in the classification. Books with rectangular title boards, no matter the color, tended to be classified as "Law" and books with a large landscape photographs tended to be "Travel". This trend continued to other categories, such as "Cookbooks, Food & Wine" with a central image of food stretching to the edges of the cover, "Biographies & Memoirs" featuring close-up shots of people, and reference and textbooks containing

solid color bands.

C. Text Matters

Another interesting design principle captured by the AlexNet is the text qualities and the font properties. The best example of this is "Mystery, Thriller & Suspense," shown in Fig. 8. Despite having a similar color pallet and image content to "Romance" and "Science Fiction & Fantasy," the common thread in many of the classified "Mystery, Thriller & Suspense" books was large overlaid sans serif text. Also shows that "Calendars" often de-emphasize the title text so the focus is on the cover image. On the other hand, the figure also shows that "Literature & Fiction" often uses expressive fonts to reveal messages about the book. The text style on the cover of a book affects the classification, revealing that relationships between text style and genre exist.

In particular, of the 30 classes, "Test Preparation" had the highest recognition rate at 68.9%, much higher than the overall accuracy. The reason behind this high accuracy is that "Test Preparation" book covers are often formulaic. They tend to have an acronym in large letters (e.g. "SAT," "GRE," "GMAT," etc.) near the top with horizontal or vertical stripes and possibly a small image of people. The large text is important because when compared to other non-fiction and reference classes, the presence of large acronyms is the most discrim- inating factor. Figure 9 shows books from other categories that were incorrectly classified as "Test Preparation." These examples follow the design rules similar to many other "Test Preparation" books, but the actual content of the text reveals the books as other classes.

6. Conclusion

In this paper, we presented the application of machine learn- ing to predict the genre of a book based on its cover image. We showed that it is possible to draw a relationship between book cover images and genre using automatic recognition. Using a CNN model, we categorized book covers into genres and the results of using AlexNet with transfer learning had an accuracy of 24.7% for Top 1, 33.1% for Top 2, and 40.3% for Top 3 in 30-class classification. The 5-layer LeNet had a lower accuracy of 13.5.7% for Top 1, 21.4% for Top 2, and 27.8% for Top 3. Using the pre-trained AlexNet had a dramatic effect on the accuracy compared to the LeNet.

However, classification of books based on the cover image is a difficult task. We revealed that many books have cover images with few visual features or ambiguous features causing for many incorrect predictions. While uncovering some of the design rules found by the CNN, we found that books can have also misleading covers. In addition, because books can be part of multiple genres, the CNN had a poor Top 1 performance. To overcome this, experiments can be done using multi-label classification.

Future research will be put into further analysis of the char- acteristics of the classifications and the features determined by the network in an attempt to design a network that is optimized for this task. Increasing the size of the network or tuning the hyperparameters may improve the performance. In addition, the book cover dataset we created can be used for other tasks as it contains other information such as title, author, and category hierarchy. Genre classification can also be done using supplemental information such as textual features alongside the cover images. We hope to design more robust models to better capture the essence of cover design.

7. References

[1] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[2] K. Chellapilla, S. Puri, and P. Simard, "High performance convolutional neural networks for document

processing," in *10th Int. Workshop Fron- tiers in Handwriting Recognition*. Suvisoft, 2006.
[3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278– 2324, 1998.
[4] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolu- tional networks," in *2014 European Conf. Comput. Vision*. Springer, 2014, pp. 818–833.
[5] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE Conf. Comput. Vision and Pattern Recognition*. IEEE, 2012, pp. 3642–3649.
[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comp. Vision and Pattern Recognition*, 2015, pp. 1–9.
[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Inform. Process. Syst.*, 2012, pp. 1097–1105.
[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *2012 IEEE Conf. Comput. Vision and Patern Recognition*. IEEE, 2009, pp. 248–255.
[10] M. Z. Afzal, S. Capobianco, M. I. Malik, S. Marinai, T. M. Breuel, A. Dengel, and M. Liwicki, "Deepdocclassifier: Document classification with deep convolutional neural network," in *Int. Conf. Document Anal. and Recognition*. IEEE, 2015, pp. 1111–1115.