Can you Judge a Book by its Cover?

Sigtryggur Kjartansson & Alexander Ashavsky
CS 231N Convolutional Neural Networks, Stanford University

Motivation & Background

- People should never judge a book by its cover, but what about machines?
- The cover is a reader's first impression with the book.
- Important for marketing and categorization.

Problem Statement

Goal: Determine a book's genre by its cover and title.
Benchmark: Top 1-3 Accuracy from Kenji et al., 2016.
Approach: Ensembles of image and text classifiers.

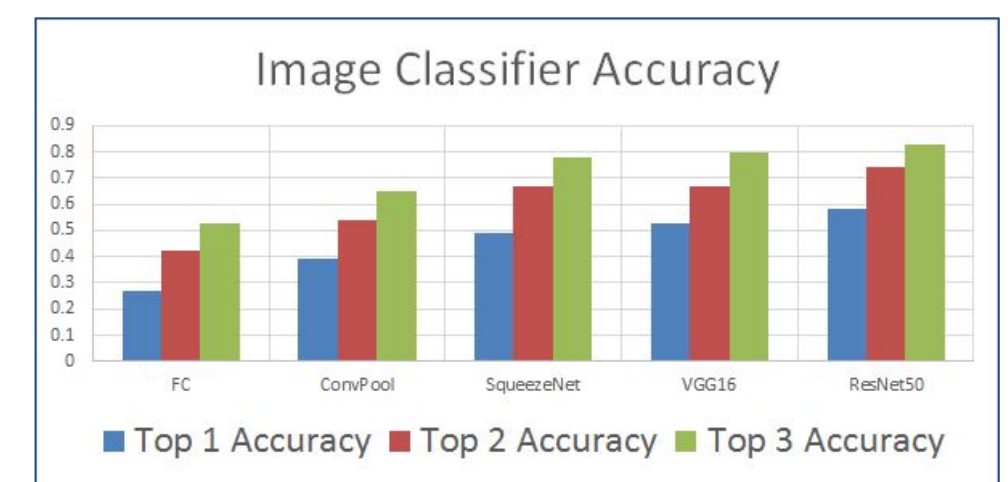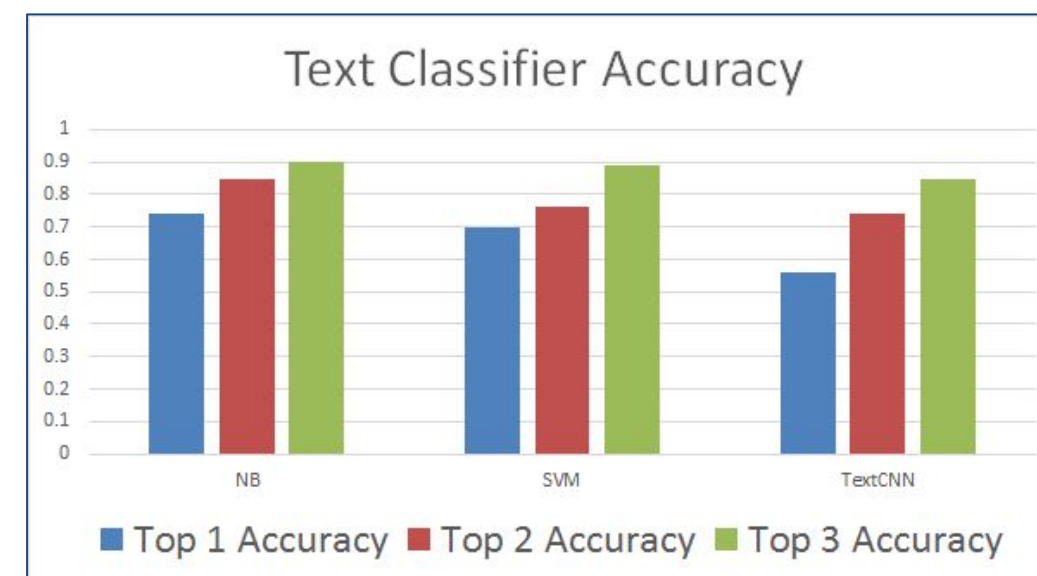The Official Guide to the GRE → Classifier → *Test Preparation*
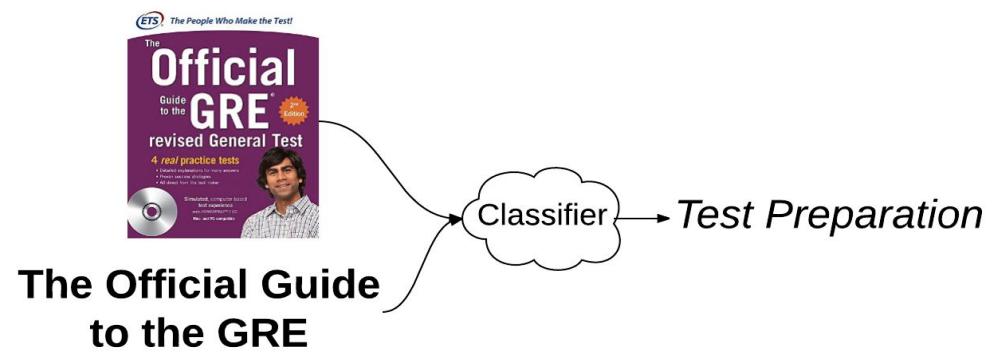
Datasets

- Amazon Marketplace book cover images and titles, with human-curated genres.
- Subsample 19,000 of these samples evenly across 10 genres:
  - *Train*: 15,391, *Dev*: 1,710, *Test*: 1,900
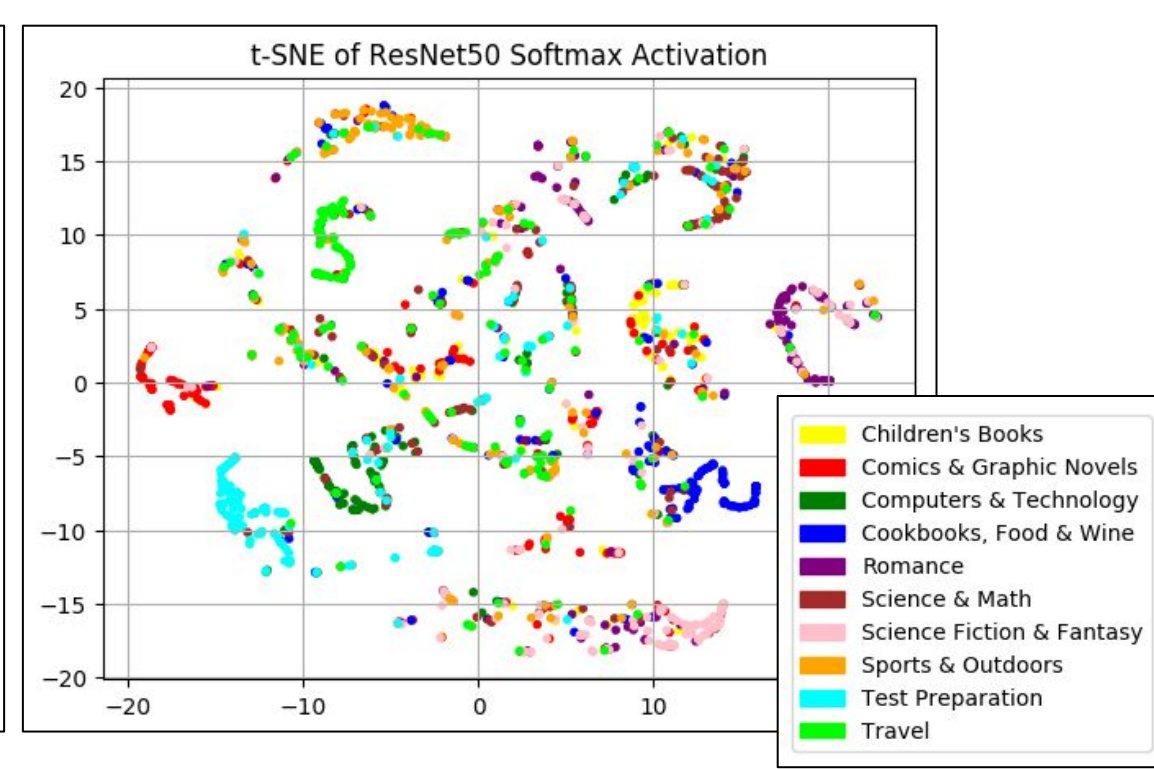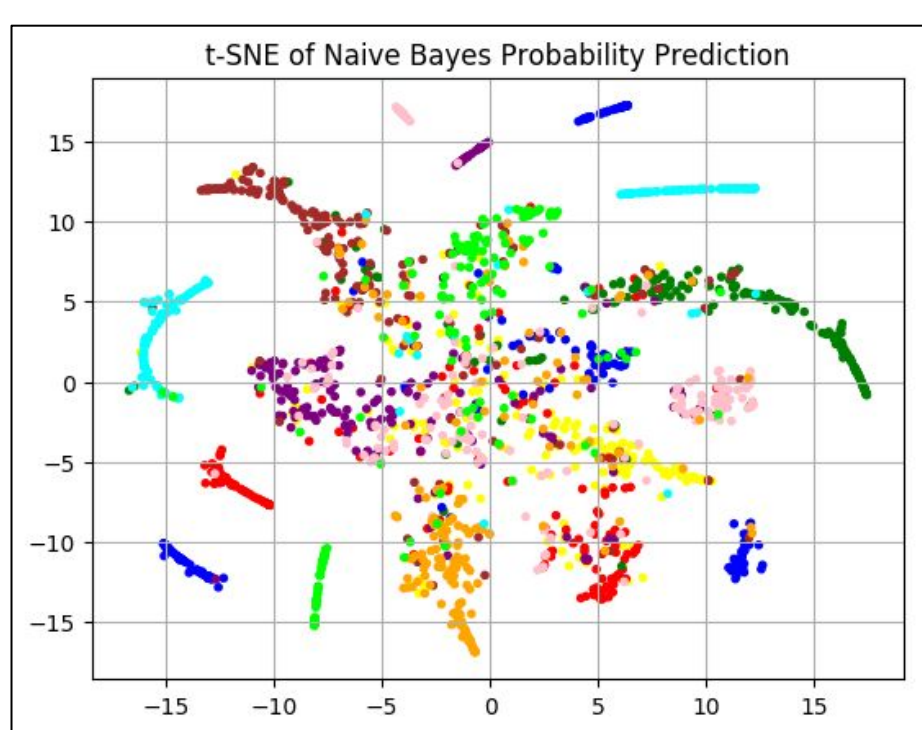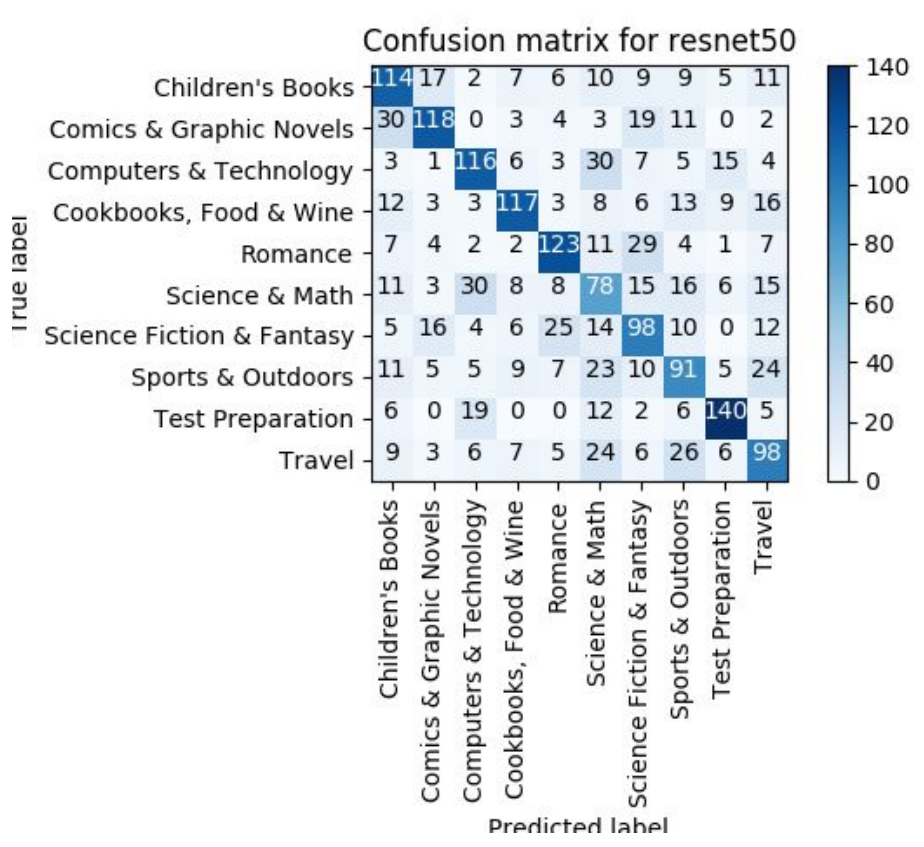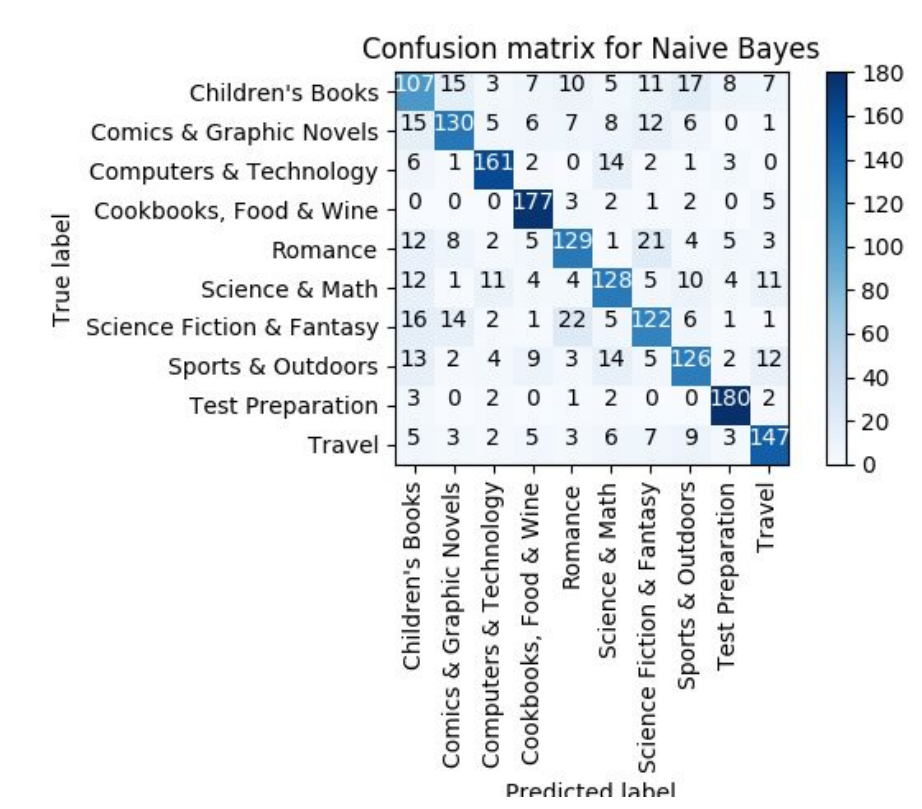- Total words: 168,498
- Total unique words: 18,555

Word Zipf Plot

Histogram of Title Length
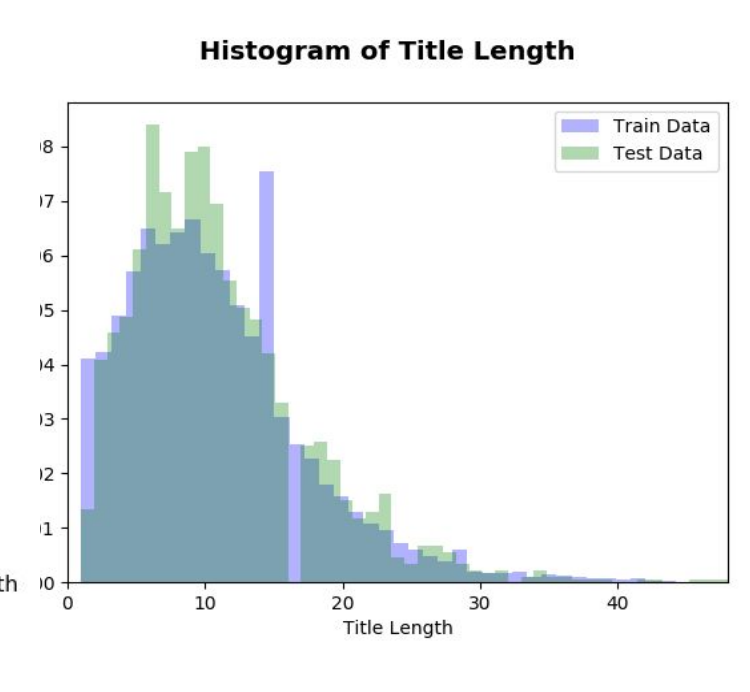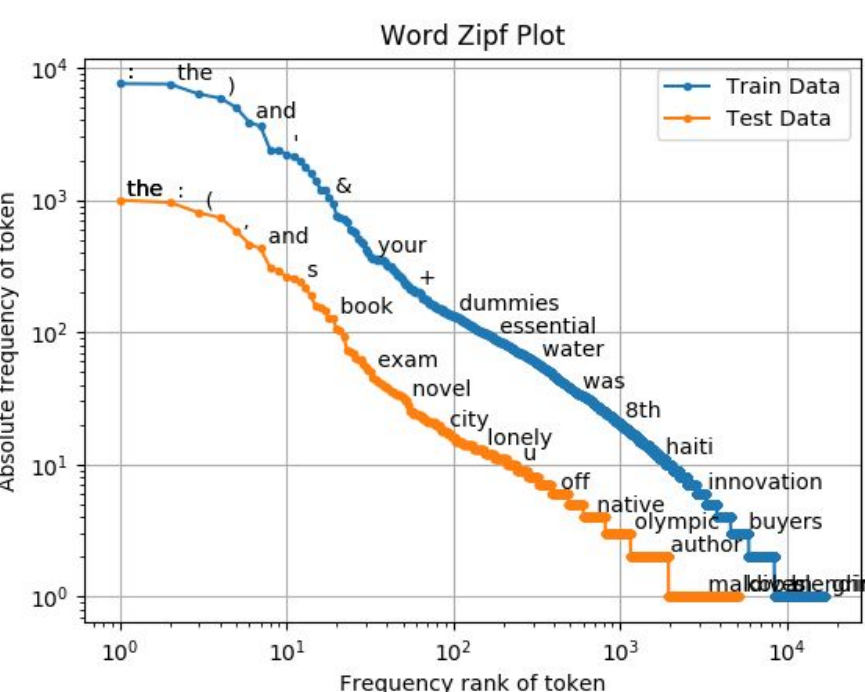
Experimental Evaluation & Results

Text Classifier Accuracy

Image Classifier Accuracy

Naive Bayes Results by Genre

Genre	Top1	Top2	Top3
Children's Books	0.53	0.71	0.81
Comics & Graphic Novels	0.68	0.79	0.89
Computers & Technology	0.85	0.93	0.97
Cookbooks Food & Wine	0.93	0.98	0.99
Romance	0.68	0.83	0.86
Science & Math	0.67	0.79	0.85
Science Fiction & Fantasy	0.64	0.83	0.86
Sports & Outdoors	0.66	0.84	0.90
Test Preparation	0.95	0.98	0.99
Travel	0.77	0.85	0.91

ResNet50 Results by Genre

Genre	Top1	Top2	Top3
Children's Books	0.60	0.76	0.84
Comics & Graphic Novels	0.62	0.78	0.88
Computers & Technology	0.61	0.75	0.83
Cookbooks Food & Wine	0.62	0.74	0.84
Romance	0.65	0.79	0.86
Science & Math	0.41	0.62	0.77
Science Fiction & Fantasy	0.52	0.75	0.82
Sports & Outdoors	0.48	0.68	0.81
Test Preparation	0.74	0.83	0.91
Travel	0.52	0.67	0.77

Confusion matrix for Naive Bayes

Confusion matrix for resnet50

t-SNE of Naive Bayes Probability Prediction

t-SNE of ResNet50 Softmax Activation

Methods

We leveraged the following models for our task:

Image Classification:
- Fully Connected Net
- Conv-Pool Net
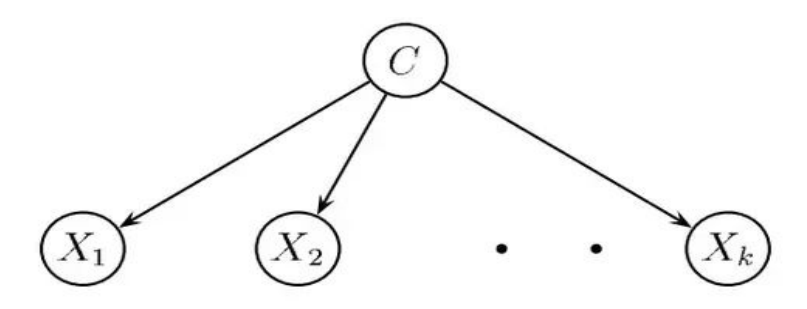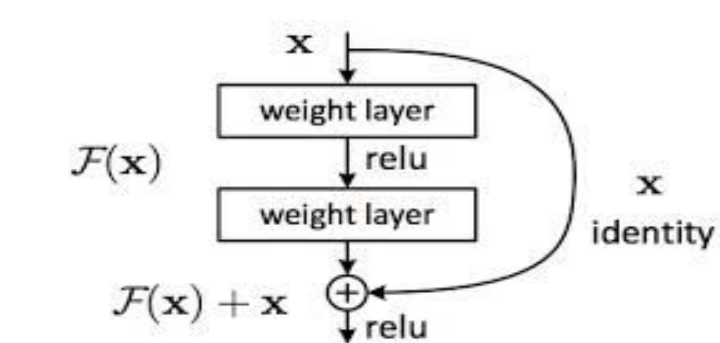- SqueezeNet*
- VGG-16*
- ResNet-50*

Text Classification:
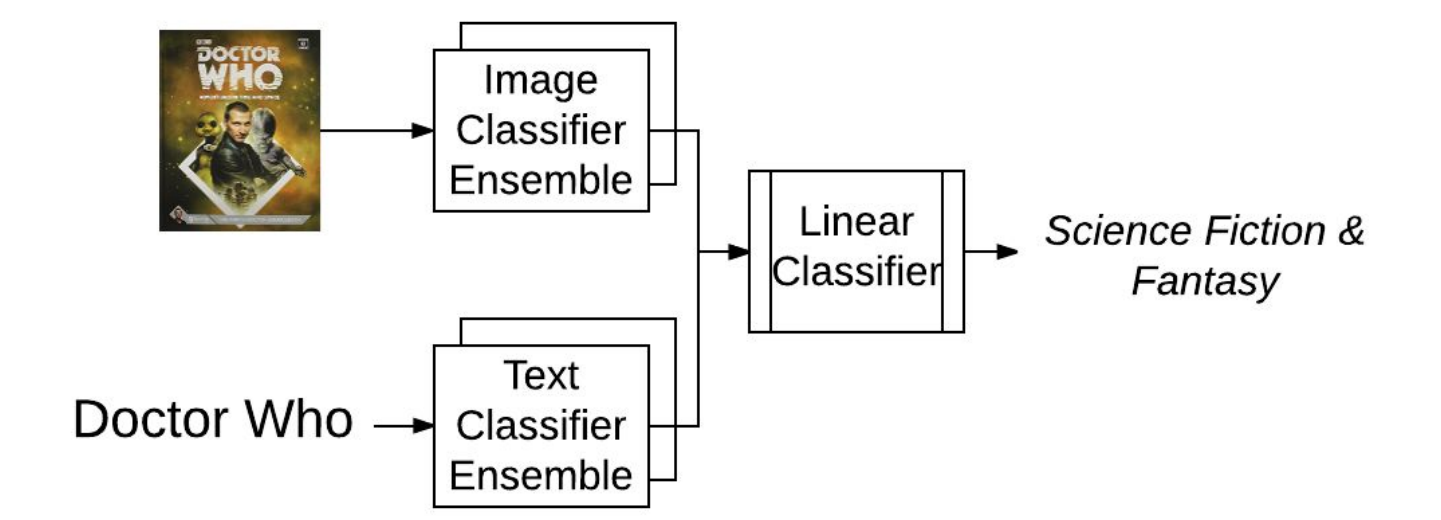- SVM
- Naive Bayes
- Text CNN

Naive Bayes Network

$$p(x_i \mid x_{i+1}, \ldots, x_n, C_k) = p(x_i \mid C_k)$$

Residual Unit

$\mathcal{F}(\mathbf{x})$ ... weight layer ... relu ... weight layer ... $\mathbf{x}$ identity ... $\mathcal{F}(\mathbf{x}) + \mathbf{x}$ ... relu

- Train ensembles of best image & text classifiers.
- Combine with a linear classifier.

Doctor Who → Image Classifier Ensemble / Text Classifier Ensemble → Linear Classifier → *Science Fiction & Fantasy*

* SqueezeNet, VGG-16, and ResNet-50 were all pre-trained on ImageNet.

Conclusion & Future Direction

- There is a relationship between book cover artwork, title, and genre.
- Image is a strong indicator, text is even stronger.
- Naive Bayes is the best text classifier.
- Future: Use an attention mechanism to ignore red herrings in artwork.
- Future: Extract title from image.
- Future: Analyze features determined by the network

[1] Kenji et al., 2016; arXiv:1610.09204
[2] He et al., 2015; arXiv:1512.03385