Classify A to Z

Jason Jung, Kate Larsen, Ethen Liu, Santiago Olivar | MSiA 490-30 Deep Learning | Spring 2017 | Northwestern University

Problem Statement

- **Goal**: Given the product's text description and image from Amazon.com, classify its category
- **Audience**: Useful for companies that sell products online and wish to classify or organize them. Unlike large companies like Amazon, who can afford to use expensive software like IBM E-Commerce to organize product information, our software can be used by smaller companies
- **Challenge**: Each category includes hundreds of subcategories which can be very different from one another. Thus, the model has to learn the abstract concept of categories
- **Similar works**: Students in Stanford University classified the similar dataset except with just text descriptions by using algorithms like Naive Bayes and tree classifiers that produced 86% accuracy

Dataset

- **Raw Data & Features Extracted**:
  - Downloaded in json format from a publicly available dataset that labeled categories for different Amazon products
  - Extracted text fields of title and descriptions and URL to download the images

LaCroix Sparkling Water, Orange, 12-Ounce Cans (Pack of 12)

- **Challenges**:
  - Encountered few preprocessing issues due to clean json format
  - Found some images were misclassified or overlapped categories
  - Removed the Sports & Outdoors category due to troublesome images even for human accuracy
- **Data Size**:
  - Balanced the classes to 35,000 per category leading to 140,000 observations in total

References and Related Work

- An Analysis Of Deep Neural Network Models For Practical Applications: https://arxiv.org/pdf/1605.07678.pdf
- Dataset: http://jmcauley.ucsd.edu/data/amazon/
- Going deeper with convolutions: https://arxiv.org/pdf/1409.4842.pdf
- LIME: https://github.com/marcotcr/lime

Technical Approach

Training

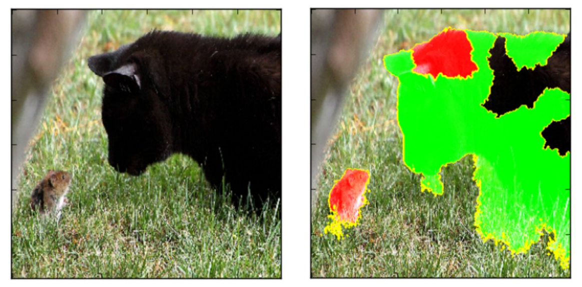
Text	Image
• Stemmed words • TF-IDF • Trained machine learning models: ○ SVM ○ Naive Bayes ○ Logistic Regression ○ Random Forest	• Utilized Keras' Inception network model that was pre-trained on the Imagenet dataset to perform transfer learning (more details in the section below)

Ensemble

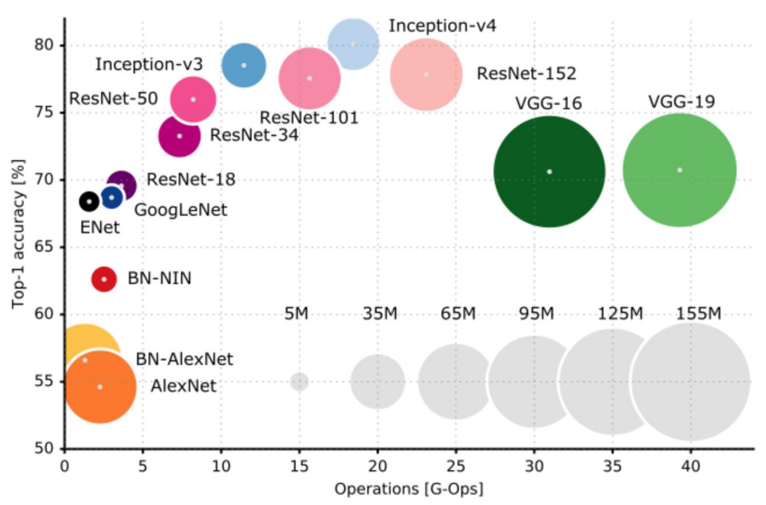
- Combined the image and text models' outputs by using a weighted average of their prediction

Explanation

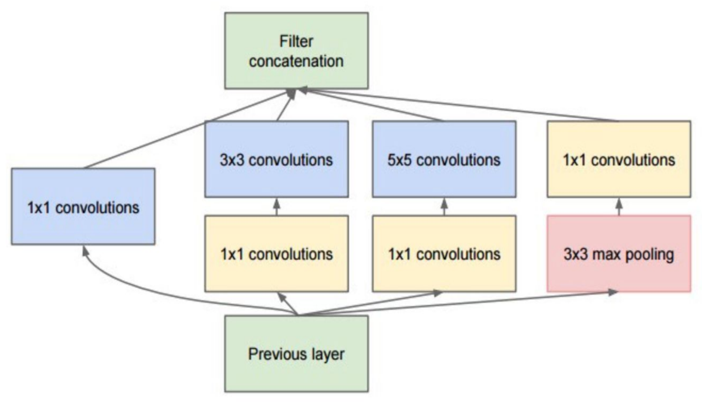
- Used LIME technique (Local Interpretable Model agnostic Explanations) to explain the predictions of the classifier



- LIME shows the model focused on the cat (green) rather than the mouse or background (red)



Network Architecture Comparison



Inception Module

Methodology Research

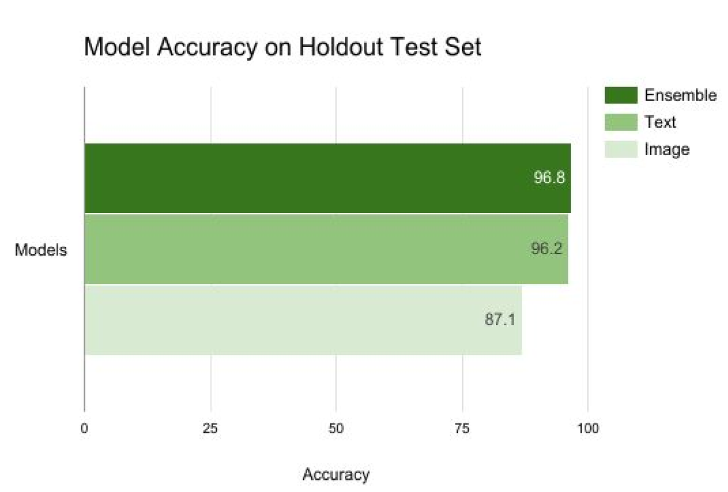
Left: State-of-art Neural Network Architecture

- The x-axis shows the number of network parameters/operations and the y-axis shows network performance on the Imagenet challenge
- This shows that Inception-v3 network strikes a good balance between model performance and complexity

Right: Inception Module

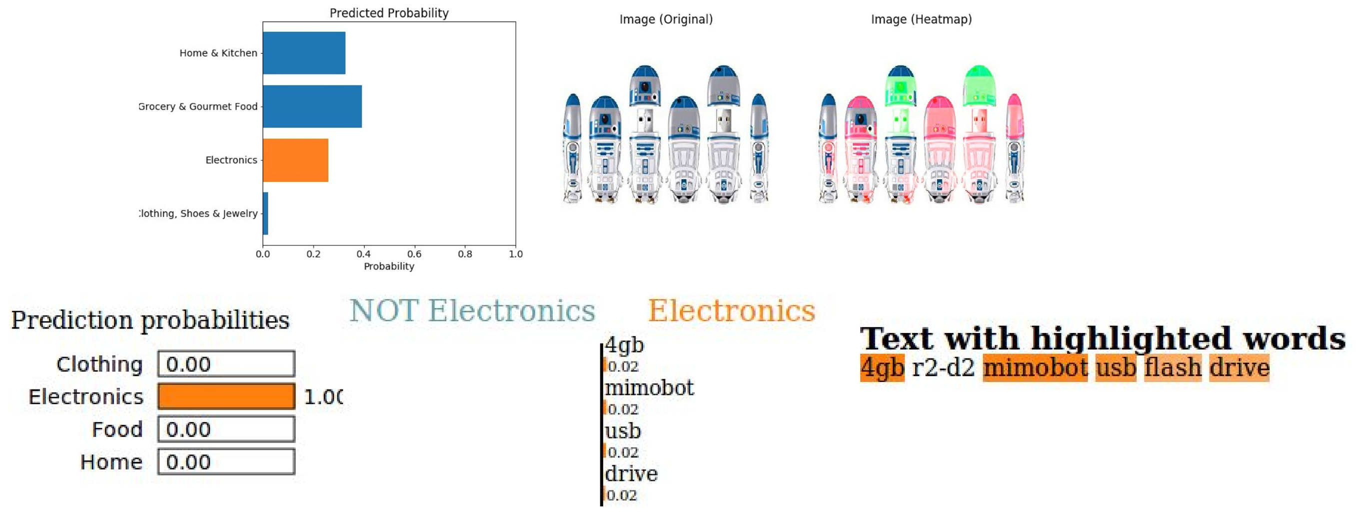
- The Inception network's success was due to the Inception module
- At a high level, because input features are correlated, redundancy can be removed by combining them appropriately with the 1x1 convolutions before feeding them to more expensive 3x3 or 5x5 convolutions

Results

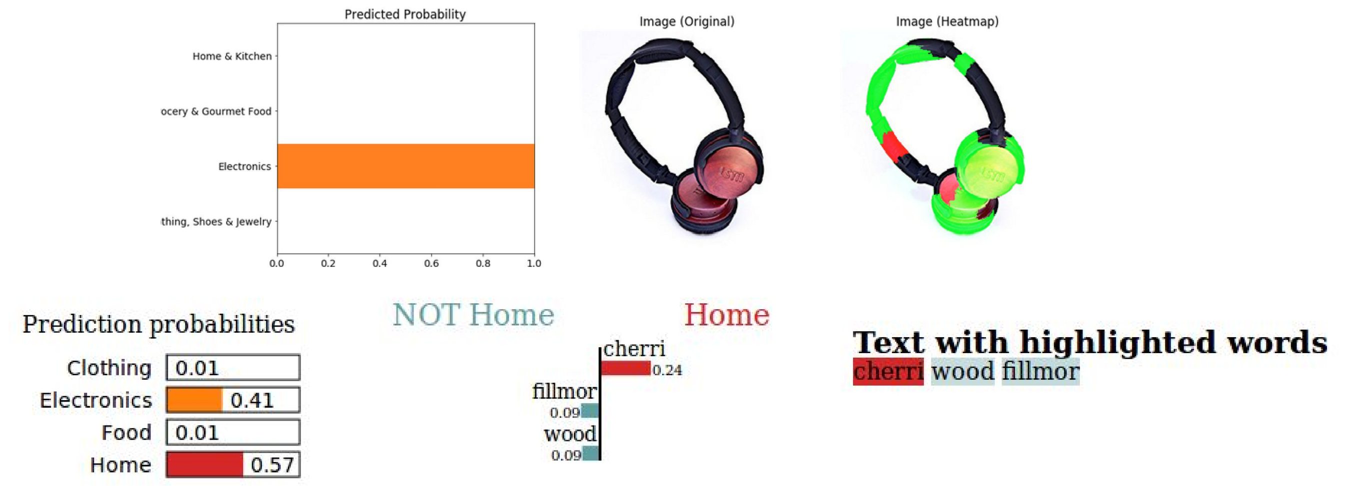


The accuracy measure on a hold out test set for the text model was 96.2% and 87.1% for the image model, while an ensemble (0.7/0.3 weight for text/image) resulted in a slight improvement of 96.8%

Strength of Text Model: The R2D2 usb drive below shows how the text model correctly classifies the product focusing on the orange words below, while the image model performs poorly by focusing on the green parts of the image



Strength of Image Model: The headphones below show how the image model performs well by focusing on earpieces, while the text model performs poorly by focusing on the stemmed word "cherri."



Conclusion

- Text alone is very powerful for predicting item category, however performance is best achieved when ensembling the text and image model with a 70% - 30% respective weight
- One limitation of our approach was that we structured our model to classify at a high level of category. Thus, a possible improvement to the project includes classifying at a more granular level such as classifying TV, cell phone as oppose to Electronics
- LIME is handy for visual explanation of black-box model's decision