

CS 512: Advanced Machine Learning

Project Final Report

April 2023

Venkata Pratheek Reddy Chavva - vchavv2

We chose VGG16, VGG19, and RESNET50 as our three pre-trained CNN models for the project. These models are the most suitable for our project as they cater to the specific requirements and differences of our project. VGG16 is preferred for its strong ability to extract meaningful features from images through its deep architecture, which consists of 16 layers. VGG16 has shown to achieve high accuracy on various benchmark datasets. VGG19 is a variant of the VGG16 CNN model. The additional layers in VGG19 enable it to learn more complex features from images, resulting in better accuracy and performance compared to VGG16 in some scenarios. ResNet50's deep architecture and skip connections enable it to effectively identify patterns and classify images into various categories, making it a robust model for image classification tasks.

BOOK GENRE PREDICTION USING CNN AND TRANSFER LEARNING - Venkata Pratheek Reddy Chavva

Book covers communicate information to potential readers, but can that same information be learned by computers? The goal is to determine if genre information can be learned based on the visual aspects of a cover created by the designer. Books come with a wide variety of book covers and styles, including nondescript and misleading covers.

To tackle this problem I used the concept of transfer learning and developed a Convolutional Neural Networks (CNN) based system for book cover genre classification. So, I used a pre-trained network on ImageNet. By pre-training VGG16, VGG19 and ResNet50 models on a very large dataset such as ImageNet, it's possible to take advantage of the learned features and transfer it to other applications.

The goal is to determine if genre information can be learned based on the visual aspects of a cover created by the designer. The difficulty of this task is that books come with a wide variety of book covers and styles. Unlike other object detection and classification tasks, genres are not concretely defined.

Data Preprocessing

I've extracted over 50,000 images from [amazon.com](https://www.amazon.com) with over 30 classes. But due to runtime constraints, I have selected only 5 categories which had more number of images. For increasing training and validation images, I used image augmentation techniques like random rotation and random horizontal flips. I have performed center cropping on the dataset and reshaped them to 224x224. This is the size the models expect. Images that are larger were truncated and smaller ones were interpolated. Also, I normalized the RGB channel values of images.

Training and Evaluation

We've trained the pre-trained model on the Book cover dataset by freezing all the convolutional layers. We added a custom classifier. We trained a classifier consisting of the following layers

1. Fully connected with ReLU activation (n_inputs, 256)
2. Dropout with 40% chance of dropping
3. Fully connected with log softmax output (256, n_classes)

To build our custom classifier, we use the `nn.Sequential()` module which allows us to specify each layer one after the other. We assign our custom classifier to the final classifier layer in the already trained network. When we add on the extra layers, they are set to `require_grad=True` by default. These will be the only layers that are trained.

Below are the training and validation loss for each model:



VGG16



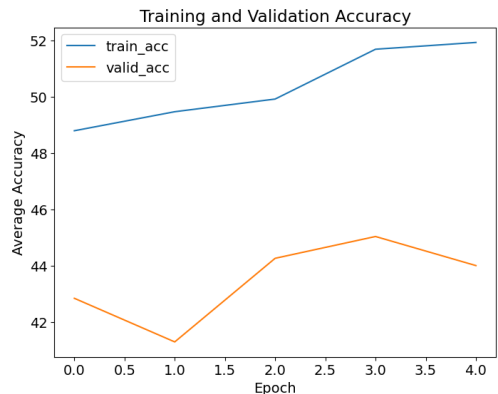
VGG19



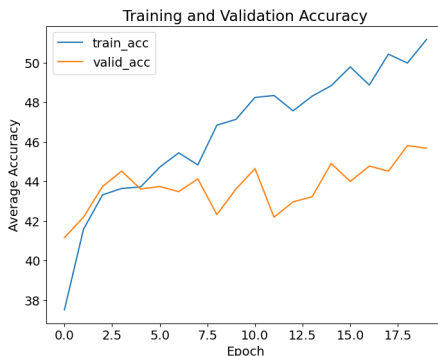
ResNet50

Training loss has reduced as the epochs increased for all the models. But, validation loss slightly increased as the number of epochs increased. This means the models have overfit the training data. Since ResNet50 is a complex model compared to VGG19 and VGG16, overfitting is more for ResNet50. Complex models tend to easily overfit for the given data.

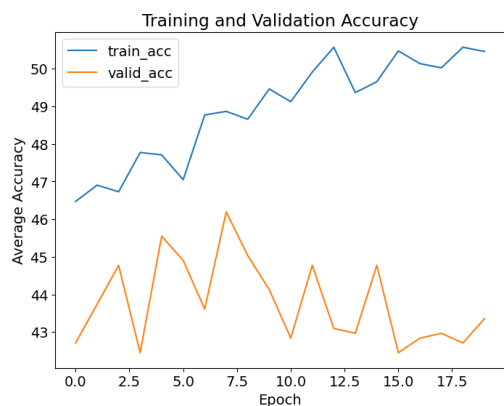
Below are the Training and validation accuracies for each model:



VGG16



VGG19



ResNet50

Similar to the losses, ResNet50 has less accuracies compared to simpler VGG models. VGG19 has the highest validation accuracy compared to other models.

Results

Test accuracy of the three models are:

Model	Accuracy
VGG16	37.5
VGG19	50
ResNet50	47.3

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. Also, some of the book covers had plain backgrounds with text of the title which is very basic for the models to train on. Hence, the accuracy of the models was not as high as anticipated.

References

Iwana, Brian & Rizvi, Syed Tahseen Raza & Ahmed, Sheraz & Dengel, Andreas & Uchida, Seiichi. (2016). Judging a Book By its Cover.

L. Yang and A. Shami, "A Transfer Learning and Optimized CNN Based Intrusion Detection System for Internet of Vehicles," ICC 2022 - IEEE International Conference on Communications, 2022, pp. 2774-2779, doi: 10.1109/ICC45855.2022.9838780.

Derrick Liu, Yushi Wang. (2017) Image Classification of Vehicle Make and Model Using Convolutional Neural Networks and Transfer Learning

3D Object Representations for Fine-Grained Categorization, Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei, 4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013 (3dRR-13). Sydney, Australia. Dec. 8, 2013.