CS 230
2018 Winter

Transfer Learning-based CNN Classification for Simpsons Characters

Yueheng Li yueheng@stanford.edu **Youtube**: https://www.youtube.com/watch?v=JSE-100dfjI

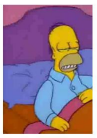
Stanford

Predicting

- The simpsons gains popularity around the world. It's interesting and important for the animation industry and fans to classify certain characters.
- Built a CNN with transfer learning by using VGG16. The input of the algorithm was 10 simpsons characters (10 classes) images, and the output was the classification of the simpsons characters.
- The model achieved 76% precision, 73% recall and 74% f1 score on average for test data.

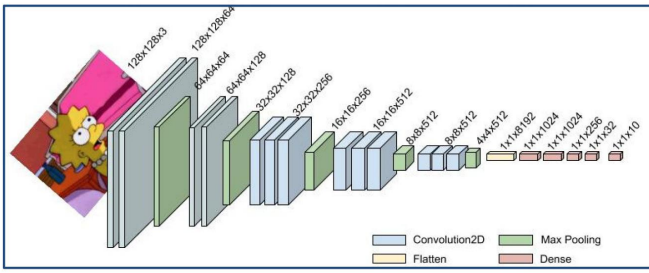
Dataset and Features

- The dataset were directly taken and labled from TV show episodes and downloaded from Kaggle. Data were labeled with ground truth. It consisted of 10 simpsons characters (10 classes). Each character had around 1000 to 2000 images.
- 10816 training images, 1999 dev images and 996 test images. Image resolution was 128x128.
- Data were preprocessed: 1) Characters with less than 1000 images were eliminated from the original dataset and each character has 1000 to 2000 images. 2) Data augmentation was used for optimization, including flipping, cropping, rotating, and etc.


Related Work

The simpsons character classification has mostly been attempted by simpsons fans on Kaggle. Methods used including training CNN from scratch, which didn't take advantage of transfer learning and the accuracy was relatively low; training using support vector machine, which has simple implementation but seems to not work very well. Other methods used were logistic regression, k neighbors classifier, decision tree classifier and etc.

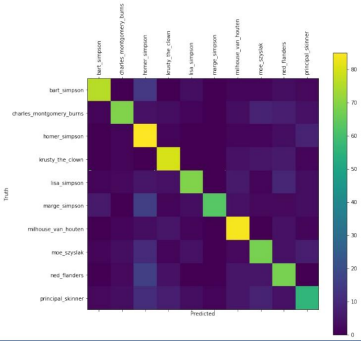


Model

- As show on the left the model used consisted of pretrained VGG16 convolution and max pooling layers, and 5 fully connected layers of neural size 1024, 1024, 256, 32 and 10 were attached after the flatten layer of VGG16. The size of the each layer is shown on the left.
- Cross entropy was used as the loss function.
- Accuracy of dev dataset was used to evaluate the training as the evaluation metric.

Results

character	precision	recall	f1-score	support
bart simpson	0.87	0.78	0.82	100
charles montgomery burns	0.84	0.75	0.79	100
homer simpson	0.52	0.86	0.65	100
krusty the clown	0.82	0.77	0.79	97
lisa simpson	0.79	0.71	0.75	100
marge simpson	1.00	0.64	0.78	100
milhouse van houten	0.76	0.83	0.79	99
moe szyslak	0.67	0.68	0.68	100
ned flanders	0.63	0.68	0.65	100
principal skinner	0.69	0.62	0.65	100
avg / total	0.76	0.73	0.74	996



Discussion

- Iteratively improved the results by 1) investigating different architectures 2) tuning hyperparameters
- As shown above, for test data, the model achieved 76% precision, 73% recall and 74% f1 score on average. Marge simpson has the highest precision 100% and Homer simpson has the lowest presision 52%. The confusion matrix showed the model performs well in general for all 10 characters.

Future

- Change the model architecture to train more characters, especially for characters with very few images
- Further increase model performance by tuning hyperparameters

Reference

- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. (2017) ImageNet classification with deep convolutional neural networks.
- James Bergstra and Yoshua Bengio. (2012) Random search for hyper-parameter optimization.