OscarNet: Using Transfer Learning to Classify Disposable Waste

Sam Kennedy, Tom Kennedy and John Knowles
[sjkenned, tkenned9, jkn0wles] @ stanford.edu

INTRODUCTION

Our project attempted to solve the problem of waste sorting, to aid in discarding personal trash.

Our team implemented a CNN classifier for sorting images into seven categories, spanning recycling, trash and non-waste.

Our model used transfer learning from the VGG-19 network trained on ImageNet. We trained a small dense neural network with a Softmax classifier.

The Facts

$30.83 spent training on Amazon

99.8% final training accuracy

1 hidden layer

PRE-TRAINING

Augmentation

Due to the small size of our dataset, we implemented preprocessing and augmentation through the Keras ImageDataGenerator. We applied shear, rotation, zoom, and shifts to the original dataset (Figure 4).

Figure 4

Hyperparameter Search

We tested:

- Batch size
- NN architecture
- Dropout rate
- Learning rate (Adam)
- Epsilon (Adam)

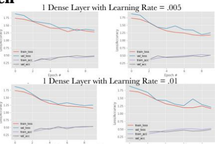
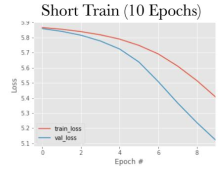


Figure 5

RESULTS

Short Train (10 Epochs)

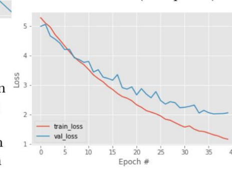


	Training Loss	Validation Loss	Training Accuracy	Validation Accuracy	AUC
15 Epochs	5.2889	4.9947	50.08%	49.68%	14
65 Epochs	1.1599	1.571	98.62%	81.8%	18.5
163 Epochs	0.8578	1.4382	99.78%	88.42%	19

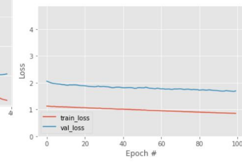
Training

We trained for 150 Epochs on batch sizes of 32 using Keras and Tensorflow. We saw consistent improvements with accuracy and loss across both training and validation sets.

Medium Train (50 Epochs)



Long Train (100 Epochs)



DATASET

Overview

We had 2415 total examples, spanning seven classes: paper, plastic, glass, cardboard, metal, trash, non-waste.

We chose to use 93.5% of our data for training, and 6.5% for validation.

Figure 1

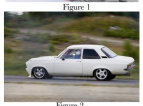


Figure 2

Sources

We used data from 3 sources, all of close up images of objects, mainly taken with Apple iPhone 6/7 cameras.

Images of waste from CS229 project by Gary Thung and Mindy Yang [2]

Objects dataset from PASCAL VOC 2012 [3]

Flowers dataset by Visual Geometry Group at the University of Oxford

Figure 3

NETWORK OVERVIEW

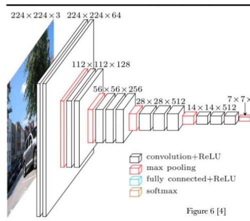


Figure 6 [4]

VGG19 (minus FC Layers)

The VGG-19 architecture is an extra-deep network with fourteen convolutional layers, divided into conv-blocks each with max-pooling layers (Figure 6). We use all layers up to the last max-pooling layer, and use the features produced from the last convolution as input to our transfer network. We retrained the last two conv-blocks in order to optimize the convolved features to our dataset.

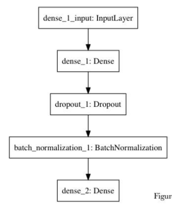


Figure 7

Transfer Network

Our network is comprised of four layers: a dense layer with ReLU activation, a dropout layer, a BatchNorm layer, and a dense output layer with a dense activation (Figure 7).

Loss Categorical Cross-entropy

Optimization Adam

FUTURE

We have four areas to explore with future work:

- Reduce variance by further regularization techniques
- Aquire more data by embedding network in mobile app
- Tighten focus to separation of recycling materials
- Create "ensemble" network using multiple CNN models feeding into a single transfer network

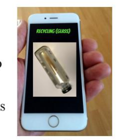


Figure 8

REFERENCES

[1] Simonyan, and Zisserman. "Very Deep Convolutional Networks for Large-scale image recognition." arXiv:1409.1556 (2014).

[2] Thung, and Yang. "Classification of Trash for Recyclability Status." CS229 Project Report 2016 (2016).

[3] Everingham, Gool, Williams, Winn, and Zisserman. "The Pascal Visual Object Classes Challenge" International Journal of Computer Vision, 303–338, 2010.

[4] https://blog.heuritech.com/2016/02/29/a-brief-report-of-the-heuritech-deep-learning-meetup-5/3