Deformable Convolutional Networks for Sketch Classification

Natasha Goenawan, Mandy Lu

Introduction

Humans have used sketches to express and record their ideas since prehistoric times. These sketches tend to be less detailed and thus harder to distinguish than photographs, causing traditional natural image classification techniques to underperform. Furthermore, while people can generally agree on what anobject should look like in a basic sense, how they ultimately visualize it can vary. Thus, a key challenge in sketch recognition is how to accommodate geometric variations and transformations in object scale, pose, viewpoint, and part deformation. Deformable convolutional layers are able to learn the offset of parts and may capture the primary features that humans remember about objects and enhance model performance.

Problem Statement

Inspired by deformable convolutions , we propose a multiclass classification model for sketches by incorporating a deformable convolutional layer on top of a CNN architecture with pool, dense, and dropout layers.

Dataset

TU-Berlin sketch dataset

Contains 250 object classes in the TU-Berlin dataset including: airplane, alarm clock, etc. On each image, objects are represented by their ground truth class labels along with index in the file name. We preprocessed the data into 64x64 single channel images.

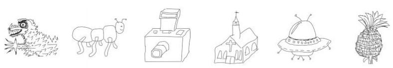


Figure 1. sample dataset sketches

We will use the TU-Berlin sketch dataset to evaluate our models, which has 16000 images in the training set, 2000 images in the dev set, and 2000 images in the test set. We used accuracy and confusion loss metrics to evaluate classification performance. To get a more intuitive understanding of the performance of our model and its refining effect on bounding box regression, we also visualized the confusion matrices at each iterative step on the original images.

Model

Baseline

Images fed through 3 layers of CNNs with 2 fully connected layers, and softmax activation.

Current Model

Images fed through 4 layers of CNNs with max pooling, a deformable convolutional layer, 2 dropout layers, and softmax activation.

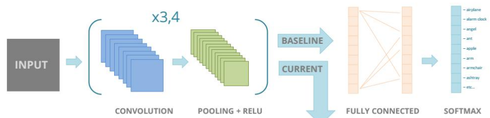
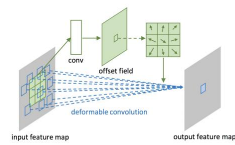
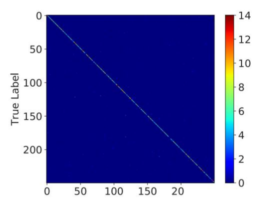


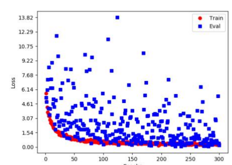
Figure 2.
Deformable convolutional layers learn the offset of parts and may capture the primary features that humans remember about objects and enhance model performance.



Evaluation

We are analyzing validation results on sketch classification during training time, taking the percentage of correctly classified sketches. However, we have found that validation accuracy is not a good metric alone as the model exhibits the imbalanced class problem, which is common in multi-class classification problems. To combat this issue, we have also incorporated the confusion matrix error in our loss function.





Results

Table 1.
Performance comparsion of our model tooth-er models.

MODEL	ACCURACY
SIFT-varient +BoF + SVM[1]	56 %
IDM + SVM[3]	71.30 %
ConvNet[4]	75.42 %
ConvNet[5]	77.69 %
Baseline **-** ConvNet	49 %
Current **-** DeformConvNet	56 %

Future

We anticipated higher accuracies with the deformable convolutional layer, but we are also working with a 250-class classifier, so our baseline accuracies start low. We have yet to incorporate any pre-trained CNN models as some of our inspiration papers did. Additionally, we will train on a larger data set by using the augmented Google Sketch dataset (~2 million. Generally, we found it very difficult to avoid overfitting with unseen people. This project is easily extendible and and poses an interesting and relatively unexplored problem that can yield insights into how humans perceive, categorize, and represent objects.

Acknowledgements

[1] Eitz, M., Hays, J., & Alexa, M. (2012). How do humans sketch objects? ACM Transactions on Graphics,31(4), 1-10. doi:10.1145/2185520.2335395

[2] Felzenszwalb, P., Girshick, R., McAllester, D., \& Ramanan, D. (2010). Object detection with discriminatively trained part based models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), 1627-1645.

[3] Yesilbek, K. T., Sen, C., Cakmak, S., & Sezgin, T. M. (2015). SVM for sketch recognition: Which hyperparameter interval to try? 2015 23nd Signal Processing and Communications Applications Conference (SIU). doi:10.1109/siu.2015.7129986

[4] Seddati, O., Dupont, S., & Mahmoudi, S. (2015). DeepSketch: Deep convolutional neural networks for sketch recognition and similarity search. 2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI). doi:10.1109/cbmi.2015.7153606

[5] Seddati, O., Dupont, S., & Mahmoudi, S. (2016). DeepSketch2: Deep convolutional neural networks for partial sketch recognition. 2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI). doi:10.1109/cbmi.2015.7153606