Freehand Sketch Recognition Using ResNet

Wayne Lu & Elizabeth Tran
Stanford University

BACKGROUND

People use sketches to express and record their ideas. However, most computer programs cannot interpret free-hand sketches. Currently, recognition systems either constrain the user's drawing style or fail to robustly handle complex input.

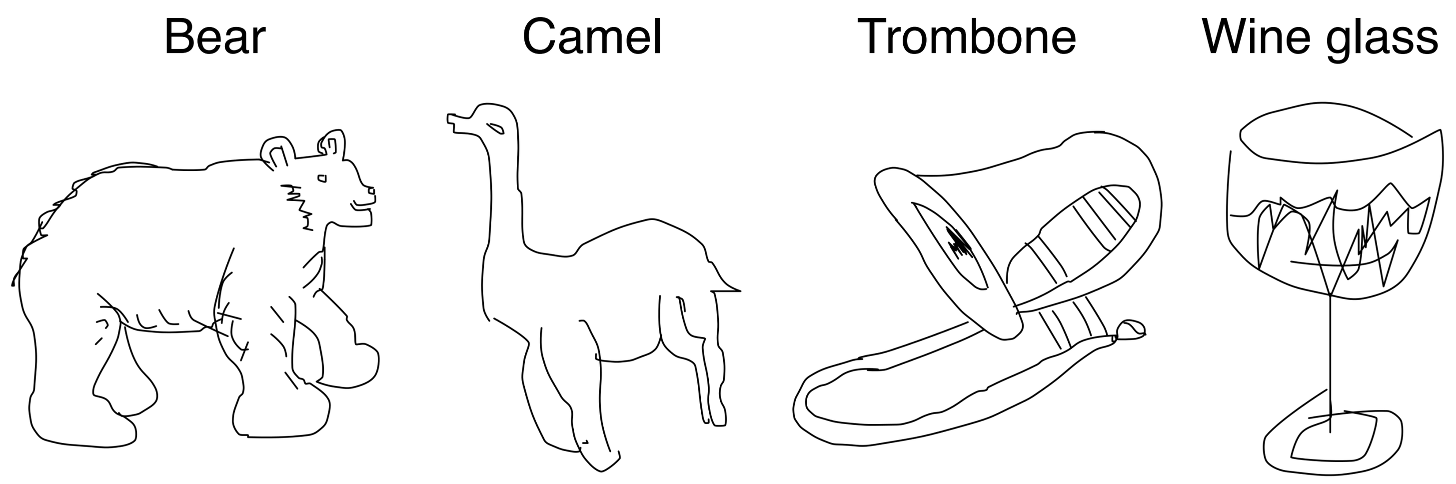

Figure 1. Example of different classes of sketches

METHODS

Using the TU-Berlin sketch dataset [1], which composes of 250 object categories gathered from 20,000 human sketches, we trained a convolutional neural network (CNN) based on the ResNet architecture [2] in order to improve performance over traditional multi-class support vector classifications. We are able to achieve a test accuracy of 65%.

Layer	Output Size
Input	128 x 128 x 1
Dropout	128 x 128 x 1
7x7 Conv, 64, /2	64 x 64 x 64
3x3 Residual Unit, 64	64 x 64 x 64
3x3 Residual Unit, 64	64 x 64 x 64
3x3 Residual Unit, 64	64 x 64 x 64
Dropout	64 x 64 x 64
3x3 Residual Unit, 128 /2	32 x 32 x 128
3x3 Residual Unit, 128	32 x 32 x 128
3x3 Residual Unit, 128	32 x 32 x 128
Dropout	32 x 32 x 128
3x3 Residual Unit, 256, /2	16 x 16 x 256
3x3 Residual Unit, 256	16 x 16 x 256
3x3 Residual Unit, 256	16 x 16 x 256
Dropout	16 x 16 x 256
3x3 Residual Unit, 512, /2	8 x 8 x 512
3x3 Residual Unit, 512	8 x 8 x 512
3x3 Residual Unit, 512	8 x 8 x 512
8x8 Average Pooling	512
Dropout	512
FC-250	250

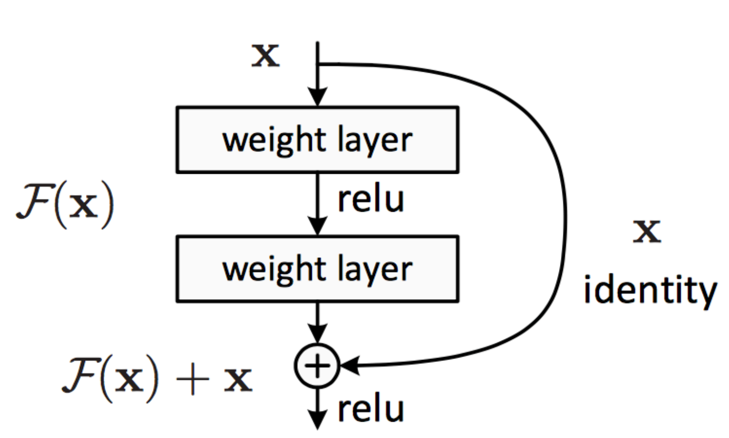

Figure 1. Architecture of our model (left) and a basic residual unit [2] (right).

To augment the training data, we generate additional training examples by horizontally flipping existing training examples. In addition, we invert all examples to better match the effects of our input dropout.

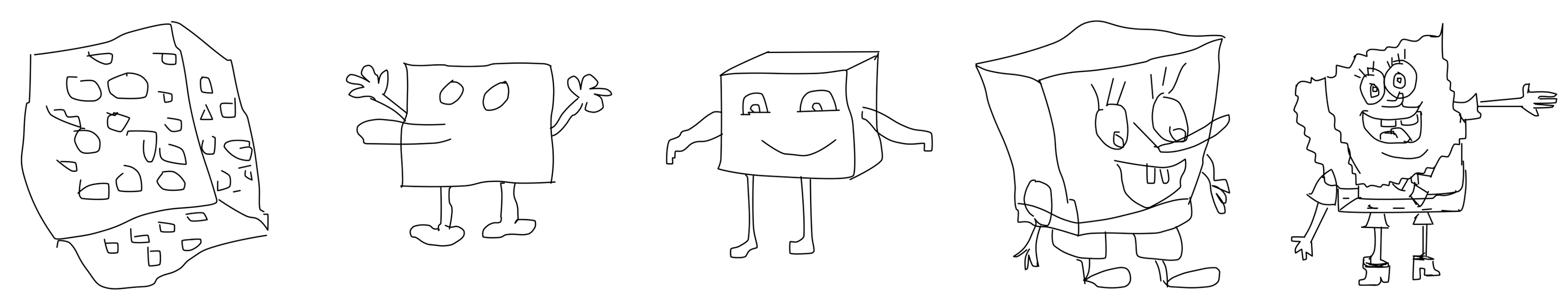

Figure 2. Examples of interclass variations of Spongebob

RESULTS

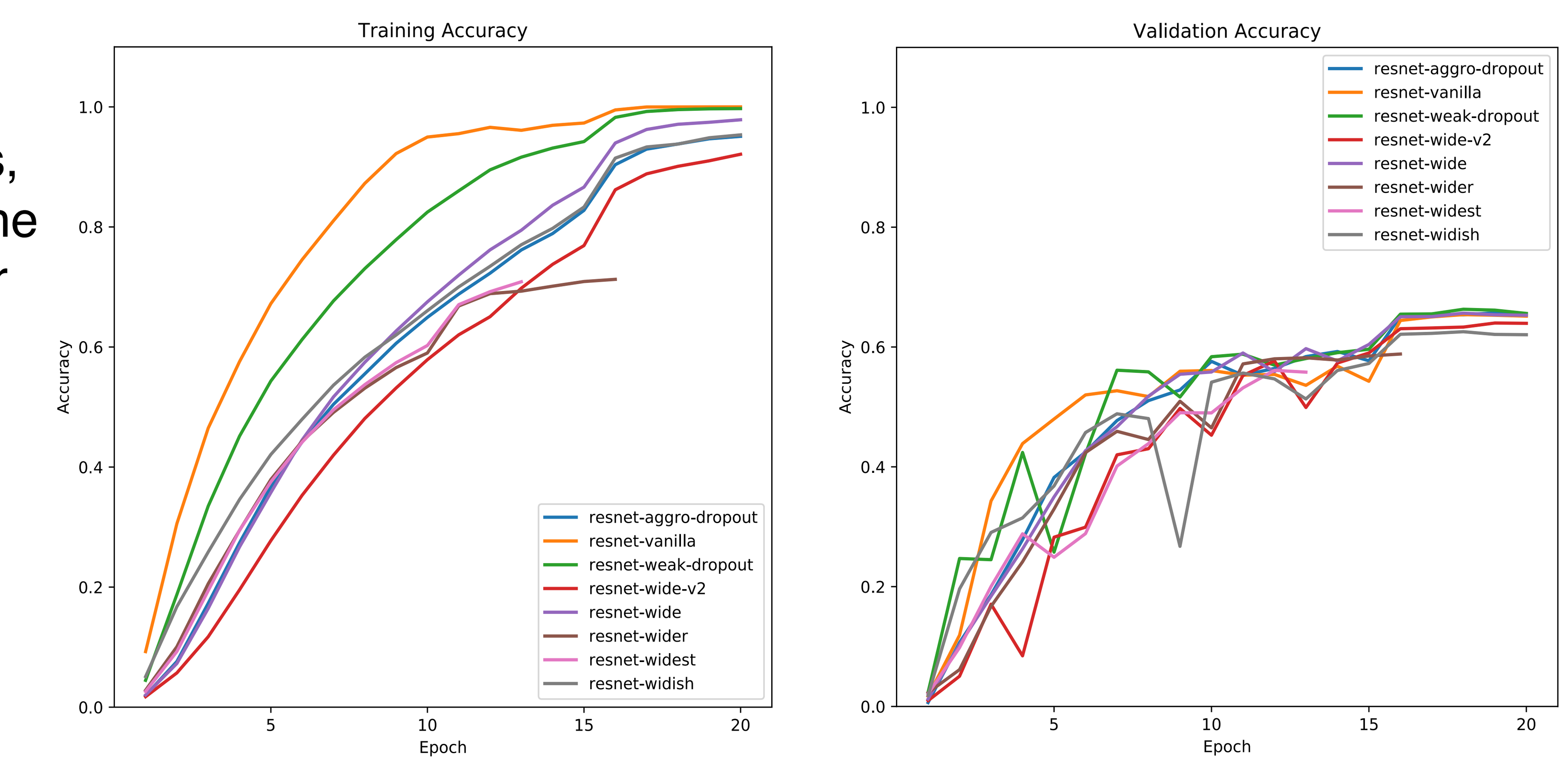

Figure 3. Performance comparison between variations of our base architecture. Experiments showed that moderate dropout with deep networks provides the best results. Overaggressive dropout and shallow networks both lead to lower accuracy. Notably, compensating for shallowness with increased width does not result in improved accuracy.

Figure 4. An example of where our model thought the airplane was a syringe. This shows the variability of human sketches.

Method	%
SIFT-varient +BoF + SVM [1]	56
IDM + SVM [3]	71.30
ConvNet [4]	75.42
ConvNet [5]	77.69
Ours – ResNet Dropout	65

Table 1. Performance comparison of our models versus other methods

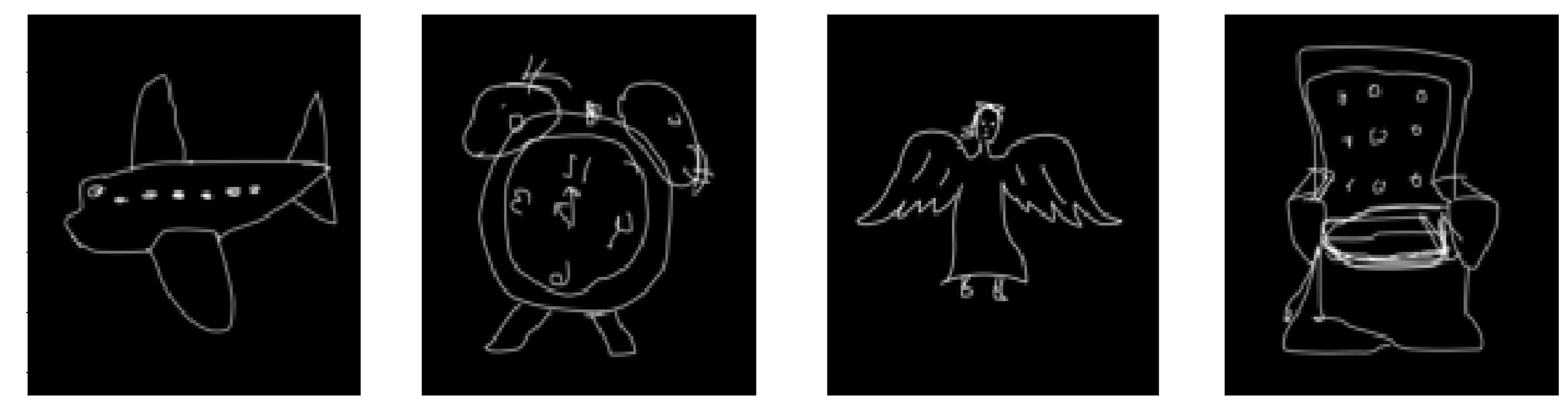

Figure 5. Examples of sketches with inverted colors values

CONCLUSION

Free-hand sketches are inherently hard to classify due to large amounts of intraclass variation and interclass overlap, along with a lack of complex visual information, in contrast to traditional image recognition which instead deals with an overabundance of visual information. Our results suggest that depth is the largest contributing factor to classification accuracy, with dropout regularization providing minor improvement. Widening layers does not result in noticeable changes, and the need to reduce memory consumption by reducing network depth in fact leads to lower accuracy.

References:
[1] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? *ACM Trans. Graph. (Proc. SIGGRAPH),* 31(4):44:1–44:10, 2012.
[2] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
[3] K. T. Yesilbek, C. Sen, S. Cakmak, and T. M. Sezgin. Svm- based sketch recognition: which hyperparameter interval to try? In *Proceedings of the workshop on Sketch-Based Interfaces and Modeling*, pages 117–121. Eurographics Association, 2015.
[4] O. Seddati, S. Dupont, and S. Mahmoudi. Deepsketch: deep convolutional neural networks for sketch recognition and similarity search. In *Content-Based Multimedia Indexing (CBMI), 2015 13th International Workshop on*, pages 1–6. IEEE, 2015.
[5] O. Seddati, S. Dupont, and S. Mahmoudi. Deepsketch 2: Deep convolutional neural networks for partial sketch recognition. In *Content-Based Multimedia Indexing (CBMI), 2016 14th International Workshop on*, pages 1–6. IEEE, 2016.