Identification and Quantifying Outlier Images and Datasets

We found outliers using the Kullback-Leibler (KL) divergence on the pixel values of images. We also rewrite $\lambda$ as a function of difference in distributions. These images were used in Outlier Exposure testing to see the effects of OOD data on training.

The KL distance is:

$$D = KL(P(Q)) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

and $\lambda$ is:

$$\lambda(D) = \tanh(D).$$

Then, $\lambda$ was further modified to increase as the model got further in training so:

$$\lambda(D, i) = \tanh(D)(1 - \cos(i/20)).$$

The table above summarizes our results from using outlier exposure when classifying based on gender. The choice for the outlier group is incredibly significant, and causes accuracy and other metrics to vary wildly.

Future work:

- Use feature norms of the images and other outputs of the CNN to sort the images and improve classification.
- Adapt the activation features code into the outlier exposure code to see if it is more accurate than using the KL divergence based on pixels.
- Alter $\lambda$ by including KL divergence dependent on the distribution between the training and outlier set.

Acknowledgements

This work is supported in part by the US NSF award DMS-2051019. We would like to thank our mentor, Dr. Nicole Yang, along with all of the mentors at the Emory University REU site.

References