A Predictive Model for Eye-Movements

Lauren Arnett (lba2138) & Chengzhi Mao (cm3797)

1 Project Statement

It is estimated that human vision demands anywhere from 30 to 50 % of brain processing. As such, optimizing images for processing by the human brain allows for better information retrieval and retention across the image. Applications for study of how humans' eyes move across images span from advertising to art. Being able to predict where humans are most likely to look provides a guideline as to where to place the most important information, what humans are attracted to in viewing art, or how an image should be cropped to feature the subject. Using an existing dataset of eye movements, we are building a predictive model to generate the most likely fixation locations on a new image.

2 Methods

We will be using one of two potential datasets of fixation locations. The first, developed at Osnabrück University, aggregates 949 images from multiple studies conducted with different batches of participants over image categories of natural and urban scenes, web sites, fractal, pink-noise, and ambiguous artistic figures. The second was developed specifically for a similar eye-movement-prediction study by researchers at MIT with data for 15 participants over 1003 natural images from Flickr and LabelMe. While the first dataset includes a wider variety of images, the second dataset provides data for many more images per viewer.

Regardless of which dataset we decide to use, we will have a relatively small number of viewers per image. Given that we have this pre-defined population, we are looking at using Bayesian methods to keep our model simple, in the spirit of Occam's razor. Should we decide to also use deep learning methods, we will be exploring adversarial training methods to try to generalize our model.

Specifically, we will combine Bayesian inference with deep learning techniques, since the number of of training samples is limited and the eye movement labels are noisy. Thus, we need Bayesian methods to perform inference under uncertainty. Additionally, the input image is high-dimensional—while traditional Bayesian methods have a difficult time handling such data, deep learning methods easily process it.

Adversarial training can be used to match the eye fixation distribution of different people, removing individual-specific information and making the model able to generalize to an unknown group of people.

Lauren will handle preprocessing of the dataset, Chengzhi will set up a testing framework, and we will be working on building the model together.

3 Evaluation

We will use the standard 80% / 20% train / test split on our dataset, reserving a subset of images for testing. Training on image data for all viewers of an image, we will predict where that population is most likely to have fixation locations on a new image. In order to validate how close we were to the actual fixation location, we will be taking the distance between the means of the actual and predicted coordinates. In addition, as the fixiation point for different people looking at the same image is diversified, we would propose a ground truth variance as the indication of uncertainty, which could further evaluate the uncertainty estimation of our Bayesian Learning model.

References

[1] Judd, T., Ehinger, K., Durand, F. and Torralba, A., (2009) "Learning to Predict Where Humans Look", IEEE International Conference on Computer Vision (ICCV),

[2] Wilming N., Onat S., Ossandón J., Acik A., Kietzmann T.C., Kaspar K., Gameiro R.R., Vormberg A., and König P., (2017) "An extensive dataset of eye movements during viewing of complex images.", Nature Scientific Data 4: 160126. https://doi.org/10.1038/sdata.2016.126