Deep learning and eye tracking

Zachary J. Cole[1], Karl M. Kuntzelman[1], Michael D. Dodd[1], & Matthew M. Johnson[1]

[1] University of Nebraska-Lincoln

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Zachary J. Cole, 238 Burnett Hall, Lincoln, NE 68588-0308. E-mail: z@neurophysicole.com

Abstract

Abstract stuff....

Keywords: deep learning, eye tracking, convolutional neural network, cognitive state, endogenous attention

Word count: X

Deep learning and eye tracking

–>

Methods

participants

% E1 - 124Ss; E9 77Ss Two separate datasets were used to develop and test the deep CNN architecture. The two datasets were collected from two separate experiments referred to as exploratory and confirmatory. The participants for both datasets consisted of college students (Exploratory $N = 124$; Confirmatory $N = 77$) from a large Midwestern university who participated in exchange for class credit. Participants who took part in the exploratory experiment did not participate in the confirmatory experiment.

materials and procedure

Each participant viewed ## scene images (see Figure x) while carrying out a search, memorization, or rating task. The same materials were used in both experiments with a minor variation in the procedures. In the confirmatory experiment, participants were directed as to where search targets might appear in the image (e.g., on flat surfaces). No such instructions were provided in the exploratory experiment. For the search task, participants were instructed to find a "Z" or "N" embedded in the image. If the letter was found, the participants were instructed to press a button which terminated the trial. For the memorization task, participants were instructed to memorize the image for a test that will take place when the task is completed. Memory was tested by asking participants to select which of two images they had seen during the task. For the rating task, participants were asked to think about how they would rate the image on a scale from 1 (very unpleasant) to 7 (very pleasant). The participants were prompted for their rating immediately after viewing the image. In both experiments, trials were presented in one mixed block, and three separate task blocks. For the mixed block, the trial types were randomly intermixed within the block.

For the three separate task blocks, each block consisted entirely of one of the three tasks (search, memorize, rate).

apparatus

While the participants viewed the scene images, their eye movements were recorded using an SR Research EyeLink II eye tracker, with a sampling rate of 1000Hz. On some of the search trials, a probe was presented on the screen at six seconds. To equate the data from all three conditions, only the first six seconds of each trial was analyzed. Trials that were missing ## data points were excluded before analysis. For both datasets, the trials were pooled across participants. After removing bad trials, the exploratory dataset consisted of 12,177 trials, and the confirmatory dataset consisted of 9,301 trials.

datasets

The trial data for both experiments were converted into plot images. The x and y coordinates and pupil size were used to plot each sample collected by the eye tracker on a scatterplot diagram (e.g., see Figure X). The coordinates were used to plot the location of the dot, and pupil size was used to determine the size of the dot. The plots were sized to match the dimensions of the computer monitor used to collect the data (1024 x 768 pixels), then were shrunk to (240 x 180 pixels) in an effort to limit the size of the data files.

To systematically assess the predictive value of the data provided by each of the four image dimensions (x-coordinates, y-coordinates, pupil size, dot color), plots were made with each of the dimensions removed. Plots were also made only using x-coordinate data, y-coordinate data, and pupil size data (see Figure X). For each of these separate sets of plots, a raw timeline dataset of the corresponding image dimensions (i.e., x-coordinate only, y-coordinate only, pupil size only) was also developed (see Table X). % when we isolated variables, did we still include time? CHECK

classification

Deep CNNs were used to classify the trials into categories of search, memorize, or rate. Each model split the data into 70% training, 15% validation, and 15% testing. Each network was run through 10 iterations of the data. The same decoding models were run on the raw timeline data, and the image data.

% ~20 different versions changing kernel size, stride rate, number of filters, number of convolutional layers.. highest performing model was 03b

% 1) reshape to 180x240; 2) convolution: 5 filters, 20x20, 2x2; 3) LeakyReLU, alpha = 0.1; 4) BatchNormalization; 5) convolution: 3 filters, 10x10, stride 3x3; 6) LeakyReLU, alpha = 0.1; 7) BatchNormalization; 8) Flatten; 9) Dense, 6 units; 10) Dropout, alpha = 0.1; 11) BatchNormalization; 12) Dense-softmax, 3 units % compiled using categorical crossentropy % How many parameters in the network? 524288 I think - for e9?

The exploratory models consisted of #-# convolutional layers, #-# fully connected layers, and #-# other layers. To maximize the accuracy of the exploratory model, numerous variations on model parameters were adjusted and tested using the exploratory dataset. Adjustments consisted of changing the kernel size, stride rate, and the number of filters. In total, 16 models were tested (see Table x). The same models were used to decode the raw timeline data and the images. The model that resulted in the highest accuracy is shown in Figure x. This model consisted of two convolutional layers, and one fully connected layer (see Figure x). This final model was tested on the confirmatory dataset.

–>

–>

–>