Your Brain on Ads: Deep Learning Approach to Neuromarketing

Abstract

Neuromarketing is an interdisciplinary area of study at the intersection of neurological science, consumer behavior, and marketing. While deep learning has been evaluated for use with techniques such as EEG, there has been less research on how deep learning may be applied to fMRI data in the context of neuromarketing. In this study we trained a convolutional neural network on fMRI data to evaluate its potential application in this area. We found that the best predictions are produced for brain regions responsible for visual perception of color, form, and motion. This suggests that style-related features of images might have a more significant effect on attention than others.

Introduction

Neuromarketing is an interdisciplinary area of study at the intersection of neurological science, consumer behavior, and marketing. Techniques used within neuromarketing include magnetic resonance imagistic (MRI), electroencephalogram (EEG), positron transmission tomography (PET), computerized axial tomography (CAT/CT) and following the line of sight (ECT).[1] A goal of neuromarketing is tying observed phenomena within the brain when exposed to advertisement to some consumer behavior. The techniques used within this field can assist

[1] Popescu, Gabriela, and Iasmina Iosim. "New Tendencies in Consumer Behaviour." *Agricultural Management / Lucrari Stiintifice Seria I, Management Agricol* 14, no. 2 (April 2012): 613–22. https://search-ebscohost-com.proxy.uchicago.edu/login.aspx?direct=true&db=bth&AN=91912514&site=bsi-live&scope=site.

marketers in understanding how a consumer's brain evaluates products and brands, especially in the context of unconscious brain functions latent within consumer preferences or attitudes.[2]

Recently, deep learning approaches have been explored as potential approaches within the field of neuromarketing. Within this, deep learning has been used in conjunction with EEG signals and evaluated against other machine learning approaches such as random forest.[3] EEG signals have further been explored within the context of evaluating deep learning for product selection and movie rating.[4] Similarly, deep learning has been found to be superior to conventional neuromarketing and can be used to leverage advancements in machine learning and the availability of relevant data sets.[5]

Therefore, within this research study, fMRI was used as a method within neuromarketing for which a deep learning approach was evaluated. fMRI has been used to study consumer behavior within contexts such as word of mouth diffusion.[6] It has further been paired with deep learning in the context of brain disorder diagnostics and classification of Alzheimer disease.[7][8] This research aims to evaluate the potential of deep learning trained on fMRI data for the purposes of neuromarketing.

---

[2] Hammou, Khalid Ait, Md Hasan Galib, and Jihane Melloul. "The contributions of neuromarketing in marketing research." *Journal of management research* 5, no. 4 (2013): 20.
[3] Aldayel, Mashael, Mourad Ykhlef, and Abeer Al-Nafjan. 2020. "Deep Learning for EEG-Based Preference Classification in Neuromarketing" *Applied Sciences* 10, no. 4: 1525. https://doi.org/10.3390/app10041525
[4] Alimardani, Maryam, and Mory Kaba. "Deep learning for neuromarketing; classification of user preference using EEG signals." In *12th Augmented Human International Conference*, pp. 1-7. 2021.
[5] Georgiadis, Kostas, Fotis P. Kalaganis, Vangelis P. Oikonomou, Spiros Nikolopoulos, Nikos A. Laskaris, and Ioannis Kompatsiaris. "Harneshing the Potential of EEG in Neuromarketing with Deep Learning and Riemannian Geometry." In *International Conference on Brain Informatics*, pp. 21-32. Cham: Springer Nature Switzerland, 2023.
[6] Melissa Yi-Ting Hsu and Julian Ming-Sung Cheng. "FMRI Neuromarketing and Consumer Learning Theory: Word-of-Mouth Effectiveness After Product Harm Crisis." *European Journal of Marketing* 52, no. 1 (2018): 199-223. doi:https://doi.org/10.1108/EJM-12-2016-0866.
[7] Yin, Wutao, Longhai Li, and Fang-Xiang Wu. "Deep learning for brain disorder diagnosis based on fMRI images." *Neurocomputing* 469 (2022): 332-345.
[8] Sarraf, Saman, and Ghassem Tofighi. "Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks." *arXiv preprint arXiv:1603.08631* (2016).

By leveraging the BOLD5000 fMRI Dataset, a large-scale, slow event-related fMRI dataset collected on 4 subjects, each observing 5,254 images over 15 scanning sessions, a convolutional neural network was trained to predict decorrelated cortical activity based on the images observed in the BOLD5000 fMRI Dataset. This is to evaluate how deep learning might be used within the context of memory or attention, two areas important to neuromarketing and that may be related to the level of encoding of visual stimuli. Images analysis was performed to understand which elements of an image may be related to observed activation levels within the fMRI data.

Background

Neuromarketing

Conventional marketing research encompasses techniques such as surveys, interviews, questionnaires, or focus groups. These methods are limited in their ability to glean information on the latent dimensions of consumer behavior. Neuromarketing, coined in 2002, combines the fields of neuroscience and marketing.[9] Neuromarketing provides a clearer, more fulsome, understanding of consumer behavior. The introduction of neuroscience to marketing has helped enrich marketing's understanding of the biological facets of the brain as it relates to consumer behavior, which has helped unearth insights that often elude traditional research methods in marketing. The approach has been adopted within industry with corporations having employed these techniques within their own marketing operations.[10]

fMRI

[9] Morin, Christophe. "Neuromarketing: the new science of consumer behavior." *Society* 48, no. 2 (2011): 131-135.
[10] Lin, Meng-Hsien, Samantha NN Cross, William J. Jones, and Terry L. Childers. "Applying EEG in consumer neuroscience." *European Journal of Marketing* 52, no. 1/2 (2018): 66-91.

FMRI experiments are characterized by minimization of recall bias, reduction of cognitive bias, integration of the psychological and cognitive processes, assessment of human behavior during unawareness and ability to distinguish between similarities and differences in the neurocognitive system.[11] These features enable fMRI experiments to be more effective than traditional approaches to marketing research in efforts to understand the unconscious factors of consumer behavior.

fMRI data study and interpretation has interdisciplinary roots ranging from neurology to machine learning. fMRI time series enables sophisticated modeling of brain functional networks, leading to new understandings of the human brain such as brain disorder diagnosis.[12] Recently, machine and statistical learning methods have been explored to predict cognitive brain states from fMRI data and build neural networks.[13] Deep learning techniques are a state of the art approach analyzing fMRI data sets and have resulted in performance improvements in diverse fMRI applications. Deep learning allows fMRI data to be considered as images, time series, or image series and image series which readily lend themselves to CNNs and RNNs.[14] The purpose of this research is to evaluate how the use of deep learning on fMRI data may be used in the context of neuromarketing. To this end, a CNN was trained on fMRI data to evaluate how the use of these techniques from other areas of study may be used in neuromarketing research.

Methods and Data

[11]Reimann, Martin, Oliver Schilke, Bernd Weber, Carolin Neuhaus, and Judith Zaichkowsky. "Functional magnetic resonance imaging in consumer research: A review and application." *Psychology & Marketing* 28, no. 6 (2011): 608-637.

[12] Yin, Wutao, Longhai Li, and Fang-Xiang Wu. "Deep learning for brain disorder diagnosis based on fMRI images." *Neurocomputing* 469 (2022): 332-345.

[13] Michele Svanera, Mattia Savardi, Sergio Benini, Alberto Signoroni, Gal Raz, Talma Hendler, Lars Muckli, Rainer Goebel, Giancarlo Valente, Transfer learning of deep neural network representations for fMRI decoding, Journal of Neuroscience Methods, Volume 328, 2019, 108319, ISSN 0165-0270,

[14] Yin, Wutao, Longhai Li, and Fang-Xiang Wu. "Deep learning for brain disorder diagnosis based on fMRI images." *Neurocomputing* 469 (2022): 332-345.

With all combinations of participants, trials, and images, our final dataset consisted of 113,220 observations. The images were our sole input feature, and unique neural regions were our outputs. The neural regions were the early visual cortex (EarlyVis), lateral occipital complex (LOC), occipital place area (OPA), parahippocampal place area (PPA), and retrosplenial cortex (RSC), each in the right hemisphere (RH) and left hemisphere (LH), for a total of 10 regions of interest (Table 1). Data was split into training and test sets with a test size of 0.2.

While voxel-level data for neural regions were provided, we did not need such granularity, as were more interested in overall activation of the region. As a result, our outputs consisted of the mean activation level across all voxels for each region. Further manipulation involved normalization using min-max scaling, setting the minimum value for each region to zero and the maximum to one. Normalization allows for better performance of the model, through both stabilizing learning and improving generalization.

As mentioned earlier, there were 5,254 unique images presented to participants. However, only 4,916 images were available in our dataset, so we had to remove all records including the missing images. After removal, 105,912 records remained. Images were converted to Numpy arrays but not transformed any further, as the pixel values were already within a standard range of bounds. Due to the size of the data, we needed to incorporate a generator while training our model. A generator allowed us to train the model on all images without running out of memory, as each image was loaded as it was needed.

Table 1

	LHEarlyVis	LHLOC	LHOPA	LHPPA	LHRSC	RHEarlyVis	RHLOC	RHOPA	RHPPA	RHRSC
Mean	0.0011	0.0007	0.0005	0.0009	0.0002	0.0011	0.0008	0.0009	0.0011	0.0007
Median	0.0014	0.0008	0.0003	0.0007	-0.0001	0.0014	0.0010	0.0009	0.0009	0.0005

St. Dev.	0.0069	0.0053	0.0062	0.0076	0.0091	0.0071	0.0057	0.0070	0.0078	0.0080
Min	-0.0333	-0.0320	-0.0426	-0.0392	-0.0580	-0.0385	-0.0298	-0.0456	-0.0361	-0.0400
Max	0.0459	0.0323	0.0384	0.0538	0.0507	0.0489	0.0407	0.0548	0.0438	0.0449

Analysis

Images are complex sources of data—in our dataset, each image is a 375x375x3 array, which results in over 400,000 unique values for each image. This combined with the ~113,000 unique observations results in an exponentially complex dataset. As a result, complex machine learning models are necessary. We chose a Convolutional Neural Network (CNN) to model our data due to its ability to handle high-dimensional spaces, the nonlinear nature of our dataset, and its optimization for image processing. Though limits in processing power restricted the variety of methods we could try, ultimately we feel that the complexity of our model adequately matched the data.

Model Training

The neural network was trained with the following layers:

Convolutional 2D layer: 32 filters, activation = 'relu', padding = 'same'
Max pooling layer
Convolutional 2D layer: 64 filters, activation = 'relu', padding = 'same'
Max pooling layer
Convolutional 2D layer: 128 filters, activation = 'relu'
Max pooling layer
Convolutional 2D layer: 256 filters, activation = 'relu'
Max pooling layer
Convolutional 2D layer: 512 filters, activation = 'relu'
Flatten layer
Dense layer: 64 filters, activation = 'relu'
Dense layer (output): 10 filters, activation = 'linear'

Due to the continuous nature of the output variables, a linear activation in the final layer was found to give the best performance. The model was trained in batches of size 10 over three epochs, as MSE plateaued at the third epoch. MAE reached 0.066 during training, and the model was evaluated at 0.066 on the test set.

The model performed differently for different brain regions (Figure 1). Peak performance was with the RHOPA, with a MAE of 0.048, while the model performed the worst on RHPPA, with a MAE of 0.072. One issue that we faced with the model is that the predicted values are in a very small range. For example, the actual values for LHLOC ranged between 0.2 and 0.8, while the predicted values ranged between 0.48 and 0.54. Not all iterations of the model had this issue, but this seemed to become more of an issue as model performance improved. One reason that this occurred could be that, as the training size of the data increased, the model converged around the mean of each of the regions. This would indicate that the model is too generalized and is unable to pick up small complexities in the data. Further analysis could attempt different methods at adjusting model complexity.

Results

To mitigate the above issues, we use-cross validation. We train the models for k = 5 and calculate the mean absolute error as our resulting metric. After running the cross-validation we get mean absolute errors for each of the regions of interest (Table 2).

Table 2

LHEarly Vis	LHLOC	LHOPA	LHPPA	LHRSC	RHEarly Vis	RHLOC	RHOPA	RHPPA	RHRSC
0.073	0.066	0.059	0.066	0.065	0.068	0.068	0.056	0.081	0.072

The resulting MAE is 0.067 after cross-validation which suggests that our initial result is robust. However, there are some changes in the metrics for separate regions. Namely, we can see that LHOPA and RHOPA remain regions with the best predictions, while LHPPA is now closer to the mean value.

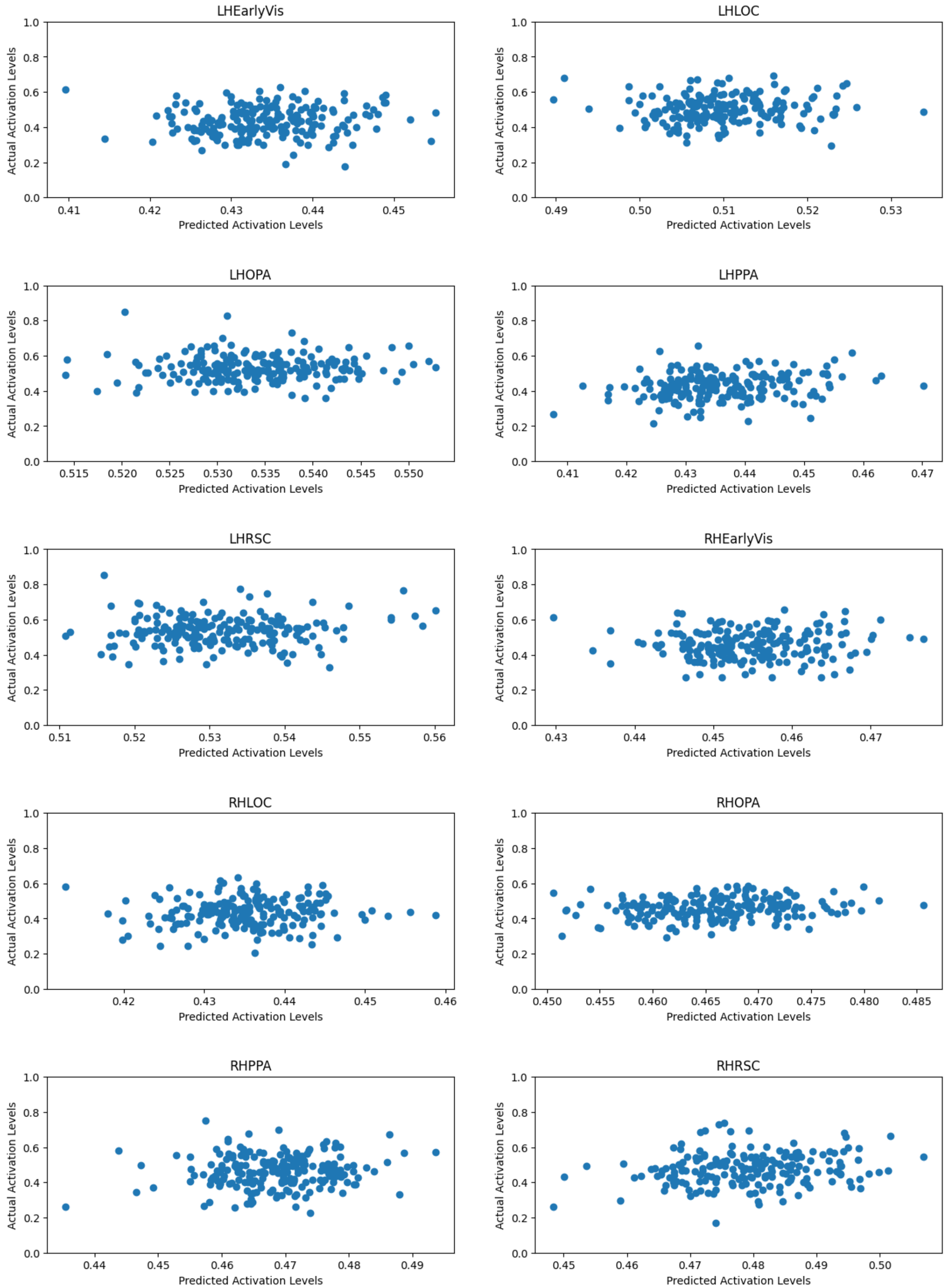
Some other attempts at dealing with a limited range of predictions included 1) changing the optimizer of the model to Adam and 2) changing the activation function of the final layer to a sigmoid. None of these modifications improve the predictive power, therefore we left the first version of the model for estimation.

Discussion

The low mean squared error obtained by the model indicates that there is general correspondence between the actual and predicted values. One of the results we observe is that some regions of interest in the brain have a more pronounced trend of predictions than others (as is shown in Figure 1) and a lower MAE. Specifically, regions RHOPA, LHOPA, and LHPPA have MAE lower than 0.06 (0.048, 0.055, and 0.057 respectively). At the same time, RHPPA has the highest MAE of 0.072. After cross-validation though, the two regions which maintain a low MAE correspond to the occipital *place area* which is responsible for **visual perception, including color, form and motion**. Unlike those, mean absolute error for LHPPA becomes larger, meaning that the *parahippocampal place area* which responds to houses, landmarks, indoor and outdoor scenes, is not explained better than the average observation. This finding suggests that our model predicts the general style of an image rather than objects or details. This can be a result of producing features which capture mostly the exposition of the pictures such as color and style, or lack of overall complexity of the model.

Another potential reason for this result is that the *Scene Images* dataset constitutes about 20% of the sample (1000 observations) which makes scenes and landscapes a significant factor to the training of the model. This is supported by the fact that after cross-validation, the region of interest responsible for response to landscape gives less accurate predictions. In this sense, the nature of the data is very specific and has to be adjusted for.

Figure 1


LHEarlyVis
LHLOC
LHOPA
LHPPA
LHRSC
RHEarlyVis
RHLOC
RHOPA
RHPPA
RHRSC
Actual Activation Levels
Predicted Activation Levels