ORIGINAL ARTICLE

A Novel Object Categorization Decoder from fMRI Signals Using Deep Neural Networks

Alireza Shakeripour [1], Zahra Bahmani [1, 2*], Poorya Aghaomidi [1], Shima Seyed-Allaei [3]

[1] *Department of Biomedical Engineering, Faculty of Electrical & Computer Engineering, Tarbiat Modares University, Tehran, Iran*

[2] *Department of Cognitive Neuroscience, Faculty of Interdisciplinary Sciences and Technologies, Tarbiat Modares University, Tehran, Iran*

[3] *School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran*

*Corresponding Author: Zahra Bahmani
Email: z.bahmani@modares.ac.ir

Received: 04 December 2023 / Accepted: 18 April 2024

Abstract

Purpose: Understanding neural mechanisms is critical for discerning the nature of brain disorders and enhancing treatment methodologies. Functional Magnetic Resonance Imaging (fMRI) plays a vital role in gaining this knowledge by recording various brain regions. In this study, our primary aim was to categorize visual objects based on fMRI data during a natural scene viewing task. We intend to elucidate the challenges and limitations of previous models in order to produce a generalizable model across different subjects using advanced deep-learning methods.

Materials and Methods: We've designed a new deep-learning model based on transformers for processing fMRI data. The model includes two blocks, the first block receives fMRI data as input and transforms the input data to a set of features called fMRI space. Simultaneously a visual space is extracted from visual images using a pre-trained inceptionv3 network. The model tries to construct the fMRI space similar to the extracted visual space. The other block is a Fully Connected (FC) network for object recognition based on fMRI space. Using transformer capabilities and an overlapping method, the proposed architecture accounts for structural changes across different voxel sizes of the subjects' brains.

Results: A unique model was trained for all subjects with different brain sizes. The results demonstrated that the proposed network achieves an impressive similarity correlation between visual space and fMRI space around 0.86 for train and 0.86 for test dataset. Furthermore, the classification accuracy was about 70.3%. These outcomes underscored the effectiveness of our fMRI transformer network in extracting features from fMRI data.

Conclusion: The results indicated the potential of our model for decoding images from the brain activities of new subjects. This unveils a novel direction in image reconstruction from neural activities, an area that has remained relatively uncharted due to its inherent intricacies.

Keywords: Functional Magnetic Resonance Imaging; Deep Learning; Object Recognition; Region Of Interest Connectivity; Brain Decoding.

DOI:

1. Introduction

In computational neuroscience, a wide range of studies seek to answer the question of how sensory stimuli are encoded in nerve cells (neurons) and the feasibility of decoding stimuli from neural information. In the last two decades, the use of machine learning for neural decoding from fMRI data has significantly increased in both quantity and quality. In the early research, the goal of decoding was to identify the categories of objects presented to individuals [1]. Numerous studies have delved into the realm of fMRI data, with a significant portion focusing on object categorization [2-9], motion direction [10, 11], perceptual imagination [12], and even memory [13]. These investigations have demonstrated the remarkable ability of classification-based machine learning techniques to decode visual features from fMRI signals. These methods learn the linear or nonlinear mapping between brain activity patterns and stimulus categories from the training dataset. Besides, some studies tried to reconstruct the full image from fMRI signals [14-17].

Understanding the functioning of various brain regions not only enhances our comprehension of brain operations but also offers potential diagnostic insights into anomalies and brain disorders when comparing outcomes across individuals. This study strives to contribute to this endeavor by enhancing the generalization capacity of existing models.

In comparison to other non-invasive data collection methods, fMRI data offers advantages, including high spatial resolution and access to data from various brain regions. Nonetheless, this data acquisition method comes with certain limitations, such as indirect recording based on factors like blood or oxygen consumption by neurons. Additionally, the smallest unit of this aggregated data reflects the activity of hundreds of thousands of neurons. This level of resolution is insufficient for reconstructing and distinguishing individual components of an image, where neurons may carry valuable information. The act of combining them results in the loss of this information. Therefore, complete image reconstruction from fMRI data is controversial, but object categorization can be decoded from fMRI signals.

A fundamental challenge in constructing a model based on data from multiple subjects lies in the variations in the structural and functional aspects of individuals' brain architecture. Although by mapping people's brains on each other, the general processing areas related to different senses coincide, this mapping causes a loss of neural activity information. In certain cases, such as decoding left and right-hand movements, where the activity or inactivity of relatively large brain areas on both hemispheres is sufficient for decoding, this mapping can be useful. However, for tasks such as image reconstruction that involve intricate patterns and details from diverse fMRI data points, this mapping may prove misleading. Moreover, the existence of differences such as blood pressure, mental state, and level of attention in different people causes statistical differences in the responses recorded in different fMRI. Consequently, existing image categorization systems are typically designed using one person's brain signals and assessed with that same individual's test data. This subject-dependent strategy significantly restricts the applicability and practicality of mind-reading models.

In this research, we have sought to take a step towards making brain decoding more operational. Some of the innovations of this study are:

• To extract information from fMRI, a new transformer deep neural network has been proposed which uses the concept of attention and tries to extract the features that are considered in different visual parts of the brain, and emphasizes the relationship between these areas.

• Methods to increase the generalization properties of models by assimilating the input data are proposed. The presented model has the ability to be used for brain decoding of new subjects without the need to register subjects' brains and match their volumes with each other.

In the subsequent sections, we will begin by introducing the utilized dataset, followed by a comprehensive description of the methodology. Subsequently, we will scrutinize the results, then engage in a detailed discussion, and finally, conclude with a summary of the key findings.

2. Materials and Methods

We used the NSD dataset, which consists of eight participants who had their fMRI signals recorded while they viewed images selected from the COCO dataset. Figure 1 shows how the test task was performed. We selected the data of the first three subjects (which had different voxel numbers in fMRI data). The numbers of voxels of the

subject's brain data were as follows, respectively, {81 * 104 * 83}, {82 * 106 * 84} and {81 * 106 * 82}. All fMRI data in NSD were collected at 7T using a whole-brain 1.8-mm 1.6-s gradient-echo EPI pulse sequence. For pre-processing, an approach aimed at preserving as much spatial and temporal detail as possible was adopted. This involved one temporal resampling to correct slice time differences and one spatial resampling to correct head motion within and across scan sessions, EPI distortion, and gradient nonlinearities. Following pre-processing, a General Linear Model (GLM) analysis of the pre-processed time-series data was performed. The approach aimed to estimate BOLD response amplitudes ('betas') for single trials, which was challenging due to the low signal-to-noise ratio. Each subject viewed 10,000 images from the COCO dataset, and each image was randomly played 3 times. These images were categorized into 80 different classes [18, 19].

Since the number of images related to each class varied significantly, particularly in the class related to humans, we divided the data into two categories: the class of humans and the class of non-humans. So, the proposed model predicted if a human existed in the image. In Figure 2, the number of images related to each class in the COCO dataset and the number of images related to each class after the new labeling are presented.

The data set includes 90000 images for all three subjects but only 77,250 images are public. To prevent overfitting and avoid duplicate data between the test and training datasets, each image was used only once, resulting in a selection of 26,888 unique images.

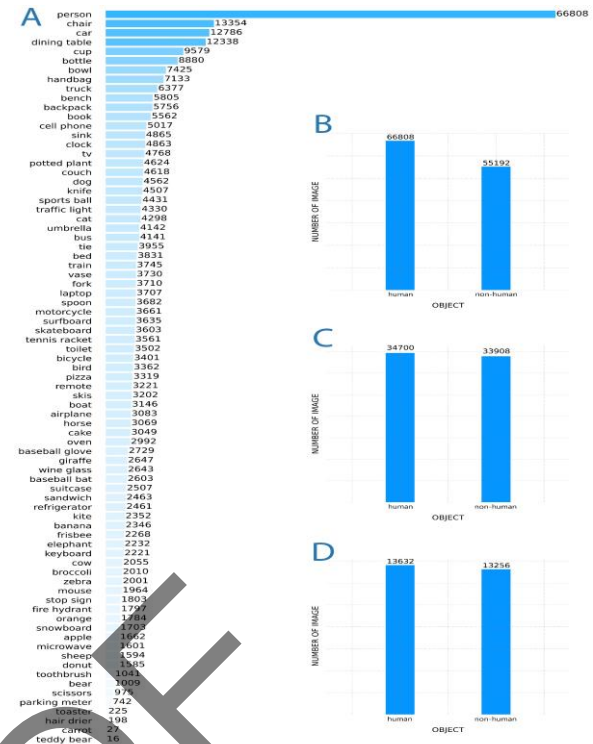

Figure 1. Number of images related to each category in total and subtotal COCO dataset: A) Number of images in all categories of total dataset of COCO, B) Number of images in human and non-human categories in the total dataset of COCO, C) Number of images in human and non-human categories which are used in the NSD task, D) The same as C for the first three subjects

As outlined in the introduction, various challenges in image reconstruction from fMRI signals limited the models. The existing methods of image reconstruction are mostly subject-dependent models which train a model on

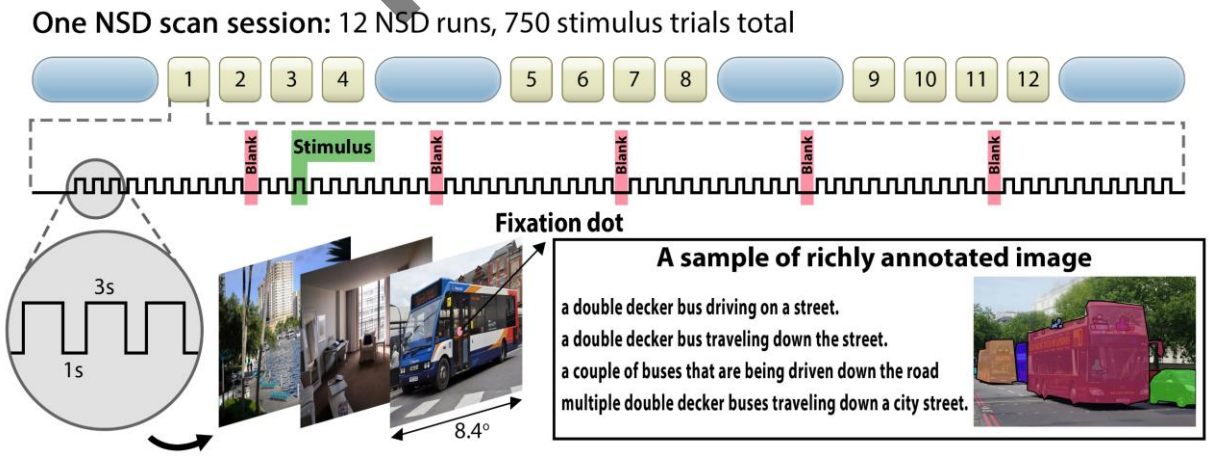

Figure 2. The NSD experiment was designed to collect data on how the human brain responds to different types of natural images, including objects, scenes, and people. The experiment consisted of 12 scan sessions, each of which lasted 5 minutes and consisted of 63 or 62 stimulus trials with randomly interspersed blank trials. During each trial, participants viewed a natural scene and judged whether they had seen the image before. The images were taken from Microsoft's COCO dataset, which is a richly annotated data set with object information

an individual's training data and then evaluate the model on that same individual's test data. In this study, we introduced a subject-independent model. However, this approach faces significant challenges due to variations in the voxel numbers of whole brain and ROIs among individuals and also a lack of generalization of models across subjects. As a result, developing a method capable of generalizing to a new subject's data was not initially considered feasible. In response to these challenges, we proposed a novel approach to address these issues and move closer to the development of a generalized model that is independent of the specific subject. Figure 3 illustrates the sequential steps involved in constructing the developed model.

In this study, we introduced a transformer neural network designed for extracting information from fMRI data. This network addresses the issue of voxel mismatch and standardizes the distribution of extracted information across various individuals. The model construction process comprises two main blocks. The first block includes two phases.

In the initial phase, we leveraged a pre-trained network based on the ImageNet image dataset. Through its image encoder, we extracted features from each image of the dataset which constructed the visual space of the image. Subsequently, we employed a decoder, which was built using an MLP network architecture. Using these extracted features, we classified the components presented within the images.

In the second phase, we established a transformer network with the goal of extracting features from the fMRI signal constructing an fMRI space. We then employed the visual space from the previous phase to classify these features into components and complete our decoding process. An important aspect to highlight is that we feed the network with our fMRI signal as a collection of regions associated with vision. This strategic approach allows our transformer network to leverage its inherent capabilities, measuring the interconnections and attention levels among these regions and the voxels. As a result, it effectively identified the active regions within each area which were relevant to the target image. To facilitate this, we employed the HCP mask, which encompasses 380 distinct brain regions. We have curated 200 regions that are particularly pertinent to visual processing in the brain which have potential information for extracting visual features and aiding in the decoding process. A comprehensive list of these selected areas is provided in Table 1.

Another crucial consideration is that we input fMRI signals from individuals with different brain sizes into the network. To accomplish this, we employ the overlapping technique without resizing or zero-padding. Instead, we facilitate compatibility among subjects by sharing information, thus effectively compensating for the differences in brain sizes. This approach allows us to perform this compensation without resorting to additional data through interpolation, extrapolation, or the removal of any part of the data. Instead, we leverage the fMRI signal itself. Consequently, we can effectively decode the fMRI signals of individuals with diverse brain sizes using the proposed network.

Another noteworthy advantage of the network is its significantly reduced number of training parameters. This efficiency enhancement expedites model training, particularly when dealing with substantial datasets. The architecture of our model is shown in Figure 4.

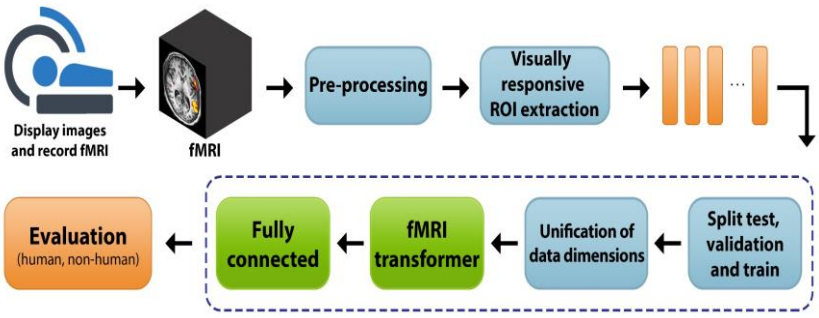

Figure 3. Block diagram illustrating the step-by-step process for constructing the developed model

Table 1. 200 selected brain areas involved in visual processing

Region Name		
V1_L&R	LO3_L&R	TGv_L&R
V2_L&R	FEF_L&R	TE1m_L&R
V3_L&R	43_L&R	PSL_L&R
V4_L&R	OP4_L&R	STV_L&R
V6_L&R	OP1_L&R	TPOJ1_L&R
V3A_L&R	OP2-3_L&R	TPOJ2_L&R
V7_L&R	PoI2_L&R	TPOJ3_L&R
IPS1_L&R	FOP4_L&R	7Pm_L&R
V3B_L&R	MI_L&R	7AL_L&R
V6A_L&R	Pir_L&R	7Am_L&R
V8_L&R	AVI_L&R	7Pl_L&R
FFC_L&R	AAIC_L&R	7PC_L&R
SFL_L&R	8Ad_L&R	V3CD_L&R
8Av_L&R	8BL_L&R	p9-46v_L&R
PGi_L&R	PIT_L&R	FOP3_L&R
PGs_L&R	VMV1_L&R	FOP2_L&R
RSC_L&R	VMV3_L&R	PoI1_L&R
POS2_L&R	VMV2_L&R	Ig_L&R
PCV_L&R	VVC_L&R	FOP5_L&R
7m_L&R	MST_L&R	PI_L&R
POS1_L&R	LO1_L&R	TF_L&R
23d_L&R	LO2_L&R	TGd_L&R
v23ab_L&R	MT_L&R	TE1a_L&R
d23ab_L&R	PH_L&R	TE1p_L&R
31pv_L&R	V4t_L&R	TE2a_L&R
DVT_L&R	FST_L&R	TE2p_L&R
PHT_L&R	PFm_L&R	s6-8_L&R
46_L&R	a9-46v_L&R	9-46d_L&R
LIPv_L&R	VIP_L&R	MIP_L&R
LIPd_L&R	AIP_L&R	PFt_L&R
PGp_L&R	IP2_L&R	IP1_L&R
IP0_L&R	PFop_L&R	PF_L&R
9p_L&R	8C_L&R	9a_L&R
	i6-8_L&R	

Next, we employed accuracy, recall, F1-score, and precision as quantitative evaluation criteria of the proposed decoder (Equations 1, 2, 3, and 4).

$$precision = \frac{TP}{TP + FP} \quad (1)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

$$cosine\ similarity = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \cdot \sum_{i=1}^{n} B_i^2}} \quad (5)$$

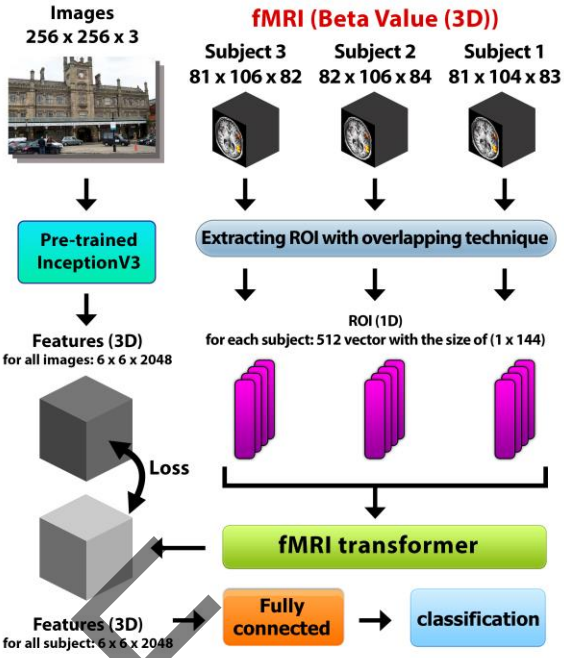

Figure 4. Model architecture; pre-trained image encoder extracts features from images. The transformer network extracts features from fMRI signals and tries to construct an fMRI space similar to visual space. Then object categorization is done using the FC network

where TP represents true positive, FP represents false positive, TN represents true negative, and FN represents false negative. Given two n-dimensional vectors of attributes, A and B, the cosine similarity, is represented using a dot product where Ai and Bi are the ith components of vectors A and B, respectively (Equation 5). We employed cosine similarity to evaluate how well our model's feature space (fMRI space) aligns with feature space obtained from images by the Inception V3 network. This metric assesses the directional similarity between vectors, making it ideal for comparing high-dimensional data. Scores range from -1 to 1, where a score of 1 indicates perfect similarity, 0 indicates no similarity, and -1 indicates perfect dissimilarity. By computing cosine similarity scores for each data point and averaging them, we obtained a comprehensive measure of alignment (Equation 6).

$$PEI = \frac{Accuracy\ (predicted\ from\ fMRI\ space)}{Accuracy\ (predicted\ from\ visual\ space)} \times 100 \quad (6)$$

The Percentage of extractable information (PEI) showed the percent of information extracted from fMRI space for object categorization with 5 respect to the total information

that was accessible in the visual space for object classification (Equation 6).

The software environment and hardware used for this study are presented in Table 3. We implemented the code in Python 3.10, and the neural network was constructed using the TensorFlow deep learning framework.

Table 3. Software environment and hardware used for this study

Item	Configuration
Operation system	Windows 11
CPU	AMD Ryzen 9 3950X 16-Core Processor
Memory	128G
GPU	NVIDIA GeForce RTX 2080 Ti
Video memory	12G
Hard disk	8TB
Software	Python3.10; TensorFlow2.10; CUDA 12.2
Compiler	Anaconda; Jupyter Notebook

3. Results

This study presented a novel subject-independent approach for image reconstruction from fMRI data. The model consists of two main blocks: a pre-trained network for extracting visual features from images and a transformer network for extracting features from fMRI signals. The model also employs an overlapping technique to accommodate different brain sizes among subjects. This network addresses the challenge of voxel mismatch and standardizes the distribution of extracted information across individuals. In this part, according to the measurement metrics that we discussed in the previous section, we examined the results of the proposed model.

We divided the dataset into three parts: the training data set (80% of the total data), the evaluation dataset (10% of the total data), and the test dataset (10% of the total data). In Table 3, the data format used for each of the networks is explained. Despite different numbers of voxels or different sizes of fMRI matrices, the proposed model proceeded similarly for different subject inputs. There was no need to apply further pre-processing techniques and group registration methods for transforming the input data to a specific size. The main advantage of such a model is invariance to the input size. Using the information of all subjects for training a model led to a subject-independent model which could decode the information of different brain subjects.

Table 2. Size and format of data used for each network

Subject	Parameter	Values
Subject 01	Size of fMRI matrix	81 × 104 × 83
	Number of beta values	27750
Subject 02	Size of fMRI matrix	82 × 106 × 84
	Number of beta values	27750
Subject 03	Size of fMRI matrix	81 × 106 × 82
	Number of beta values	21750
Total (fMRI transformer)	Number of beta values	77250
Total (FC)	Number of beta values	26888
3 subjects	ROI Patch size	512 × 144

Training parameters of both blocks of the proposed network are shown in Table 4 and Table 5. Table 4 demonstrates the parameters of the fMRI transformer and Table 5 introduces the parameters of the FC network.

The results obtained for our fMRI transformer network, using the Cosine similarity metric, were 86.8 for the training dataset, 86.5 for the evaluation dataset, and 86.95 for the test dataset (Table 6).

Table 4. Training parameters of fMRI transformer

Training Parameters	Setting Values
Backbone	fMRI transformer
Loss	Cosine similarity
Optimizer	Adam
Batch size	32
Epochs	50 + 100
Learning rate	0.0001
Dropout rate	0.2
Features size	73728 (6 × 6 × 2048)
Number of Features	77250
Total parameter	2,164,608

Additionally, the results obtained for the accuracy of the FC network, which utilized features extracted from the Inception V3 network during training, were 72.4 for the training dataset, 67.2 for the evaluation dataset, and 67.5 for the test dataset. Furthermore, the accuracies obtained from the FC network for features extracted from the fMRI transformer network were 71.1, 65.6, and 66.5 for the training, evaluation, and test dataset (Table 7). The dataset of two categories was balanced, however, we measured the other performance metric to ensure a fair decision of proposed networks. Furthermore, the precision, recall, F1-score, and AUC obtained from the FC network for features extracted from the fMRI transformer network were 72.2, 55.3, 62.62, and 61.59 for the test dataset. The fMRI transformer has to construct an fMRI space that is similar to the visual space extracted from the inception V3 network. The accuracy of object categorization using fMRI data was almost the same as the accuracy of object categorization based on visual space extracted from inception v3. This performance underscored the efficacy of the fMRI transformer network in feature extraction and fMRI space construction. In other words, the fMRI transformer was capable of extracting as much information as possible about the categories of images presented to the subjects using fMRI data. Sample results of the proposed model are shown in Figure 5. The second image was a mobile phone with a part of a finger. This small part of the human body was the reason for the label of the image as a category of person. This detection was hard for both inception V3 and fMRI transformer to capture. However, the first image was correctly categorized as a person class, and the third image was correctly categorized as a non-person class.

Table 5. Training parameters of the FC network

Training Parameters	Setting Values
Backbone	Inception V3 & fMRI transformer
Loss	Binary Cross Entropy
Optimizer	Adamax
Batch size	512
Epochs	100
Learning rate	0.0001
Dropout rate	0.2
Features size	73728 ($6 \times 6 \times 2048$)
Number of Features	26888
Image size	$256 \times 256 \times 3$
Number of Class	2

Table 6. Loss and similarity metrics of fMRI transformer

	Result	Values
Loss	Train	-0.868
	Validation	-0.865
	Test	-0.8695
Cosine similarity	Train	0.868
	Validation	0.865
	Test	0.8695

Table 7. Performance metrics of FC network

Result		Values	
		Inception V3 features	Prediction features
Loss	Train	0.5894	1.8061
	Validation	0.6366	0.7117
	Test	0.6333	0.7156
accuracy	Train	0.724	0.711
	Validation	0.672	0.656
	Test	0.675	0.665
PEI	Train	98.20%	
	Validation	97.61%	
	Test	98.51%	

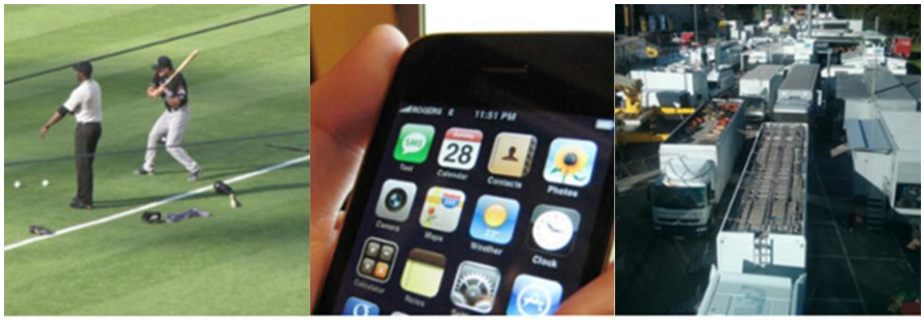

Figure 5. Sample results of proposed model

4. Discussion

This study emerges in the context of computational neuroscience's longstanding quest to decode sensory stimuli from neural information. In the past two decades, machine learning's role in decoding brain activity from fMRI data has substantially expanded in both its scale and quality. Initially, researchers focused on identifying object categories presented to individuals, and these pioneering studies laid the foundation for decoding stimulus categories.

Building on this foundation, recent fMRI studies have demonstrated that a wide array of visual features can be decoded from fMRI activity patterns, encompassing aspects such as orientation, spatial frequency, motion direction, object categorization, perceptual imagination, dreams, and even memory. The crux of these studies has been the utilization of classification-based machine learning methods to map the intricate relationship between brain activity patterns and stimulus categories, with a primary focus on the identification of stimulus categories.

However, a fundamental challenge arises when transitioning from experimental settings to real-world applications, where individuals vary significantly in terms of mental states, fatigue, and levels of attention. A substantial limitation arises from the need to construct individualized models, as structural and functional differences in people's brain architecture necessitate the establishment of subject-specific models. This presents a significant roadblock to the practical application of mind-reading models and demonstrates the need for subject-independent models that are able to process the various data of different subjects.

Notwithstanding these challenges, the application of fMRI remains a powerful tool. Its high spatial resolution and ability to capture data from diverse brain regions set it apart as a non-invasive method for comprehending neural activity.

Our research was built upon a robust foundation, utilizing the NSD dataset, which encompassed eight participants whose fMRI signals were recorded as they viewed images drawn from the COCO dataset. The choice of dataset was strategic, as it allowed us to operate within a real-world context, replicating the challenges and diversity encountered in practical scenarios. This dataset served as a critical testing ground for our methodology and demonstrated its viability in tackling real-world complexities.

To maximize the effectiveness of our approach, we focused our training on the data of the first three participants. These participants exhibited varying data sizes and voxel dimensions, mirroring the heterogeneity found in real-world applications. This approach was instrumental in addressing the challenge of voxel mismatch and served as a cornerstone in building a model capable of generalized application.

We addressed the challenge of data duplication between test and training datasets by adopting a prudent approach. Each image was utilized only once out of the 77,250 images, resulting in a final selection of 26,888 unique images. This rigorous approach not only prevented overfitting but also closely replicated the challenges posed by the practical usage of brain decoding models.

The heart of our innovation lies in the development of a transformer neural network. This network, designed specifically for fMRI data analysis, has the unique capability to address the voxel mismatch problem while standardizing information extracted across individuals. Our approach comprises two key components: the initial phase, where we employ a pre-trained network to extract features from each image constructing the visual space, and the subsequent phase, involving the use of a transformer network to extract features from fMRI signals constructing the fMRI space.

In addressing the challenge of different brain sizes among individuals, we've employed the overlapping technique, avoiding resizing or zero-padding. Our approach focuses on ensuring compatibility among the categorized areas by sharing information, effectively compensating for differences in brain size. This innovative technique is key to our ability to decode fMRI signals from individuals with diverse brain sizes.

Of particular note is our strategic decision to input the fMRI signal as a collection of regions related to the visual system of the brain. This approach leverages the transformer network's inherent capabilities, enabling it to measure the interconnections and attention levels among these regions and the voxels. This, in turn, effectively identifies the active regions within each area relevant to the target image. The adoption of the HCP mask, consisting of 380 distinct brain regions, facilitates this process. We've thoughtfully curated 200 regions from

this mask, each specifically pertinent to visual tasks, making them ideal for the extraction of visual features and enhancing the decoding process.

In sum, our methodology represents a significant leap forward in the field of fMRI data analysis and brain decoding. This research strives to make significant strides in rendering brain decoding more operationally viable. Our contributions, such as the transformer neural network and innovative model generalization methods, promise to facilitate the translation of computational neuroscience into real-world applications. By bridging the gap between laboratory-controlled experiments and diverse, real-world scenarios.

5. Conclusion

This study marks a substantial advancement in the field of computational neuroscience, aiming to make brain decoding more practically applicable. Over the last two decades, machine learning's role in decoding brain activity from fMRI data has grown significantly, particularly in object categorization. Recent research has shown the potential of fMRI data to decode various visual features. Yet, challenges arise in transitioning from controlled experiments to real-world applications, where individual variations like mental states are difficult to measure.

To address these challenges, our novel approach revolved around a transformer neural network designed for fMRI data analysis. This network tackled the voxel mismatch problem, offering the potential to decode images from the brain activity in a subject-independent manner. The proposed model with very few training parameters facilitated its use in scenarios with large datasets. The proposed network constructed an fMRI space, based on information from Bold signals, which was similar to the visual space of images extracted from inception V3. The cosine similarity was about 0.86 for the test dataset. Using constructed fMRI space and a FC network, 98.5% of accessible information about object categorization was extracted (PEI=98.5%). In essence, this study stands as a pioneering step toward making mind-reading models more accessible and applicable. This offers exciting prospects for the future of image reconstruction research and the broader field of neuroscience.

Acknowledgment

We would like to express our gratitude to the Institute of Communication and Information Technology and Dr. Mohammad Shahram Moin, which has supported this thesis according to contract number 9912 dated 09/02/2023.

References

1- Bing Du, Xiaomu Cheng, Yiping Duan, and Huansheng Ning, "fmri brain decoding and its applications in brain–computer interface: A survey." *Brain Sciences*, Vol. 12 (No. 2), p. 228, (2022).

2- David D Cox and Robert L Savoy, "Functional magnetic resonance imaging (fMRI)"brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex." *Neuroimage*, Vol. 19 (No. 2), pp. 261-70, (2003).

3- James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini, "Distributed and overlapping representations of faces and objects in ventral temporal cortex." *Science*, Vol. 293 (No. 5539), pp. 2425-30, (2001).

4- Wei Huang *et al.*, "Long short-term memory-based neural decoding of object categories evoked by natural images." *Human Brain Mapping*, Vol. 41 (No. 15), pp. 4442-53, (2020).

5- Alexander G Huth, Tyler Lee, Shinji Nishimoto, Natalia Y Bilenko, An T Vu, and Jack L Gallant, "Decoding the semantic content of natural movies from human brain activity." *Frontiers in systems neuroscience*, Vol. 10p. 81, (2016).

6- Alexander G Huth, Shinji Nishimoto, An T Vu, and Jack L Gallant, "A continuous semantic space describes the representation of thousands of object and action categories across the human brain." *Neuron*, Vol. 76 (No. 6), pp. 1210-24, (2012).

7- Tom M Mitchell *et al.*, "Predicting human brain activity associated with the meanings of nouns." *Science*, Vol. 320 (No. 5880), pp. 1191-95, (2008).

8- Sutao Song, Zhichao Zhan, Zhiying Long, Jiacai Zhang, and Li Yao, "Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data." *PloS one*, Vol. 6 (No. 2), p. e17191, (2011).

9- Chong Wang *et al.*, "'When'and 'what'did you see? A novel fMRI-based visual decoding framework." *Journal of Neural Engineering*, Vol. 17 (No. 5), p. 056013, (2020).

10- John-Dylan Haynes and Geraint Rees, "Predicting the orientation of invisible stimuli from activity in human

primary visual cortex." *Nature neuroscience*, Vol. 8 (No. 5), pp. 686-91, (2005).

11- Yukiyasu Kamitani and Frank Tong, "Decoding the visual and subjective contents of the human brain." *Nature neuroscience*, Vol. 8 (No. 5), pp. 679-85, (2005).

12- Leila Reddy, Naotsugu Tsuchiya, and Thomas Serre, "Reading the mind's eye: decoding category information during mental imagery." *Neuroimage*, Vol. 50 (No. 2), pp. 818-25, (2010).

13- Bradley R Postle, "The cognitive neuroscience of visual short-term memory." *Current opinion in behavioral sciences*, Vol. 1pp. 40-46, (2015).

14- Guy Gaziv *et al.*, "Self-supervised natural image reconstruction and large-scale semantic classification from brain activity." *Neuroimage*, Vol. 254p. 119121, (2022).

15- Sikun Lin, Thomas Sprague, and Ambuj K Singh, "Mind reader: Reconstructing complex images from brain activities." *Advances in Neural Information Processing Systems*, Vol. 35pp. 29624-36, (2022).

16- Guohua Shen, Kshitij Dwivedi, Kei Majima, Tomoyasu Horikawa, and Yukiyasu Kamitani, "End-to-end deep image reconstruction from human brain activity." *Frontiers in computational neuroscience*, Vol. 13p. 21, (2019).

17- Zijin Gu, Keith Jamison, Amy Kuceyeski, and Mert Sabuncu, "Decoding natural image stimuli from fMRI data with a surface-based convolutional network." *arXiv preprint arXiv:2212.02409*, (2022).

18- Emily J Allen *et al.*, "A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence." *Nature neuroscience*, Vol. 25 (No. 1), pp. 116-26, (2022).

19- Tsung-Yi Lin *et al.*, "Microsoft coco: Common objects in context." in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, (2014): *Springer*, pp. 740-55.