

Article - Engineering, Technology and Techniques

A Novel Approach of Dynamic Vision Reconstruction from fMRI Profiles Using Siamese Conditional Generative Adversarial Network

Rathi Karuppasamy^{1*}

<https://orcid.org/0000-0003-4000-6338>

Gomathi Velusamy¹

<http://orcid.org/0000-0003-3639-485X>

Raja Soosaimarian Peter Raj²

<http://orcid.org/0000-0002-7216-2207>

¹National Engineering College, Department of Computer Science and Engineering, Kovilpatti, Thoothukudi, Tamilnadu, India; ²Vellore Institute of Technology, School of Computer Science and Engineering, Vellore, Tamilnadu, India.

Editor-in-Chief: Alexandre Rasi Aoki

Associate Editor: Alexandre Rasi Aoki

Received: 04-May-2022; Accepted: 24-Mar-2023.

*Correspondence: rathikaruppasamy@gmail.com; Tel.: +91-9994768305 (R.K.).

HIGHLIGHTS

- We combined Siamese Neural Network with CGAN to create a SCGAN framework.
- Visual perception reconstructed by Conditional Generative Adversarial Network (CGAN).
- Reducing image loss, the linear version learns to be expecting the latent area out of Blood Oxygen Level Dependent.
- Reconstruction on video fMRI dataset was objectively identifiable.

Abstract: This paper aims to improve the quality of reconstructed visual stimuli and reduce the computational complexity of the visual stimuli reconstruction processes in the form of functional Magnetic Resonance Imaging (fMRI) profiles. The preceding work envisions the non-cognitive contents of brain activity vain to integrate visual data of diverse hierarchical levels. Existing approaches such as Deep Canonically Correlated Auto Encoder detect the significant challenges of reconstructing visual stimuli from brain activity: fMRI noise, large dimensionality of a limited number of data instances, and complex structure of visual stimuli. In this activity, we will also analyze the scope for utilizing the spatiotemporal data to resolve the neural correlates of visual stimulus representations and reconstruct the resembling visual stimuli. The purpose of this work is to manipulate those suffering from developmental disabilities. A novel Siamese conditional Generative Adversarial Network (ScGAN) approach is proposed to resolve these significant issues. The key features of ScGAN are as follows: 1. Siamese Neural Network (SNN) is a dimensionality reduction approach that takes as visual stimulus information alloy component and its goals to discover each of them effectively. It shows the critical component of visual stimuli. 2. In a conditional Generative Adversarial Network, the labels portrayan expansion to a latent variable to better generate and discriminate visual stimuli. Experiments on four fMRI datasets prove that our technique can reconstruct visual stimuli precisely. The performance metrics are evaluated by Mean Squared Error (MSE), Accuracy, Pearson Correlation Coefficient (PCC), Losses,

Structural Similarity Index (SSIM), Computational Time, etc. It proves that the proposed method yields better outcomes in terms of accuracy.

Keywords: Vision reconstruction; fMRI; visual cortex; encoding and decoding; CGAN; SNN.

INTRODUCTION

Scientists and intellectuals have been trying to understand and decode the cognitive process that enables humans to perceive and explore the visual world. We now explore how the human brain communicates active visual information with the outside environment and whether or not brain activity may be used to reconstruct what someone is seeing. These questions about brain decoding and encoding[1] were mostly about using artificial visual stimuli.[2-3] Such methods expose the underlying computing vision in a very narrow-centred way. A different approach that can navigate the complexities of natural vision and locate and interpret the visual representation of assigned brain activity is desired.

However, even as we can decipher the human brain activity estimated via fMRI into deep neural network features across more than one layer of the network[4], the enormous length of the visual stimuli features and the lack of normalization in the regression approach contribute to the low deciphering accuracy. Consequently, the reconstructed visual stimulus is slightly similar to the unique one. Some version improves the reconstructed visual stimuli to be much like real ones without using the categorical data of the visual stimuli. Another activity specializes in reconstructing a specific form of visual stimulus [5-6], which has enhanced information but lacks generality.

This article suggests SCGAN, a brand new version that reconstructs significant naturalistic visual stimuli from fMRI. Our approach comprises a Siamese and conditional generative adversarial network that allows a voxel-wise visual stimuli generation. In our work, a neural encoder based on the Siamese Neural Network[7] extracts a visual feature of an input visual stimulus. The fMRI decoder determines the mapping out of the fMRI information to the extracted visual features. In the SCGAN, the reconstruction results in the coarse visual stimuli of the decoded visual features. The CGAN creates a significant naturalistic visual stimulus by a coarse one.

Through quantitative analysis of the visual stimuli reconstruction using our approach, we have determined an improvement of the reconstructed visual stimuli quality among different approaches.

Related work

There are only a few experiments briefing perceived visual stimuli reconstruction in the published writing of human brain decoding. An amalgamation of multi-scale local stimuli bases with predetermined shapes to reconstruct lower-order data, namely binary contrast patterns.[8] Used Deep belief networks[9-10] and a simple linear Gaussian method[11] to reconstruct the handwritten digits and characters. Researchers are constructing a reconstruction method in which visual stimuli bases can be naturally evaluated using Bayesian canonical correlation analysis (BCCA).[12] Furthermore, work is being done to rebuild natural movie data.[13-15] Convolution at a deep level stacking deconvolution and convolution layers improves the generative adversarial network (GAN) design.[16]

The BigBiGAN method allows for the reconstruction of naturalistic visual stimulus. Because of BigBiGAN's latent space, this method generates conspicuous semantic data. It extracts visually appealing visual stimuli data from fMRI signal data.[17-19] By combining GAN with Bayesian learning, the GAN-based Bayesian Visual Reconstruction Model (GAN-BVRM) aims to improve a reconstructed visual stimulus quality from a finite data collection fusion.[18] The Deep Generative Multi-view Model (DGMM) demonstrated better reconstruction accuracy of visual images.[20] The Siamese Reconstruction Network (SRN) method uses the limited amount of training examples available. The several samples in the training data can be increased from n to $2n$ pairs using these methods.[21] The DVAE/GAN (Dual-Variational Auto Encoder/GAN) design successfully reduced visual stimulus clutter and noise and overpowered the modality gap.[22]

Table 1. Characteristics table of the algorithms. (KL-KL divergence, MSE-Mean Squared Error, Adv-Adversarial Loss, MAE-Mean Absolute Error, E2E-End-to-End Training)

Approach	Datasets	Loss	E2E Training	Pre-training
DCCGAN	BRAINS, GOD, vim-1	MSE, MAE	No	AlexNet based Comparator trained in ImageNet
DNN	Natural Images, Alphabetical Letters	MSE	No	AlexNet and VGG-19 based pre-trained on ImageNet
GAN	Natural Images, Artificial Shapes	MSE, Adv	Yes	CaffeNet based comparator pre-trained on ImageNet
GAN-BVRM	Vim-1	MSE	No	Generator of BigGAN pre-trained on ImageNet
DVAE/GAN	6-9, BRAINS, GOD	KL, Adv	No	Model pre-trained on external data from ImageNet
SSGAN	Natural Images	MAE, Adv	No	-
BigBiGAN	GOD	Adv	No	BigBiGAN pre-trained on ImageNet

Motivation and justification

- The preceding work envisions the non-cognitive contents of brain activity vain to integrate visual data of diverse hierarchical levels.
- Recent research initiatives have revealed the possibility of establishing the neurological correlates of voxel information with their associated visual image.
- In this activity, we will also analyze the scope for utilizing the spatiotemporal data to resolve the neural correlates of visual stimulus representations and reconstruct the resembling visual stimuli via the Deep Learning approach.

Contributions to the work

- The voxel-wise encoding approaches exhibited various single-voxel portrayals and revealed their category representations.
- The prior work shows that the reconstruction tasks are evaluated only by accuracy and losses. This paper evaluates the existing datasets of reconstruction tasks using ten metrics and numerous parameters.
- The current work shows various approaches to reconstructing natural visual stimuli such as images, movies, etc. This work uses a unique algorithm to reconstruct multiple data sets.

Outline of the work

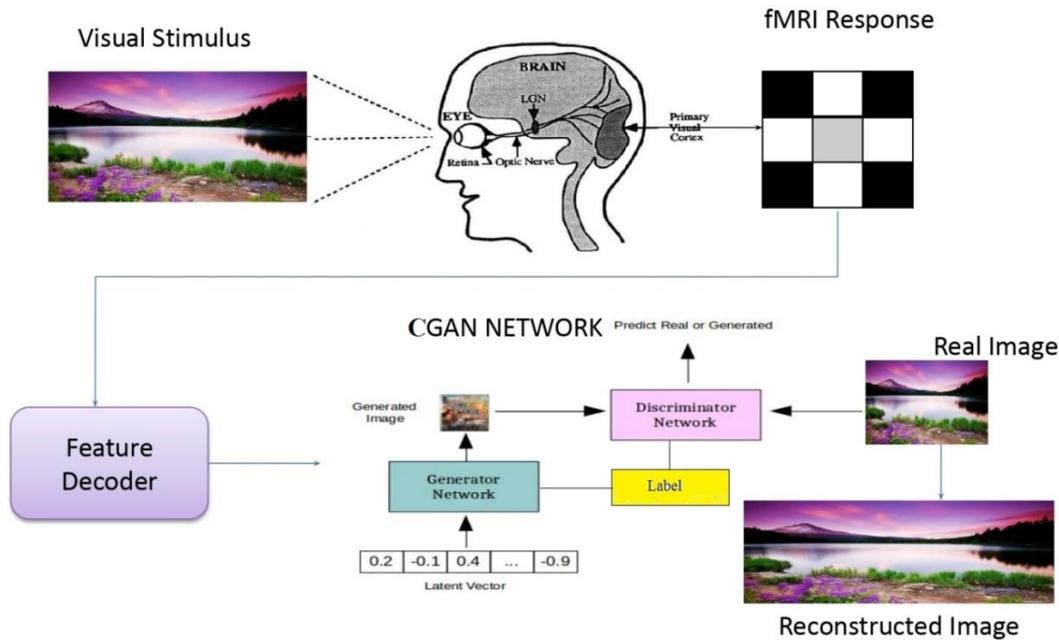


Figure 1. Framework Diagram for Siamese conditional Generative Adversarial Network

Organization of the paper

The remaining paper is organized as follows: The approach to the visual stimulus reconstruction challenge is briefly discussed in Section 2. The experimental setup of the suggested approach for natural movie reconstruction using fMRI profiles is presented in Section 3. Section 4 explains the suggested approach's performance analysis utilizing fMRI activity, and Section 5 draws a conclusion and identifies areas for further research.

MATERIAL AND METHODS

Siamese Neural Network (SNN)

The SNN[Figure 2] is a category of neural network framework in which two or more sub-networks are identical. "Identical" in this context relates to a similar configuration, comprising equal weights and parameters. Improving parameters is mimicked over sub-networks, and it is utilized to compare visual stimuli feature vectors to discover similarities between two visual stimuli. The Siamese Neural Network encodes specific features of a visual stimulus. The key benefits of the Siamese networks are more resistance to unbalanced classes; it nice to have a group with the best classifier; semantic similarity as a source of information. The disadvantages are that it takes longer to train than traditional networks and cannot generate probabilities. The mean squared error loss and binary cross entropy loss are incorporated in the training phase. As a result, the objective function in the Siamese network model is,

$$L_s(x_i^a, x_i^t) = y(x_i^a, x_i^t) \log p(x_i^a, x_i^t) + (1 - y(x_i^a, x_i^t)) \log p(1 - y(x_i^a, x_i^t)) \dots (1)$$

Where, x_i^a it stands for the feature vector and x_i^t stands for the voxel vector, it uses trained samples to calculate the Siamese loss function of the reconstruction process.

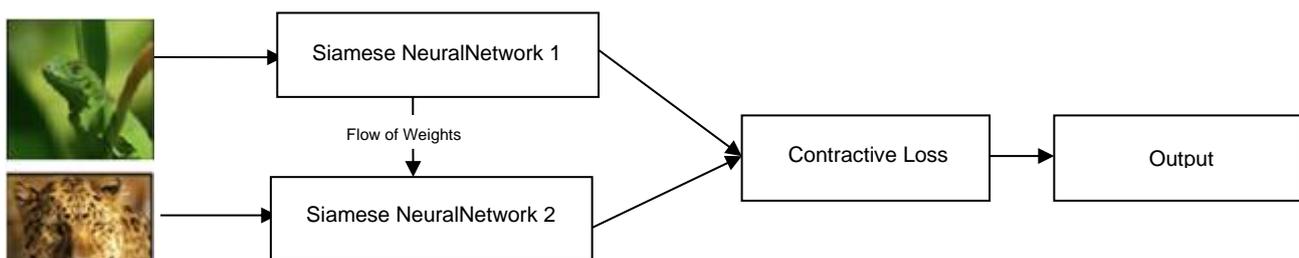


Figure 2. Siamese Neural Network Architecture

Generative Adversarial Network (GAN)

Supervised learning is built on enormous volumes of information; however, we occasionally run across data shortages. In this case, a generative model is utilized to convert supervised learning to semi-supervised learning. The GAN network is proposed by Ian Goodfellow. [23] Generative Adversarial Network concurrently trains two networks. The Discriminator (D) net accepts an input visual stimulus of a training set and returns a scalar indicating the likelihood that the input is in the training set. The Generator (G) generates visual images from an image vector taken at random with simple Gaussian distribution. The generator net learns to create visual stimuli to deceive the discriminator, while the discriminator learns to differentiate between genuine and produced data. For this reason, Generative Adversarial Network[Figure 3] is “adversarial”. In another way, the discriminator and generator play a min-max game over the value function $V(G; D)$ as follows:

$$\min_G \max_D V(D, G) = \min_G \max_D \{ E_{x \sim P_{data}(x)} [\log(D(X))] + E_{z \sim P_Z(z)} [\log(1 - D(G(z)))] \} \tag{2}$$

$P_{data}(x)$ denotes the actual distribution of instances, and $P_Z(z)$ denotes a latent space distribution computed over the random vector z .

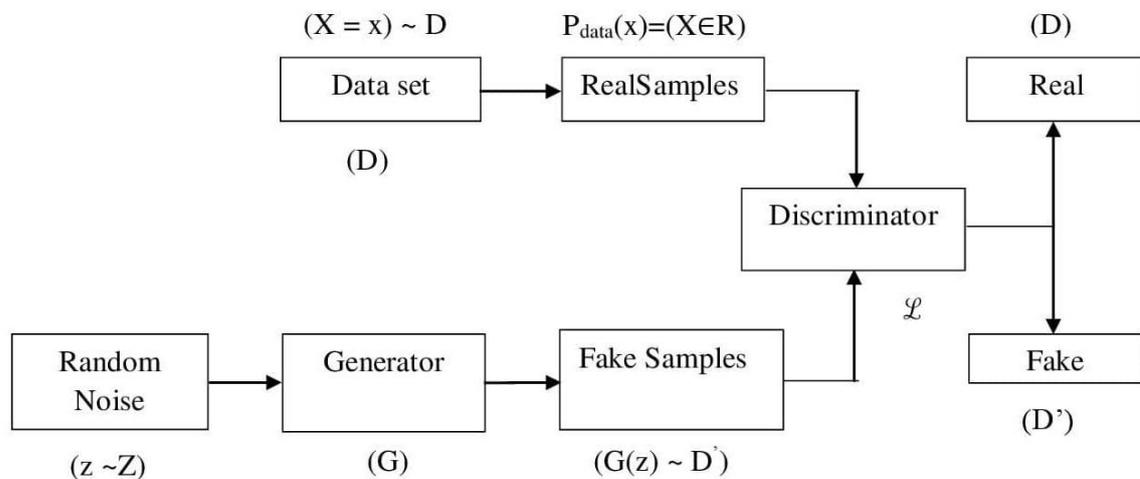


Figure 3. GAN Architecture

Conditional Generative Adversarial Network (CGAN)

GAN was able to produce some better samples of data points, but it was unable to create data points with the goal label, and the dataset produced lacked diversity. Mirza [24] introduced the CGAN framework[Figure 4], changed the architecture by adding additional label information y to the generator’s input and attempted to create the appropriate data point. It also adds additional information to the discriminator input to make it easier to discern between original and duplicate data.

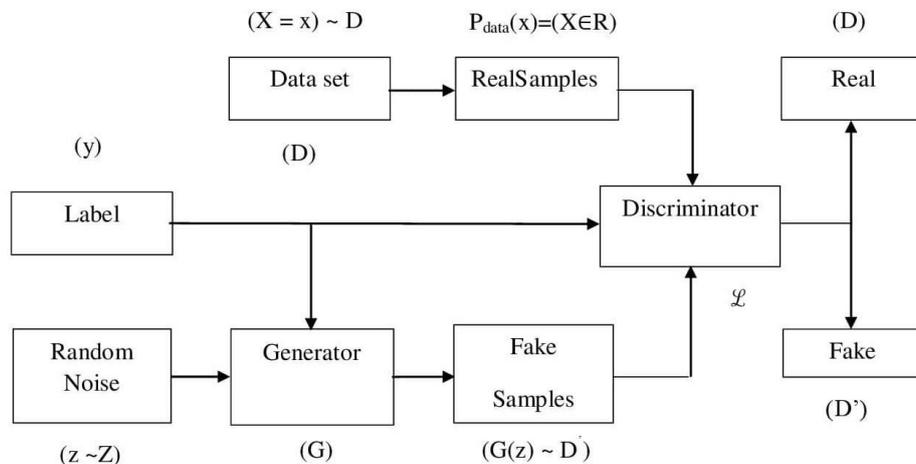


Figure 4. Architecture of CGAN

The joint hidden representation in this framework combines the random input noise Z with the label y , and the GAN training phase provides plenty of flexibility in how it accepts input. The discriminator receives the data points X and y and the generative output $G(z)$, the same as in the vanilla GAN network. The conditional generative adversarial network loss function is comparable to the GAN:

$$\min_G \max_D V(D, G) = \{E_{x \sim P_{data}(x)} [\log(D(x | y))] + E_{z \sim P_Z(z)} [\log(1 - D(G(z | y)))]\} \quad (3)$$

Proposed Method – Siamese Conditional Generative Adversarial Network (SCGAN)

This segment defines the novel method for reconstructing the visual images from brain activity profiles as the Siamese conditional generative adversarial network model.

The training set comprises visual images and fMRI activity patterns, which are denoted as I and F , respectively. SCGAN [Figure 1] is an approach for mapping between the data space F and latent space Z . In this algorithm, an encoder network is a Siamese Neural Network (SNN), and a decoder network is a Conditional Generative Adversarial Network (CGAN). The shared latent variables are treated as the following Gaussian prior distribution,

$$p(Z) = \prod_{i=1}^N \mathcal{N}(z_i | 0, I) \quad (4)$$

SNN is made up of two symmetric neural network sub-networks with equal weights. In our work, we used one convolutional layer succeeded by three fully connected layers to create each identical neural network. All layers were activated using the rectified linear units (ReLU) nonlinearity, and the adaptive moment estimation (ADAM) optimizer was utilized to modify the learning rate. In this case, the feature vectors of that original data are denoted as f_{vi} , and their output is indicated as 'o'. It produces new input (I_n) data with $(N+1)$ dimensional space.

$$I_n = (f_{vi}^T, o^T) \quad (5)$$

Normalize each feature can be expressed as follows,

$$N = (f_{vi} - m_i) / 2\sigma_i \quad (6)$$

This normalized feature information is fed into the decoding network of the conditional generative adversarial network. The encoder function is $E: I \rightarrow Z$.

The decoder network, the generator, and the discriminator of the generative models are conditioned on some additional information (a). For generating the equivalent data, a new additional parameter 'a' is added to the generator in CGAN. It can execute the conditioning by additional data 'a' offers an extra input layer to the generator (G) and the discriminator (D).

The preceding primary noise $P_Z(Z)$ of input data and additional data 'a' are mixed in combined hidden representations in the generator. The adversarial training paradigm provides a great deal of flexibility in the composition of this hidden representation. The learning function of the generator network is trained alongside two discriminator networks named D_N and D_Z , which learn to distinguish between the original and reconstructed data and latent points. The decoder function is $D: Z \rightarrow I$.

The SCGAN objective function is mathematically expressed as,

$$\min_G \max_D V(D, G) = \min_G \max_D \left\{ E_{I \sim P_{data}(I)} \left[\log \left(D \left(\frac{I}{a} \right) \right) \right] + E_{z \sim P_Z(z)} \left[\log \left(1 - D \left(G \left(\frac{z}{a} \right) \right) \right) \right] \right\} \quad (7)$$

From the CGAN network,

Next, find the best discriminator among all the discriminators that are possible for this particular optimization. There may be multiple discriminators that can satisfy this optimization, but how to find the best discriminator D . For Generator (G) fixed, the optimal discriminator D obtained as,

$$D_G^* = \frac{P_{data}(x)}{P_{data}(x) + P_g(x)} \quad (8)$$

The training criterion for the discriminator D , given any generator G , is to maximize the loss function equation. Hence, the optimal discriminator for a given G denoted as,

$$D_G^* = \arg \max_D V(D, G) \quad (9)$$

So, D_G^* is the optimal value as well as the maximum value.

The role of the generator is in reverse of that of D so that the optimal value of G could minimize the loss function occurs, when $D = D_G^*$. So, the optimal solution G^* as,

$$G^* = \arg \min_G V(D_G^*, G) \quad (10)$$

At this point, the optimization problem stated in (A) has a unique solution G^* also, this solution satisfies $P_g = P_{data}$. so, the optimal D_G^* is,

$$D_G^* = \frac{P_{data}(x)}{P_{data}(x) + P_g(x)} \quad (11)$$

Natural Image reconstruction procedure

Algorithm for SCGAN

Require: Reconstruction of Visual Stimuli (I).

Input: Visual Stimuli $I \in \mathbb{R}$, fMRI Pattern $F \in \mathbb{R}$.

Output: actual samples $P_{data}(I)$, label a, latent space variable Z, reconstructed sample I.

Initialize:

- 1) Set the visual image training set to I and fMRI activity as F. Here, $I \in \mathbb{R}$ and $F \in \mathbb{R}$. Set the latent variables indicated as $Z \in \mathbb{R}$. \mathbb{R} : Coefficient value of the Image.
- 2) Elect the voxels that address the visual stimuli. Such voxel activity is assessed using the coefficient of determination implied as \mathbb{R}^2 .
- 3) Enumerate a \mathbb{R}^2 value through 5-fold cross-validation.
- 4) Extract the features of stimuli implied as $P_{data}(I)$ via SNN.
- 5) For each training recurrence
 - a. Discriminator training:
 - i. Select the stochastic mini batch of actual samples $P_{data}(I)$ and their equivalent labels (a) and it indicated as (I, a).
 - ii. Calculate D (I, a) as a mini-batch and back propagate the loss to amend $\theta^{(D)}$ use to reduce a loss function.
 - iii. Choose the mini batch of latent space variable (z) and additional information (a) and it is implied as (z, a). It produces a counterfeit sample of visual stimulus. $G(z, a) = I^* | a$
 - iv. Evaluate D ($I^* | a, a$) as the mini-batch and back propagate the loss to modify $\theta^{(D)}$ using the loss function minimization.
 - b. Generator training:
 - i. Select the mini batch of the random noise vector (z) and additional data (a) and it indicated as (z, a). It creates a counterfeit instance of the visual stimulus. $G(z, a) = I^* | a$.
 - ii. Calculate D ($I^* | a, a$) as the mini-batch, and the loss should be back propagated to amend $\theta^{(G)}$ used to minimize the loss function.

EXPERIMENTAL RESULTS ANALYSIS & DISCUSSIONS

Dataset Description

This section provides an overview of publicly accessible datasets utilized in Deep Learning-based visual stimuli reconstruction out of fMRI activity profiles. At the same time, many datasets are available for reconstructing the visual stimulus, such as natural movies. Table 1 shows several distinguishing aspects of datasets.

Natural Movies:[14] It carries 374 video clips out of random video blocks and YouTube videos at a pixel resolution of 800*600, and the number of voxel information in this dataset is 10214. The fMRI activity pattern taken from the Region of interest is V1, V2, V3, V4, LOC, PPA, FFA, TPJ, LIP, FEF, and PEF.

Table 2.The dataset description of Natural movies dataset.

Dataset	No.of.Instances	Pixel Resolution	No.of.Voxels	ROIs
Natural Movies	374	800*600	10214	V1, V2, V3, V4, LOC, PPA, FFA, TPJ, LIP, FEF, and PEF.

Voxel Preferences

Most voxels do not acknowledge visual images; voxel selection is crucial for brain decoding techniques. One popular strategy is to select the voxels correlated with a visual stimulus during the training section. We select voxels for which approach gives improved predictability in encoding operation.

It validates our idea that the visual stimulus's best-predicted voxels should be involved in the decoding model. The \mathbb{R}^2 (Coefficient of Determination), which reflects a percentage of variance described with the model, quantifies the model fit between data predictions and computes the voxel activity. In this task, we first estimated the voxel on training phase data using fivefold cross-validation and chose the voxel information with a positive coefficient of determination for further study.

Performance Metrics

The performance of visual stimuli reconstruction of Natural movies data is calculated using the succeeding performance metrics. The description and formula of each metric are shown in Table 3.

Table 3.Performance Metrics Description

S.No	Metric	Formula	Description
1.	Mean Squared Error	$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2$	Mean Squared Error (MSE) is the most basic statistic for evaluating the quality of visual stimulus reconstruction. Given X_i and Y_i , the original stimuli and the reconstructed stimuli are flattened representations of one-dimensional space.
2.	Pearson's Correlation Coefficient	$PCC(x, y) = \frac{\sum(x - \mu_x)(y - \mu_y)}{\sqrt{\sum(x - \mu_x)^2 \sum(y - \mu_y)^2}}$	Pearson's Correlation Coefficient (PCC) is a popular statistical tool for determining the linear correlation amid two variables.
3.	Structural Similarity Index	$SSIM(p, q) = \frac{(2\mu_p\mu_q + C1)(2\sigma_{pq} + C2)}{(\mu_p^2 + \mu_q^2 + C1)(\sigma_p^2 + \sigma_q^2 + C2)}$	Structural Similarity Index (SSIM) is a popular similarity metric of images for extracting structural information from visual images. The likeness of spatially close information of pixels among the original and reconstructed visual stimulus is measured by SSIM. Here, μ and σ indicate the mean and variance of the p and q values; σ_{pq} indicates the covariance value of p and q; C1 and C2 are constants.
4.	Propensity Score Matching	$d(x, y) = \sum_t \frac{1}{H_t W_t} \sum_{h,w} \ w_t \cdot (f_x^t - f_y^t)\ _2^2$	The conditional likelihood of assignment to a specific treatment specified by a vector of observable covariates is a propensity score matching. Where, w_t is denoted as channel-wise activation, $d(x,y)$ denotes the distance amid the reconstructed (y) and original visual stimuli(x).

Cont. Table 3

5.	Maximum Mean Discrepancy	$M_k(P_r, P_g) = E_{x, x' \sim P_r} [k(x, x')] - 2E_{x \sim P_r, y \sim P_g} [k(x, y)] + E_{y, y' \sim P_g} [k(y, y')]$	The Maximum Mean Discrepancy (MMD) method compares two probability distributions, Pr and Pg, by utilizing samples selected randomly from each distribution.
6.	Inception Score	$IS = \exp \left(E_{x \sim p} \frac{D_{KL}(p\{y x\} p(y))}{KL \text{ Divergence}} \right)$	In adversarial networks, Kullback-Leibler divergence is calculated by comparing the conditional and marginal label distributions over the generated data.
7.	Frechet Inception Distance	$FID = \ \mu_X - \mu_Y\ ^2 + Tr(\Sigma_X + \Sigma_Y - 2\sqrt{\Sigma_X \Sigma_Y})$	The Wasserstein-2 distance between Gaussian fitted multi-variants data stored in a feature space. Where μ is the mean value, and Σ is the covariance of the real (X) and false (Y) generated images.
8.	Accuracy	$Accuracy = \frac{\text{Number of accurate prediction}}{\text{Total number of prediction}}$	One of the measures used to measure accuracy is the percentage of correct predictions made from the total number of input samples.
9.	Loss	$E_x[\log D(x)] + E_z[\log(1 - D(G(z)))]$	The min-max function is utilized in this paper. The CGAN network's generator seeks to minimize loss, but the discriminator seeks to maximize loss function.

Results and Discussion

The evaluation of visual stimulus reconstruction tasks based on performance metrics is presented below:

Performance analysis of Reconstruction task based on Performance Metrics

Table 4 to Table 13 show a performance evaluation of the natural movie reconstruction task based on performance metrics discussed in Table 2.

Table 4. Performance Evaluation based on PCC (mean±std)

Dataset	Work	Approach	PCC(mean±std)
Natural Movies	Existing:	Existing:	
	Wen et.al.[15]	CNN	.252±.141
	Ian et.al. [23]	GAN	.278±.146
	Qiao et.al. [18]	GAN-BVRM	.309±.152
	Jiang et.al. [21]	SRN	.327±.157
	Ziqi et.al. [22]	DVAE/GAN	.352±.168
	Proposed:	Proposed:	
Ours	SCGAN	.358±.174	

Table 5. Performance Evaluation based on MSE (mean±std)

Dataset	Work	Approach	MSE(mean±std)
Natural Movies	Existing:	Existing:	
	Wen et.al.[15]	CNN	.092±.127
	Ian et.al. [23]	GAN	.088±.120
	Qiao et.al. [18]	GAN-BVRM	.070±.116
	Jiang et.al. [21]	SRN	.064±.110
	Ziqi et.al. [22]	DVAE/GAN	.054±.107
	Proposed:	Proposed:	
Ours	SCGAN	.046±.104	

Table 6. Performance Evaluation based on SSIM (mean±std)

Dataset	Work	Approach	SSIM(mean±std)
Natural Movies	Existing:	Existing:	
	Wen et.al.[15]	CNN	0.180±.170
	Ian et.al. [23]	GAN	0.188±.153
	Qiao et.al. [18]	GAN-BVRM	0.192±.145
	Jiang et.al. [21]	SRN	0.200±.138
	Ziqi et.al. [22]	DVAE/GAN	0.303±.135
	Proposed:	Proposed:	
	Ours	SCGAN	0.410±.126

Table 7. Performance Evaluation based on PSM (Standardized difference between real and reconstruct samples)

Dataset	Work	Approach	PSM
Natural Movies	Existing:	Existing:	
	Wen et.al.[15]	CNN	0.986
	Ian et.al. [23]	GAN	0.976
	Qiao et.al. [18]	GAN-BVRM	0.968
	Jiang et.al. [21]	SRN	0.949
	Ziqi et.al. [22]	DVAE/GAN	0.943
	Proposed:	Proposed:	
	Ours	SCGAN	0.937

Table 8. Performance Evaluation based on MMD (Difference along with the samples feature mean)

Dataset	Work	Approach	MMD
Natural Movies	Existing:	Existing:	
	Wen et.al.[15]	CNN	0.95
	Ian et.al. [23]	GAN	0.88
	Qiao et.al. [18]	GAN-BVRM	0.85
	Jiang et.al. [21]	SRN	0.81
	Ziqi et.al. [22]	DVAE/GAN	0.78
	Proposed:	Proposed:	
	Ours	SCGAN	0.75

Table 9. Performance Evaluation based on IS (mean ± std)

Dataset	Work	Approach	IS (mean ± std)	
			Real Data	Ours
Natural Movies	Existing:	Existing:		
	Wen et.al.[15]	CNN	28.45± .22	16.38± .30
	Ian et.al. [23]	GAN		17.54± .28
	Qiao et.al. [18]	GAN-BVRM		18.90± .25
	Jiang et.al. [21]	SRN		19.99± .21
	Ziqi et.al. [22]	DVAE/GAN		20.72± .19
	Proposed:	Proposed:		
	Ours	SCGAN		21.8± .18

Table 10. Performance Evaluation based on FID (mean ± std)

Dataset	Work	Approach	FID (mean ± std)	
			Real Data	Ours
Natural Movies	Existing:	Existing:		
	Wen et.al.[15]	CNN	34.32± .96	22.82± .80
	Ian et.al. [23]	GAN		23.54± .82
	Qiao et.al. [18]	GAN-BVRM		25.90± .85
	Jiang et.al. [21]	SRN		26.74± .81
	Ziqi et.al. [22]	DVAE/GAN		27.25± .84
	Proposed:	Proposed:		
	Ours	SCGAN		28.80± .86

Table 11. Performance Evaluation based on Accuracy (%)

Dataset	Work	Approach	Accuracy (%)
Natural Movies	Existing:	Existing:	
	Wen et.al.[15]	CNN	40
	Ian et.al. [23]	GAN	35
	Qiao et.al. [18]	GAN-BVRM	35
	Jiang et.al. [21]	SRN	38
	Ziqi et.al. [22]	DVAE/GAN	42
	Proposed:	Proposed:	
Ours	SCGAN	44	

Table 12. Performance Evaluation based on Losses up to 5000 epochs

Dataset	Work	Approach	Loss	
			Generator	Discriminator
Natural Movies	Existing:	Existing:		
	Wen et.al.[15]	CNN	2.5310	0.7445
	Ian et.al. [23]	GAN	2.4962	0.9524
	Qiao et.al. [18]	GAN-BVRM	2.5463	1.1289
	Jiang et.al. [21]	SRN	2.4264	1.3221
	Ziqi et.al. [22]	DVAE/GAN	2.4519	1.0815
	Proposed:	Proposed:		
Ours	SCGAN	2.4803	1.0926	

Table 13. Performance Evaluation based on Computational Complexity

Dataset	Work	Approach	Time (s)
Natural Movies	Existing:	Existing:	
	Wen et.al.[15]	CNN	168
	Ian et.al. [23]	GAN	150.8
	Qiao et.al. [18]	GAN-BVRM	162.4
	Jiang et.al. [21]	SRN	145.6
	Ziqi et.al. [22]	DVAE/GAN	130.1
	Proposed:	Proposed:	
Ours	SCGAN	120.9	

Experimental Result

To test the SCGAN model, we have used the natural movies dataset, a high-grade quality dataset, with the fMRI profile recording of three subjects while presented with the images. The dataset consists of the BOLD activity of images from voxels (10214) in visual cortex areas of the brain, such as V1, V2, V3, V4, LOC, PPA, FFA, TPJ, LIP, FEF, and PEF. Actual images in the work use data information from subject 1.

The reconstructed natural movies by the proposed SCGAN algorithm are displayed in Figure 5. The initial row indicates the actual visual stimulus, and the next row indicates reconstructed images of the visual stimulus. The accuracy rate of reconstructed images is 44%.

**Figure 5.** Visual Stimulus reconstruction for Natural Movies dataset.

DISCUSSION

This research provides a growing body of knowledge on using deep learning approaches to model and understand human brain activity representations of visual stimuli. It builds on the previous finding with static and dynamic visual stimuli by extending the suggested method (SCGAN) to characterize and interpret fMRI activity as dynamic and static visual images. These findings support the hypothesis that the process of feed-forward directed at visual stimuli identification has a significant impact on the brain responses underlying dynamic images, not just for a ventral stream but also, to a lesser extent, for a dorsal stream. It provides information about the visual representation of the dorsal stream.

Despite the lack of repetition or neural feedback links, the SCGAN allows the construction of a fully quantifiable prediction approach of brain activity responses to every visual input. The voxel-by-voxel encoding method for depicting single-voxel depictions demonstrates the distinct activities of different brain areas during vision.

It also develops a high throughput for synthesizing brain activities to visual stimuli, enabling brain mapping of class portrayal and selection without using fMRI trials. Furthermore, this study enables decoding brain fMRI responses in semantic and visual spaces, allowing for real-time visual stimulus reconstruction.

CONCLUSION

To solve the challenge of visual stimulus reconstruction, we introduced a novel approach using a Siamese conditional generative adversarial network model. Through latent variables, we can find various relationships between fMRI activity pattern voxels and visual stimuli pixels. We also created a prediction distribution that successfully recreated visual stimuli using fMRI activity patterns. Although we focused on the difficulty of visual stimuli reconstruction in this paper, we can also use our method to solve encoding problems. The superiority of the offered approach has been proven by extensive experimental testing.

In the future, we will have two demanding and hopeful options. In our method, we can first study the reconstruction of active vision using several GAN models. Second, we may investigate multi-subject decoding by using each subject's fMRI signals as a single entity.

Funding : This research received no external funding.

Conflicts of Interest : The authors declare no conflict of interest.

REFERENCES

1. Naselaris T, Kay KN, Nishimoto S, Gallant JL. Encoding and decoding in fMRI. *NeuroImage*. 2011; 56(2):400–10.
2. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nat Neurosci*. 2005; 8:679–85.
3. Haynes JD, Rees G. Decoding mental states from brain activity in humans. *Nat Rev Neurosci*. 2006; 7:523–34.
4. Shen G, Horikawa T, Majima K, Kamitani Y. Deep image reconstruction from human brain activity. *PLOS Computational Biology*. 2019;15(1):e1006633. Available from: <https://doi.org/10.1371/journal.pcbi.1006633>
5. VanRullen R, Reddy L. Reconstructing Faces from fMRI Patterns using Deep Generative Neural Networks. *Communication Biology* 2.2019; 193. Available from: <https://doi.org/10.1038/s42003-019-0438-y>
6. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 July 21–26; Honolulu, HI, USA. IEEE; 2017. p. 5967–5976.
7. Jiang L, Qiao K, Wang L, Zhang C, Chen J, Zeng L, et al. Siamese Reconstruction Network: Accurate Image Reconstruction from Human Brain Activity by Learning to Compare. *Appl. Sci*. 2019;9: 4749.
8. Miyawaki Y, Uchida H, Yamashita O, Sato M, Morito Y, Hiroki CT, et al. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*. 2008; 60(5):915–29.
9. Marcel AJ, Van Gerven, Botond Cseke, Floris P De Lange, Tom Heskes. Efficient bayesian multivariate fmri analysis using a sparsifying spatio-temporal prior. *NeuroImage*. 2010; 50(1):150–61.
10. Marcel AJ Van Gerven, Floris P, De Lange, Tom Heskes. Neural decoding with hierarchical generative models. *Neural Comput*. 2010; 22(12):3127–42.
11. Schoenmakers S, Barth M, Heskes T, Van Gerven M. Linear reconstruction of perceived images from human brain activity. *NeuroImage*, 2013. 83:951–61.
12. Fujiwara Y, Miyawaki Y, Kamitani Y. Modular encoding and decoding models derived from bayesian canonical correlation analysis. *Neural Comput*. 2013; 25(4):979–1005.

13. Nishimoto S, Naselaris T, Benjamini Y, An T Vu, Bin Yu, Gallant JL. Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 2011; 21(19):1641–1646.
14. Wen H, Shi J, Zhang Y, Lu KH, Cao J, Liu Z. Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision. *Cerebral Cortex.* 2018; 28(12):4136-4160. Available from: <https://doi.org/10.1093/cercor/bhx268>
15. Seeliger K, Guclu U, Ambrogioni L, Gucluturk Y, vanGerven MAJ. Generative " adversarial networks for reconstructing natural images from brain activity. *NeuroImage.* 2018; 181: 775–785. Available from: DOI: 10.1016/j.neuroimage.2018.07.043
16. Mozafari M, Reddy L, VanRullen R. Reconstructing Natural Scenes from fMRI Patterns using BigBiGAN. In 2020 International Joint Conference on Neural networks(IJCNN);2020 July 19-24;Glasgow, United Kingdom.IEEE; 2020.p 1-8.
17. Qiao K, Chen J, Wang L, Zhang C, Tong L, Yan B. BigGAN-based Bayesian Reconstruction of Natural Images from Human Brain Activity. *Neuroscience.* 2020; 444: 92–105. Available from: <https://doi.org/10.1016/j.neuroscience.2020.07.040>
18. Donahue J, Simonyan K. Large Scale Adversarial Representation Learning. In 2019 3rd Conference on Advances in Neural Information Processing Systems (NeurIPS), 2019 Vancouver, Canada.vol. 32.Available from: <https://doi.org/10.48550/arXiv.1907.02544>
19. Du C, Du C, He H. Sharing deep generative representation for perceived image reconstruction from human brain activity. *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017 May 14-19; Anchorage, AK, USA.IEEE; 2017. p. 1049-1056. Available from: DOI: 10.1109/IJCNN.2017.7965968
20. Jiang L, Qiao K, Wang L, Zhang C, Chen J, Zeng L, Bu H, Yan B. Siamese Reconstruction Network: Accurate Image Reconstruction from Human Brain Activity by Learning to Compare. *Appl. Sci.* 2019; 9: 4749.
21. Ren Z, Li J, Xue X, Li X, Yang F, Jiao Z, Gao X. Reconstructing seen image from brain activity by visually-guided cognitive representation and adversarial learning, *NeuroImage.* 2021; 228:117602. ISSN 1053-8119. Available from: <https://doi.org/10.1016/j.neuroimage.2020.117602>
22. Goodfellow IJ, Abadie JP, Mirza M, Xu B, Farley DW, Ozair S, Courville A, Bengio Y. Generative Adversarial Networks. In *NeurIPS.* 2014; 2672-2680.
23. Mirza M, Osindero S. Conditional Generative Adversarial Nets. *ArXiv.* 2014. Available from: <https://doi.org/10.48550/arXiv.1411.1784>



© 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).