2018 4th International Conference on Computing Communication and Automation (ICCCA)

Human Vision Reconstruction using Brain Activity Profiles

K.Rathi
Research Scholar,Dept.of.CSE
National Engineering College
Kovilpatti,TamilNadu,India
rathikaruppasamy@gmail.com

Dr.V.Gomathi
Prof&Head,Dept.of.CSE
National Engineering College
Kovilpatti,TamilNadu,India
vgcse@nec.edu.in

***Abstract*— The visual cue understandings of the human frequently happen at the occipital lobe of the brain with a fantastic mapping process. Nowadays, the ability of reconstruction of visual cues from the human cerebrum activities using functional Magnetic Resonance Imaging (fMRI) is an active research field under neuropsychological research. This research would bring scope for those suffering on the autism spectrum disorder (ASD) or those affected by psychotic disorders. This capability is also called a theory of mind (ToM). Also, this ability has evolved a long time, especially in primates, because of selection pressure brought about by increasing social complexity. This speculation is called "social brain hypothesis." Few researchers have suggested mapping techniques that can reconstruct a digital image of what a person is picturing in their mind, by merely considering their electrical activity of the brain as fMRI voxel information. During the gathering of voxel information for experiments, it usually is required to display the participants the MNIST digit images, faces of many individuals, one-by-one, on the computer screen, in parallel to recording event of their brain activity. There are open research challenges to develop efficient reconstruction models to predict the potential space of mapping the human vision from voxel data. Primarily, this mapping algorithm identified the pattern of brain signals related to the mental image and reproduced it as a digital image. The primary goal is to propose a new stimuli recreation technique, in which the voxel data and corresponding pixel information will correlate with an in-depth learning approach based on latent-variable distributions. This reconstruction method mainly depends on the observed brain activity patterns in the form of physiological modalities fMRI or EEG. In this work, we will also study the scope for exploiting the spatiotemporal EEG data to decide the neural associates of visual scene portrayals and to recreate the presence of the comparing stimuli.**

Keywords*— *fMRI, Cognitive Neuroscience, Encoder and Decoder, DGMM, BPFGS, Recurrent Neural Network, Long Short Term Memory.

I. INTRODUCTION

Cognitive neuroscience is an interdisciplinary study area of psychology and neuroscience. An exciting research field in this domain is building the mathematical model for how the psychological activities correlate to the physiological neural circuitry of a human.EEG is a traditional and non-intrusive method for checking electrical movement of the cerebrum by following the Intl.10-20 System. The EEG signals usually recorded using individual EEG sensors [EMOTIV wireless Kit, Cognionics wearable EEG Cap, Bio Semi] with essential temporal resolutions concerning data sampling rate (min.128 – 512 samples/sec) and positioning of minimum 2 to 256 electrodes as the EEG recording channels.

Considerable efforts will devote by the researchers working with EEG data model to affect the domain, Cognitive Neuro-feedback system and solving motor imagery-related tasks. Their works mainly focus the human emotion analysis, cognitive / brain disorders, linguistic modeling, etc. In addition to that, the functional magnetic resonance imaging measures the brain activity profile by changing the blood flow in the form of voxels, to recreate the recognized images rightly from the fMRI activity.

In visual decoding, the functional magnetic resonance imaging (fMRI) is a useful device for decoding the activity of the brain. Most of the research produce actualized the characterization of image group, dreams, memory, and imagination using multivoxel pattern analysis. Multiscale local images with predefined shapes were used to reconstruct the lower order information of binary contrast pattern[1]. The handwritten characters constructed by a straightforward linear Gaussian approach[2].In the proposal of the reconstruction model, the visual image reconstruction has limited representation power [3].

Improve the reconstruction accuracy of this process; the posterior regularization is helping to constrain the testing examples and are near their neighbors from the preparation set[4]. A nonlinear extension of the BCAA was formulated using a DGMM [5]. Technical innovations of deep neural networks are helping to know about the hierarchical visual processing in computational neuroscience [6]. The fMRI activity patterns to the DNN features of viewed images predicted by the developed decoders [7]. Encoding and decoding models are the basic approaches for reconstructing the image (low base image or exemplary image) from the human brain activity. It is not suitable for combined the multiple hierarchical level features even though sophisticated decoding and encoding models. So its need todevelop [8].

Instead of hierarchical neural representations of a human visual system, the DNN visual features used in reconstructing an image from the human brain activity. In this process, the fMRI pattern decoded into DNN features and it also produces a similar output [9]. The early visual cortex of lower BOLD signal is the response to faces the dissension see had been as of now displayed than for the new faces [10].fMRI is used to restrict areas in monkey cerebrum. And its produced the stronger response to face compared to other objects, so this region preferred for the electrophysiological analysis [11].

The privilege ATL and the fusiform gyrus is the arrangement of ventral stream locales identified by the bold response (same face with different expression) after averaging together. It has data about individual pictures of faces[12]. Inquiry of face distinguishing proof by the valuable, attractive reverberation imaging it's a homologous investigation, so it's the main reason for the cortical wellspring of this data ascribed to fusiform gyrus. Fusiform base face space visual features used for facial image reconstruction. And these processes are not considered as a temporal aspect of a face processing [13].

II. MATERIALS AND METHODS

A. Experimental Data

Handle these experiments on two public fMRI datasets. Dataset1 contains a hundred manually written dark scale digits(equal number of 9s) at a 28*28 pixel goals taken from the preparation set of MNIST database and the fMRI information from visual areas(V1, V2, V3). The fMRI information of V1, V2 taken from three subjects at a 56*56 pixel goals of 360 dark scale manually written characters (B, R, A, I, N, S) displayed in Dataset2. For every aspect, 60 singular examples midway exhibited amid the test. The outlined subtle elements of both the dataset in Table 1:

TABLE.I:SUMMERIZED DETAILS OF DATASET

Datasets	**No.of Instances**	**No. of pixels**	**No. of voxels**	**No. of ROIs**	**No. of Training**
Dataset1	100	784	3092	V1,V2,V3	90
Dataset2	360	784	2420	V1,V2	330

To keep members sharpness, they were formally asked for a center on the obsession point and also to react with a catch squeeze when the obsession area changed shading. The obsession area changed shading once every five upgrades by and large. Changes were introduced aimlessly yet equally spread over the length of the examination. The analysis went on for 40 min with an individual-managed relax period in the center. Later the investigation, a supplementary output was constructed.

The useful pictures were gathered with a Siemens Trio 3 T MRI framework with an EPI arrangement utilizing a 32 channel loop. Head improvement restricted with froth pads and a tight bit of tape over the brow. Useful imaging, an auxiliary output obtained. In an alternate session, the useful localizer data was picked up, again using an EPI gathering During securing, a visual tracker was used to affirm if members were focusing their look.

With the utilization of SPM8 programming, the utilitarian volumes were remade, changed to the first sweep of the session and cut time revised. Members moved under mm across the sessions. For every unique image, which was exhibited two times to the subject, the reaction of each voxel to the image was processed utilizing a direct model.

III. RECURRENT NEURAL NETWORK

A Recurrent Neural Network (RNN) is associations between hubs outline a planned graph along a grouping. The RNNs is to make use of the following information. In a traditional neural network, all inputs and outputs are independent of each other. A typical RNN looks like:

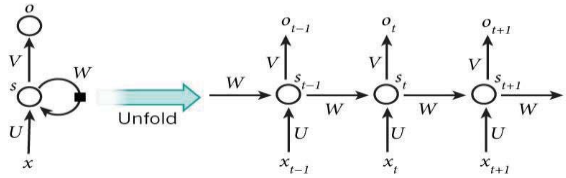


FIG I. RNN ARCHITECTURE

Algorithms:

Step 1: A solitary time venture of the input (x_t) provided to the system.

Step2: Calculate its current state (h_t) using a utilizing a mix of the current info and the Previous state.

$$h_t = f(h_{t-1}, x_t)$$

Step 3: The initiation work is tanh, the weight at the repetitive*,* w_{hh} ,and the weight at the info neuron is neuron is w_{xh}.

$$h_t = \tanh(w_{hh} h_{t-1} + w_{xh} x_t)$$

Step 4: Once all the time steps are finished the last current state is utilized to figure to figure the yield y_t.

$$y_t = w_{hy} h_t$$

Step 5: The output (y_t) contrasted with the actual yield (y_t)and the error created.

$$E_t(y_t, y_t) = -y_t \log(y_t)$$
$$E_t(y, y) = -\sum y_t \log(y_t)$$

A. Recurrent Neural Network – Encoder – Decoder

In the traditional study, the linearizing feature space model is used to predict the activity of the human brain. But in this model used for specific semantic categories. In this paper, we use the RNN Encoding model instead of the linearizing model. Encoding model is more touchy than the different model, and it makes a non-direct mapping between the visual stimuli and also the brain activity signal. Two recurrent neural networks used for RNN Encoder – Decoder. One is used to encode the visual stimuli into the vector representations, and another one has used to decodes the representation into the visual stimuli.

Use hidden units to improve the memory capacity of this model. During the training sequence, the RNN learn a probability distribution and each time step of output is the conditional distribution. The Hidden state has the updated and the reset gates and its motivated by the Long short-term memory. To maximize the accuracy of the brain voxel activity, we use the optimal feature for every voxel. RNN

Encoder – Decoder to maximize the conditional log-likelihood.

$$\max_{\theta} \frac{1}{N}\sum_{n=1}^{N} \log p_{\theta}(y_n \mid x_n)$$

Without the gate units, the model is not able to get the result meaningful.

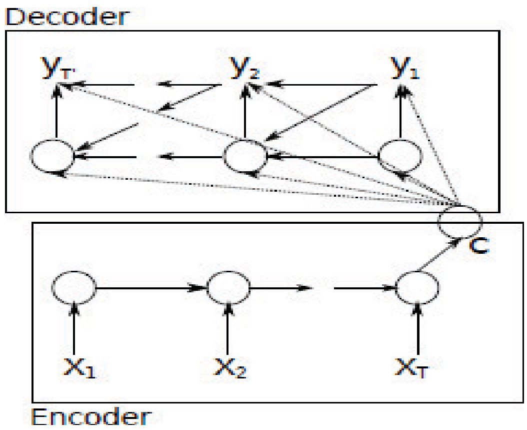


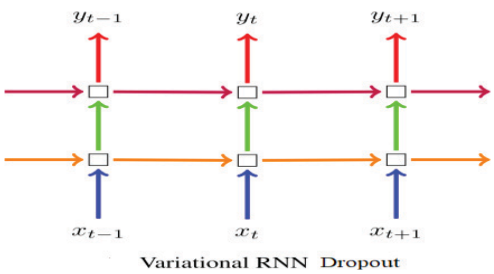
Fig 2: RNN ENCODER – DECODER

B. Long-Short Term Memory

Long haul dependency learned by this techniques. In this model, the recurrent module has a different chain structure. It is well suited for processing and classification task. LSTM units are input gate, memory cell gate, output gate and forget barrier.

C. Dropout in Recurrent Neural Network

Every time step the dropouts created in recurrent networks. In RNN, the dropout is used only with input and output data. In case, all the various layers including input and output layer use the same dropout in each time step, and its create the variational inference.



Now, assume a simple Recurrent Neural Network, the input sequence denoted as x, and hidden state denoted as h_t with the time step, and weight denoted as w.

$$h_t = f_h(x_t, h_{t-1}) = \sigma(x_t w_h + h_{t-1} U_h + b_h)$$

D. Dropout in LSTM

Four available gates used in LSTM, such as input, output, forget and input modulation.

$$o = sigm(h_{t-1}U_o + x_t w_o)$$

$$g = \tanh(h_{t-1}U_g + x_t w_g)$$

$$c_t = f^{\circ} c_{t-1} + i^{\circ} g$$

$$i = sigm(h_{t-1}U_i + x_t w_i)$$

$$f = sigm(h_{t-1}U_f + x_t w_f)$$

$$h_t = o^{\circ} \tanh(c_t)$$

IV. MATHEMATICAL-MODEL

A. Deep Generative Model [14]:

A generative model is a data distribution using unsupervised learning. Whereas in statistics, a generative model is a model for generating all values in the phenomenon.

Visual images and fMRI activity pattern denoted as x and y respectively and also introduced the shared latent variable z.

$$p(z) = \prod_{i=1}^{N} \mathrm{N}(z_i \mid 0, I)$$

When noises observed in the image with zero mean and diagonal covariance matrix in voxel activation. Then the Gaussian distribution function is given by

$$p_{\theta}(x \mid z) = \prod_{i=1}^{N} \mathrm{N}(x_i \mid \mu_x(z_i), diag(\sigma_x^2(z_i)))$$

In fMRI activity, the non-linear transformation is more powerful, and it’s used to suppress the noise and predict the information. Activity pattern of fMRI has a projection matrix and covariance matrix, the likelihood function is,

$$p(y \mid z) = \pi_{i=1}^{N} \mathrm{N}(y \mid B^{\tau} z_i, \Psi)$$

In this case, fMRI voxels are highly correlated. Inferring high dimensional covariance matrix Ψ, introduce the auxiliary latent variable Ā, the low-rank assumption model to decrease the computational complexity.

$$P(\bar{A}) = \prod_{i=1}^{N} \mathrm{N}(\bar{A} \mid 0, I)$$

Rewriting,

$$P(y \mid z, \bar{A}) = \prod_{i=1}^{N} \mathrm{N}(y \mid B^{\tau} z_i + H^{\tau}\bar{A}, \gamma^{-1} I)$$

KL Divergence is to calculate the distinction between two likelihood circulations over similar factors. In variational circulation idea, the KL dissimilarity equation is,

$$\max_{\theta} \frac{1}{N}\sum_{n=1}^{N} \log p_{\theta}(y_n \mid x_n)$$

From KL divergence

$$P(z \mid x) = P(x \mid z)P(z) / \int_z P(x,z)dz$$

$$D_{KL}(Q \,||\, P) = \sum_z Q(z)[(\log Q(z) / P(z \mid x)) + \log P(x)]$$

$$\log P(x) = D_{KL}(Q \,||\, P) - \sum_z Q(z)[\log Q(z) / \log P(z \mid x)]$$

$$\log P(x) = \mathrm{E}[-\log Q(z) / \log P(z,x)] + D_{KL}(Q \,||\, P)$$

$$\log P(x) = \mathrm{E}[-\{\log Q(z) - \log P(z,x)\}] + D_{KL}(Q \,||\, P)$$

$$\log P(x) = \mathrm{E}[\log P(z,x) - \log Q(z)] + D_{KL}(Q \,||\, P)$$

Continuous version of KL Divergence is,

$$D_{kl}(P(x) \,||\, Q(x)) = \int_{-\infty}^{\infty} P(x) \ln \left[\frac{p(x)}{q(x)}\right] dx$$

Prediction distribution of visual images denoted as $x_{pred}$and the brain activity is y_*, the posterior conveyance is written by,

$$p(x_{pred} \mid y_*) = \int p(x_{pred} \mid z_*) p(z_* \mid y_*) dz_*$$

B. Linear Reconstruction Model [15]:

The Gaussian decoding model, a parameter is estimated in the presence of the different regularization techniques.

Let (x, y) signify an upgrade reaction combine in Gaussian deciphering,

$$X = \arg\max(p(x \mid y))$$

At that point the forward encoding model in multivariate Gaussian with zero mean and covariance matrix(R) is,

$$p(y \mid x) \alpha \exp(-\frac{1}{2} x^T R^{-1} x)$$

The canonical form of multivariate Gaussian distribution,

$$X = (R - RB(\sum 1 + B^T RB)^{-1} B^T R) B \sum{}^{-1} y$$

Here, B is the regression coefficients.

The covariance matrix for the earlier is written by,

$$R = \frac{1}{N-1}(\sum_n z^n (z^n)^T)$$

Regression coefficient b_k to solve the minimization problem,

$$b_k = \arg\min \left\{ \frac{1}{2N} || y^k - Xb ||_2^2 + R_a(b) \right\}$$

Relapse coefficients for all voxels in shut shape utilizing,

$$\hat{B} = (X^T X + \tilde{G})^{-1} X^T Y$$

For each voxel k, explained variance was figured in understanding to,

$$\hat{r}_k = (\mathrm{var}(y_k) - \mathrm{var}(y_k - \hat{y}_k)) / \mathrm{var}(y_k)$$

C. Visual image reconstruction using local image decoders [16] :

This approach provides perceptual states from cerebrum action while finding data portrayal in multivoxel design

To anticipate the mean differentiation of every nearby picture components. A discriminant capacity of different class k in a neighborhood decoder communicated as,

$$y_{wk}(r) = \sum_d^D w_k^d r^d + w_k^D$$

Using softmax function,

$$p_w(k \mid r) = \frac{\exp[y_{wk}(r)]}{\sum_j^k \exp[y_{wj}(r)]}$$

The weight parameter has zero mean Gaussian distribution with a change, whose backward treated as hyperparameter,

$$p(w_k^d \mid a_k^d) = \mathrm{N}(0, \frac{1}{a_d^k})$$

p $(a_d^k) = \frac{1}{a_d^k}$ is treated as a random variable. The output of the local image decoder is given by,

$$\hat{I}(x \mid r) = \sum_m^M \lambda_m c_m(r) \phi_m(x)$$

D. DNN feature decoding [17] :

Deep Neural Network (DNN) has a certain level of complexity and to process data complexly, using sophisticated mathematical modeling. It is a feed forward network with many hidden layers. In this method, sparse linear regression algorithm is used to select the vital voxels for decoding.In single DNN layer reconstruction is given below and it reduced the optimization problem.

$$v^* = \arg\min \frac{1}{2} \sum_{l \in L} ||\phi^{(l)}(v) - y^{(l)}||_2^2$$

Combine DNN feature with multiple layers is given by,

$$v^{*} = \arg\min \frac{1}{2} \sum_{l \in L} \beta_{l} || \phi^{(l)}(v) - y^{(l)} ||_2^2$$

This cost function minimized by Limited Memory BFGS algorithm. This algorithm was solving unconstrained values in non-optimization problems.

V.COMPARATIVE ANALYSIS

The deep generative model implemented the perceived image reconstruction problem, and it helps to derive the predictive distribution to recreate the visual stimuli from cerebrum action and also deal with encoding tasks. This method has high computational complexity.

Whereas, Linear reconstruction model, provide the high quality of reconstruction from the stimuli, by inverting the properly encoding reconstruction model. During encoding and decoding task, the performances analyzed with regression.

Perform the constraint-free visual image reconstruction, discriminant functions are used, which deals with the restoration of the optical image at multiple scales. This method provides information about the activity which discovered from the human brain activity.

The recent method, DNN feature decoding was used to minimize the cost function of layers using the algorithm of LMBFGS, and it has solved the non-optimization problem on reconstruction with unconstraint values.

VI CONCLUSION

This paper gives a new technique for reconstructing the visual images from fMRI signals based on Recurrent Neural Network. Feature visualization technique to create a problem in image reconstruction. RNN Encoder-Decoder is a novel way to investigate the visual representations, and its fixed to every individual voxel and visualize the patterns. The intricate images are predicted accurately using the RNN Encoding model. In ancient technique has a high computational complexity. So, in future our team members to work with Graphics Processing Unit (GPU) for reducing the time for reconstructing the stimuli.

REFERENCES

[1] Yoichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa-aki Sato, Yusuke Morito, Hiroki CTanabe, NorihiroSadato, and Yukiyasu KamitanI. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron,60(5):915-929, 2008.

[2] SanneSchoenmakers, Markus Barth, Tom Heskes, and Marcel van Gerven. Linear reconstruction of perceived images from human brain activity.NeuroImage, 83:951-961,2013.

[3] Yusuke Fujiwara, Yoichi Miyawaki, and YukiyasuKamitani. Modular encoding and decoding models derived from Bayesian canonical correlation analysis. Neural computation, 25(4):979-1005, 2013.

[4] Jun Zhu, Ning Chen, and Erica P Xing, Bayesian inference with posterior regularization and applications to infinite latent svms. Journal of Machine Learning Research, 15(1):1799-1847, 2014.

[5] Diederik P Kingma and Max Welling. Auto-encoding variationalbayes.In ICLR, 2014.

[6] Yamins, D.L.,&Dicarlo, J.J. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 19, 356-365,2016.

[7] Nguyen, A., Dosovitskiy, A., Yosinski, J., Brox, T., &Clune, J. Synthesizing the preferred inputs for neurons in neural networks via deep generator networks. Adv. Neural Inf. Process. Syst. 29,3387-3395, 2016.

[8] Horikawa,T., &Kamitani,Y. Generic decoding of seen and imagines objects using hierarchical visual features. Nat. Commun. 8,15037,2017.

[9] Mur M, Ruff DA, Bodurka J, Bandettini PA, Kriegeskorte N. Face-identity change activation outside the face system: “Release from adaptation” may not always indicate neuronal selectivity, Cereb Cortex , 209: 2027-2042, 2010.

[10] FreiwaldWA,TsaoDY.Functionalcompartmentalization and viewpoint generalization within the macaque face-processing system, Science, 30:845-851, 2010.

[11] Nestor A, Plaut DC, Behrmann M. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis, ProcNatlAcadSci USA , 108 : 9998-10003, 2011.

[12] Anzellotti S, Fairhall SL, Caramazza A(2014) Decoding representations of face identity that are tolerant to rotation. Cereb Cortex 24:1988-1995. CrossRef Medline.

[13] Nestor A, Plaut DC, Behrmann M(2016) Feature-based face representations and image reconstruction from behavioral and neural data. ProcNatlAcadSci USA 113:416-421. CrossRef Medline.

[14] Changde Du,Changying Du,Huiguang He : Sharing deep generative representation for perceived image reconstruction from human brain activity.

[15] Sanne Schoenmakers, Markus Barth, Tom Heskes, Marcel Van Gerven: Linear reconstruction of perceived images from human brain activity. NeuroImage.YNIMG- 10695.pg:11;4C:4,6,7,8,9.(2013).

[16] Hajime Uchida, Okito Yamashita, Masa-aki sato : Visual image reconstruction from human brain activity using a combination of multiscale local image decoders.Neuron 60,pg:915-929.[2008].

[17] Guohua Shen, Tomoyasu Horikawa, Kei Majima,and Yukiyasu Kamitani : Deep image reconstruction from human brain activity.BioRxiv.pg:1-19.(2017).