AN EVENT-CONTRASTIVE CONNECTOME NETWORK FOR AUTOMATIC ASSESSMENT OF INDIVIDUAL FACE PROCESSING AND MEMORY ABILITY

Wan-Ting Hsieh[1,2], *Hao-Chun Yang*[1,2], *Fu-Sheng Tsai*[1,2], *Chon-Wen Shyi*[3], *Chi-Chun Lee*[1,2]

[1]Department of Electrical Engineering, National Tsing Hua University, Taiwan
[2]MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
[3]Department of Psychology, National Chung Cheng University, Taiwan

ABSTRACT

Human adapt their behaviors by continuously monitoring one another to function socially in our society. The ability to process face identity from memory is a crucial basic capability. In this work, we propose an event-contrastive connectome network (E-cCN) in representing brain's functional connectivity with contrastive loss to handle layers of fMRI data variabilities exists under different controlled stimuli events to achieve improved automatic assessing of an individual's face processing and memory ability. Our proposed connectome network achieves an overall recognition accuracy of 80.20% and 82.05% in binary classification of separating high versus low scoring subjects on tasks of Taiwanese Face Memory Test (TFMT) and component inverse efficiency score (cIE) respectively. Further, our network embedding representation demonstrate distinct connectivity patterns in key face processing brain regions (ROIs) when comparing between high and low face processing and memory ability.

Index Terms— contrastive loss, face processing and memory, fMRI, connectome embedding

1. INTRODUCTION

Studies in social psychology has proposed an ABC model of attitude stating that humans rely on three basic capabilities, i.e., affect, behavior, and cognition, in order to achieve successful social interactions with one another in our society [1, 2]. Among these capabilities, the ability to express and interpret social behaviors that convey demands toward social engagement is critical in carrying out smooth in-person interactions. These social intents are often communicated through eye contact, face perception, and language expression. In fact, the ability to recognize face identity is one of the key components, studies have shown that people who are better at face identity learning tend to build better interpersonal relationships and blend into society more easily [3]. On the other hand, the ones suffer from neurodegenerative disorder that causes deficits in face identity and memory (e.g., prosopagnosia frontotemporal dementia and schizophrenia) further demonstrate symptoms of social deficits [4, 5]. Quantitative assessment of face processing and memory skill is usually carried out through a series of psychological testings, e.g., Cambridge Face Memory Test (CFMT) [6], component task, and configural task [7].

Many studies have already brought neuroscientific insights into human brain's functional connectivity and the mechanism of an individual's face processing and memory. For example, Mehrnam et al. measured electroencephalography (EEG) to performance differentiation between subjects of innocent from guilty who participated in concealed face recognition test [8], and Lynn et al. found that increased face perception impairment between autistic adolescents and adulthood is due to abnormal connectivity between fusiform face area (FFA) and other brain regions through fMRI studies [9]. Other studies have also dedicated in understanding this relationship using fMRI while subjects going through various design of face recognition tasks [10]. While extensive works have brought evidences into the brain functional networks when processing face and memory, there is no computational work on predicting individual face and memory processing ability using the collected fMRI data directly.

In this work, we propose an event-contrastive connectome network (E-cCN) to robustly represent the brain's functional connectivity for discriminative task. The E-cCN is learned based on Node2Vec graph embedding approach [11] jointly optimized with contrastive loss criterion (enhancing intra-event compactness and inter-event dispersion) [12] across the three different experimental events (object, neutral face, expressive face) conventially used in the fMRI experimental protocol. We evaluate our framework in representing subject's brain functional connectivity to perform automatic assessment of their corresponding Taiwanese Face Memory Test (TFMT) and component inverse efficiency score (cIE) [13]. Our framework achieves an unweighted accuracy of 80.2 % and 82.05 % in 2-class TFMT and cIE tasks respectively. The use of contrastive loss improves the recognition rates by 6.48% and 1.71% relative on TFMT and cIE respectively over Node2vec. Further, we demonstrate there exists distinct pattern of brain network structure embedded between high performing face processing and memory subjects compared to those with poor performance using our E-cCN.

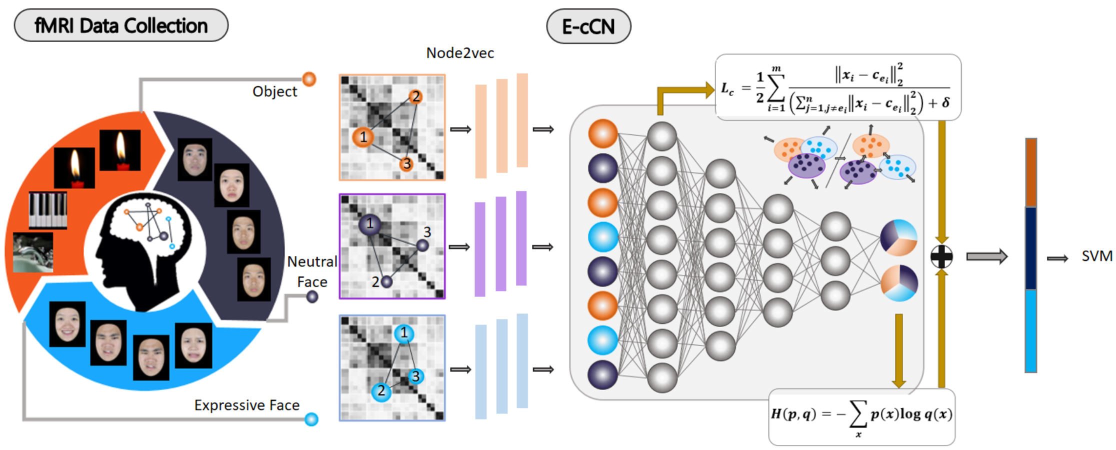

Fig. 1. *A schematic of our proposed E-cCN architecture in performing automatic face recognition ability decoding.*

Table 1. *Summary of 19 face-related ROIs*

Right
IFGOR(#1), Amygdala(#3), ATL(#5), Calcarine(#7), FFA(#9), Hippocampus(#11), mSTS(#13), OFA(#15), OP Junction(#16), Precuneus(#17), pSTS(#19)
Left
Amygdala(#2), ATL(#4), Calcarine(#6), FFA(#8), Hippocampus(#10), mSTS(#12), OFA(#14), pSTS(#18)

2. RESEARCH METHODOLOGY

2.1. fMRI Data Collection and Preprocessing

We recruited 44 subjects conducting one-back working memory (WM) test. During fMRI scanning, subjects were stimulated by a series of face and object images in separate blocks, and they were instructed to press the button when they identified that the current image was the same as the previous one. There were three different **events** chosen for the experimental manipulation, i.e., *object*, *neutral face*, and *expressive face*.

The face images were collected from a previous study by Shyi et al. [14], while the images of object were gathered from the internet. The WM task began and ended each with a fixation block. There were 18 image blocks with 17 fixation blocks in between (each block lasted for12s). fMRI scanning was performed on 3T scanner with 8-channel phase-array head coil. A high-resolution T1-weighted 3D-SPGR anatomical scan was acquired for co-registration between structural and functional images (TR/TE=2000/33ms, voxel size=$3 * 3 * 3mm^3$, 40 slices). We performed all necessary pre-processing steps using SPM8 [15]. In this work, we selected 19 face processing ROIs as our mask (table 1) [16].

2.2. Face Perception and Memory Tests Assessments

In this work, our goal is automatic assessing face processing and memory ability by modeling the subjects fMRI data. Each subject underwent two different assessment tasks after fMRI scanning. The two tests used here were the Taiwanese Face Memory Test (TFMT) and the component task.

2.2.1. Taiwanese Face Memory Test (TFMT)

In this test, participants were asked to accurately identify a given target face within the three faces displayed during each trial. The whole test comprised of 30 trials of novel-image stage and 24 trials of novel-image with varying degrees of noise (54 total trials for each subject). The TFMT score for each subject was calculated as total number of accurate identification divided by the number of trials (a score between 0 to 1).

2.2.2. Component Test

There were a total of 48 trials in the component task, each trial presented a pair of faces that could be identical or different from each other. When they were not same, they differed in the *components* of the face, e.g., mouth or eyes. The test was used to assess how fast an subject can accurately answer whether the pair of faces were distinct. Assessment score of component inverse efficiency score (cIE) was used to quantify speed-accuracy tradeoffs [17].

$$\textit{Inverse Efficiency Score (IES)} = \frac{RT}{1 - PE} \tag{1}$$

where RT and PE are reaction time and proportion of error. Finally, we binarize both TFMT and cIE into high versus low as our recognition label using the median value as cut-off.

2.3. Event-Contrastive Connectome Network (E-cCN)

Figure 1 depicts schematic of our proposed event-contrastive connectome network (E-cCN). The E-cCN is learned by jointly optimizing contrastive-loss across three fMRI experimental manipulations (**events**). It is a two-stage procedure: 1) perform Node2Vec on functional connectivity matrix of different blocks, and 2) contrastive loss embedded network trained after Node2Vec graph embedding.

2.3.1. Graph Embedding of Functional Connectivity

We conduct Node2Vec on the functional connectivity matrix that models the connectome through local and global random walks from a brain region [11]. First, we calculate Pearson correlation coefficients between nodal voxel-averaged time

Table 2. *It presents the binary face recognition ability classification results of our proposed E-cCN and other modeling techniques. E1, E2 and E3 denote 3 different experimental events, object, neutral face and expressive face. The accuracy is measured in UAR(%).*

	TFMT				**cIE**			
	E1-Object	E2-Nface	E3-Eface	Fusion	E1-Object	E2-Nface	E3-Eface	Fusion
ICA	55.79	55.68	60.95	62.32	60.47	65.38	59.40	61.32
PCA	64.32	60.32	62.32	64.95	49.16	58.55	54.91	59.40
Graph-SVM	64.32	64.32	64.32	67.57	61.32	61.32	62.39	65.17
DeepWalk	53.05	51.58	63.47	62.95	61.32	70.73	68.80	68.80
LINE	55.78	45.16	51.79	59.07	58.33	53.21	64.74	64.32
Node2Vec	58.55	72.65	69.02	73.72	63.68	78.42	63.68	80.34
E-cCN	74.95	71.58	62.95	**80.20**	79.27	76.71	79.49	**82.05**

series of each experimental event, i.e., object, neutral face and expressive face. Then each correlation matrix can be turned into a vectorized representation serves as a brain connectome representation for each *event* block after Node2Vec embedding. The related parameter settings are as below: there is one iteration including 500 random walks with a length of 8 steps. The output dimension is 50, and the window size is 4. The return parameter p is 1, and in-out parameter q is 1.6.

2.3.2. Contrastive Loss Embedded Network

Functional connectivity obtained from fMRI images under each experimental event includes composite variability, e.g., when stimulated with a face, it may create a composite reaction of *seeing* something and *observing* faces in the measured brain activity. This composite variability leads to undesirable effect in the recognition task. It is further confounded with other brain activity, e.g., experimentation tasks, emotional states and psychological conditions.

We propose to enhance the discriminative power of the Node2Vec feature using an event-condition connectome network (E-cCN). The E-cCN network is learned by simultaneously optimizing with contrastive loss criterion [12] and cross entropy loss, L_{CE}. The contrastive loss, L_c, is embedded to the first layer of E-cCN by explicitly centering the representation with respect to each event-center in the following form:

$$L_c = \frac{1}{2}\sum_{i=1}^{m}\frac{\|x_i - c_{e_i}\|_2^2}{(\sum_{j=1,j\neq e_i}^{m}\|x_i - c_j\|_2^2) + \delta} \tag{2}$$

where m is the number of events, x_i is the data under condition i and c_j is the center of j event determined by the mean value (c_{e_i} denotes to the similar meaning of event e_i). δ is set as 1 preventing the denominator equal to 0. Finally, our total loss L_{total} of the complete network is:

$$L_{total} = L_c + L_{CE} \tag{3}$$

2.4. TFMT and cIE Classification

We derive our fMRI feature representation by extracting E-cCN output from the first layer in order to compare with representation extracted from Node2Vec only. Since each participant undergoes 6 blocks per type of event resulting in 6 different representations per event, we use mean pooling to obtain per-*event* fMRI representation as input to the linear-kernel support vector machine for TFMT and cIE classification.

3. EXPERIMENTAL SETUP AND RESULTS

We carry out high versus low classification for both TFMT and component cIE task. The evaluation scheme is done via leave-one-person-out cross-validation, and the accuracy is measured in unweighted average recall (UAR). Univariate feature selection is also carried out in the training set.

3.1. Experiment Setup

The E-cCN architecture is composed of 5 fully-connected layers with node dimensions at each layer to be 950-512-256-64-2. The total loss function is composed of contrastive loss on 950-dimension representation and cross entropy loss for the discriminative task. The batch size and epoch are set 64 and 50 respectively. Network trained on contrastive loss and cross entropy loss using Adam with lr = 0.001 and 0.01, respectively. The activation function is ReLU with batch normalization. We extract the first layer (950-dimension) as subject's feature representation.

We compare our framework with the following method to derive fMRI feature representation:

- **ICA**: Perform independent component analysis on BOLD signal.
- **PCA**: Perform principal component analysis on BOLD signal.
- **Graph-SVM**: Perform functional correlation matrix of without neural network embedding.
- **DeepWalk**: Perform DeepWalk on functional correlation matrix [18].
- **LINE**: Perform LINE embedding on functional correlation matrix [19].
- **Node2Vec**: Perform Node2Vec embedding on functional correlation matrix without contrastive loss embedding [11].

These features are then fed into final TFMT and cIE classification procedure.

3.2. Experiment Results and Discussions

Table 2 summarizes our complete experimental results. The proposed E-cCN obtains the best accuracy compared to all other baseline systems, especially when using representation learned from the *object*-event; it obtains 74.95% recognition rate in TFMT and 79.27% in cIE. We further perform feature-level fusion by concatenating representation from three different events, denoted as 'Fusion' in Table 2. The E-cCN Fusion improves the accuracy to 80.2% and 82.05% (5.25% and 2.78% relative improvement over object only). When comparing with all of the baseline systems, performance obtained by our proposed E-cCN is significantly better than all systems indicating the contrastive-based joint optimization network can indeed help uncover the discriminative information in the subject's brain functional connectivity embedding.

In general, by introducing the graph embedding approach Node2Vec, which considers global walks across brain network providing a better brain connectivity representation, the performance improves 6.15% in TFMT and 15.17% in cIE from Graph-SVM. Besides, the contrastive loss embedding(E-cCN) further helps improve 6.48% and 1.71% in the recognition accuracy when comparing to Node2Vec. In addition, we observe that specifically in face condition, contrastive loss of E-cCN may slightly negatively impacts the results, which could indicate that the original Node2Vec representation learned directly from face-condition is sufficiently *cleaned* to perform well on the classification. In summary, the proposed contrastive loss are robust across all events, and the fusion improves the recognition rates of face and memory ability over 80% on both measures of TFMT and cIE.

3.3. Brain Connectome Visualization

To further understand how the brain network function appeared under different stimuli condition between two different groups of subjects (high-scoring versus low-scoring of TFMT and cIE), we compute the cosine similarity by averaging E-cCN representations separately in two groups under given events, neutral faces and emotional faces. In this work, we present a visualization on the brain network formed by at least 5 ROIs where one of them is chosen as a central point. The chosen networks for subjects when stimulated with two events are presented in figure 2. The dots on the plot stand for brain regions. The connected lines between dots indicate the cosine similarity measures of corresponding pair of regions. The darker line color means a higher similarity between the two ROIs, and the red dots means the ROIs that have a stronger links with others, i.e., can be imagined as forming a functional network.

According to prior works, rOFA(#15) and rFFA(#9) are necessary for normal face processing [20]. On the other hand, when being exposed to emotional faces, rFFA and other expressive-related regions such as ATL(#4, #5), Amygdala(#2, #3), IFGOrb(#1), pSTS(#18, #19) have been identified to function together [21]. In our work, for subjects that

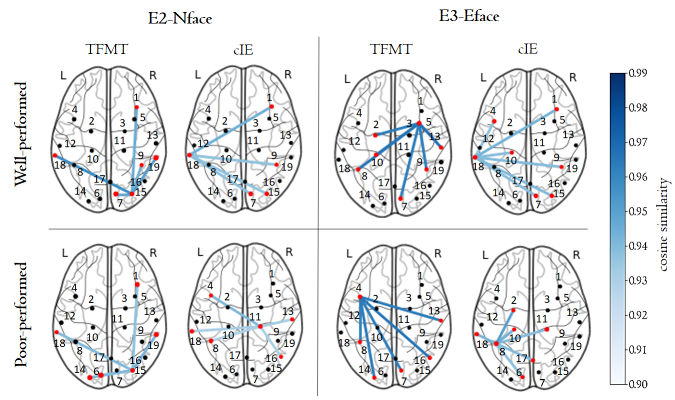

Fig. 2. *A figure of chosen networks in high-scoring and low-scoring group stimulated with the events of neutral and expressive faces during fMRI scanning.*

are in the high-performing group meet the expectation, i.e., rOFA and rFFA show up together demonstrating a normal face-perception process, while the low-performing group either misses key region, e.g., rFFA, or involves with high level ROIs lATL(#4) and hippocampus(#11) in neutral-face event, indicating that the reduce in face processing memory function may be attributed to the poor connectivity functioning in key face processing and memory brain regions.

4. CONCLUSION AND FUTURE WORK

In this work, we present a novel fMRI feature embedding framework for automatic assessing individual face processing and memory ability, i.e., the TFMT and cIE. Specifically, we propose an E-cCN to model the participants brain functional connectivity and obtain a promising accuracy of over 80% in classifying between high versus low performing group. The proposed approach outperforms existing state-of-the-art approaches in modeling brain connectivity for discriminative task. Our model provides enhanced representation by jointly optimizing three experimental manipulations using contrastive loss embedding. Our results demonstrate that it not only provides promising modeling power, but also through visualization of the learned connectivity, we show differential brain functions between subjects with high or low scoring ability in face identification from memory.

There are several future goals to pursue. One immediate way is to incorporate event-conditioned constraint in learning Node2Vec, such that the complete architecture can be tuned end-to-end. Secondly, the additional use of structural data such as diffusion tensor imaging (DTI), diffusion weighted imaging (DWI) to construct a multi-view model with fMRI, which has been shown useful in hub detection task [22], could be integrated as additional brain-related representation. Lastly, we would like to apply our method in other brain-related disease, such as autism and Bipolar, which will hopefully bring further insights in understanding the brain network differences between health controls and patients.

5. REFERENCES

[1] Icek Ajzen, “Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology,” in *Advances in experimental social psychology*, vol. 20, pp. 1–63. Elsevier, 1987.

[2] Tim RV Davis and Fred Luthans, “A social learning approach to organizational behavior,” *Academy of Management Review*, vol. 5, no. 2, pp. 281–290, 1980.

[3] James V Haxby, Elizabeth A Hoffman, and M Ida Gobbini, “Human neural systems for face recognition and social communication,” *Biological psychiatry*, vol. 51, no. 1, pp. 59–67, 2002.

[4] Rosalind Hutchings, Romina Palermo, Olivier Piguet, and Fiona Kumfor, “Disrupted face processing in frontotemporal dementia: a review of the clinical and neuroanatomical evidence,” *Neuropsychology review*, 2017.

[5] Tamsyn E Van Rheenen, Nicole Joshua, David J Castle, and Susan L Rossell, “Configural and featural face processing influences on emotion recognition in schizophrenia and bipolar disorder,” *Journal of the International Neuropsychological Society*, 2017.

[6] Brad Duchaine and Ken Nakayama, “The cambridge face memory test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants,” *Neuropsychologia*, vol. 44, no. 4, pp. 576–585, 2006.

[7] Daphne Maurer, Richard Le Grand, and Catherine J Mondloch, “The many faces of configural processing,” *Trends in cognitive sciences*, vol. 6, pp. 255–260, 2002.

[8] AH Mehrnam, AM Nasrabadi, Mahrad Ghodousi, A Mohammadian, and Sh Torabi, “A new approach to analyze data from eeg-based concealed face recognition system,” *International Journal of Psychophysiology*, vol. 116, pp. 1–8, 2017.

[9] Andrew C Lynn, Aarthi Padmanabhan, Daniel Simmonds, William Foran, Michael N Hallquist, Beatriz Luna, and Kirsten O’Hearn, “Functional connectivity differences in autism during face and car recognition: underconnectivity and atypical age-related changes,” *Developmental science*, vol. 21, no. 1, pp. e12508, 2018.

[10] Nancy Kanwisher, Josh McDermott, and Marvin M Chun, “The fusiform face area: a module in human extrastriate cortex specialized for face perception,” *Journal of neuroscience*, vol. 17, pp. 4302–4311, 1997.

[11] Aditya Grover and Jure Leskovec, “node2vec: Scalable feature learning for networks,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016, pp. 855–864.

[12] Ce Qi and Fei Su, “Contrastive-center loss for deep neural networks,” *arXiv preprint arXiv:1707.07391*, 2017.

[13] Ya-Hsin Cheng, Gary Chon-Wen Shyi, and Kuan-Hao Cheng, “Age differences in face memory and face processing between younger and older adults in taiwan,” *Chinese Journal of Psychology*, 2016.

[14] GCW Shyi, STT Huang, and CY Yeh, “Taiwan corpora of chinese emotions and relevant psychophysiological data-A college-student database of facial expression for basic emotions,” *Chinese Journal of Psychology*, vol. 55, pp. 455–475, 2013.

[15] William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols, *Statistical parametric mapping: the analysis of functional brain images*, Elsevier, 2011.

[16] Christopher J Fox, Giuseppe Iaria, and Jason JS Barton, “Defining the face processing network: optimization of the functional localizer in fmri,” *Human brain mapping*, vol. 30, no. 5, pp. 1637–1651, 2009.

[17] James T Townsend and F Gregory Ashby, “Methods of modeling capacity in simple processing systems,” *Cognitive theory*, vol. 3, pp. 200–239, 1978.

[18] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena, “Deepwalk: Online learning of social representations,” in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 701–710.

[19] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei, “Line: Large-scale information network embedding,” in *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015, pp. 1067–1077.

[20] Bruno Rossion, Roberto Caldara, Mohamed Seghier, Anne-Marie Schuller, Francois Lazeyras, and Eugene Mayer, “A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing,” *Brain*, vol. 126, no. 11, pp. 2381–2395, 2003.

[21] Jessica A Collins and Ingrid R Olson, “Beyond the ffa: the role of the ventral anterior temporal lobes in face processing,” *Neuropsychologia*, 2014.

[22] Guixiang Ma, Chun-Ta Lu, Lifang He, S Yu Philip, and Ann B Ragin, “Multi-view graph embedding with hub detection for brain network analysis,” in *ICDM, IEEE International Conference on Data Mining*. IEEE, 2017.