



Application of Convolutional Recurrent Neural Network for Individual Recognition Based on Resting State fMRI Data

Lebo Wang¹, Kaiming Li², Xu Chen³ and Xiaoping P. Hu^{1,2,3*}

¹ Department of Electrical and Computer Engineering, University of California, Riverside, Riverside, CA, United States,

² Department of Bioengineering, University of California, Riverside, Riverside, CA, United States, ³ Center for Advanced Neuroimaging, University of California, Riverside, Riverside, CA, United States

OPEN ACCESS

Edited by:

Amir Shmuel,
McGill University, Canada

Reviewed by:

Wenqin Wang,
Tianjin Polytechnic University, China
Xiaoyun Liang,
Australian Catholic University,
Australia
Hanbo Chen,
Tencent Holdings Limited, China

*Correspondence:

Xiaoping P. Hu
xhu@engr.ucr.edu

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 09 December 2018

Accepted: 15 April 2019

Published: 01 May 2019

Citation:

Wang L, Li K, Chen X and Hu XP
(2019) Application of Convolutional
Recurrent Neural Network
for Individual Recognition Based on
Resting State fMRI Data.
Front. Neurosci. 13:434.
doi: 10.3389/fnins.2019.00434

In most task and resting state fMRI studies, a group consensus is often sought, where individual variability is considered a nuisance. None the less, biological variability is an important factor that cannot be ignored and is gaining more attention in the field. One recent development is the individual identification based on static functional connectome. While the original work was based on the static connectome, subsequent efforts using recurrent neural networks (RNN) demonstrated that the inclusion of temporal features greatly improved identification accuracy. Given that convolutional RNN (ConvRNN) seamlessly integrates spatial and temporal features, the present work applied ConvRNN for individual identification with resting state fMRI data. Our result demonstrates ConvRNN achieving a higher identification accuracy than conventional RNN, likely due to better extraction of local features between neighboring ROIs. Furthermore, given that each convolutional output assembles in-place features, they provide a natural way for us to visualize the informative spatial pattern and temporal information, opening up a promising new avenue for analyzing fMRI data.

Keywords: functional magnetic resonance imaging, individual identification, recurrent neural network, convolutional neural network, visualization

INTRODUCTION

Mainstream fMRI studies have been focusing on deriving population consensus using group analysis. A group analysis in neuroimaging, albeit important, commonly neglects individual-to-individual variations. The importance of individual variability in neurobiological research has drawn increasing attention (Mohr and Nagel, 2010). Using task-fMRI, significant individual differences in brain activation were identified, reflecting alterations in cognitive function and behavior (Barch et al., 2013). Individual variability in functional connectivity (FC) has been successfully used to identify subjects from a large group. More specifically, static connectivity patterns throughout the brain were shown to be subject specific and distinctive across scan sessions and conditions, providing powerful features for individual identification (Finn et al., 2015). Therefore, exploring the individual uniqueness of the brain connectivity points to a new avenue to study the brain.

Although the static FC achieved decent accuracy, it required a sufficiently long data set (600 frames, 7.2 min) and considered only the spatial pattern through the temporal correlation without

taking temporal features into full account. The performance degraded with short clips of fMRI data, probably due to temporal variability (or dynamics) in the resting-state fMRI data, which leads to high variability in the FC derived from a short window. On the other hand, the dynamic information of resting state activity, if taken into account, could provide additional features for individual identification, improving the accuracy with the short time series.

In the application of time sequence modeling, recurrent neural networks (RNNs) have shown outstanding promise in a broad range of applications, including video classification, machine translation, and biomedical image segmentation (Sharma et al., 2015; Chen et al., 2016; Vaswani et al., 2017; Gao et al., 2018). For fMRI data analysis, RNN was able to model the dynamics of brain activity in response to sensory stimuli, providing accurate estimates of hemodynamic response with temporal dynamics (Güçlü and van Gerven, 2017). RNNs have also been implemented to incorporate temporal information along with spatial features from resting-state fMRI data instead of merely spatial pattern in the connectome (Dvornek et al., 2017; Chen and Hu, 2018). Furthermore, a convolution-based RNN was introduced to make full use of features in both spatial and temporal domains, consistently outperforming fully connected RNNs (Shi et al., 2015). Therefore, combining the local features between adjacent ROIs by the convolutional structure and sequence modeling capability of RNN may lead to a better approach to extract spatiotemporal features for individual identification on resting-state fMRI data.

In the meantime, it is also valuable to visualize the underlying features in the trained convolutional models. Although deep learning is becoming a panacea in almost every domain, it has been criticized due to its poor interpretability as being a black-box. While many attempts have been made to provide an interpretation and an intuitive understanding of trained networks, our understanding of how these networks work and what is important behind their performance have not kept up with the pace of the development of neural networks. While dedicated deep learning models have achieved amazing performance by end-to-end learning through huge volumes of data, better comprehension of the success of these models can uncover fundamental principles of deep neural networks and reveal important features within the data.

In this work, we adopted convolutional RNN or ConvRNN for individual identification using resting-state fMRI data. The convolutional recurrent model was able to achieve individual identification with shared convolutional weights capturing local coactivation features. In-place visualization of the informative area by ConvRNN also opened up a new avenue for understanding fMRI data based on individual differences.

MATERIALS AND METHODS

Dataset and Preprocessing

The resting-state fMRI data for 100 subjects from the Human Connectome Project (HCP) (54 females, age: 22–36, and TR = 0.72 s) was used in this work. Each subject had four

resting-state fMRI sessions, 1200 volumes for each session, leading to 4800 volumes per subject in total (Van Essen et al., 2013). The fMRI data was preprocessed by the HCP minimal preprocessing pipeline (Glasser et al., 2013) and denoised by ICA-FIX (Salimi-Khorshidi et al., 2014), for the removal of spatial artifact/distortion and motion-related fluctuations. Surface-based registration was performed with the MSM-ALL template (Robinson et al., 2014). To decrease the computation complexity, two hundred and thirty-six regions of interest (ROIs) over the cerebral cortex, as shown in **Figure 1** based on meta-analysis (Power et al., 2011), were used for subsequent analysis. BOLD signals within each ROI (10 mm diameter sphere) were averaged spatially. We ordered our ROIs-based data in a 1D array and preserved the order of ROIs according to the Power Atlas, where ROIs having similar functional connectivities are close to each other. They were also demeaned and scaled to unit variance over the temporal axis. For each fMRI session, fMRI data with 1200 volumes was divided into twelve 100-frame clips as inputs of ConvRNN. Data from Day 1 was used as the training dataset. The two sessions from Day 2 were used as validation and testing datasets, respectively. The best model was decided based on the validation dataset and the final performance was assessed on the testing dataset.

Convolutional Recurrent Neural Network

The architecture of the ConvRNN is given in **Figure 2**, along with its unrolling version. In contrast to conventional RNN, convolution was applied in both the input-to-state and state-to-state transitions, in place of the Hadamard product. There were two stacked convolutional layers, with the first convolutional layer containing 8 filters and the second convolutional layer having 16 filters. The kernel size of all convolutional filters was 2. Padding was used in all convolutional layers such that the outputs from each layer had the same spatial dimension as the original input, which is very important for the subsequent visualization of the in-place features. Batch normalization layers were used before the non-linear activation layers of Rectified linear unit (ReLU) to reshape the distribution of convolutional layer outputs in order for better convergence and easy training

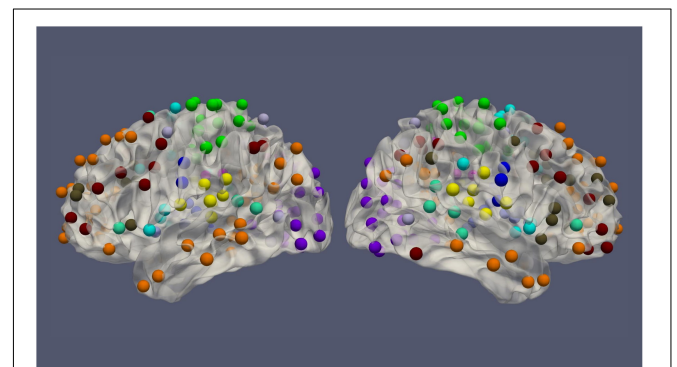
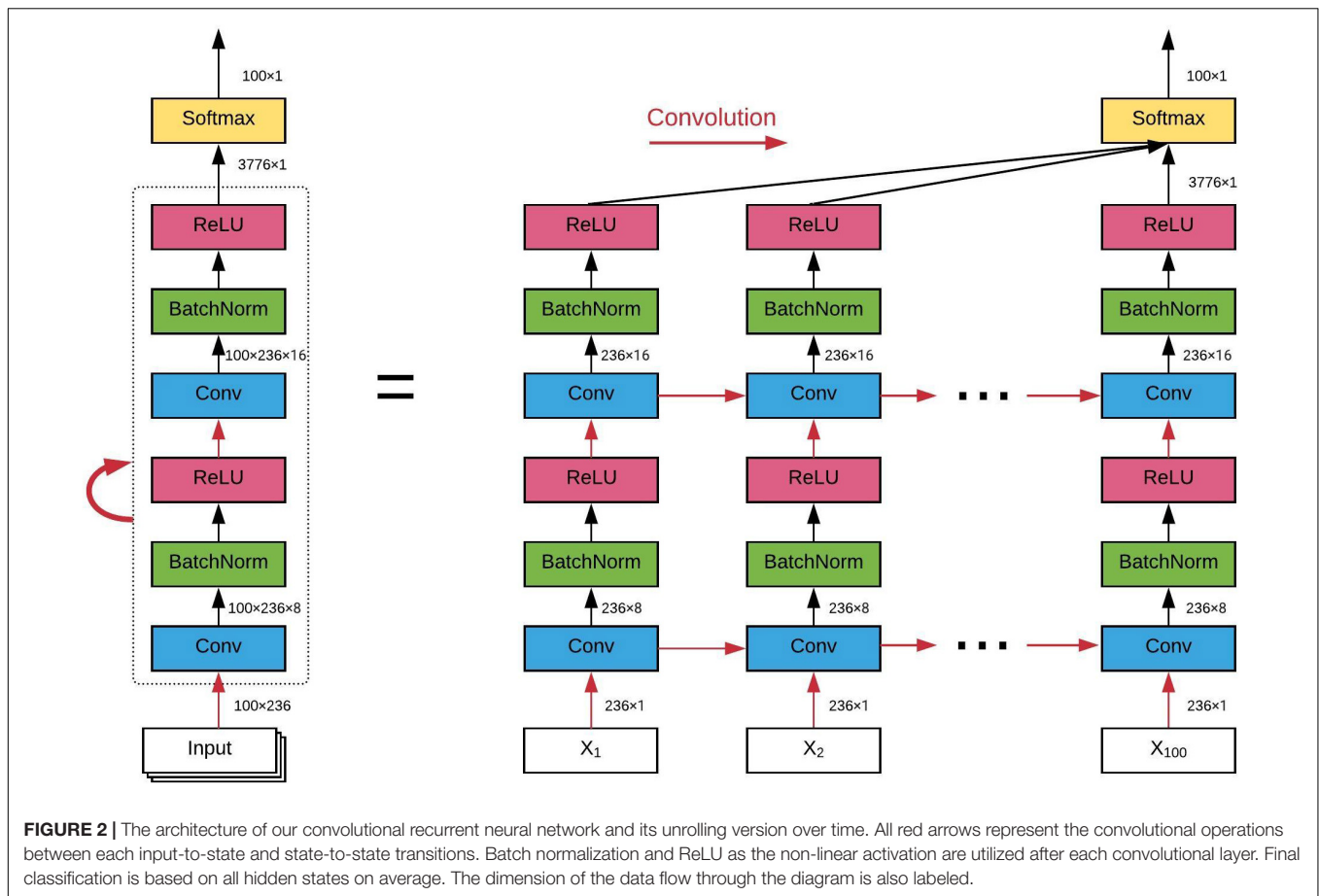


FIGURE 1 | The spatial distribution of 236 ROIs over the cerebral cortex. Voxels within the 10 mm diameter sphere were averaged to get the value for each ROI.



(Ioffe and Szegedy, 2015). The final Softmax layer with 100-category outputs was used for classification based on averaged outputs from all hidden states. No temporal or spatial pooling layer was employed to keep the spatial and temporal resolutions of the original fMRI data. All kernel weights were initialized by the Xavier uniform initializer (Glorot and Bengio, 2010), and recurrent weights were initialized as random orthogonal matrices (Saxe et al., 2013).

Training of the Neural Network

Our implementation of ConvRNN was carried out in Keras (Chollet, 2015) with the Tensorflow backend (Abadi et al., 2016). Considering the limited number of frames for each subject, we chose 100 frames of fMRI data as inputs during training and validation, which is the tradeoff between the number of fMRI clips and the number of frames for each clip. Shuffled minibatches of training data as inputs were fed into the ConvRNN with the batch size of 128. Adam optimizer (Kingma and Ba, 2014) was applied for training with the initial learning rate set to 0.001, and reduced if the validation accuracy stopped increasing. Dropout layer with 50% was utilized before the final classification to avoid overfitting only during the training (Srivastava et al., 2014). After each training epoch, the model was evaluated on the validation dataset and saved only if better validation accuracy was achieved. Finally, the performance of the best model was

measured on the testing dataset, which was never involved during training or validation.

It is well known that RNN is difficult to train properly, even though it is a powerful model for time series modeling. The main reasons are vanishing and exploding gradient issues of Backpropagation Through Time (BPTT) on the unrolling version of RNN (Bengio et al., 1994). Therefore, advanced architectures with gating mechanism to overcome the vanishing and exploding gradient problem, such as the Long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) and the Gated recurrent unit (GRU) (Cho et al., 2014), have gained a lot of popularity in practice to model long-term dependencies. In this work, LSTM with convolutional structure was applied. For training techniques, we used L2 regularization for recurrent weights, along with the gradient clipping strategy as a simple and computationally efficient method, effectively addressing the issue of exploding gradients (Pascanu et al., 2013). In the present work, the clipping norm of the gradient was set to 1. Different L2 values (0.1, 0.01, 0.001, and 0.0001) on recurrent kernel weights were tested to achieve the best validation accuracy.

Visualization of the Individual Identification

Our ConvRNN first performed feature extraction through two convolutional recurrent layers and then fed the features into

the Softmax layer for the 100-category classification. Original data was projected to a high-dimensional feature space, which was easily separated by the classification layer. In the feature space, fMRI data from the same subject are expected to be close to each other and cluster tightly. In consideration of the single classification layer, the identification accuracy of our ConvRNN relies heavily on the performance of feature extraction by convolutional recurrent layers. In order to ascertain and visualize the performance of convolutional recurrent layers in low dimensional space, t-Distributed Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton, 2008) was applied to map datapoints in high-dimensional feature space onto a two-dimensional representation.

To visualize and understand informative areas related to individual identification, intermediate outputs from convolutional layers were examined. Output patterns were obtained from convolutional layers after non-linear activation and mapped onto the cortical surface (20 mm radius sphere). As all regions are considered equal in our convolutional model during the training, but they are of different importance to the final classification. We also used the occlusion method to visualize informative areas (Zeiler and Fergus, 2014). More specifically, in order to ascertain the contribution of ROIs with regard to individual identification, input of each ROI was zeroed out, and the subsequent performance decrease with the same model configuration was considered as the contribution of this ROI to the final classification.

RESULTS

We carried out the supervised classification task to identify each subject from a group of 100 subjects. First, the identification accuracy of different models was assessed on the testing dataset with 100 frames of fMRI data as inputs. As seen in **Table 1**, our ConvRNN model was able to achieve 98.50% accuracy on the testing dataset, where the best performance was obtained with the L2 value of 0.001 during training. The test accuracy for the traditional RNN with average temporal pooling was 94.43% (Chen and Hu, 2018). In order to exclude the influence of temporal averaging, we trained another RNN without the temporal averaging and achieved an identification accuracy of 95.33%. Furthermore, we evaluated the performance of these models using different window sizes on the testing dataset. With the pre-trained models, we adopted different number of frames (from 1200 frames to single frame) as inputs from the testing dataset and evaluated the identification performance. Testing results with different number of frames are plotted in **Figure 3**. As shown in the figure, ConvRNN outperformed conventional RNN (no temporal averaging) in all cases except with 1200 frames or with less than 10 frames. In contrast, FC could achieve over 90% accuracy with 600 frames of fMRI data. But the individual identification accuracy drops to 70% on average when only a short period of fMRI data (100 frames) is used (Finn et al., 2015).

To visualize convolutional outputs on low-dimensional space, we applied t-SNE on intermediate outputs of our ConvRNN before the classification layer. There were 16 convolutional filters

TABLE 1 | The accuracy of different models on the testing dataset and their number of model weights.

Architecture	# Parameters (feature extraction)	Test accuracy
RNN (Chen and Hu, 2018)	405K (380K)	94.43%
RNN w/o temporal pooling	405K (380K)	95.33%
ConvRNN	382K (3.8K)	98.50%

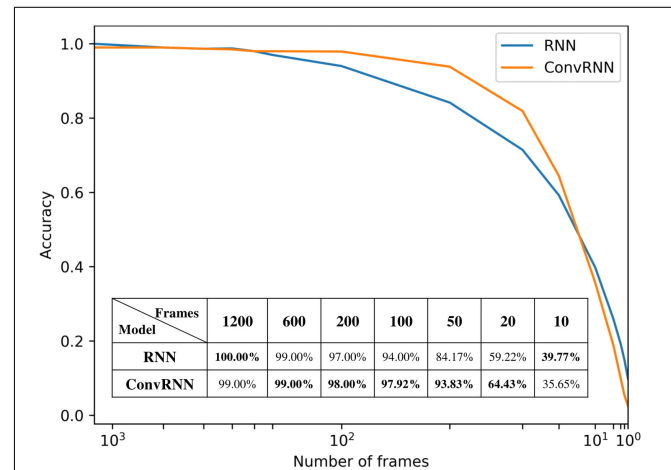


FIGURE 3 | The relationship between identification accuracy and the window size. We evaluated pre-trained models on testing dataset. Our ConvRNN outperformed RNN except with 1200 frames or with less than 10 frames.

in the second convolutional recurrent layer of our ConvRNN. With proper padding, the output of the layer was made to have the same spatial dimension as the input. The feature space with 3776 dimensions was then mapped to a 2D space in **Figure 4**. It is clear that 100 subjects (12 clips with 100 frames for each subject) in the testing dataset appear as non-overlapping cliques in different colors with each clique representing one subject. This figure clearly indicates that spatiotemporal features, capable of individual identification, were successfully obtained by convolutional recurrent layers.

To visualize intermediate outputs of ConvRNN, average patterns from first and second convolutional layers of ConvRNN are shown in **Figures 5, 6**, respectively. Most patterns from the first convolutional layer were quite similar (except Filter 6) with large distinctive areas, which could be considered as the ubiquitous low-level features from fMRI data. While high-level patterns from the second convolutional layer had diverse informative regions, which were sparse and localized inside the area of those low-level features generated by the first convolutional layer. Meanwhile, when ROIs were individually occluded, performance degradation was observed when some ROIs were occluded, while the occlusion of some ROIs led to negligible degradation of the performance. In **Figure 7**, the absolute value of the performance degradation, normalized to reflect the contribution of each ROI is shown. It is evident that the informative area generated by alternative occlusion was similar as the patterns from first convolutional layer of ConvRNN.

DISCUSSION

While most resting-state fMRI studies have relied on group averages, this study employed individual differences for

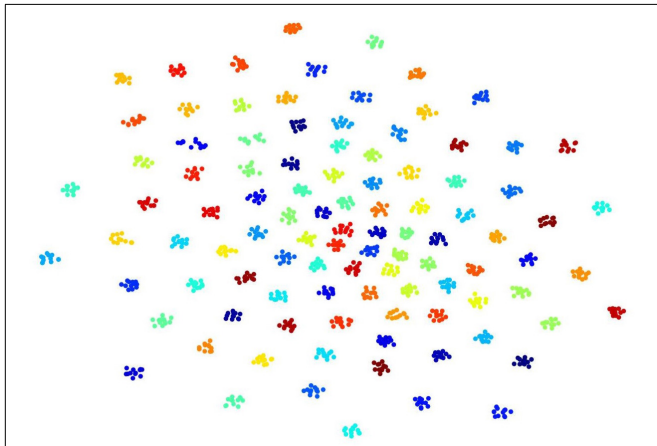


FIGURE 4 | t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization of 2nd convolutional recurrent layer outputs based on 100-subject testing dataset. Twelve hundred 100-frame testing data from 100 subjects were fed into ConvRNN with outputs being obtained before the classification layers and projected to 2D space by t-SNE. Projections for different subjects are in different colors.

individual identification. Unlike the first study of individual identification employing static FC (Finn et al., 2015), we incorporated both temporal and spatial features from the fMRI data. As an improvement of our previous work employing the recurrent architecture (Chen and Hu, 2018), we applied a convolutional recurrent neural network which led to a significant improvement in performance and a straightforward means to visualize in-place features. **Figure 3** shows that ConvRNN is better than conventional RNN for the majority of the time windows. The performance of ConvRNN was slightly worse than conventional RNN with 1200 frames, probably due to the small number of testing data when the performance was evaluated on fMRI clips with 1200 frames. On the other hand, since our ConvRNN was trained with the fixed number of frames (i.e., 100), it is not optimized for short clips of data with less than 10 frames, and its performance with frames less than 10 is therefore worse than that of conventional RNN.

Our ConvRNN has the same number of parameters compared with the conventional RNN, indicating that both models have comparable model complexity. Apart from the different types of recurrent unit, our earlier work (Chen and Hu, 2018) employed a temporal pooling layer to reduce the temporal resolution. For a fair comparison with this work, another conventional RNN was applied without the temporal averaging layer. The accuracy of the conventional RNN without the temporal averaging layer was 95.33%, which was 0.9% higher than that with the temporal averaging. This improvement

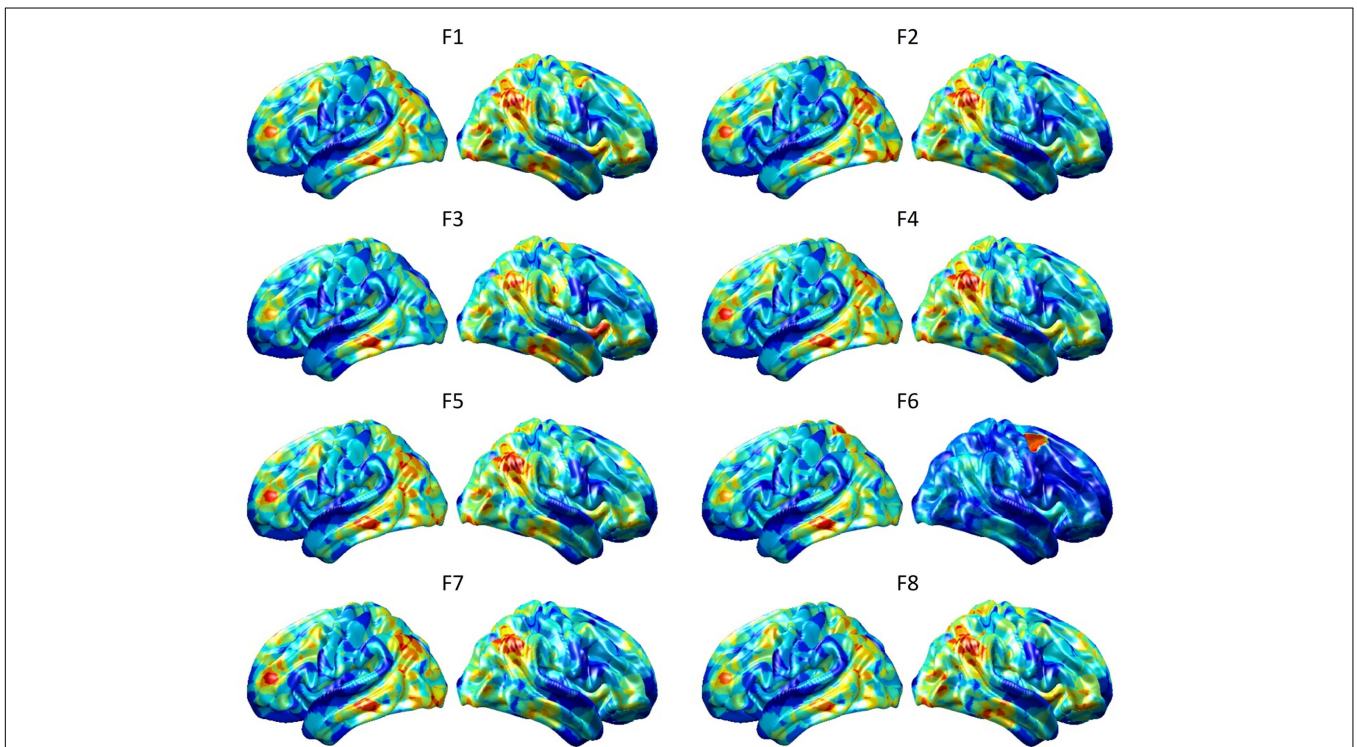
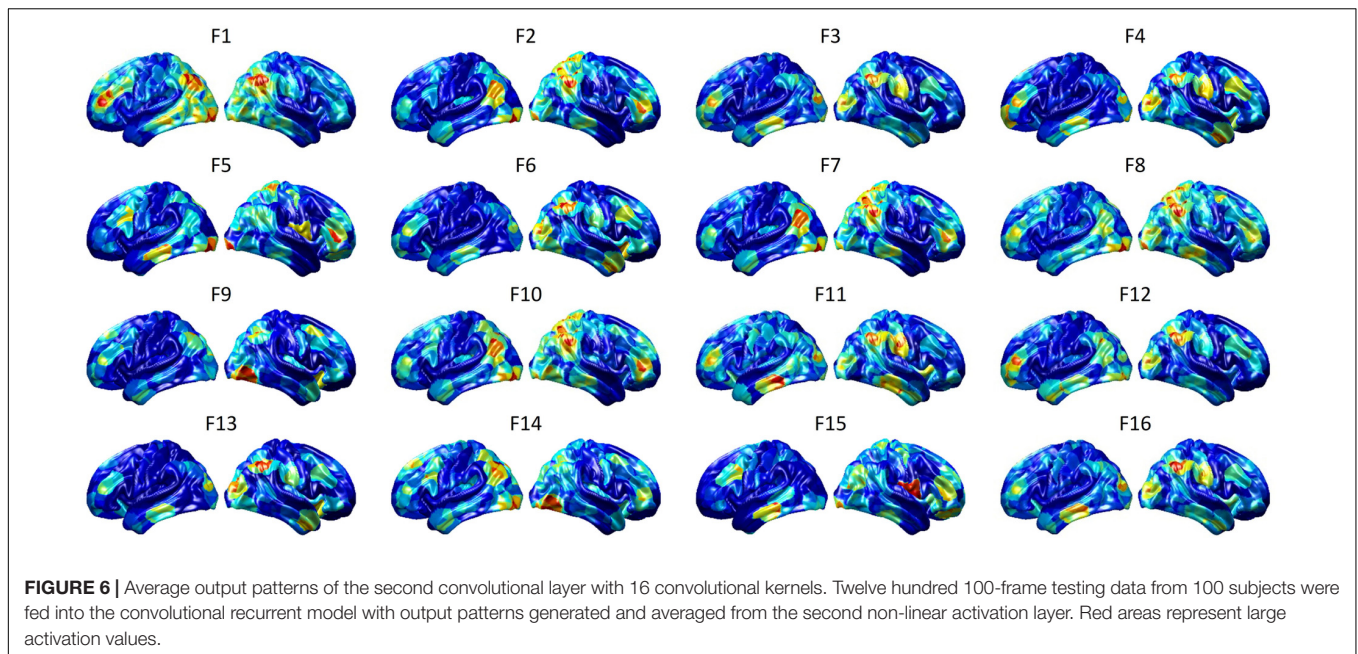
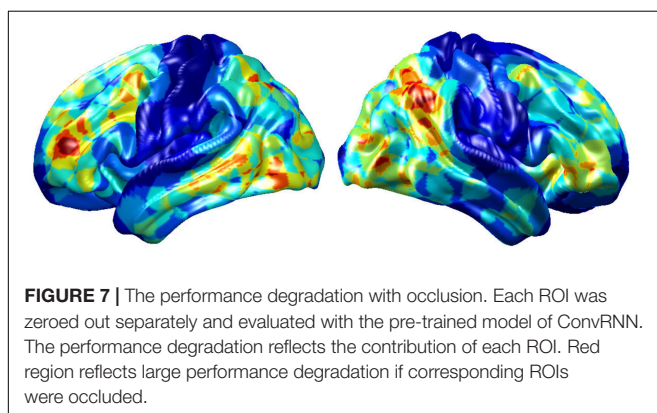


FIGURE 5 | Average output patterns of the first convolutional layer with 8 convolutional kernels. Twelve hundred 100-frame testing data from 100 subjects were fed into the convolutional recurrent model with output patterns generated and averaged from the first non-linear activation layer. Red areas represent large activation values.



due to increased temporal resolution is significantly smaller than the improvement achieved with the adoption of the convolutional structure, indicating that the latter is the main contributor to the performance enhancement. Both the spatial and temporal features were fully utilized by convolutional kernels in the ConvRNN with unprecedented identification accuracy. Also, feature extraction layers of ConvRNN showed strong discriminating power for 100 subjects, where pre-trained layers could be applied for transfer learning on new subjects or semi-supervised learning on partially labeled data. On the other hand, only one hundredths of the parameters in ConvRNN were from the convolutional recurrent layers. Convolution with shared weights in spatial and temporal spans makes it more robust against overfitting during training. Given the reduced number of trainable weights, increasing the depth and width of the model is possible without significantly increasing the model complexity, possibly capturing more sophisticated features in both spatial and temporal domains.



Furthermore, convolutional kernels with shared weights sweep across ROIs and frames. Different from the indecomposable matrix multiplication in conventional RNN, ConvRNN generates in-place features with exactly spatial correspondence as the original data. Furthermore, ConvRNN accumulates temporal information related to evolving features in the hidden state. Therefore, it is possible to examine the hidden states to have an in-place visualization and understanding of hidden features from ConvRNN. It is also clear that informative regions from two convolutional layers are totally different, in agreement with the conclusion drawn from convolutional neural networks for image classification (Yosinski et al., 2015). Beside the direct visualization of the hidden state, the occlusion of ROIs served as an indirect method for visualizing significant regions under resting state for identification. Two visualization approaches came to the same conclusion of informative ROIs in terms of individual identification using the resting-state fMRI data. In terms of resting state networks (RSNs) in the literature (Holmes et al., 2011; Lee et al., 2012), the informative area identified by our ConvRNN mainly contained frontoparietal network (FPN), default mode network (DMN), as well as visual network (VN). Our result is consistent with a previous study, which concluded that the most distinguishing network was FPN, with significant improvement achieved through the combination of multiple RSNs (Finn et al., 2015). In contrast, occlusion of language network (LN) and somatosensory motor network (SMN) did not cause much reduction in performance. One possible explanation is that there was little explicit or individual-specific language or motor activity during the acquisition of the resting-state fMRI datasets used here.

Although DMN and FPN stood out as the most important networks for individual identification, other networks also contributed to individual identification (Finn et al., 2017). It is likely that contributions from most networks may be

needed with more subjects to be identified. In addition, a hierarchical approach, incorporating all networks, might be the most appropriate and robust approach. Furthermore, a recent study demonstrated that task induced changes in FC provided better prediction of individuals, whereas resting-state fMRI data failed to capture the full range of individual differences (Greene et al., 2018). Such changes in FC could also be incorporated into our model to further improve the identification accuracy.

Several limitations should be noted for further work. First, the present study adopted only 236 ROIs within 10 mm diameter spheres on average, which was enough for an accurate identification for a group of 100 subjects. Inadequate power of identification could be present on new subjects beyond existing subjects. Possible reasons are the limited feature extraction of our model and the high variability of the fMRI data. It is likely that smaller and more ROIs are needed for the identification of more subjects. But more advanced models with good generalization on fMRI data should be explored. Second, current visualization of individual identification depicted in **Figure 7** was based on group average of ROIs' performance. While this highlights areas that are most important for individual identification, it does not explicitly depict individual features. Such features will be the focus of our future studies. Third, visualization of the spatial pattern is easy to understand, but the remarkable performance achieved by RNN suggests that a substantial amount of information is coming from temporal features. Visualizing and understanding temporal features are still necessary to gain a deeper insight into

the brain dynamics. Furthermore, other popular architectures (e.g., Siamese network) and pre-trained models should be applied and compared with our approach in terms of classification performance and training efficiency in future work.

CONCLUSION

In this paper, we described the application of the convolutional recurrent neural network for individual identification based on resting-state fMRI data. To explore the dynamics in the resting-state fMRI data, the convolutional architecture with recurrent structure was implemented to extract and incorporate features in both spatial and temporal domains. Compared to conventional RNN model, our ConvRNN model exhibited better identification performance, with local features between neighboring ROIs being modeled by convolutional kernels. Moreover, visualization based on the ConvRNN model provides a direct understanding of the success of identification; this could lead to a promising alternative for analyzing fMRI data.

AUTHOR CONTRIBUTIONS

LW designed and performed experiments and wrote the manuscript. KL analyzed data and reviewed the manuscript. XC collected the data. XH designed research and reviewed the manuscript.

REFERENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "TensorFlow: a system for large-scale machine learning," in *OSDI'16 Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation*, (Savannah, GA), 265–283.
- Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., et al. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* 80, 169–189. doi: 10.1016/j.neuroimage.2013.05.033
- Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* 5, 157–166. doi: 10.1109/72.279181
- Chen, J., Yang, L., Zhang, Y., Alber, M., and Chen, D. Z. (2016). "Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation," in *Proceedings of the Advances in Neural Information Processing Systems*, (Cambridge, MA), 3036–3044.
- Chen, S., and Hu, X. P. (2018). Individual identification using functional brain fingerprint detected by recurrent neural network. *Brain Connect.* 8, 197–204. doi: 10.1089/brain.2017.0561
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv:1406.1078 [Preprint].
- Chollet, F. (2015). *Keras. GitHub*. Available at: <https://github.com/fchollet/keras> (accessed March 7, 2018).
- Dvornek, N. C., Ventola, P., Pelpheyre, K. A., and Duncan, J. S. (2017). "Identifying autism from resting-state fMRI using long short-term memory networks," in *Proceedings of the International Workshop on Machine Learning in Medical Imaging*, (Granada), 362–370. doi: 10.1007/978-3-319-67389-9_42
- Finn, E. S., Scheinost, D., Finn, D. M., Shen, X., Papademetris, X., and Constable, R. T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? *Neuroimage* 160, 140–151. doi: 10.1016/j.neuroimage.2017.03.064
- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., et al. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* 18:1664. doi: 10.1038/nn.4135
- Gao, Y., Phillips, J. M., Zheng, Y., Min, R., Fletcher, P. T., and Gerig, G. (2018). "Fully convolutional structured LSTM networks for joint 4D medical image segmentation," in *Proceedings of the Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*, (Piscataway, NJ), 1104–1108.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127
- Glorot, X., and Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, (Sardinia), 249–256.
- Greene, A. S., Gao, S., Scheinost, D., and Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nat. Commun.* 9:2807. doi: 10.1038/s41467-018-04920-3
- Güçlü, U., and van Gerven, M. A. (2017). Modeling the dynamics of human brain activity with recurrent neural networks. *Front. Comput. Neurosci.* 11:7. doi: 10.3389/fncom.2017.00007
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780.
- Holmes, A. J., Hollinshead, M. O., O'keefe, T. M., Petrov, V. I., Fariello, G. R., Wald, L. L., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Sci. Data* 2:150031.
- Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167 [Preprint].
- Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. arXiv:1412.6980 [Preprint].

- Lee, M. H., Smyser, C. D., and Shimony, J. S. (2012). Resting-state fMRI: a review of methods and clinical applications. *Am. J. Neuroradiol.* 34, 1866–1872. doi: 10.3174/ajnr.a3263
- Maaten, L. V. D., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Machine Learn. Res.* 9, 2579–2605.
- Mohr, P. N., and Nagel, I. E. (2010). Variability in brain activity as an individual difference measure in neuroscience. *J. Neurosci.* 30, 7755–7757. doi: 10.1523/jneurosci.1560-10.2010
- Pascanu, R., Mikolov, T., and Bengio, Y. (2013). “On the difficulty of training recurrent neural networks,” in *Proceedings of the International Conference on Machine Learning*, (Atlanta, GA), 1310–1318.
- Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron* 72, 665–678.
- Robinson, E. C., Jbabdi, S., Glasser, M. F., Andersson, J., Burgess, G. C., Harms, M. P., et al. (2014). MSM: a new flexible framework for multimodal surface matching. *Neuroimage* 100, 414–426. doi: 10.1016/j.neuroimage.2014.05.069
- Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* 90, 449–468. doi: 10.1016/j.neuroimage.2013.11.046
- Saxe, A. M., McClelland, J. L., and Ganguli, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. arXiv:1312.6120 [Preprint].
- Sharma, S., Kiros, R., and Salakhutdinov, R. (2015). Action recognition using visual attention. arXiv:1511.04119 [Preprint].
- Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-C. (2015). “Convolutional LSTM network: a machine learning approach for precipitation nowcasting,” in *Proceedings of the Advances in Neural Information Processing Systems*, (Cambridge, MA), 802–810.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Machine Learn. Res.* 15, 1929–1958.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., et al. (2013). The WU-Minn human connectome project: an overview. *Neuroimage* 80, 62–79. doi: 10.1016/j.neuroimage.2013.05.041
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Proceedings of the Advances in Neural Information Processing Systems*, (Cambridge, MA), 6000–6010.
- Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., and Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv*
- Zeiler, M. D., and Fergus, R. (2014). “Visualizing and understanding convolutional networks,” in *Proceedings of the European conference on computer vision*, (Munich), 818–833. doi: 10.1007/978-3-319-10590-1_53

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang, Li, Chen and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.