Infant Cognitive Scores Prediction with Multi-stream Attention-based Temporal Path Signature Features ⋆

Xin Zhang[1], Jiale Cheng[1*], Hao Ni[2], Chenyang Li[1], Xiangmin Xu[1], Zhengwang Wu[3], Li Wang[4], Weili Lin[3], Dinggang Shen[3], and Gang Li[3]

[1] School of Electronic and Information Engineering, South China University of Technology
[2] Department of Mathematics, University College London
[3] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
[4] University of North Carolina

Abstract. There is stunning rapid development of human brains in the first year of life. Some studies have revealed the tight connection between cognition skills and cortical morphology in this period. Nonetheless, it is still a great challenge to predict cognitive scores using brain morphological features, given issues like small sample size and missing data in longitudinal studies. In this work, for the first time, we introduce the path signature method to explore hidden analytical and geometric properties of longitudinal cortical morphology features. A novel BrainPSNet is proposed with a differentiable temporal path signature layer to produce informative representations of different time points and various temporal granules. Further, a two-stream neural network is included to combine groups of raw features and path signature features for predicting the cognitive score. More importantly, considering different influences of each brain region on the cognitive function, we design a learning-based attention mask generator to automatically weight regions correspondingly. Experiments are conducted on an in-house longitudinal dataset. By comparing with several recent algorithms, the proposed method achieves the state-of-the-art performance. Additionally, the relationship between morphological features and cognitive abilities is also presented.

Keywords: Path signature feature · Infant brain development · Cognitive ability.

1 Introduction

With the advancement of magnetic resonance imaging (MRI) and image processing techniques, early structural development of the human brain is attracting

---
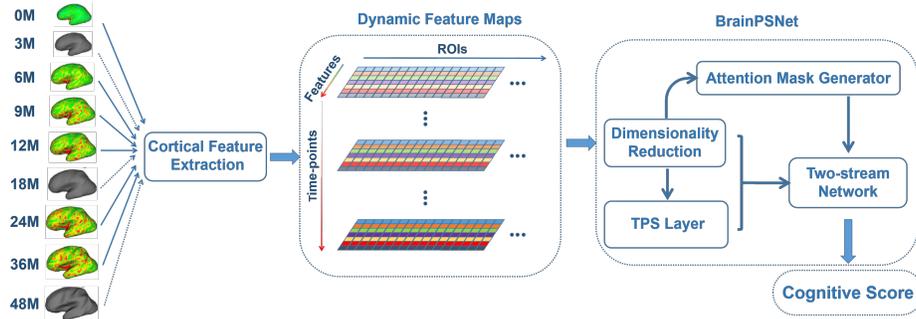
Fig. 1. Illustration of BrainPSNet. Inputs are longitudinal MRI data (grey brains indicating missing data) and output is the predicted cognitive scores.

more and more attention [10, 23]. However, only few of work related the infant brain cognitive scores to the cortical morphology. In fact, in the first year of life, the cortical structure is tightly connected to the acquisition and refinement of information processing as well as visual and language skills [12]. Thus, understanding the quantitative relationship between cognitive skills and morphological features of infant cerebral cortex is of immense importance .

To this end, in this paper, we aim to learn a representation for each infant by using longitudinal brain MRI data to predict cognitive development. Specifically, given longitudinal brain MR scans from infants, we can compute multiple biologically meaningful cortical measurements [16]. Meanwhile, we have five Mullen Scales of Early Learning (MSEL) [5] at 48 months of age to measure the cognition skills of each infant comprehensively. Hence, our goal is to build a machine learning method to predict these cognition scales using longitudinal morphological cortical features. However, there are three major challenges, including the small sample size, high dimensionality of data and missing scans. Recently, several methods are proposed to address these problems. The Bag-of-Words (BoW) based method was employed to slash the overlarge dimensionality of neuroimaging data [1]. In [28], authors generated a latent representation for each subject leveraging the complementary information among different time-points and introduced a set of indicator variant to eliminate the loss brought by incomplete data. They both achieved encouraging performance, but their optimization strategy is too complicated. Also, they can not analyze the correspondence between cerebral regions and cognitive scores since their representations for each subject did not preserve brain structural information. More importantly, the simple linear formulation they used cannot explore the temporal relationship sufficiently.

In this work, we introduce the path signature (PS) method for the first time as descriptors of dynamic dependencies in longitudinal infant cortical structure. The path signature originated from Chen's study [7] as an essential characteristic of piece-wise smooth paths in rough path theory. Lyons used it to make sense of the solution to differential equations driven by very rough signals [9]. Recently,

there is an emerging research area, which combines path signature feature with machine learning and achieves state-of-the-art results [4, 13, 15, 19, 21, 27].

Based on above discussions, we conclude our contributions as follow. First, we propose BrainPSNet with a temporal path signature (TPS) layer. To the best of our knowledge, this is the first work to apply the path signature method into longitudinal brain analysis and generate informative representations of multiple time points and temporal granules. Second, considering different brain regions have different influence on cognitive functions during infancy, we propose an effective network to exploit information from raw features and PS features separately and automatically generate learning-based attention masks for weighting groups of data. Third, by testing on a longitudinal infant dataset, our method achieves state-of-the-art performance and explores the quantitative relationship between cognitive skills and morphological features.

2 Preliminaries of Path Signature

Suppose a path $X : [a, b] \rightarrow R^d$ is a continuous mapping of finite length from interval $[a, b]$ to a d-dimensional vector space. For any $t \in [a, b]$, $X_t = (X_t^1, X_t^2, \cdots, X_t^d)$, where X_t^i denotes the ith coordinate of X_t and $i \in \{1, 2, \cdots, d\}$. Before introducing the signature, let us introduce the kth fold iterated integral of a path X, denoted by $S_k(X)_{a,b}$

The 1th iterated integral of X along the ith coordinate, denoted by $Sig(X)_{a,b}^i$, is $Sig(X)_{a,b}^i = \int_{a<t_1<b} dX_{t_1}^i$, which equals the increment of X at i^{th} coordinates, i.e. $X_b^i - X_a^i$. The 1st fold iterated integral is the collection of $Sig(X)_{a,b}^i$ for $i \in \{1, 2 \cdots, d\}$, i.e.

$$S_1(X)_{a,b} = \int_{a<t_1<b} dX_{t_1}. \tag{1}$$

Notably, $t \mapsto Sig(X)_{a,t}^i$ is still a real-valued path defined within $t \in [a, b]$. Then, the 2nd iterated integral indexed by (i_1, i_2) is denoted by $Sig(X)_{a,b}^{i_1,i_2}$ and defined as integral of $Sig(X)_{a,.}^{i_1}$ against $X^{.i_2}$:

$$Sig(X)_{a,b}^{i_1,i_2} = \int_{a<t<b} Sig(X)_{a,t}^{i_1} dX_t^{i_2} = \int_{a<t_1<t_2<b} dX_{t_1}^{i_1} dX_{t_2}^{i_2}. \tag{2}$$

Similarly, the 2nd fold iterated integral of X is the collection of all 2nd iterated integrals of X with possible index (i_1, i_2), i.e. $\left(Sig(X)_{a,b}^{i_1,i_2}\right)_{i_1,i_2 \in \{1,\cdots,d\}}$, which can be written as the tensor form as follows:

$$S_2(X)_{a,b} = \int_{a<t_1<t_2<b} dX_{t_1} \otimes dX_{t_2}. \tag{3}$$

$Sig(X)_{a,b}^{i_1,i_2} - Sig(X)_{a,b}^{i_2,i_1}$ is equal to the area enclosed by the curve (X^{i_1}, X^{i_2}) and a chord connecting the ending and the starting point of the path. In general, the k-th fold iterated integral of X, $S_k(X)_{a,b}$, is defined to be as follows:

$$S_k(X)_{a,b} = \int_{a<t_1<\cdots<t_k<b} dX_{t_1} \otimes dX_{t_2} \cdots \otimes dX_{t_k}. \tag{4}$$

Here the dimension of the k-th fold iterated integrals of the path X is d^k.

The signature of a path is a graded infinite series, which contains all the k fold iterated integrals. In practice, we may truncated the signature up to the finite degree. Let $Sig^k(X)$ denoted the truncated signature of X up to degree k as follows:

$$Sig_k(X)_{a,b} = (1, \ S_1(X)_{a,b}, \ S_2(X)_{a,b}, \cdots, \ S_k(X)_{a,b}). \tag{5}$$

By convention, the 0th iterated integral is equal to 1. The dimension of the truncated signature in Equation (5) is $(d^{k+1} - 1)/(d - 1)$.

If a path $X : [a, b] \to \mathbb{R}^d$ is linear, then the signature of $X_{[a,b]}$ can be computed explicitly as follows:

$$Sig(X)_{a,b}^{i_1,i_2,\cdots,i_k} = \frac{1}{k!} \prod_{j=1}^{k} (X_b^{i_j} - X_a^{i_j}). \tag{6}$$

The signature of a piecewise path can be computed by Chen's identity [7]. In practice, we often observe the discrete time series, which can be embedded in the path space by the linear interpolation. The corresponding signature of the embedded path can be used as a non-linear feature of the time series data.

The signature of a path has many algebraic and analytic proprieties, which make it an effective feature set of the streamed data. First of all, the signature of path uniquely determines the path up to time re-parameterization [3,9]. It means that the signature captures the information on the path trajectory while removing infinite-dimensional noise caused by the speed variation. Secondly, the signature feature is an universal, which implies that any continuous functions on the unparamertized path can be well approximated by the linear functional on the signature [14]. Intuitively, the signature of a path plays a role as the non-commutative polynomial on the path space. Further, in practice, the signature feature set can be used to handle time series of variable length, and variation caused by the time re-parameterization. It is a global descriptor of the sequential data in terms of its effect, which can be often useful for dimension reduction. Interested readers can refer to [14] and [8] for more details.

3   Dataset and Feature Extraction

In this study, 23 normal infants with their T1w and T2w MR images were collected at 9 different time points (i.e., 0, 3, 6, 9, 12, 18, 24, 36 and 48 months after birth). Since not all participants are able to show up at all scheduled time points, there are missing scans, as illustrated in Fig. 1. For feature extraction, we processed MR images by following an infant MRI computational pipeline [16] and computed 7 different morphological cortical measurements at each vertex

of the reconstructed cortical surfaces, including cortical thickness (THI), local gyrification index (LGI), mean curvature (CUR), vertex area (ARE), vertex volume (VOL), sulcal depth in Euclidean distance (SDE) and sulcal depth in string distance (SDS) [17, 20]. These cortical features are the most commonly used measurements to quantify brain development [18]. Afterwards, we parcellated the cerebral cortex into 70 anatomically meaningful regions of interest (ROIs) with an infant cortical surface atlas [26] for reducing the feature dimensionality. In each ROI, feature values of the same type of all vertices are averaged (for THI, LGI, CUR, SDE, SDS) or summarized (for ARE, VOL), thus forming a 7-dimension feature. Finally, for each available scan at a timepoint, we can extract a feature map whose width and depth equal to the number of ROIs ($N = 70$) and cortical measurements ($d = 7$) respectively. By concatenating these feature maps along the time axis, we can get a cohort of dynamic feature maps as shown in Fig. 1. Five Mullen cognitive scores are estimated at 48 months age for each participant, i.e., Visual Receptive Scale (VRS), Fine Motor Scale (FMS), Receptive Language Scale (RLS), Expressive Language Scale (ELS) and Early Learning Composite (ELC) which are firmly correlated to the morphological attributes mentioned above [5].

4   Network architecture

To accurately predict the cognitive scores according to the cortical measurements, we propose a novel BrainPSNet consisting of three major components: a temporal path signature layer, a two-stream network and an attention mask generator as shown in Fig. 2. At the beginning, considering limited sample size, a 1×1 convolutional layer is introduced to further decrease the feature dimension of each ROI.

Temporal Path Signature Layer. A temporal path signature layer is then proposed to extract dynamic information and generate discriminative representations, shown in Fig. 2(b) in detail. For the first step, 70 paths are defined along the time axis and split by the overlapping sliding window with the size W and a sliding stride $s = 1$. Consequently, for each path, $\widetilde{T} = 9 - (W - 1)$ sub-paths are obtained to further explore local temporal properties. For every sub-path, we employ Equation (5) and (6) to compute its corresponding path signature features with a receptive field of W and denote the output dimension as $n_{PS} = (d^{k+1} - 1)/(k - 1)$. Afterwards, an 1×1 convolutional layer is introduced to conduct a feature transformation from n_{PS} to $d' = 8$.

Two-stream Network. Inspired by [15], a two-stream network is proposed to process raw data and PS separately believing that each of them represents a kind of temporal information aggregated to a certain level. Considering ROIs influence cognition abilities differently along time, we introduce group fully connected layers in both two streams (surrounded by blue and pink in Fig. 2(a)) regarding features from each time point as a group. Then, group-specific fully connected layers are applied to encode cortical structures at corresponding stages of brain development. At the bottom of BrainPSNet, we concatenate and fuse the

(a)

(b)
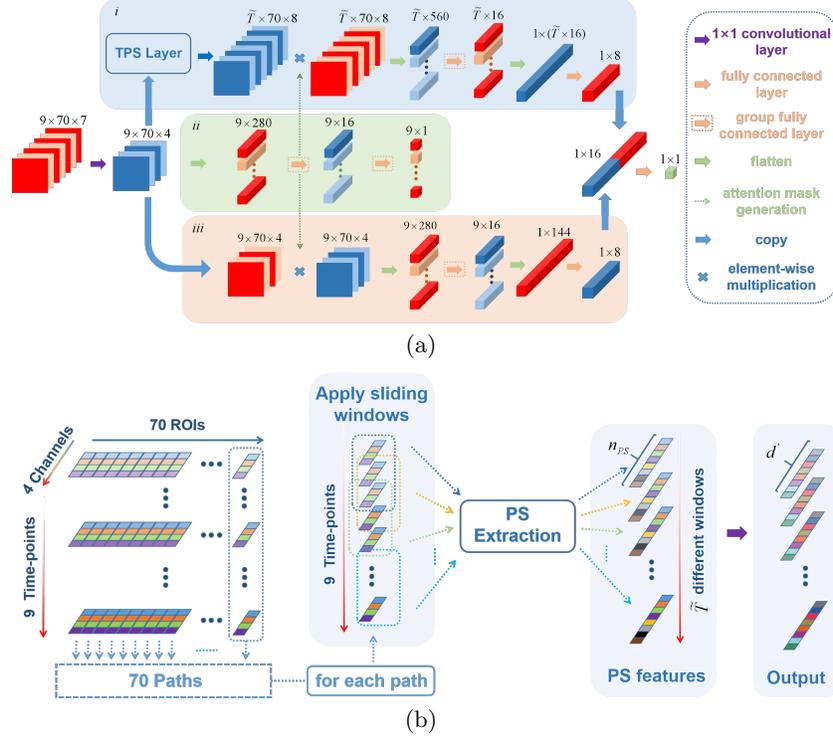
Fig. 2. The illustration of TPS layer and network structure of BrainPSNet. (a) presents a two-stream network with an attention mask generator surrounded by blue, pink and green areas separately; (b) shows the procedure of TPS layer implemented in (a).

informative vectors produced by these two streams and output a final cognitive score y.

Attention Mask Generator. With the aim of emphasizing the most influential region in each stage, an attention mask generator is constructed in the middle of Fig. 2(a). Group fully connected layers are applied sequentially to output an intermediate cognitive score $y_i, i \in \{1, 2, \cdots, 9\}$ for each group of input data. Afterwards, we sum over parameters of fully connected layers which green dash arrows depart from along input channels and generate nine 1×70 vectors corresponding to nine developmental stages and 70 ROIs. Element-wise multiplications are conducted between groups of features and corresponding attention masks to weight ROIs differently along time. Notably, we calculate moving average on attention masks to fit in $\widetilde{T}$ groups of features in PS stream. In this work, the intermediate output $\hat{y} = (y_1, y_2, \cdots, y_9)$ is just used to assist generating attention masks with the loss function defined as:

$$Loss = \lambda ||\hat{Y} - \hat{y}||_{l_1} + |Y - y| \tag{7}$$

We denote the ground truth corresponds to y and $\hat{y}$ as Y and $\hat{Y} = (Y_1, Y_2, \cdots, Y_9)$ respectively. It is noteworthy that $\hat{Y}$ is a duplication of Y at nine time points. λ is introduced to balance these two different losses.

5 Experiments

Table 1. Performance comparison between the TPS layer and sequence models (in terms of RMSE). In the last column, we calculate the total time cost for 1200 epochs of our method with different substitutes.

Methods	VRS	FMS	RLS	ELS	ELC		AVE	Time
Transformer	0.076	0.090	0.134	0.066	0.075		0.088	578 s
LSTM	0.084	0.084	0.120	**0.037**	0.081		0.081	2147 s
TPS layer	**0.046**	**0.075**	**0.095**	0.063	**0.057**		**0.067**	405 s

Table 2. Performance comparison and the ablation study of BrainPSNet (in terms of RMSE)

Methods	VRS	FMS	RLS	ELS	ELC		AVE
NN	0.219	0.259	0.165	0.196	0.182		0.204
MtJFS [2]	0.276	0.273	0.189	0.214	0.134		0.217
RMTL [6]	0.146	0.200	0.178	0.188	0.137		0.170
TrMTL [11]	0.279	0.276	0.192	0.217	0.136		0.220
LPMvRL [28]	0.162	0.189	0.139	0.165	0.138		0.158
BrainPSNet(*w/o attention*)	0.092	0.108	0.162	0.077	0.103		0.108
BrainPSNet(*w/o PS*)	0.059	0.100	0.103	0.066	0.089		0.084
BrainPSNet	**0.046**	**0.075**	**0.095**	**0.063**	**0.057**		**0.067**

Configuration. We conduct experiments on an acquired in-house dataset which has been illustrated previously in section 3. Note that based on the available data, average interpolation has been applied to missing data in both training and testing sets. Following [28], we perform leave-one-out validation and calculate root mean squared error (RMSE) between the predict values and ground truth for all five scores. Sliding window size W, truncated level k and *lambda* are fixed at $W = 4$, $k = 2$, $\lambda = 0.1$ respectively. We tune learning rate in $\{10^{-3}, 10^{-4}, 10^{-5}\}$ with Adam as optimizer. The non-linear activation for hidden neurons is ReLU. The number of epochs is at a maximum of 400 for all experiments. Notably, we normalize five cognitive scores with their maximum and minimum values separately with in a $[0, 1]$ range to have a unified comparison setting. Our code will be avalilable at ???.

Comparison. We first run out method with different sequence models to illustrate the effectiveness of PS feautres. LSTM and Transfomer [25] are selected as our substitutes. In practice, we replace the TPS layer in our model (Fig. 1) with single layer LSTM or Transformer encoder. In Table 1, AVE stands for the average RMSE for five cognitive functions. We also calculate average R-squared metric for these three methods, which are 0.803, 0.785 and 0.634 for TPS layer, LSTM and Transformer separately. Thus, the proposed TPS layer-based method shows clear advantages over the other two substitutes in various metrics.

To validate the performance of our method, several recent algorithms are selected as baselines, including: 1) NN (nearest neighbour); 2) MtJFS (Multi-Task Learning with Joint Feature Selection) [2]; 3) RMTL (Robust Multi-Task Feature Learning) [6]; 4) TrMTL (Trace-Norm Regularized Multi-Task Learning) [11] and 5) LPMvRL (Latent Partial Multi-view Representation Learning) [28]. To make these methods comparable, we applied the same preprocessing method including normalization and interpolation. From Table 2, we find that **BrainPSNet** outperforms the other algorithms under the same settings. Additionally, an ablation study is conducted to explore the ability of path signature features and attention vectors respectively. It is observed that both of them bring improvements to the final result which proves the effectiveness of our method.

Result analysis. In this section, we try to investigate which morphological features or ROIs are more important in early postnatal period utilizing testing data and the trained models. Motivated by [24], we compute the gradient of testing data w.r.t. the loss by backpropagation and regard features with higher gradient are paid more attention by the network.

First, we compare the importance of seven anatomical measurements w.r.t. five cognitive scores. As Fig. 3(a) illustrates, curvature collects more attentions for most tasks. In rest subfigures, the x-axis denotes indices of different ROIs, which can be found in surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation. In practice, we sum up the importance coefficient of corresponding ROIs in left or right hemispheres. It can be seen that the postcentral gyrus (ROI 22) and supramarginal gyrus (ROI 31) are more important for FMS and RLS respectively, while VRS and ELS are both concentrated on pars triangularis (ROI 20). An interesting observation is that for VRS, FMS and ELC, the medial orbitofrontal cortex (ROI 3) is relatively important.

6   Conclusion

In this paper, we propose a novel model, BrainPSNet, to predict cognitive scores using longitudinal cortical structures during infancy. For the first time, the path signature feature is introduced to explore hidden anatomical and geometric properties of the cortical developmental trajectories by a TPS layer. Based on path signature features and raw features, a multi-stream model is constructed to combine information of various granules and generate informative representation for each participant. Furthermore, considering different ROIs' influence on cognition abilities along time, we propose an attention generator to produce learning-based
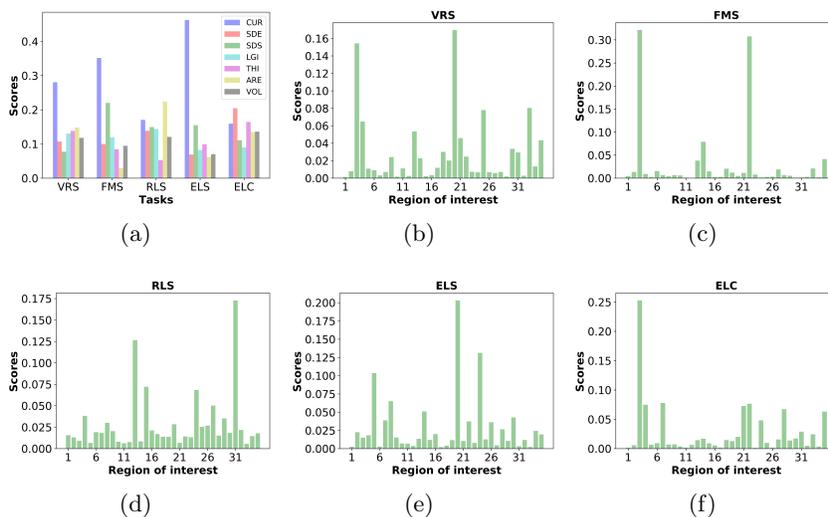
Fig. 3. The illustration of influence of anatomical features and ROIs w.r.t. five Mullen scores. (a) depicts the importance of seven anatomical features, while the rest show the importance distribution among different regions w.r.t. five cognitive scores respectively.

attention masks to weight ROIs at different developmental stages. Experiments and ablation study show that our method outperforms all baselines and achieves the state-of-the-art performance. Both path signature features and attention masks contribute to the final result.

References

1. Adeli, E., Meng, Y., Li, G., Lin, W., Shen, D.: Multi-task prediction of infant cognitive scores from longitudinal incomplete neuroimaging data. NeuroImage **185**(April 2018), 783–792 (2019)
2. Argyriou, A., Evgeniou, T., Pontil, M.: Multi-task feature learning. In: Advances in neural information processing systems. pp. 41–48 (2007)
3. Boedihardjo, H., Geng, X., Lyons, T., Yang, D.: The signature of a rough path: Uniqueness. Advances in Mathematics **293**, 720–737 (2016)
4. Bonnier, P., Kidger, P., Arribas, I.P., Salvi, C., Lyons, T.: Deep signatures. arXiv preprint arXiv:1905.08494 (2019)
5. Braaten, E.B.: Mullen Scales of Early Learning. The SAGE Encyclopedia of Intellectual and Developmental Disorders **34**(4), 379–382 (2018)
6. Chen, J., Zhou, J., Ye, J.: Integrating low-rank and group-sparse structures for robust multi-task learning. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 42–50 (2011)
7. Chen, K.T.: Integration of Paths–A Faithful Representation of Paths by Noncommutative Formal Power Series. Transactions of the American Mathematical Society **89**(2), 395 (1958)

8. Chevyrev, I., Kormilitzin, A.: A Primer on the Signature Method in Machine Learning (2016)
9. Hambly, B., Lyons, T.: Uniqueness for the signature of a path of bounded variation and the reduced path group. Annals of Mathematics **171**(1), 109–167 (2010)
10. Hazlett, H.C., Gu, H., Munsell, B.C., Kim, S.H., Styner, M., Wolff, J.J., Elison, J.T., Swanson, M.R., Zhu, H., Botteron, K.N.: Early brain development in infants at high risk for autism spectrum disorder. Nature **542**(7641), 348–351 (2017)
11. Ji, S., Ye, J.: An accelerated gradient method for trace norm minimization. In: Proceedings of the 26th International Conference On Machine Learning, ICML 2009. pp. 457–464 (2009)
12. Kagan, J., Herschkowitz, N., Herschkowitz, E.: A young mind in a growing brain, vol. 9781410613. Psychology Press (2005)
13. Lai, S., Jin, L., Yang, W.: Toward high-performance online HCCR: A CNN approach with DropDistortion, path signature and spatial stochastic max-pooling. Pattern Recognition Letters **89**, 60–66 (2017)
14. Levin, D., Lyons, T., Ni, H.: Learning from the past, predicting the statistics for the future, learning an evolving system. arXiv preprint arXiv:1309.0260 (2013)
15. Li, C., Zhang, X., Liao, L., Jin, L., Yang, W.: Skeleton-Based Gesture Recognition Using Several Fully Connected Layers with Path Signature Features and Temporal Transformer Module. Proceedings of the AAAI Conference on Artificial Intelligence **33**, 8585–8593 (2019)
16. Li, G., Wang, L., Shi, F., Gilmore, J.H., Lin, W., Shen, D.: Construction of 4D high-definition cortical surface atlases of infants: Methods and applications. Medical Image Analysis **25**(1), 22–36 (oct 2015)
17. Li, G., Wang, L., Shi, F., Lyall, A.E., Lin, W., Gilmore, J.H., Shen, D.: Mapping longitudinal development of local cortical gyrification in infants from birth to 2 years of age. Journal of Neuroscience **34**(12), 4228–4238 (2014)
18. Li, G., Wang, L., Yap, P.T., Wang, F., Wu, Z., Meng, Y., Dong, P., Kim, J., Shi, F., Rekik, I.: Computational neuroanatomy of baby brains: A review. NeuroImage **185**, 906–925 (2019)
19. Liu, M., Jin, L., Xie, Z.: PS-LSTM: Capturing Essential Sequential Online Information with Path Signature and LSTM for Writer Identification. In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR. vol. 1, pp. 664–669. IEEE (2017)
20. Lyall, A.E., Shi, F., Geng, X., Woolson, S., Li, G., Wang, L., Hamer, R.M., Shen, D., Gilmore, J.H.: Dynamic Development of Regional Cortical Thickness and Surface Area in Early Childhood. Cerebral Cortex **25**(8), 2204–2212 (2015)
21. Lyons, T., Ni, H., Oberhauser, H.: A feature set for streams and an application to high-frequency financial tick data. In: ACM International Conference Proceeding Series. vol. 04-07-Augu, pp. 1–8 (2014)
22. Meng, Y., Li, G., Gao, Y., Lin, W., Shen, D.: Learning-based subject-specific estimation of dynamic maps of cortical morphology at missing time points in longitudinal infant studies. Human Brain Mapping **37**(11), 4129–4147 (2016)
23. Rekik, I., Li, G., Yap, P.T., Chen, G., Lin, W., Shen, D.: Joint prediction of longitudinal development of cortical surfaces and white matter fibers from neonatal MRI. NeuroImage **152**, 411–424 (may 2017)
24. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization pp. 618–626 (2017)

25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems **2017-Decem**(Nips), 5999–6009 (2017)
26. Wu, Z., Wang, L., Lin, W., Gilmore, J.H., Li, G., Shen, D.: Construction of 4D infant cortical surface atlases with sharp folding patterns via spherical patch-based group-wise sparse representation. Human brain mapping **40**(13), 3860–3880 (2019)
27. Yang, W., Lyons, T., Ni, H., Schmid, C., Jin, L., Chang, J.: Leveraging the Path Signature for Skeleton-based Human Action Recognition (2017)
28. Zhang, C., Adeli, E., Wu, Z., Li, G., Lin, W., Shen, D.: Infant Brain Development Prediction With Latent Partial Multi-View Representation Learning. IEEE Transactions on Medical Imaging **38**(4), 909–918 (2019)