Conditional Recurrent Flow: Conditional Generation of Longitudinal Samples with Applications to Neuroimaging

University of Pittsburgh · The University of Wisconsin Madison · University of Texas Arlington

Seong Jae Hwang Zirui Tao Won Hwa Kim Vikas Singh

Motivation: Longitudinal Neuroimaging Analysis

Goal: Understand the *progression of longitudinal neuroimaging measures* (e.g., PET) of people with various *covariate progressions* (e.g., "Abnormal": High→Low cognition vs. "Normal": High→High cognition)

Challenge: Difficult to perform strong statistical analysis with small sample size of longitudinal neuroimaging datasets

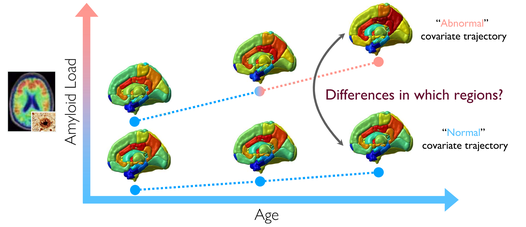


Figure: **Longitudinal neuroimaging analysis of Alzheimer's disease (AD) pathology.** Understanding how amyloid (AD pathology) accumulates differently in brain regions between people of varying conditions may help us to better understand the underlying disease mechanism.

Objective: Generation of Neuroimaging Measures

Solution: Conditional generation of longitudinal neuroimaging measures via Conditional Recurrent Flow (CRow)

Example of Figure below with a trained CRow model:

1) Given: A sequential condition of decreasing cognition (i.e., a memory test score sequence $\mathbf{y}_i^1 \rightarrow \mathbf{y}_i^2 \rightarrow \mathbf{y}_i^3$ indicating High→Medium→Low Cognition performance).

2) Model: Conditional Recurrent Flow (CRow).

3) Generate: A sequence of brain image progression $\mathbf{x}_i^1 \rightarrow \mathbf{x}_i^2 \rightarrow \mathbf{x}_i^3$ corresponding to the given cognition progression (i.e., brain regions with high (red) and low (blue) disease pathology).

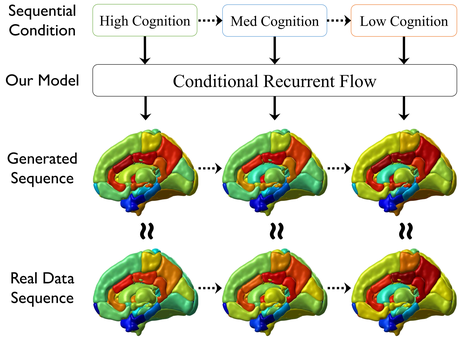


Figure: **Conditional sequence generation illustration.** The Generated Sequence follows the trend of the Real Data Sequence (i.e., similar (≈) to the real brain image progression) from the subjects with similarly decreasing cognition scores.

Preliminary 1: Normalizing Flow and Coupling Layer

Normalizing Flow: Maps a sample $\mathbf{x}$ to a latent variable $\mathbf{z} = f(\mathbf{x})$ where $\mathbf{z}$ is from a standard normal distribution $\mathbf{Z}$:

$$p_{\mathbf{X}}(\mathbf{x}) = p_{\mathbf{Z}}(\mathbf{z})/|J_{\mathbf{X}}|, \quad |J_{\mathbf{X}}| = \left|\frac{\partial[\mathbf{x} = f^{-1}(\mathbf{z})]}{\partial \mathbf{z}}\right| \tag{1}$$

where $|J_{\mathbf{X}}|$ is a Jacobian determinant.

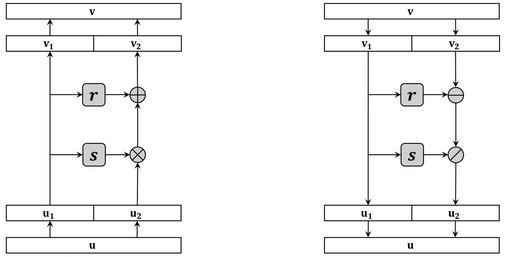
Coupling Layer [Dinh et al., 2016]: Allows an exactly invertible mapping $\mathbf{u} \leftrightarrow \mathbf{v}$ with subnetworks r and s.

1. Forward map (Fig. (a)):

$$\mathbf{v}_1 = \mathbf{u}_1, \quad \mathbf{v}_2 = \mathbf{u}_2 \otimes \exp(s(\mathbf{u}_1)) + r(\mathbf{u}_1) \tag{2}$$

2. Inverse map (Fig. (b)):

$$\mathbf{u}_1 = \mathbf{v}_1, \quad \mathbf{u}_2 = (\mathbf{v}_2 - r(\mathbf{v}_1)) \oslash \exp(s(\mathbf{v}_1)) \tag{3}$$



(a) Forward map (b) Inverse map

Figure: Coupling layer in normalizing flow.

Advantage: Easy Jacobian determinant computation via triangular Jacobian matrix $J_{\mathbf{v}}$:

$$|J_{\mathbf{v}}| = \left|\frac{\partial \mathbf{v}}{\partial \mathbf{u}}\right| = \begin{vmatrix} \frac{\partial \mathbf{v}_1}{\partial \mathbf{u}_1} & \frac{\partial \mathbf{v}_1}{\partial \mathbf{u}_2} \\ \frac{\partial \mathbf{v}_2}{\partial \mathbf{u}_1} & \frac{\partial \mathbf{v}_2}{\partial \mathbf{u}_2} \end{vmatrix} = \begin{vmatrix} I & 0 \\ \frac{\partial \mathbf{v}_2}{\partial \mathbf{u}_1} & diag(\exp s(\mathbf{u}_1)) \end{vmatrix} = \exp(\sum_i (s(\mathbf{u}_1))_i) \tag{4}$$

Coupling Block: Apply the transformation on both partitions.

1. Forward map:

$$\mathbf{v}_1 = \mathbf{u}_1 \otimes \exp(s_2(\mathbf{u}_2)) + r_2(\mathbf{u}_2), \quad \mathbf{v}_2 = \mathbf{u}_2 \otimes \exp(s_1(\mathbf{v}_1)) + r_1(\mathbf{v}_1) \tag{5}$$

2. Inverse map:

$$\mathbf{u}_2 = (\mathbf{v}_2 - r_1(\mathbf{v}_1)) \oslash \exp(s_1(\mathbf{v}_1)), \quad \mathbf{u}_1 = (\mathbf{v}_1 - r_2(\mathbf{u}_2)) \oslash \exp(s_2(\mathbf{u}_2)) \tag{6}$$

Preliminary 2: Conditional Sample Generation

Conditional Invertible Neural Network [Ardizonne et al., 2019]: A *conditional* invertible mapping between $\mathbf{x} \leftrightarrow [\mathbf{y}, \mathbf{z}]$ using *Coupling Layer*.

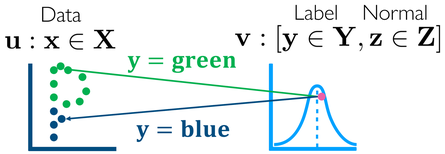


Figure: **Conditional Sample Generation.** (1) Sample $\mathbf{z} \sim \mathbf{Z}$, (2) Assign label $\mathbf{y}$ (i.e., color), (3) Inverse map to generate $\mathbf{x} = f^{-1}([\mathbf{y}, \mathbf{z}])$ in the appropriate location.

Method: Conditional Recurrent Flow (CRow)

Conditional Recurrent Flow (CRow) [Hwang et al., 2019]:

Given: Sequential data $\mathbf{x}^t \in \mathbf{X}$ and label/covariate $\mathbf{y}^t \in \mathbf{Y}$.
At each t with given $[\mathbf{u}_1^t, \mathbf{u}_2^t] = \mathbf{u}^t \leftarrow \mathbf{x}^t$ and $[\mathbf{v}_1^t, \mathbf{v}_2^t] = \mathbf{v}^t \leftarrow [\mathbf{y}^t, \mathbf{z}^t]$,

1. Forward Map:

$$\mathbf{v}_1^t = \mathbf{u}_1^t \otimes \exp(q_{s_2}(\mathbf{u}_2^t, \mathbf{h}_2^{t-1})) + q_{r_2}(\mathbf{u}_2^t, \mathbf{h}_2^{t-1}), \quad \mathbf{v}_2^t = \mathbf{u}_2^t \otimes \exp(q_{s_1}(\mathbf{v}_1^t, \mathbf{h}_1^{t-1})) + q_{r_1}(\mathbf{v}_1^t, \mathbf{h}_1^{t-1})$$

2. Inverse Map:

$$\mathbf{u}_2^t = (\mathbf{v}_2^t - q_{r_1}(\mathbf{v}_1^t, \mathbf{h}_1^{t-1})) \oslash \exp(q_{s_1}(\mathbf{v}_1^t, \mathbf{h}_1^{t-1})), \quad \mathbf{u}_1^t = (\mathbf{v}_1^t - q_{r_2}(\mathbf{u}_2^t, \mathbf{h}_2^{t-1})) \oslash \exp(q_{s_2}(\mathbf{u}_2^t, \mathbf{h}_2^{t-1}))$$

where $q(\cdot)$ is a recurrent subnetwork (e.g., GRU).

Advantage: Longitudinal *and* conditional data generation along with density estimation.

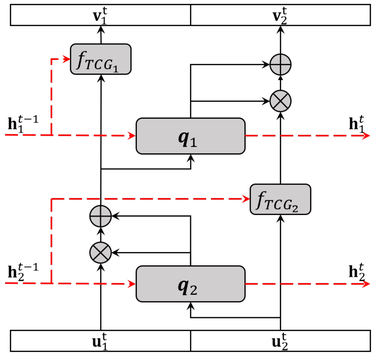


Figure: The CRow model.

Temporal Context Gating (TCG): Context-based (i.e., history) transformation of the "non-transformed partition" (e.g., $\mathbf{u}_1$ in Eq. (2)).

$$f_{\text{TCG}}(\alpha^t, \mathbf{h}^{t-1}) = \alpha^t \otimes cgate(\mathbf{h}^{t-1}) \quad \text{(forward)}, \quad f_{\text{TCG}}^{-1}(\alpha^t, \mathbf{h}^{t-1}) = \alpha^t \oslash cgate(\mathbf{h}^{t-1}) \quad \text{(inverse)}$$

Advantage: Additional recurrent power while preserving the triangular Jacobian matrix for fast Jacobian determinant computation:

$$|J_{\mathbf{v}}| = \left|\frac{\partial \mathbf{v}}{\partial \mathbf{u}}\right| = \begin{vmatrix} diag(cgate(\mathbf{h}^{t-1})) & 0 \\ \frac{\partial \mathbf{v}_2}{\partial \mathbf{u}_1} & diag(\exp s(\mathbf{u}_1)) \end{vmatrix} = [\prod_j cgate(\mathbf{h}^{t-1})_j] * [\exp(\sum_i (s(\mathbf{u}_1))_i)]$$

Experiment 1: Generate Moving MNIST Sequences

Q: Can we generate new data sequences given new sequential conditions?

(1) Train with label sequences $\mathbf{y}$: 5→5→5→5→5→5 and $\mathbf{y}$: 9→9→9→9→9→9

(2) Generate data sequences given a newly seen label sequences $\mathbf{y}$ which changes label midway: 5→5→9→9→9→9

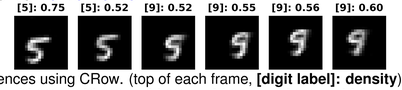


Figure: Generated sequences using CRow. (top of each frame, **[digit label]: density**)

Figure: Examples of generated Moving Fashion MNIST sequences using CRow.

Experiment 2: Neuroimaging for Alzheimer's Disease

Q: Can we detect the differences in the Alzheimer's disease (AD) pathology progression between people with "Normal" and "Abnormal" covariate progressions? (Idea illustrated in Fig. of Motivation Sec.)

Dataset: $N = 276$ Amyloid PET images (AV45) of $T = 3$ time points from Alzheimer's Disease Neuroimaging Initiative (ADNI) in 82 Desikan Atlas regions
⇒ 1. Longitudinal region-wise amyloid measures: $\mathbf{x}^t \in \mathbb{R}^{82}$ for $t = 1, 2, 3$
2. Longitudinal covariates: $\mathbf{y}^t \in \mathbb{R}$ for $t = 1, 2, 3$ (e.g., cognition)

Analysis Setup: For each covariate type,
(1) *Training*: Train with all $N = 276$ subjects
(2) *Generation*: Generate (i) Group A: 100 samples of $\mathbf{x}^t$ given "Abnormal" $\mathbf{y}^t$ and (ii) Group B: 100 samples of $\mathbf{x}^t$ given "Normal" $\mathbf{y}^t$
(3) *Statistical Analysis*: At $t = 3$, perform a group difference test between Group A and Group B in *each* region. Count significantly differences regions.

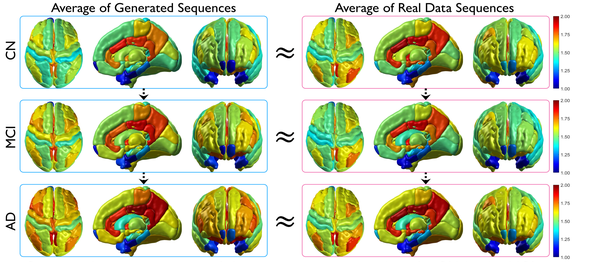


Figure: Generated sequences vs. real data sequences comparison for CN (top)→MCI (middle)→AD (bottom). **Left (blue frames):** The average of the 100 generated sequences conditioned on CN→MCI→AD. **Right (pink frames):** The average of the real samples with CN→MCI→AD in the dataset. Red/blue indicate high/low AV45. ROIs are expected to turn more red as CN→MCI→AD. The generated samples show magnitudes and sequential patterns similar (≈) to those of the real samples from the training data.

How well did CRow improve the statistical analysis?

	# of Statistically Significant ROIs (# of ROIs after type-I error correction)				
Covariates	Diagnosis	ADAS13	MMSE	RAVLT-I	CDR-SB
Control	CN→CN→CN	10→10→10	30→30→30	70→70→70	0→0→0
Progression	CN→MCI→AD	10→20→30	30→26→22	70→50→30	0→5→10
cINN (N_1=100 / N_2= 100)	11 (4)	5 (2)	5 (0)	3 (0)	7 (0)
Ours (N_1=100 / N_2= 100)	25 (11)	24 (12)	19 (**2**)	15 (**2**)	18 (7)
Ours + TCG (N_1=100 / N_2= 100)	**28** (**12**)	**32** (**14**)	**31** (**2**)	**19** (**2**)	**25** (**9**)
Control	CN→CN→CN	10→10→10	30→30→30	70→70→70	0→0→0
Early-progression	CN→MCI→MCI	10→13→16	30→28→26	70→60→50	0→2→4
cINN (N_1=150 / N_2= 150)	2 (0)	2 (2)	2 (0)	0 (0)	1 (0)
Ours (N_1=150 / N_2= 150)	**6** (2)	6 (4)	11 (**4**)	**5** (**1**)	2 (0)
Ours + TCG (N_1=150 / N_2= 150)	**6** (**4**)	**8** (**5**)	**12** (**4**)	**5** (**1**)	**5** (**1**)

Table: Number of ROIs identified by statistical group analysis using the generated measures with respect to various covariates associated with AD at significance level $\alpha = 0.01$ (type-I error controlled result shown in parenthesis). Each column denotes sequences of disease progression represented by diagnosis/test scores.

How similar are the generated sequences to the real sequences?

	Cohen's d of Gen. vs. Real of Progressions					Cohen's d of Gen. vs. Real of Early-progressions				
Covariates	Diagnosis	ADAS13	MMSE	RAVLT-I	CDR-SB	Diagnosis	ADAS13	MMSE	RAVLT-I	CDR-SB
cINN	1.255	1.596	1.149	1.894	1.551	1.065	1.498	0.948	1.843	1.454
Ours	0.419	0.556	0.348	0.711	0.645	0.359	0.561	0.295	0.613	0.625
Ours+TCG	**0.282**	**0.391**	**0.167**	**0.588**	**0.377**	**0.234**	**0.5248**	**0.090**	**0.544**	**0.499**

Table: Difference between the generated sequences and the real sequences at $t = 3$. Lower the effect size (Cohen's d), smaller the difference between the comparing distributions.

International Conference on Computer Vision (ICCV) 2019

Research supported by NIH (R01AG040396, R01EB022883, R01AG062336, R01AG059312), UW CPCP (U54AI117924), UW CIBM (T15LM007359) NSF CAREER Award (1252725), USDOT Research and Innovative Technology Administration (69A3551747134), and UTA Research Enhancement Program (REP).