Surface Vision Transformers:
Attention-Modelling applied to Cortical Analysis

King's College London
Imperial College London
Washington University in St. Louis
MeTrICS

Simon Dahan
Abdulah Fawaz
Logan Z. J. Williams
Chunhui Yang
Timothy S. Coalson
Matthew F. Glasser
A. David Edwards
Daniel Rueckert
Emma C. Robinson

MIDL Zürich 2022

1. Summary

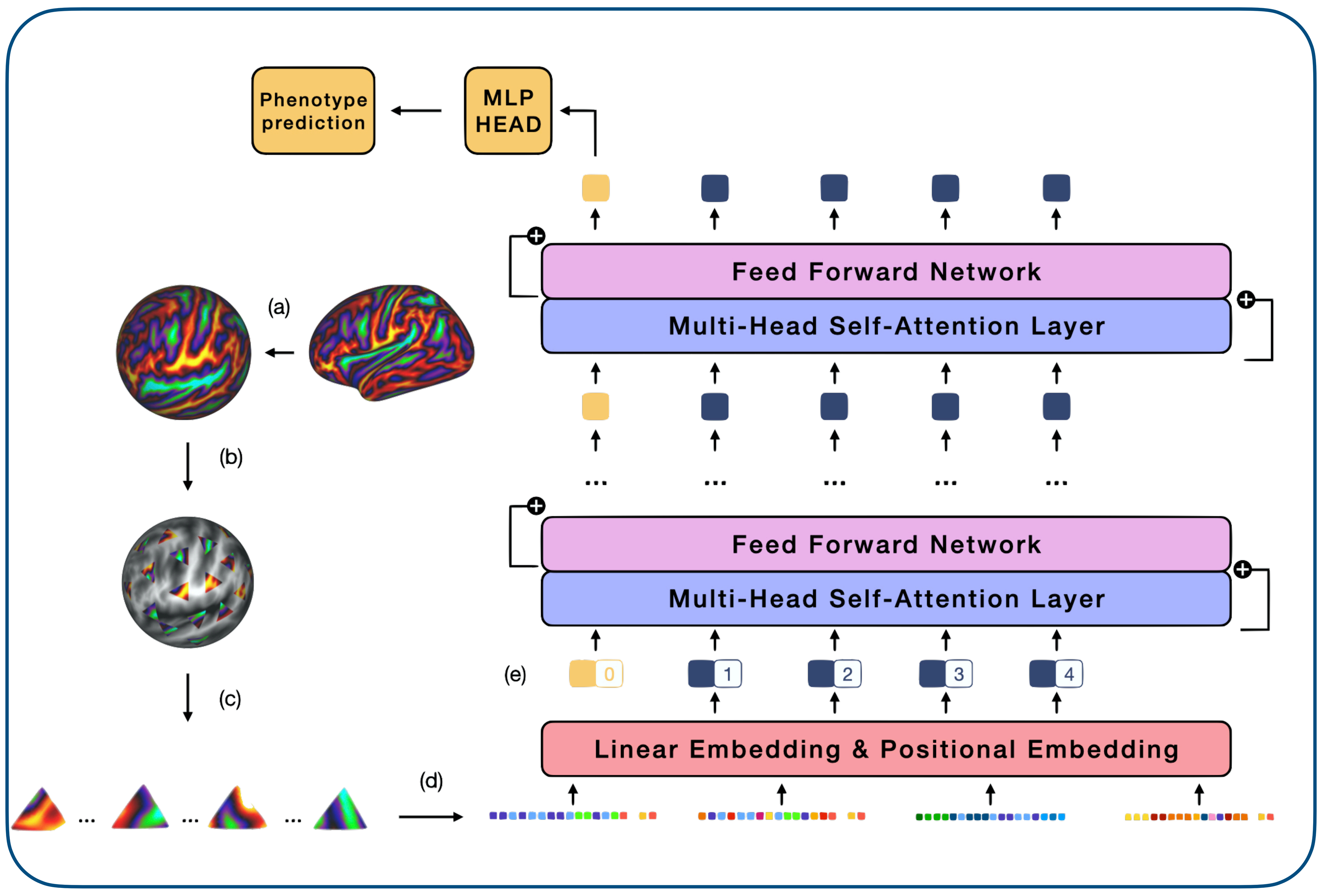
- To overcome the shortcomings of surface CNNs for studying biomedical surfaces with deep learning, we extend the methodology of **sequence modelling with vision transformers to surfaces.**
- Surface patching is realised by **projecting surface data onto a regularly tessellated icosphere** and **extracting a sequence of non-overlapping triangular patches.**
- Results show that Surface Vision Transformers **(SiT) improve performances compared to 5 geometric deep learning methods (gDL)** for neurodevelopmental phenotype predictions tasks while demonstrating **robustness to spatial transformation.**

2. Motivations

- Studying the cerebral cortex is critical to **capture rich features related to neuropsychiatric disorders and brain development from MRI.**
- **Deep learning** can overcome the limitations of image analysis techniques based on registration that fail to capture **the heterogeneity of structural and functional cortical organisation between individuals and across time.**
- However, there is **no generic geometric deep learning framework** to study surfaces:
  - most surface CNNs have shown trade-offs between complexity, expressivity, and equivariance.
  - or failing to extract long-range spatiotemporal dependencies.

3. Methods

Model: **Surface Vision Transformer (SiT)**



Patching strategy: Cortical surface data (32k vertices) is resampled into a 6th-order icosphere (40k vertices) and regularly divided into a sequence of triangular patches **- 320 patches of 153 vertices.**

Data: 4 cortical surface metrics derived from MRI scans for **588 neonatal subjects** (419 term & 169 preterm) as part of the **dHCP dataset.**

Two neurodevelopmental tasks: **Age at scan and Birth age regression** and benchmark against **5 gDL methods:** *S2CNN, ChebNet, GConvNet, Spherical UNet and MoNet.*

paper   code

SD3004   @DahanSimon   @lab_metrics

4. Results

Improving phenotype predictions results

- SiT-small **consistently outperform 3 gDL methods**
- SiT-small achieves best performances on both tasks
  - Scan-age: 0.59 **M**ean **A**verage **E**rror (in weeks) similar to MoNeT
  - Birth-age: 1.12 MAE vs 1.44 MAE for S2CNN (best gDL)
- Demonstrate **robustness to spatial transformation** between template-aligned (registered) and native (un-registered) data.
- BERT-like **pre-training strategy** (mask patch prediction) improves greatly performances of the SiT.

Methods	Pretraining	Scan age			Birth age			
		Template	Native	Avg	Template	Native	Avg	Avg
S2CNN	✗	0.63±0.02	0.73±0.25	0.68	1.35±0.68	1.52±0.60	1.44	1.06
ChebNet	✗	0.59±0.37	0.77±0.49	0.68	1.57±0.15	1.70±0.36	1.64	1.16
GConvNet	✗	0.75±0.13	0.75±0.26	0.75	1.77±0.26	2.30±0.74	2.04	1.39
Spherical UNet	✗	0.57±0.18	0.87±0.50	0.72	**0.85±0.17**	2.16±0.57	1.51	1.11
MoNet	✗	0.57±0.02	**0.61±0.05**	**0.59**	1.44±0.08	1.58±0.06	1.51	1.05
SiT-tiny	✗	0.63±0.01	0.77±0.03	0.70	1.37±0.03	1.66±0.06	1.52	1.11
SiT-tiny	✓	0.58±0.01	0.64±0.06	0.61	1.18±0.07	1.61±0.03	1.39	1.00
SiT-small	✗	0.60±0.02	0.76±0.03	0.68	1.14±0.12	**1.44±0.03**	**1.29**	0.99
SiT-small	✓	**0.55±0.04**	0.63±0.06	**0.59**	1.13±0.02	1.47±0.08	1.30	**0.95**

Visualisation of Attention Maps

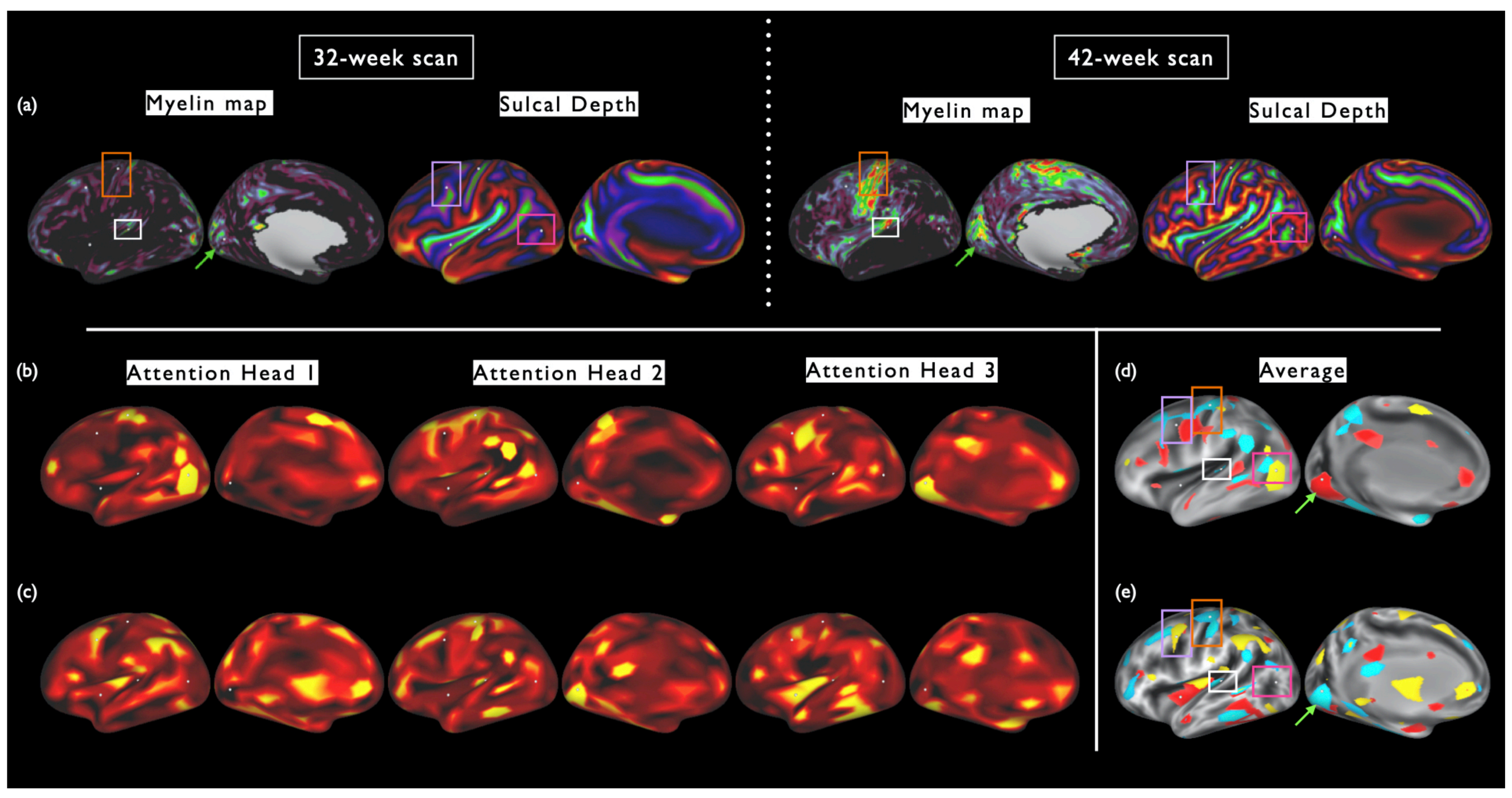


Fig: Comparison of attention maps for a preterm neonate scanned at birth and term-equivalent age. Attention maps capture the well-characterised spatio-temporal patterns of perinatal brain development.