University of Chinese Academy of Sciences

一种用于形变医学图像配准的无监督学习模型

An Unsupervised Learning Model for Deformable Medical Image Registration

Guha Balakrishnan
MIT
balakg@mit.edu

Amy Zhao
MIT
xamyzhao@mit.edu

Mert R. Sabuncu
Cornell University
msabuncu@cornell.edu

John Guttag
MIT
guttag@mit.edu

Adrian V. Dalca
MIT and MGH
adalca@mit.edu

小组成员：吴文霞，肖安琪，王国帅，蒲 尧

中国科学院大学
University of Chinese Academy of Sciences

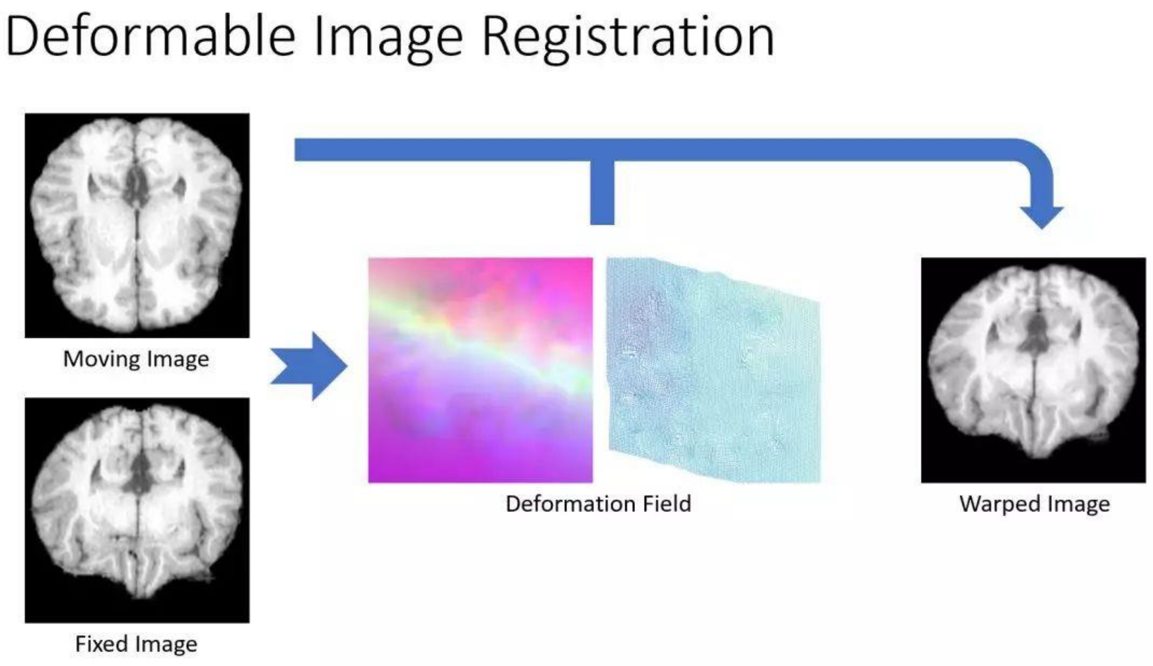
目录
Contents

01 研究背景

◎配准问题 ◎配准类型 ◎相关工作

1）配准问题

- **配准问题**：
- 给定一个浮动图像（moving image）和一张固定图像（fixed image）。预测一个位移场，进而得到形变场（deformation field），即从浮动图像到固定图像的映射，使得配准后的浮动图像（warped image）和固定图像尽可能相似。



2）配准类型

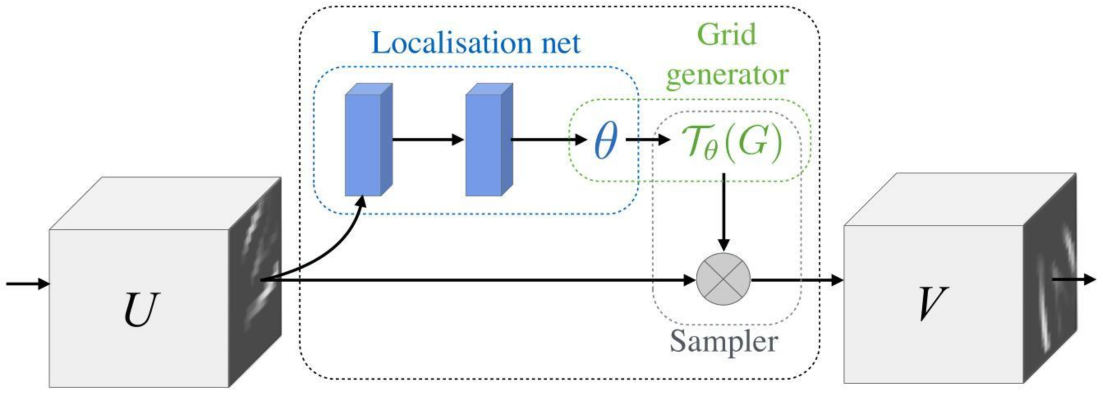
- **非学习配准**：基于数学优化的方法
- 优点：通常效果比较好且稳定
- 缺点：对每个图像进行迭代优化，耗费时间较长
- 典型模型：
  - 弹性模型（elastic-type models）
  - 统计参数映射（statistical parametric mapping）
  - b样条自由变形（free-form deformations with b-splines）
  - 麦克斯韦妖（maxwell's demons）
- **学习型配准**：通过神经网络训练的配准方法，利用大量的数据来训练一个模型，然后用这个训练好的模型对一个新的图像进行配准。这些方法大多依赖于ground truth扭曲区域或分割
- 优点：训练过程缓慢，配准过程快很多
- 缺点：训练数据较少，没传统方法稳定

- **有监督**：
  - 基于代理的行动学习实现非刚性配准
  - 利用形状匹配学习可变形图像配准
  - 基于多尺度三维卷积神经网络的非刚性图像配准
  - 快速预测图像配准-一种深度学习方法
- **无监督**：
  - 基于卷积神经网络的端到端无监督变形图像配准
  - 基于深度自监督全卷积网络的非刚性图像配准

中国科学院大学
University of Chinese Academy of Sciences

3）相关工作

- **光流估计**：
- 光流算法返回一个密集的**位移矢量场**，描绘了二维图像对之间的**小位移。**
- 传统的光流方法通常用变分方法解决优化问题。能更好地处理大位移或外观的戏剧性变化，包括基于特征的匹配和最近邻区域的匹配。
- Flownet: Learning optical flow with convolutional networks.
- Deepflow: Large displacement optical flow with deep matching.
- 缺点：需要在训练期间进行ground truth标注。



Spatial Transformer

空间变换层：空间转换网络会根据位移场生成一个归一化后的采样网格，然后用该网络对图像进行采样，就得到了配准后的图像。

优点：使神经网络能够在**不需要监督标签**的情况下执行全局参数2D图像对准。

在本文的工作中，作者将空间变换扩展到3D场景。

中国科学院大学 University of Chinese Academy of Sciences

02 研究方法

◎整体Pipeline

◎网络结构

◎空间变换函数

◎损失函数

1) 整体Pipeline

- Moving 3D → Fixed 3D
  - 使用CNN网络学习配准的体素空间变形场φ
  - 空间变换保证映射后的位置落在标准像素点上
  - 用φ平滑loss损失

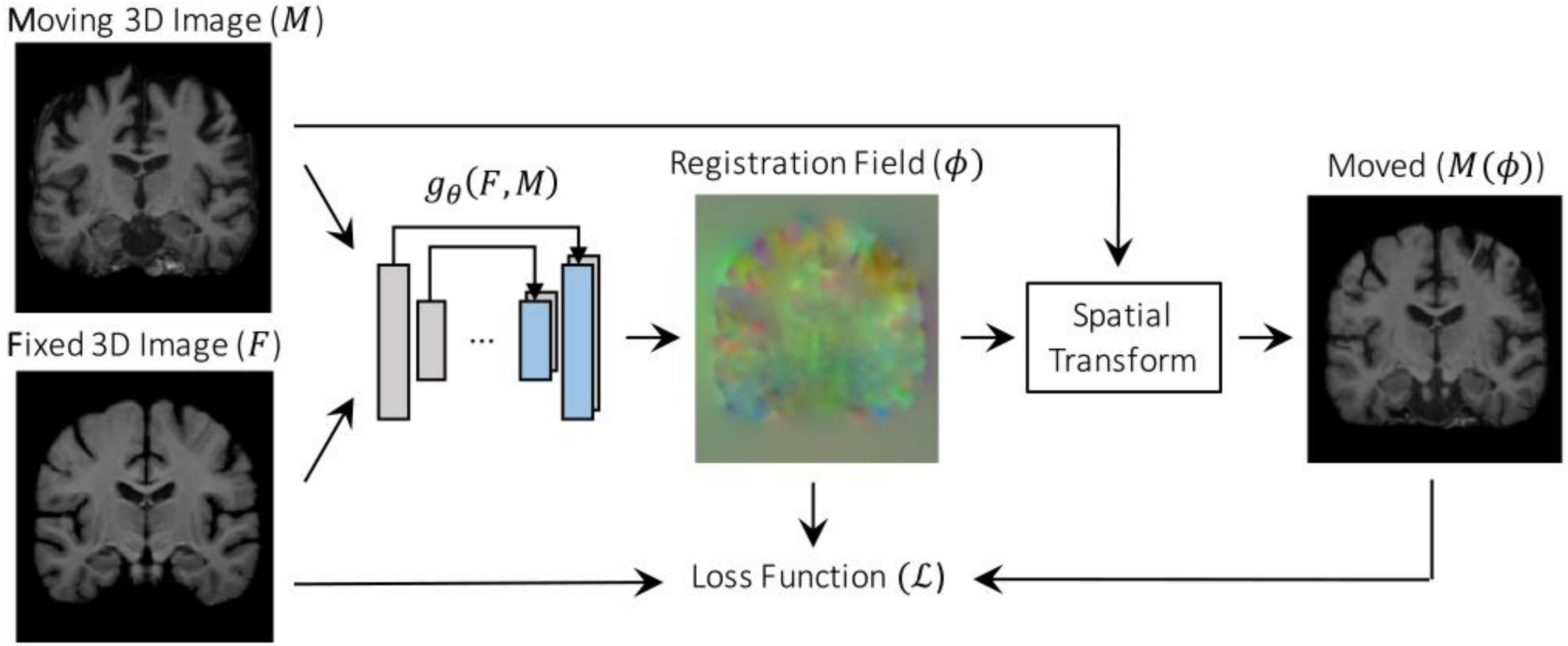


Figure 2: Overview of our method. We learn parameters for a function g that registers one 3D volume (M) to a second, fixed volume (F). During training, we warp M with ϕ using a spatial transformer function. Our loss compares M_ϕ and F and enforces smoothness of ϕ.

2） 网络结构

- 网络结构
  - 整体类似3D U-Net
  - VoxelMorph-1 & VoxelMorph-2
    - 平衡**<u>准确性</u>**和**<u>计算时间</u>**
  - VoxelMorph-2的结构更复杂
    - 卷积层更多 + 通道数更多
    - 更耗时 → 更准确
  - Leaky ReLU激活
    - 相比ReLU不会出现“坏死”现象

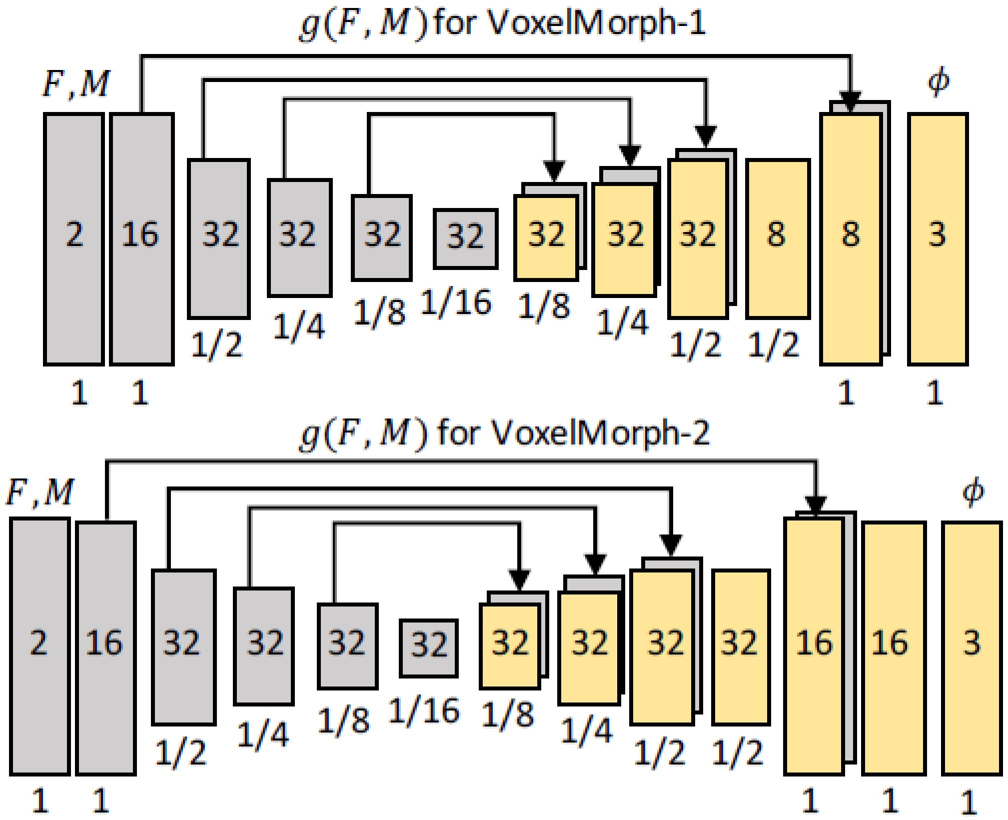
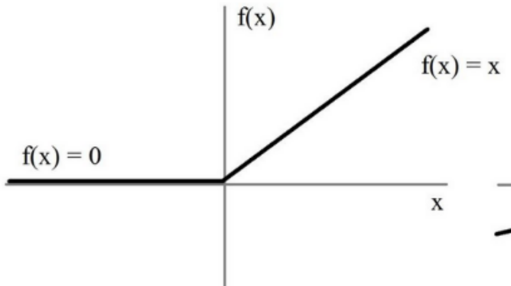
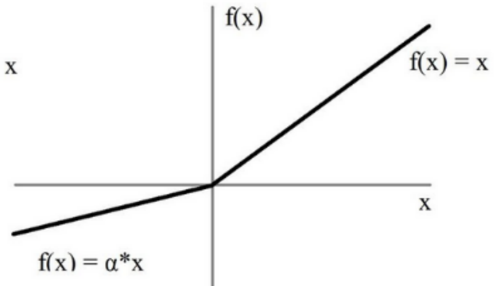


Figure 3: Proposed convolutional architectures implementing $g(F, M)$. Each rectangle represents a 3D volume. The number of channels is shown inside the rectangle, and the spatial resolution with respect to the input volume is printed underneath. VoxelMorph-2 uses a larger architecture, using one extra convolutional layer at the output resolution, and more channels for later layers.





2）网络结构

- 网络结构
  - 跳跃连接
    - 直接融合编-解码器间的特征
    - 帮助网络学习特征表示
  - 输入160x192x224x**2**
    - 2通道分别为M和F
  - 输出160x192x224x**3**
    - 为空间变形场φ
    - 3通道分别为x/y/z三个方向位移
  - 卷积核3x3x3
  - stride=2实现下采样
  - upsampling上采样

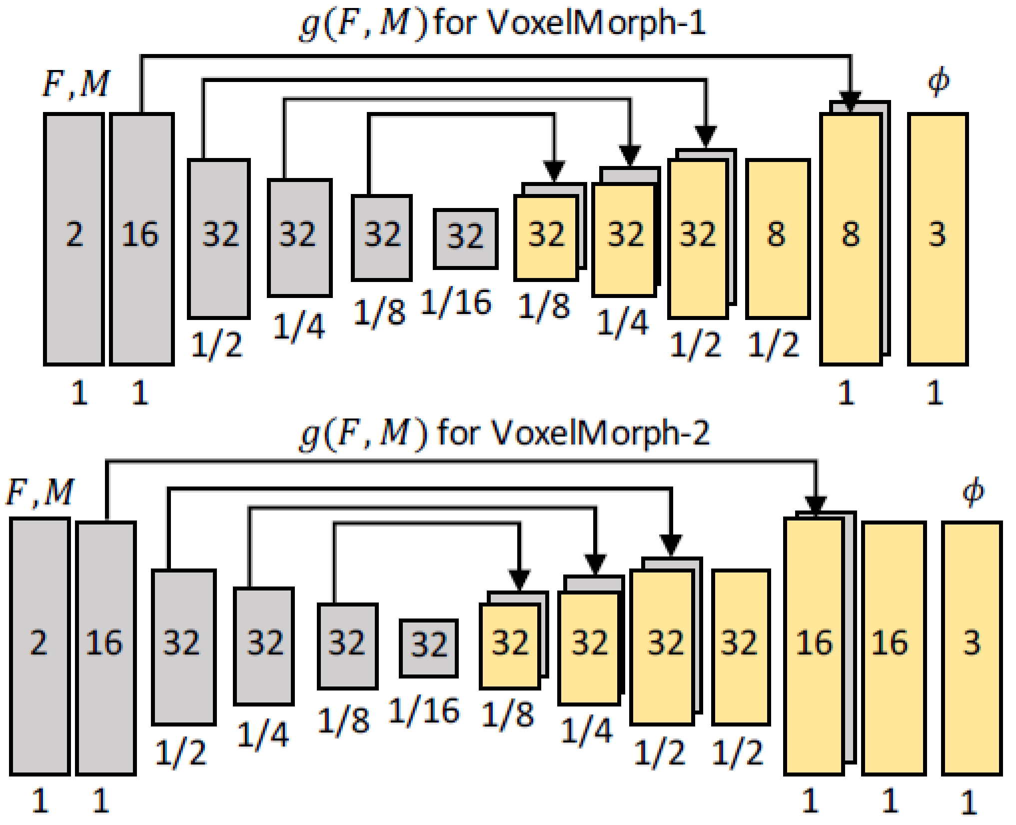


Figure 3: Proposed convolutional architectures implementing $g(F, M)$. Each rectangle represents a 3D volume. The number of channels is shown inside the rectangle, and the spatial resolution with respect to the input volume is printed underneath. VoxelMorph-2 uses a larger architecture, using one extra convolutional layer at the output resolution, and more channels for later layers.

3）空间变换函数

- 解决φ映射后位置不在标准像素点上的问题
  - 保证训练能够按照**梯度反传**的方式进行
  - 计算一个在M上的体素p经过φ映射后φ(p)的位置M(φ(p))

$$M(\phi(p)) = \sum_{q \in \mathcal{Z}(\phi(p))} M(q) \prod_{d \in \{x,y,z\}} (1 - |\phi_d(p) - q_d|),$$

  - **线性插值法**
  - Z(φ(p))是φ(p)的8-邻域体素

中国科学院大学
University of Chinese Academy of Sciences

4）损失函数

- 基于灰度值最大化一个关于图像间匹配程度的目标函数（无监督）
- 整体形式

$$\mathcal{L}(F, M, \phi) = \mathcal{L}_{sim}(F, M(\phi)) + \lambda \mathcal{L}_{smooth}(\phi)$$
$$= -CC(F, M(\phi)) + \lambda \sum_{p \in \Omega} \|\nabla \phi(p)\|^2,$$

- 具体
  - **相似性测度L_{sim}**——M(φ)和F的负的局部**交叉互相关**CC
    - 值越大表示对齐越好
    - 对灰度强度变化更鲁棒

$$CC(F, M(\phi)) = \sum_{p \in \Omega} \left[\frac{\left(\sum_{q \in \mathcal{N}(p)} \hat{F}(q) \hat{M}(\phi(q)) \right)^2}{\left(\sum_{q \in \mathcal{N}(p)} \hat{F}(q)^2 \right) \left(\sum_{q \in \mathcal{N}(p)} \hat{M}(\phi(q))^2 \right) + \epsilon} \right], \quad (4)$$

  - **空间平滑性约束L_{smooth}**——CNN空间梯度正则项
    - 最大化相似性测度时容易引起网络产生不连续的变形场
    - 防止获得的φ不是连续的

$$\mathcal{L}_{smooth}(\phi) = \sum_{p \in \Omega} \|\nabla \phi(p)\|^2.$$

University of Chinese Academy of Sciences

03 相关实验

◎数据集

◎结果分析

◎特定群体测试和超参敏感性分析

1）数据集

- 实验选用了7829张T1加权的脑部MRI数据集，主要来自8个公开数据集ADNI , OASIS, ABIDE , ADHD200 , MCIC , PPMI , HABS , and Harvard GSP， 每个数据集的采集细节、受试者年龄范围和健康状况都不同。
- 在预处理阶段先将图像**重采样**为256×256×256大小，进行**仿射空间正则化**，并使用FreeSurfer工具**提取脑部（去除头骨）并获取分割结果**，再将结果图**裁剪**到**<u>160×192×224</u>**大小。训练集、验证集和测试集包含的图像数分别为7329,250和250。
- 采用**Dice Score**进行评价，可以衡量图像的重叠情况，k=1~29，表示29个解剖结构，S表示对应的数据集

$$\text{Dice}(S^k_{M(\phi)}, S^k_F) = 2 * \frac{S^k_{M(\phi)} \cap S^k_F}{|S^k_{M(\phi)}| + |S^k_F|}.$$

2）结果分析

基于图谱的（atlas-based）配准实验，图谱表示一个参考图像或平均图像，通常是通过联合和反复校准MRI图像数据集并将它们平均在一起来构建的。它和fixed image的作用相同，但是fixed image只是一张图片，而**atlas是图像数据集的平均。**

所有描绘大脑的图像都是二维冠状切片，仅用于可视化目的。所有的配准都是3D的。

下表展示了采用不同模型：仅仿射变换，ANTs（实现了对称归一化(SyN)方法），VoxelMprph-1,2**的平均DICE值**和**运行时间**：

Method	Avg. Dice	GPU sec	CPU sec
Affine only	0.567 (0.157)	0	0
ANTs	0.749 (0.135)	-	9059 (2023)
VoxelMorph-1	0.742 (0.139)	0.365 (0.012)	57(1)
VoxelMorph-2	0.750 (0.137)	0.554 (0.017)	144 (1)

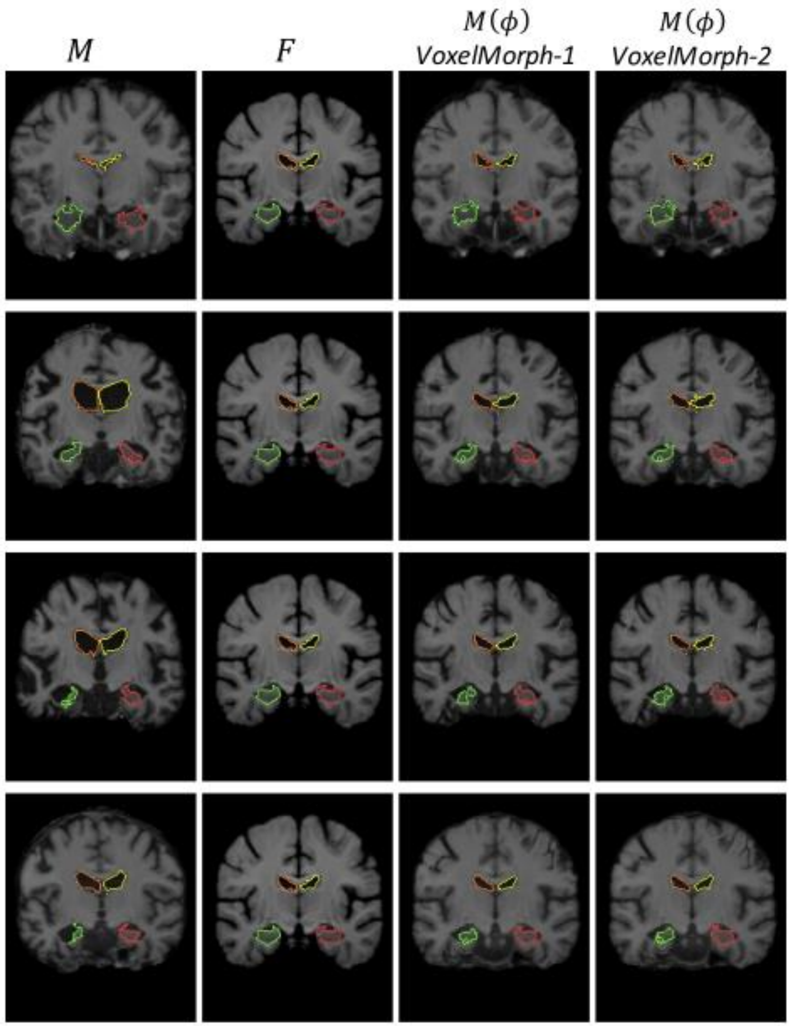
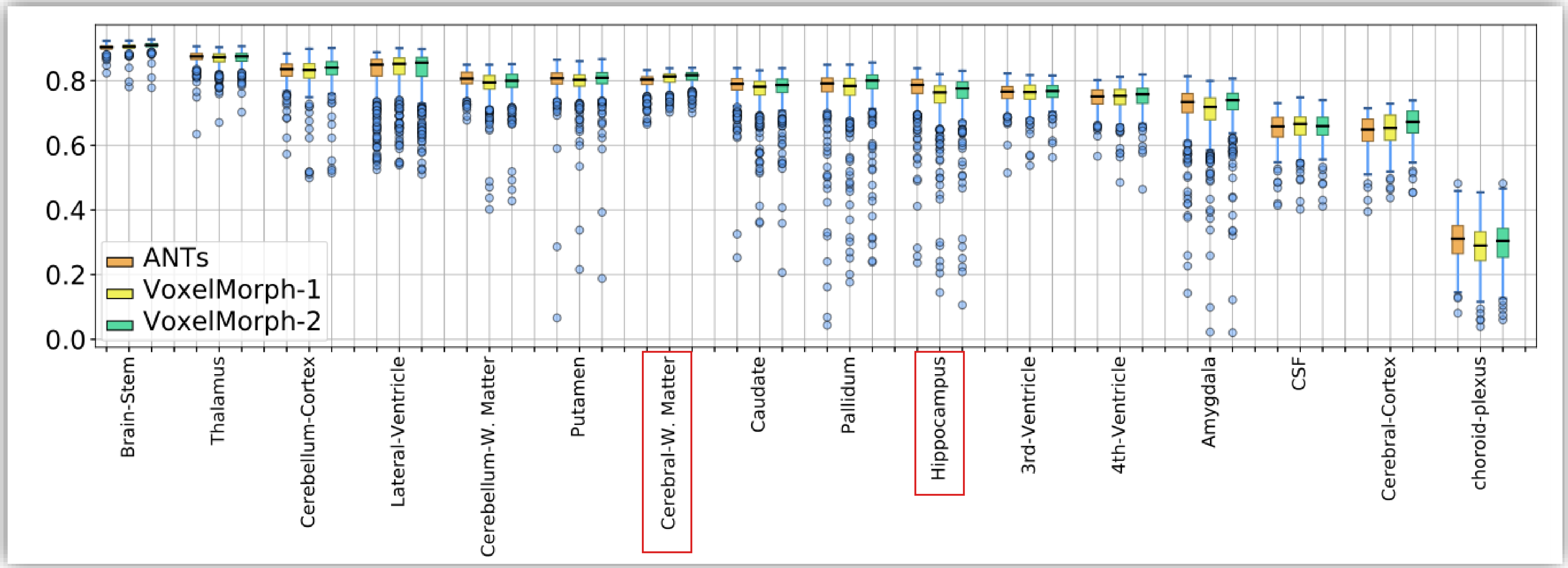


Figure 4: Example MR coronal slices extracted from input pairs (columns 1-2), and resulting $M(\phi)$ for VoxelMorph-1 and VoxelMorph-2, with overlaid boundaries of the ventricles (yellow, orange) and hippocampi (red, green). A good registration will cause structures in $M(\phi)$ to look similar to structures in F. Our networks handle large changes in shapes, such as the ventricles in row 2 and the left hippocampi in rows 3-4.

2）结果分析

采用不同配准方法，脑部不同结构配准效果的DICE值分布可视化



VoxelMorph模型在所有结构上都达到了与ANTs（实现了对称归一化(SyN)方法）相当的DICE测量值，在某些结构(如大脑白质)上的表现略好于ANTs，而在其他结构(如海马)上的表现则稍差。

中国科学院大学 University of Chinese Academy of Sciences

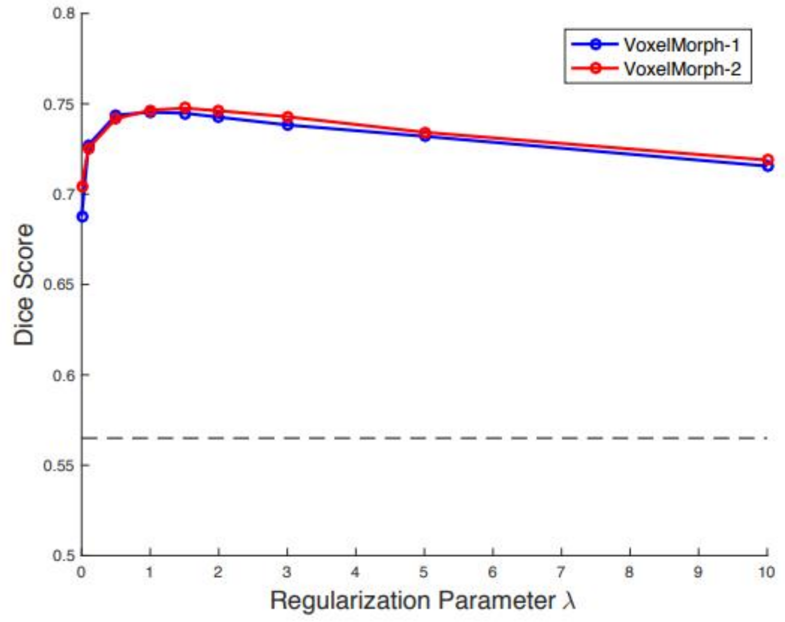
3）特定群体测试和超参敏感性分析

实验2：ABIDE数据集是有关自闭症群体的一个数据集，针对特定的ABIDE数据集测试**特定群体**方法的精度，可以看出DICE值提高了1.5%。

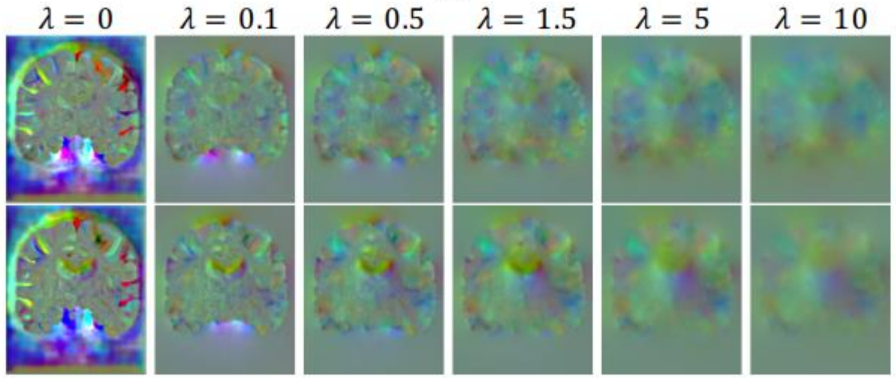
Table 2: Average Dice scores on ABIDE scans, when trained on all datasets (column 2) and ABIDE scans only (column 3). We achieve roughly 1.5% better scores when training on ABIDE only.

Method	Avg. Dice (Train on All)	Avg. Dice (Train on ABIDE)
VoxelMorph-1	0.715(0.140)	0.729(0.142)
VoxelMorph-2	0.718(0.141)	0.734(0.140)

实验3：右图表示，当Voxelmorph1 λ = 1, Voxelmorph2 λ = 1.5时，DICE得分最优。随着λ的增大，配准区域在结构边界处变得更加平滑。



(a)



中国科学院大学
University of Chinese Academy of Sciences

04 总结展望

◎前后对比 ◎后期工作

1）前后对比

- 不基于学习是指传统的配准方法，每次配准都要对度量函数进行优化，**参数不共享**；
- 基于学习就是指经过神经网络训练的(训练就是学习)，训练出来的参数函数是共享的（**只需训练一次得到参数，以后配准都使用这些参数**）。

- 由于传统的配准方法是**对每一个图像对进行优化**，所以**速度非常慢**。有监督的神经网络训练的方式虽然提升了速度，但**需要大量的标注信息**；
- 无监督的神经网络训练具有**速度快**、**不需要标注信息**的特点。

中国科学院大学
University of Chinese Academy of Sciences

2）后期工作

VoxelMorph Papers

If you use voxelmorph or some part of the code, please cite (see bibtex):

- For the atlas formation model:

  Learning Conditional Deformable Templates with Convolutional Networks
  Adrian V. Dalca, Marianne Rakic, John Guttag, Mert R. Sabuncu
  NeurIPS 2019. eprint arXiv:1908.02738

- For the diffeomorphic or probabilistic model:

  Unsupervised Learning of Probabilistic Diffeomorphic Registration for Images and Surfaces
  Adrian V. Dalca, Guha Balakrishnan, John Guttag, Mert R. Sabuncu
  MedIA: Medial Image Analysis. 2019. eprint arXiv:1903.03545

  Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration
  Adrian V. Dalca, Guha Balakrishnan, John Guttag, Mert R. Sabuncu
  MICCAI 2018. eprint arXiv:1805.04605

- For the original CNN model, MSE, CC, or segmentation-based losses:

  VoxelMorph: A Learning Framework for Deformable Medical Image Registration
  Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, Adrian V. Dalca
  IEEE TMI: Transactions on Medical Imaging. 2019. eprint arXiv:1809.05231

  An Unsupervised Learning Model for Deformable Medical Image Registration
  Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, Adrian V. Dalca
  CVPR 2018. eprint arXiv:1802.02604

https://github.com/voxelmorph/voxelmorph

2）后期工作

- 在新的工作中，作者提出了一个**概率模型**和**有效的学习策略**，可以**产生通用模板或条件模板**，并结合一个神经网络提供**有效的对齐图像到这些模板。**
- 这对于，不存在现有模板，或者使用传统方法创建一个新的模板可能非常昂贵，的临床应用，特别有用。

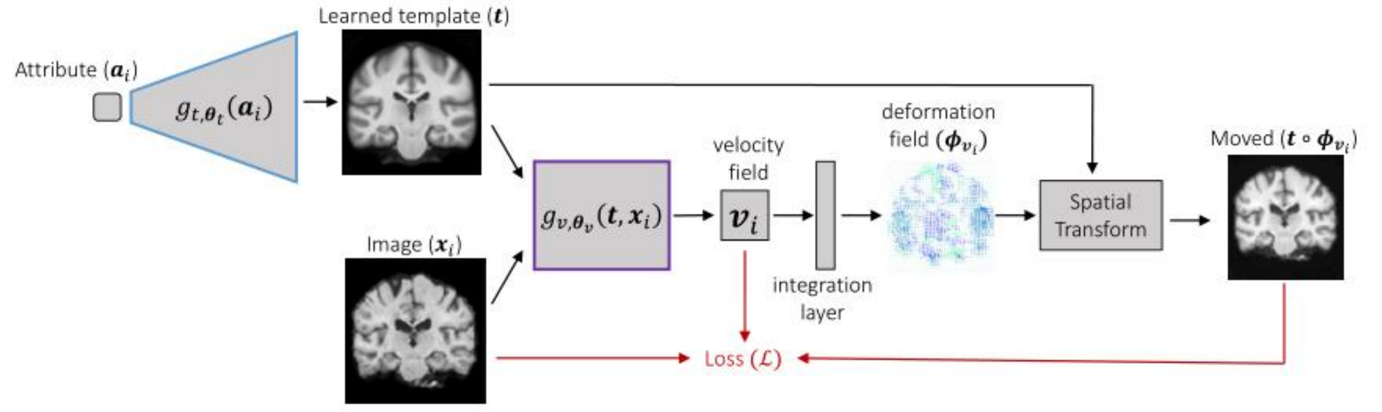


Figure 2: **Overview.** The network takes as input an image and an optional attribute vector. The upper network $g_{t,\theta_t}(\cdot)$ outputs a template, which is then registered with the input image by the second network $g_{v,\theta_v}(\cdot)$. The loss function, derived from the negative log likelihood of the generative model, leverages the template warped into $\boldsymbol{t} \circ \phi_{v_i}$.

We optimize the neural network parameters $\boldsymbol{\theta}$ using stochastic gradient algorithms, and minimize the negative maximum likelihood (1) for image $\boldsymbol{x}_i$:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_v; \boldsymbol{v}_i, \boldsymbol{x}_i, \boldsymbol{a}_i) &= -\log p_\theta(\boldsymbol{v}_i, \boldsymbol{x}_i; \boldsymbol{a}_i) \\
&= -\log p_\theta(\boldsymbol{x}_i | \boldsymbol{v}_i; \boldsymbol{a}_i) - \log p_\theta(\boldsymbol{v}_i) \\
&= -\frac{1}{2\sigma^2}\|\boldsymbol{x}_i - g_{t,\theta_t}(\boldsymbol{a}_i) \circ \boldsymbol{\phi}_{v_i}\|^2 - \gamma\|\bar{\boldsymbol{u}}\|^2 - \lambda_d \frac{d}{2}\sum_i \|\boldsymbol{u}_i\|^2 + \frac{\lambda_a}{2}\sum_i \|\nabla \boldsymbol{u}_i\|^2 + \text{const}, \quad (6)
\end{aligned}
$$

where $g_{t,\theta_t}(\boldsymbol{a}_i)$ yields the template at iteration i, and $\boldsymbol{v}_i = g_{v,\theta_v}(\boldsymbol{t}_{\theta_t,i}, \boldsymbol{x}_i)$.

2）后期工作

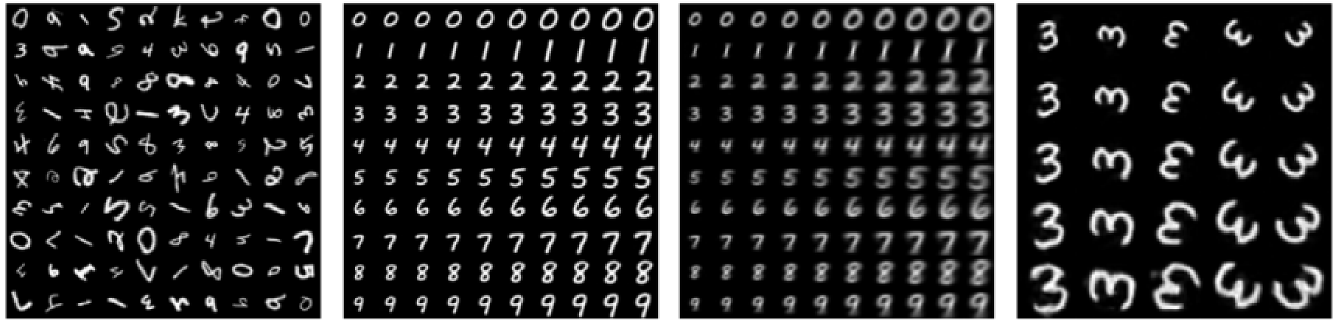


Figure 3: **MNIST examples** (1) MNIST digits from `D-scale-rot`; (2) templates conditioned on class (vertical axis) and scale (horizontal axis) on MNIST `D-scale`, learned with our model, and (3) with a decoder-only baseline model; (4) conditional templates learned with our model on the MNIST `D-class-scale-rot` dataset for the digit 3 and a variety of scaling and rotation values.

Experiment 1
on Benchmark Datasets：
MNIST & QuickDraw

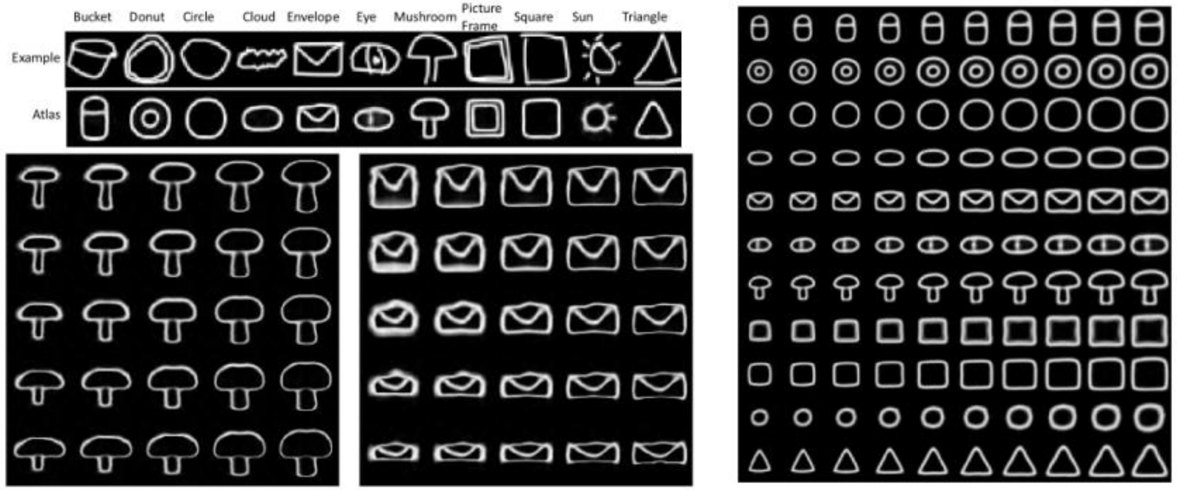


Figure 14: **Quickdraw example templates.** Left: example and learned atlases for the `D-class` QuickDraw dataset, and below variability examples similar to Figure 7-left. Right: templates for different scales and classes learned using D-class-scale simulations.

2）后期工作

Experiment 2: Neuroimaging

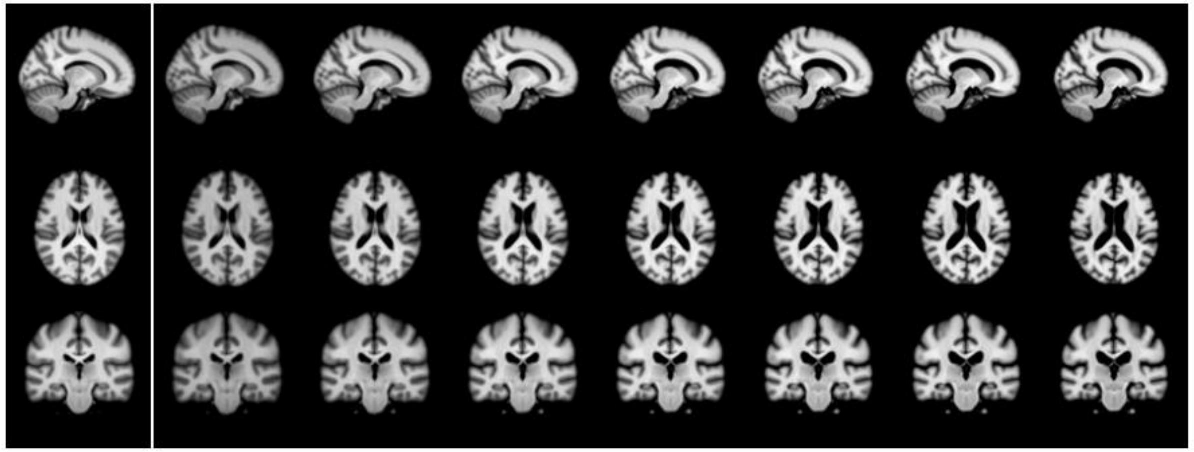


Figure 10: **Slices from Learned 3D Brain MRI templates.** Left: single unconditional template representing the entire population. Right: conditional age templates for brain MRI for ages 15 to 90, illustrating, for example, growth of the ventricles, also evident in a supplementary video.

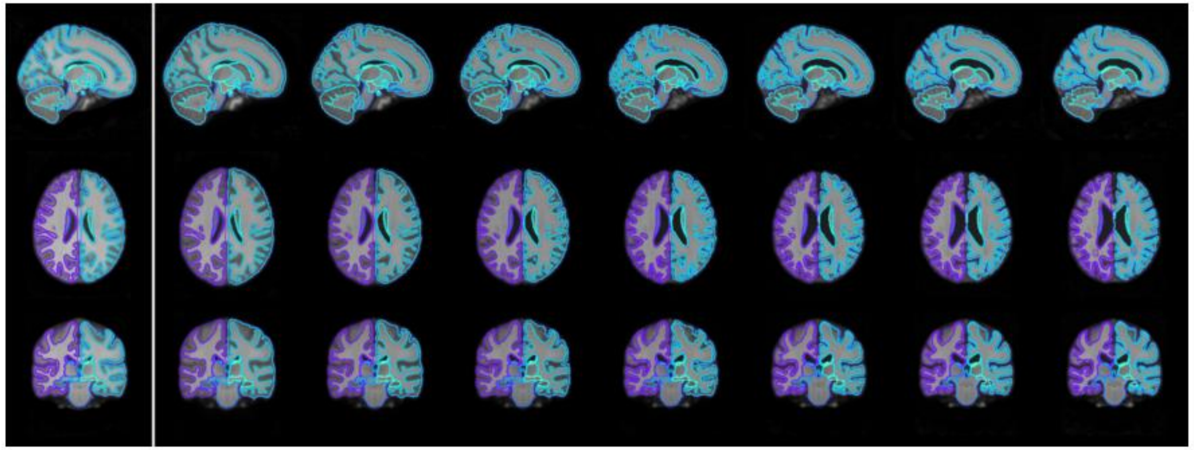


Figure 11: **Segmentations.** Example segmentations overlayed with different brain views for our unconditional template (left) and conditional templates (right) varying by age.

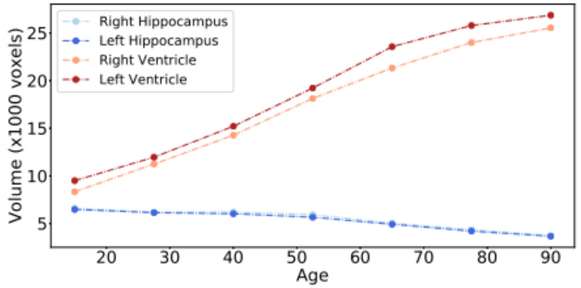


Figure 13: **Volume trends.** Change in volume of ventricles and hippocampi of the age-conditional brain templates.

参考文献

- [1] Balakrishnan G , Zhao A , Sabuncu M R , et al. An Unsupervised Learning Model for Deformable Medical Image Registration[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018.
- [2]iek, zgün, Abdulkadir A , Lienkamp S S , et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation[J]. Springer, Cham, 2016.
- [3] Dalca A V , Rakic M , Guttag J , et al. Learning Conditional Deformable Templates with Convolutional Networks[J]. 2019.

中国科学院大学
University of Chinese Academy of Sciences


中国科学院大学
University of Chinese Academy of Sciences
感谢！