

# Medical Image Generation and Analysis using Bayesian Generative Models

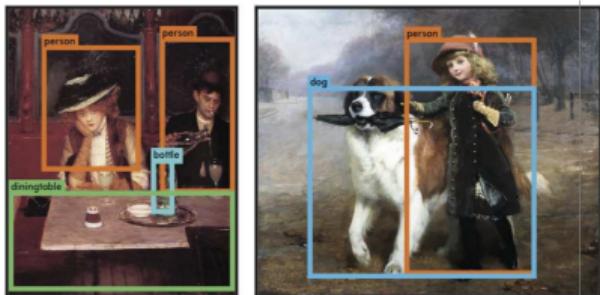
Răzvan V. Marinescu

Massachusetts Institute of Technology



# Machine Learning algorithms have achieved impressive milestones

Object detection (YOLO)



Text-to-Image Generation (DALL-E)



prompt: "an armchair in the shape of an avocado"

Image Generation (StyleGAN2)



Text generation (GPT-3)

Title: United Methodists Agree to Historic Split  
Subtitle: Those who oppose gay marriage will form their own denomination  
Article: After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.

## Diagnose with unprecedented accuracy



## Augment doctors



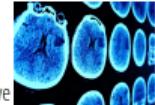
### Top 12 Ways Artificial Intelligence Will Impact Healthcare

Artificial intelligence is poised to become a transformational force in healthcare. How will providers and patients benefit from the impact of AI-driven tools?



### How Artificial Intelligence Improves Medical Imaging in Hospitals

Deep learning software, such as artificial intelligence, can improve



However, for many medical applications, these promises have not been fulfilled

## Prediction of clinical variables not always working

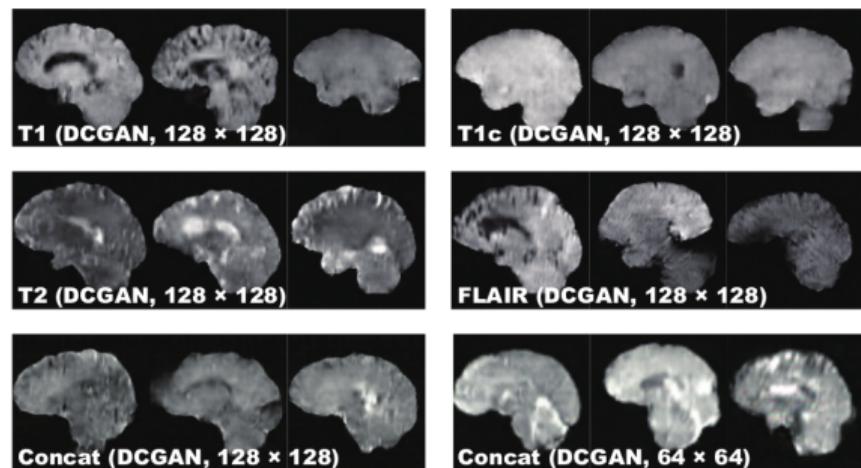
No algorithm/33 could predict cognitive scores in Alzheimer's (TADPOLE Challenge, Marinescu 2020)



- Forecasts were very good for clinical diagnosis and ventricle volume -- on the other hand, predicting ADAS turned out to be very difficult -- no team was able to regenerate forecasts that were significantly better than random guessing

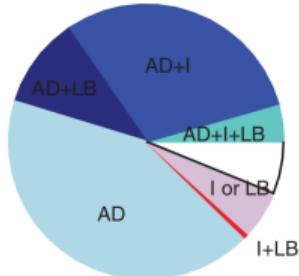
## Generated images are crude, not high-resolution, mostly 2D

Brain MRI generation (Han, 2018)



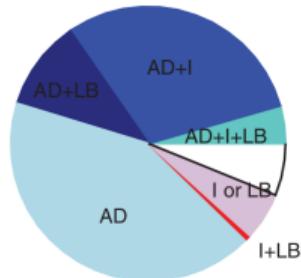
## Lack of good labels

- Alzheimer's diagnosis accuracy just 42%

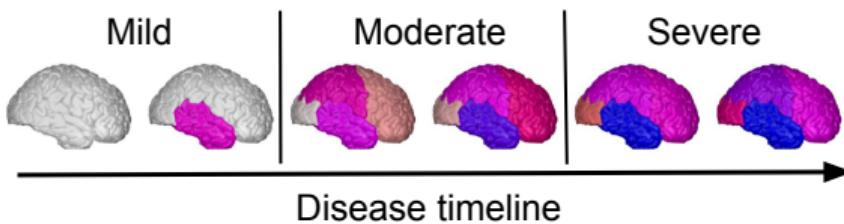


## Lack of good labels

- ▶ Alzheimer's diagnosis accuracy just 42%



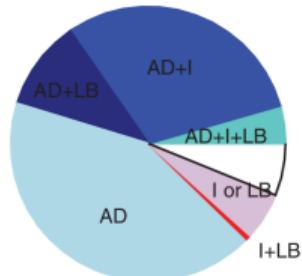
- ▶ Labels are categorical instead of continuous



# Why are Machine Learning models not working on medical applications?

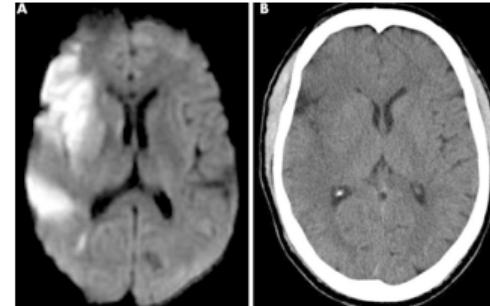
## Lack of good labels

- Alzheimer's diagnosis accuracy just 42%

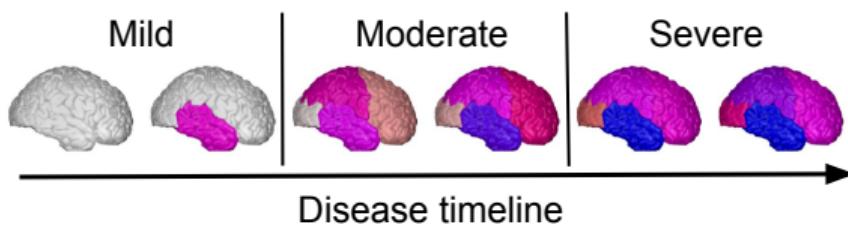


## Lack of good input data/signal

- Limited contrast



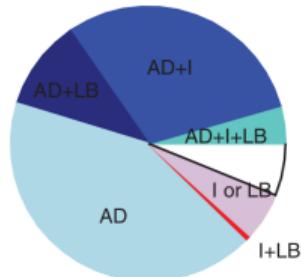
- Labels are categorical instead of continuous



# Why are Machine Learning models not working on medical applications?

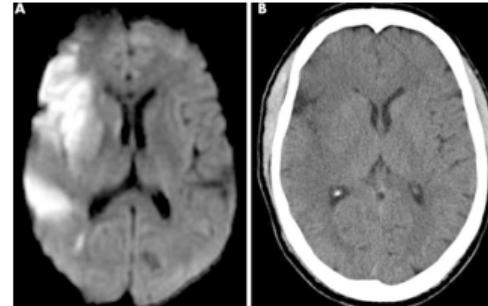
## Lack of good labels

- Alzheimer's diagnosis accuracy just 42%



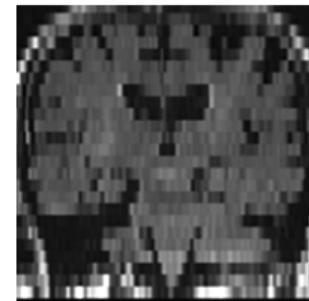
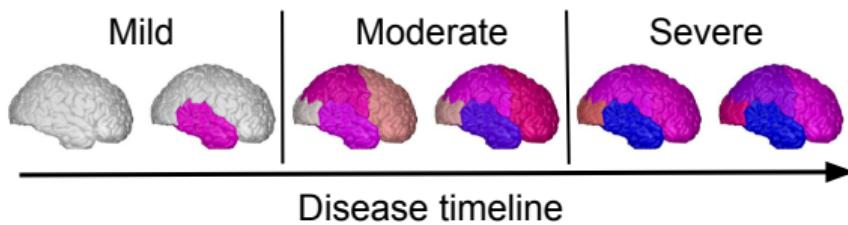
## Lack of good input data/signal

- Limited contrast



- Low-resolution

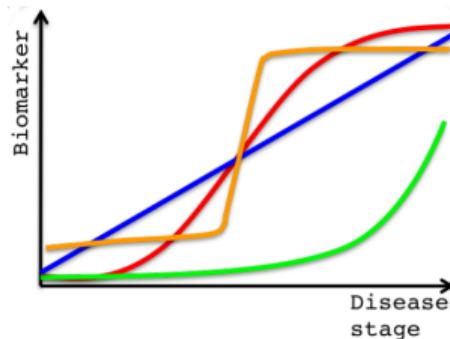
- Labels are categorical instead of continuous



What can we do?

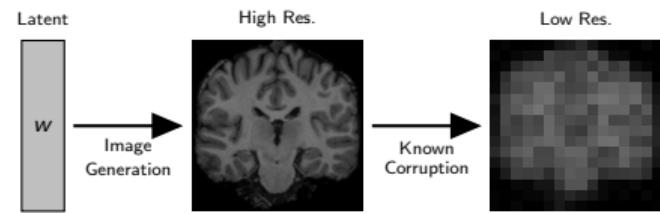
### Lack of good labels

Solution: Unsupervised Learning of Continuous Dynamics  
= Disease Progression Modelling



### Lack of good input data/signal

Solution: Image Reconstruction using Deep Generative Models

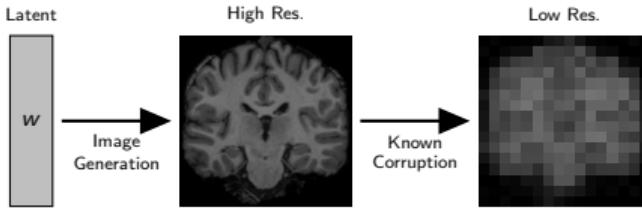


# Outline

1. Disease progression modelling of Alzheimer's disease
  - 1.1 Towards unsupervised clustering of biomarker trajectories



2. Image Reconstruction using Deep Generative Models



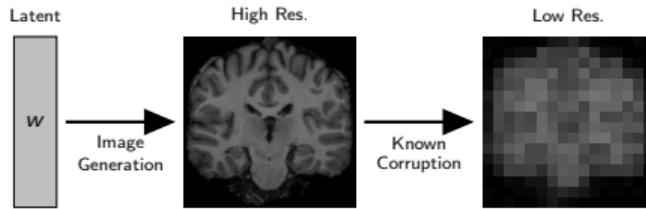
3. Future work towards brain anatomy simulators

## 1. Disease progression modelling of Alzheimer's disease

### 1.1 Towards unsupervised clustering of biomarker trajectories



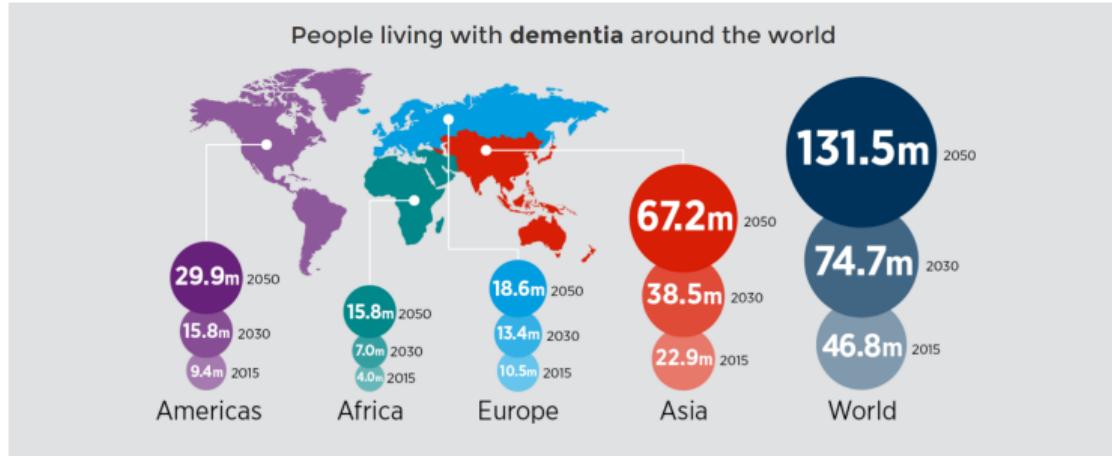
## 2. Image Reconstruction using Deep Generative Models



## 3. Future work towards brain anatomy simulators

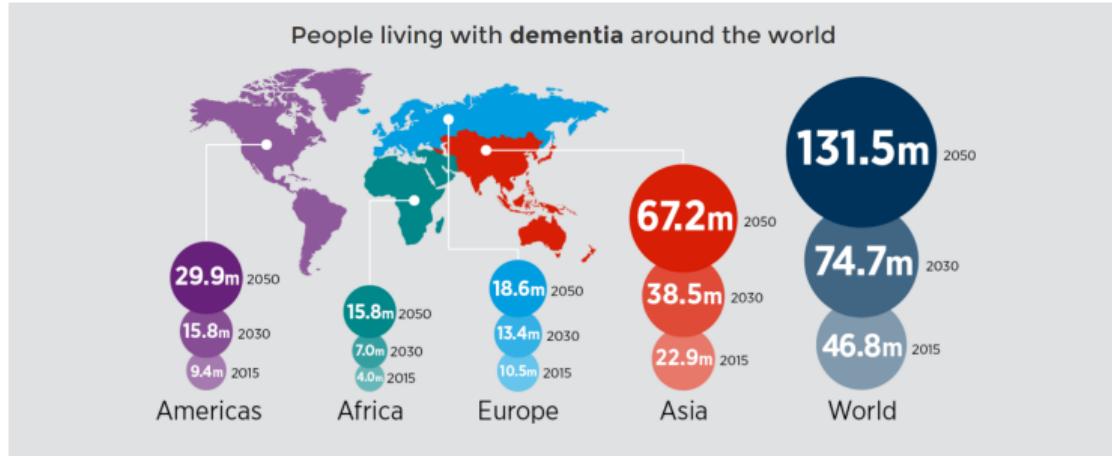
# Alzheimer's Disease is a Devastating Disease

- 46 million people affected worldwide



# Alzheimer's Disease is a Devastating Disease

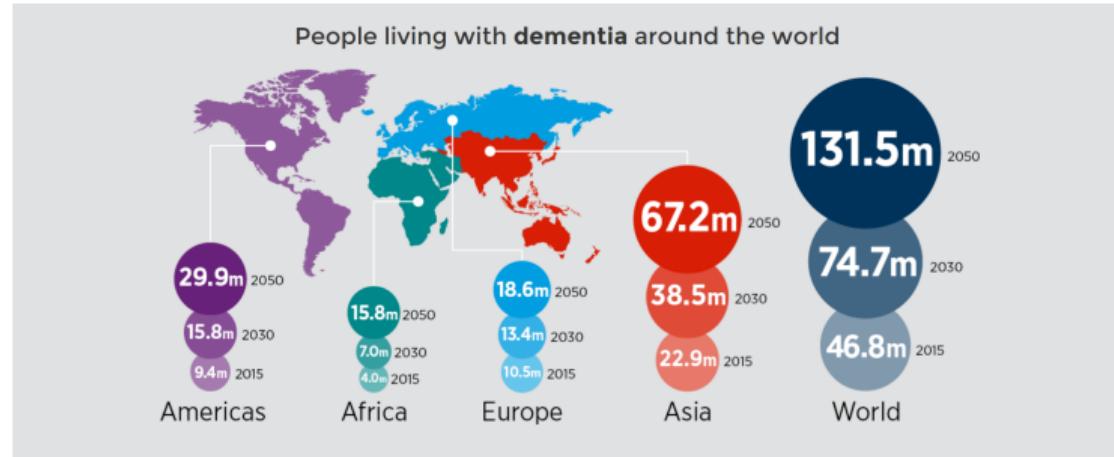
- ▶ 46 million people affected worldwide



- ▶ No treatments available that stop or slow down cognitive decline
- ▶ Q: Why did clinical trials fail? A: Treatments were not administered early enough

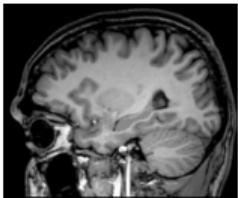
# Alzheimer's Disease is a Devastating Disease

- ▶ 46 million people affected worldwide

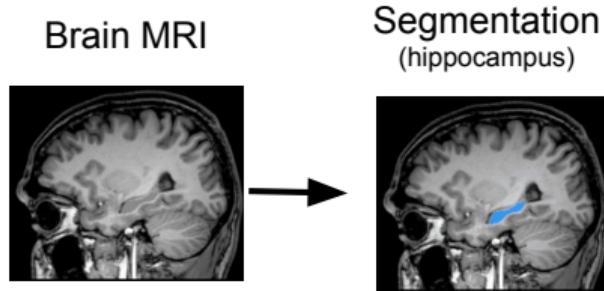


- ▶ No treatments available that stop or slow down cognitive decline
- ▶ Q: Why did clinical trials fail? A: Treatments were not administered early enough
- ▶ Q: How can we then identify subjects **early** in order to administer treatments?
- ▶ A: Disease progression model ...

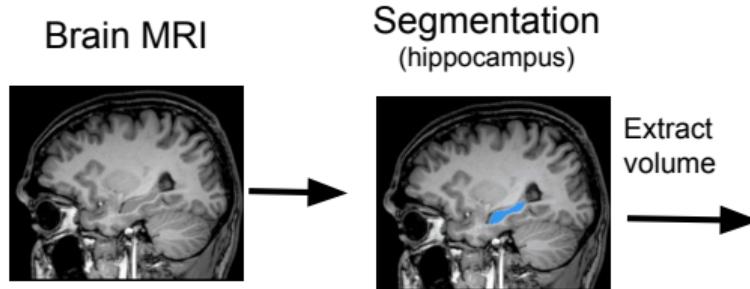
## Brain MRI



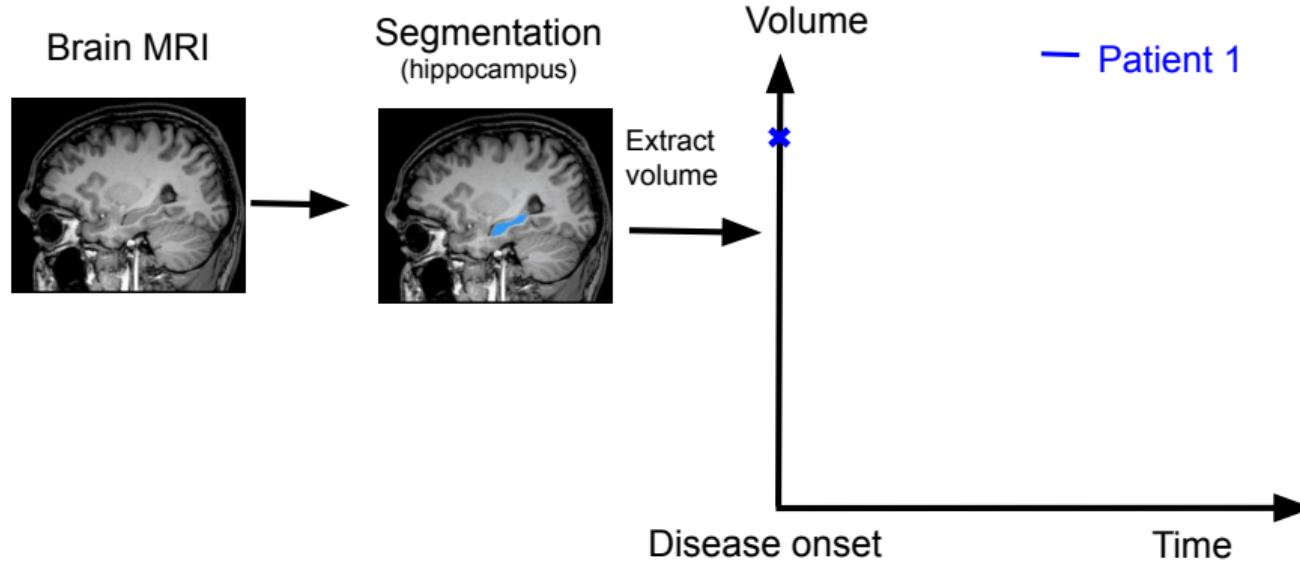
# Building a Disease Progression Model



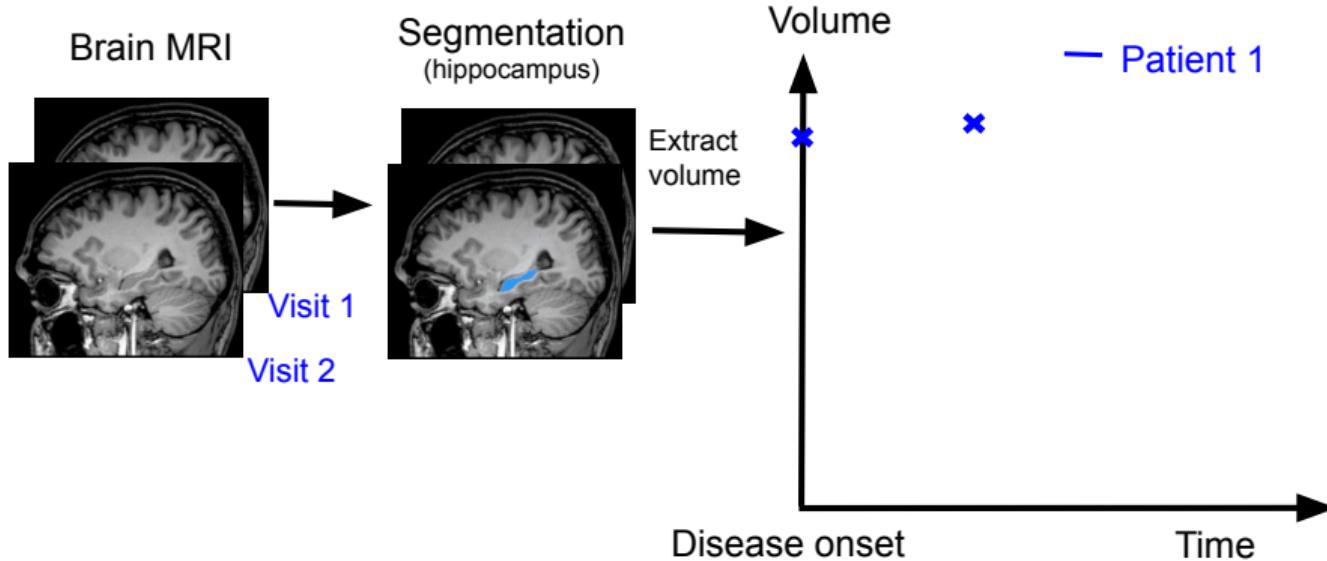
# Building a Disease Progression Model



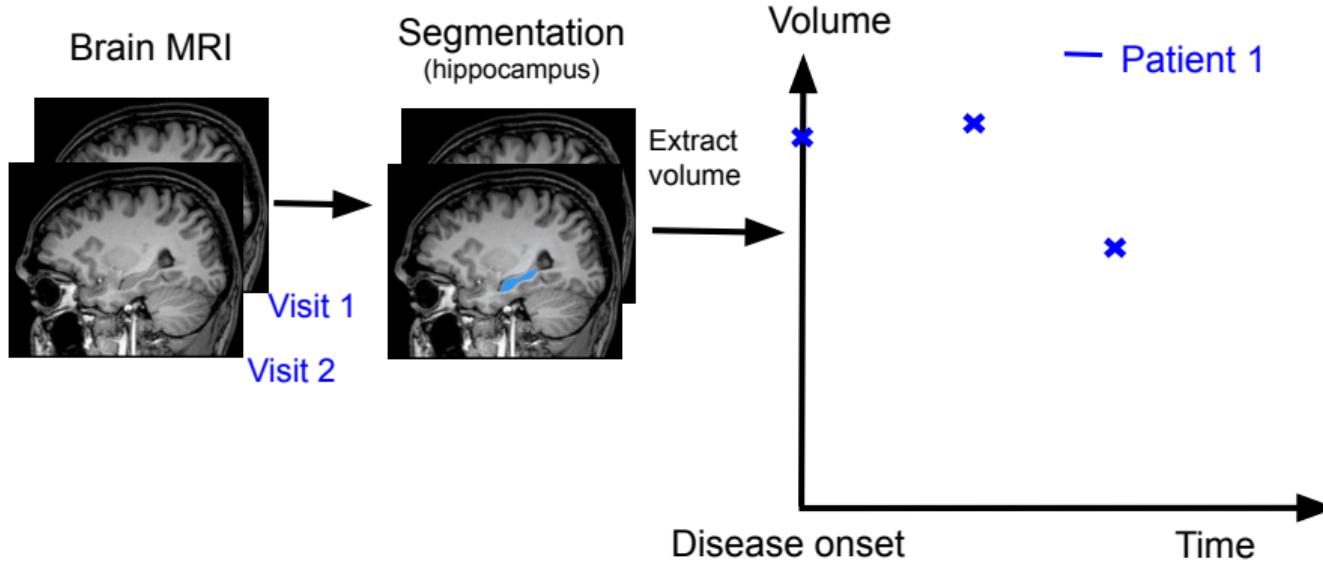
# Building a Disease Progression Model



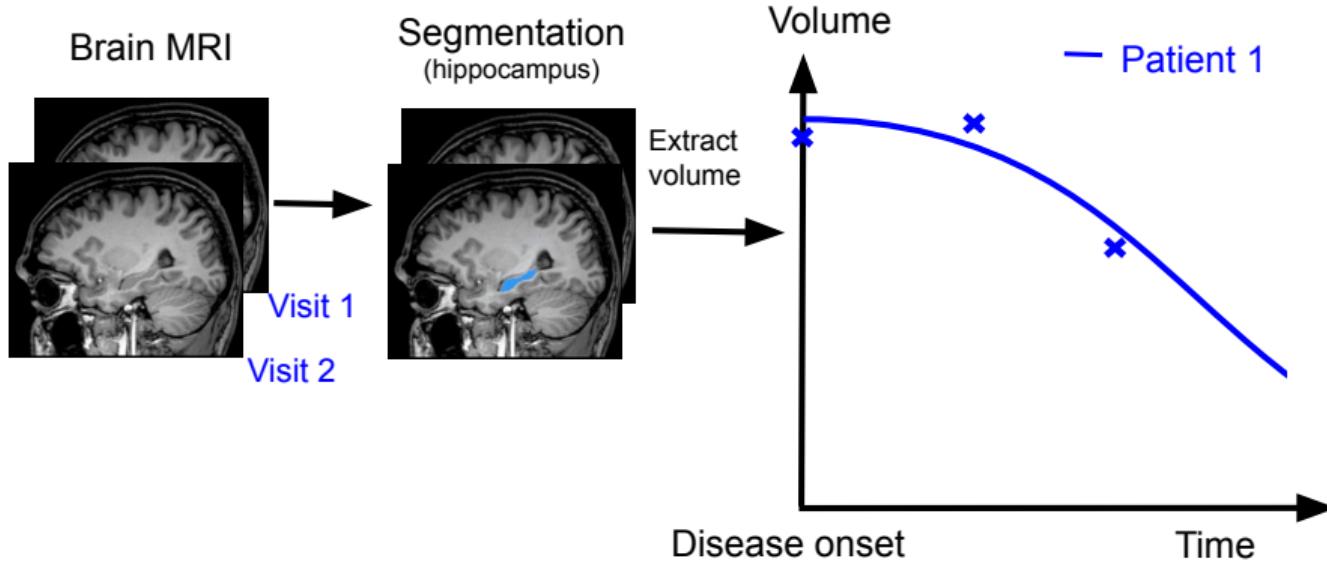
# Building a Disease Progression Model



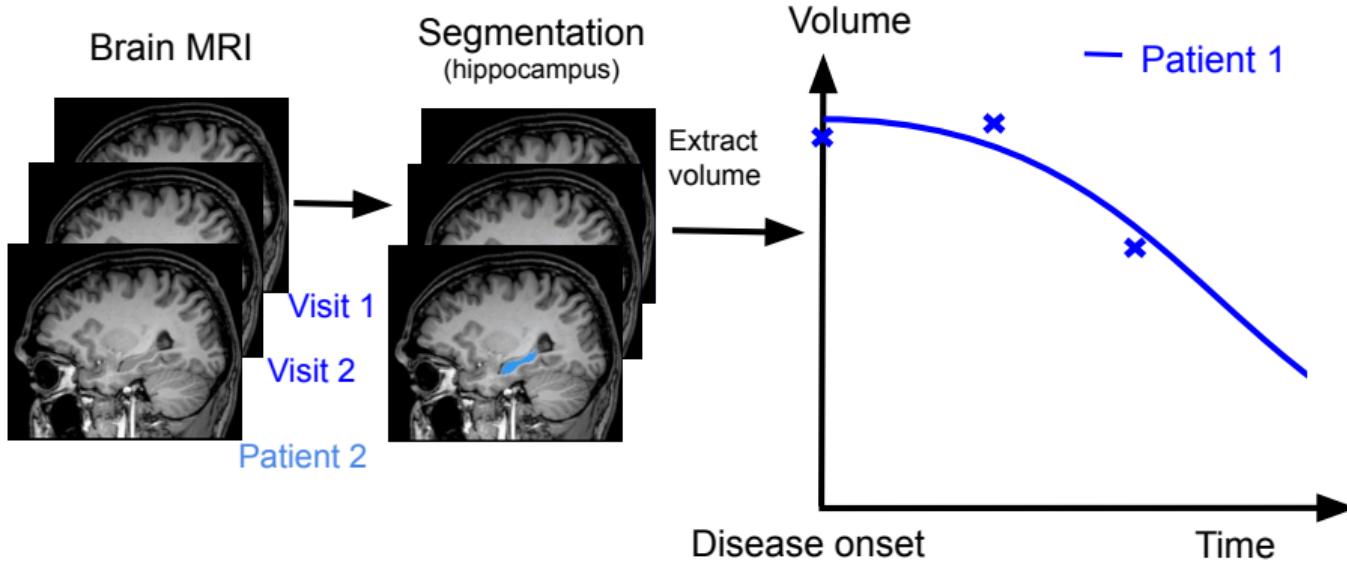
# Building a Disease Progression Model



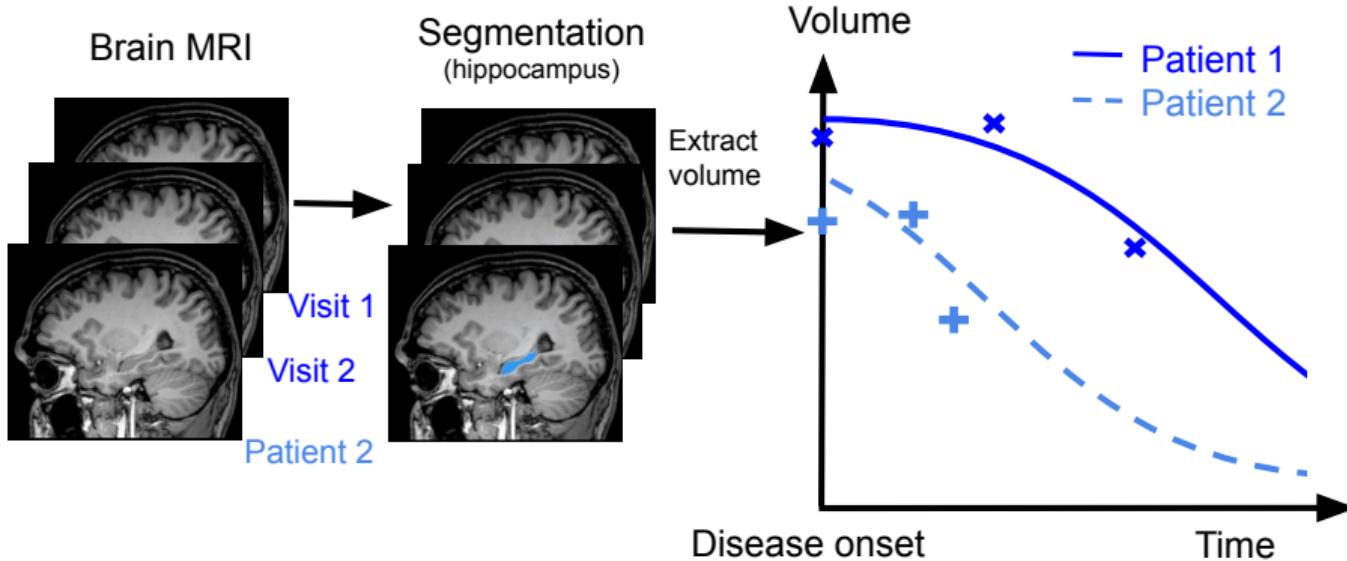
# Building a Disease Progression Model



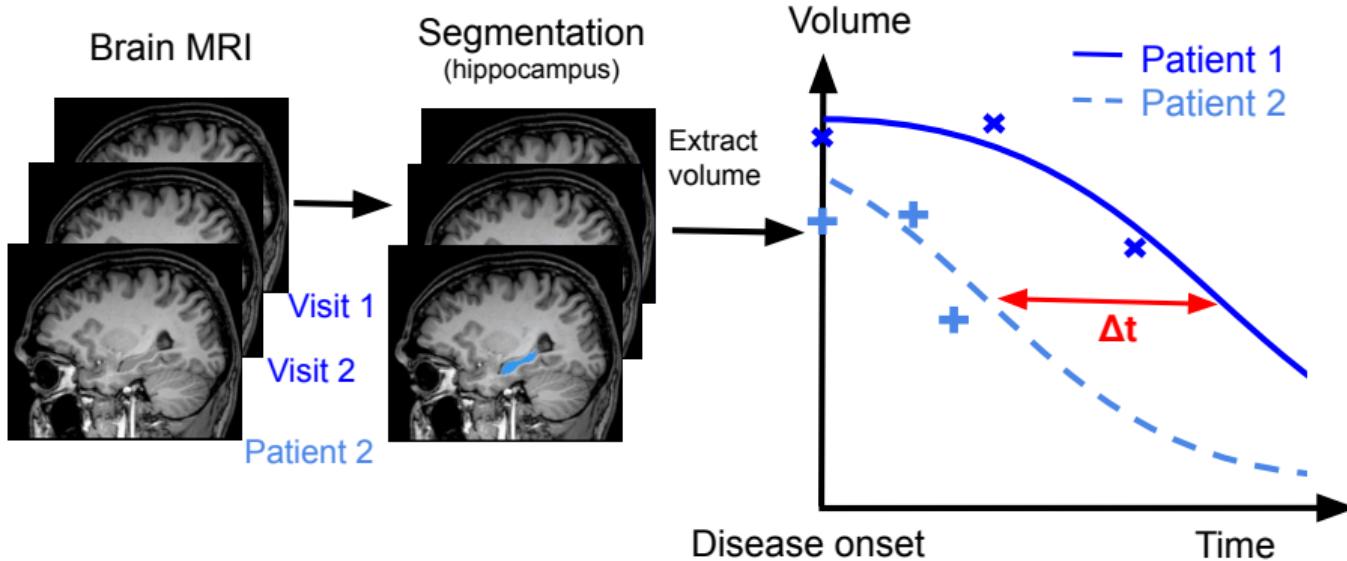
# Building a Disease Progression Model



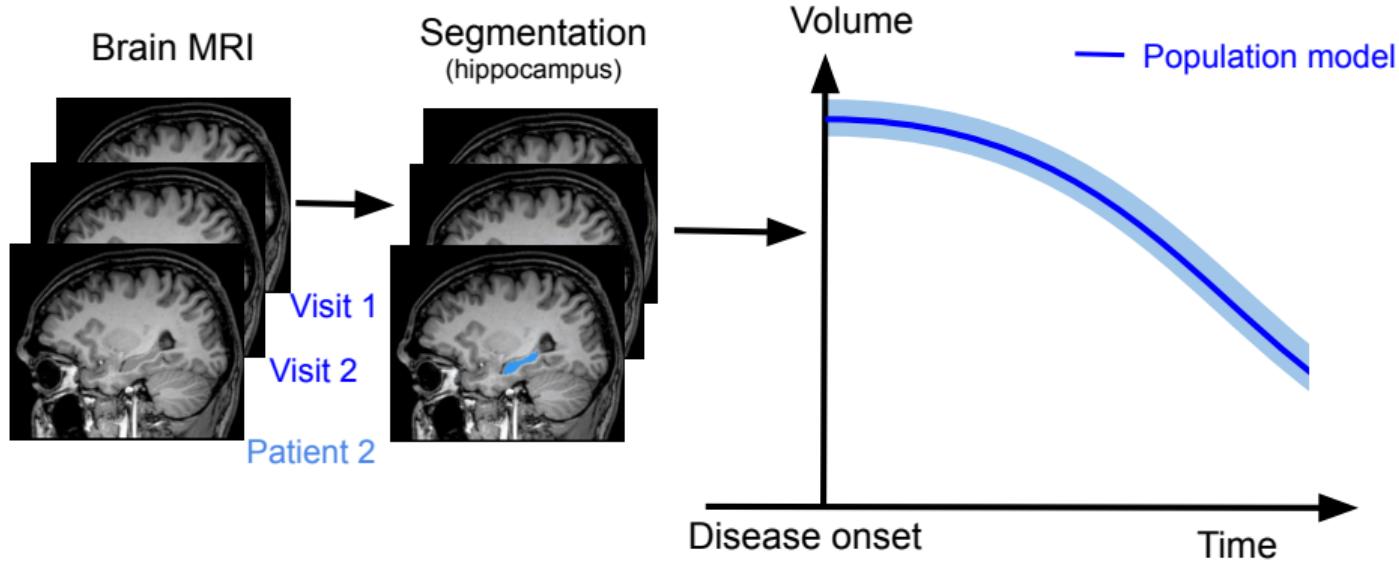
# Building a Disease Progression Model



# Building a Disease Progression Model

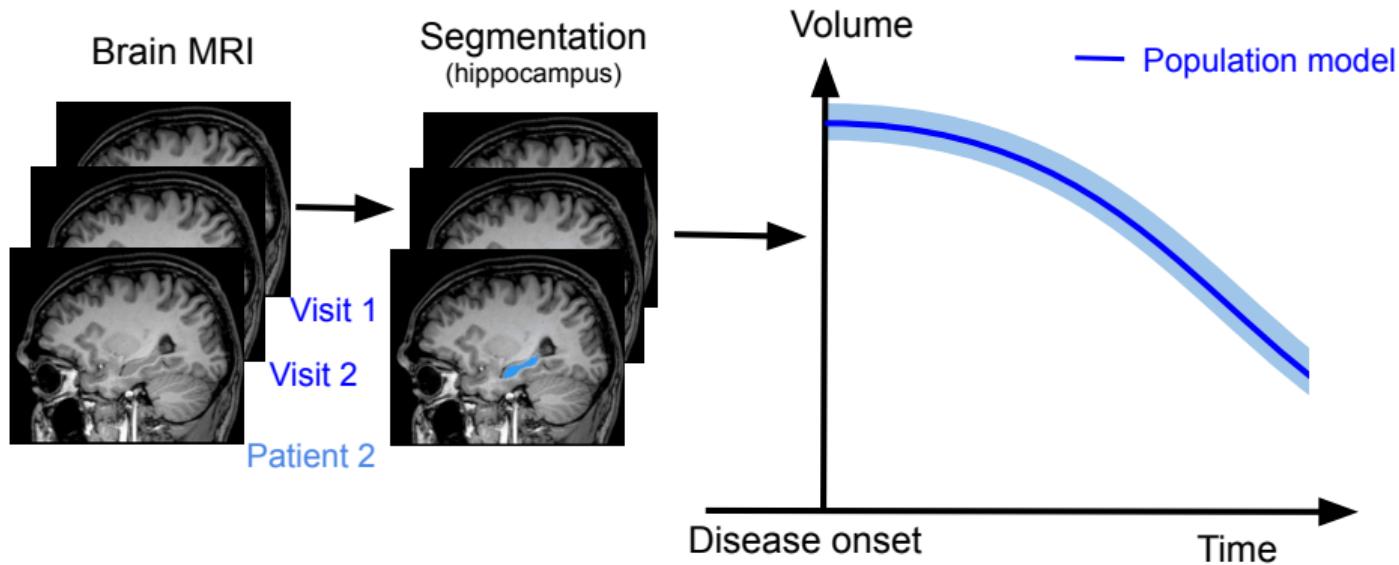


# Building a Disease Progression Model



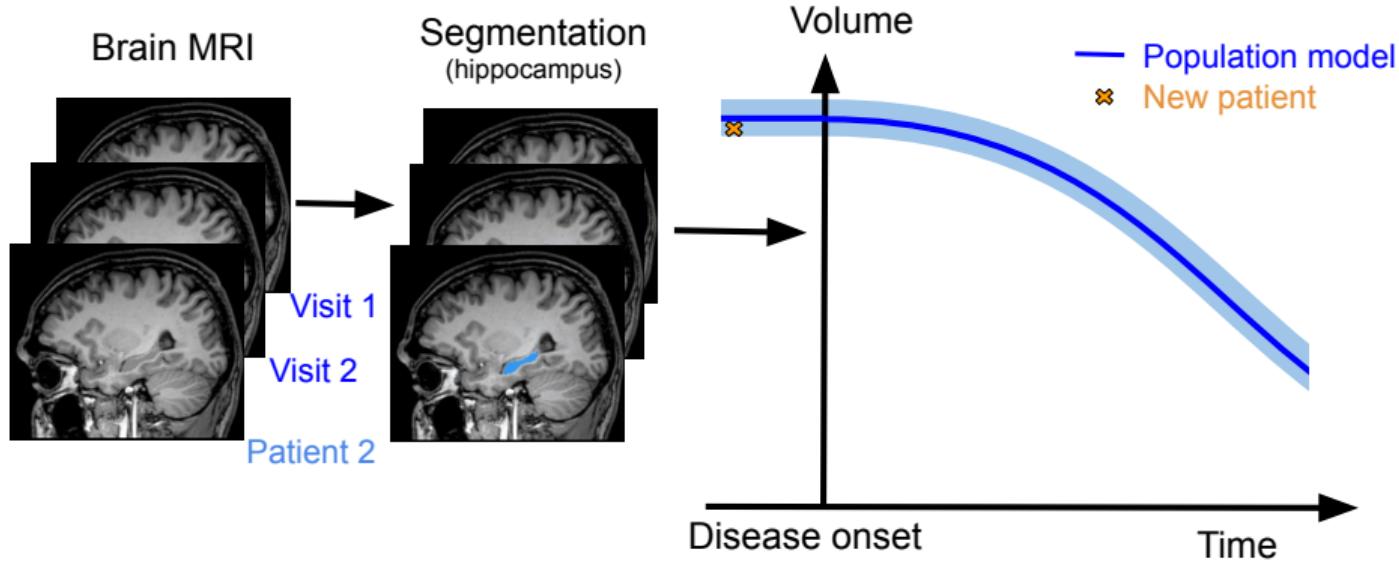
- Can now build population model

# Building a Disease Progression Model



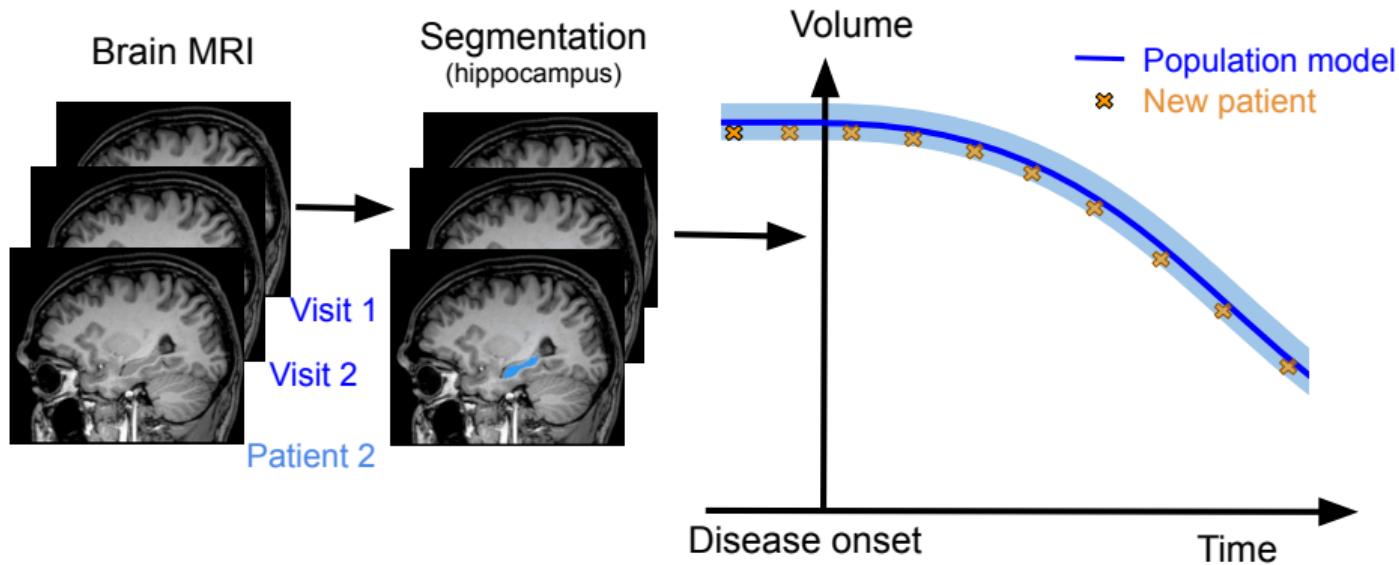
- ▶ Can now build population model
- ▶ Early diagnosis

# Building a Disease Progression Model



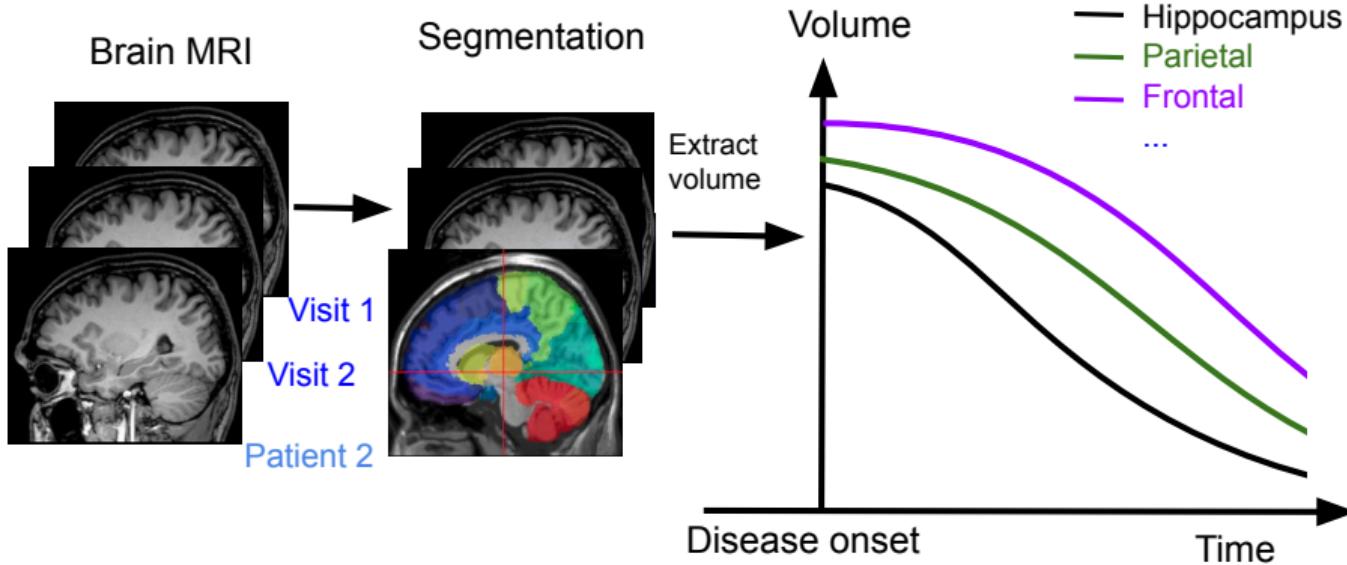
- ▶ Can now build population model
- ▶ Early diagnosis

# Building a Disease Progression Model



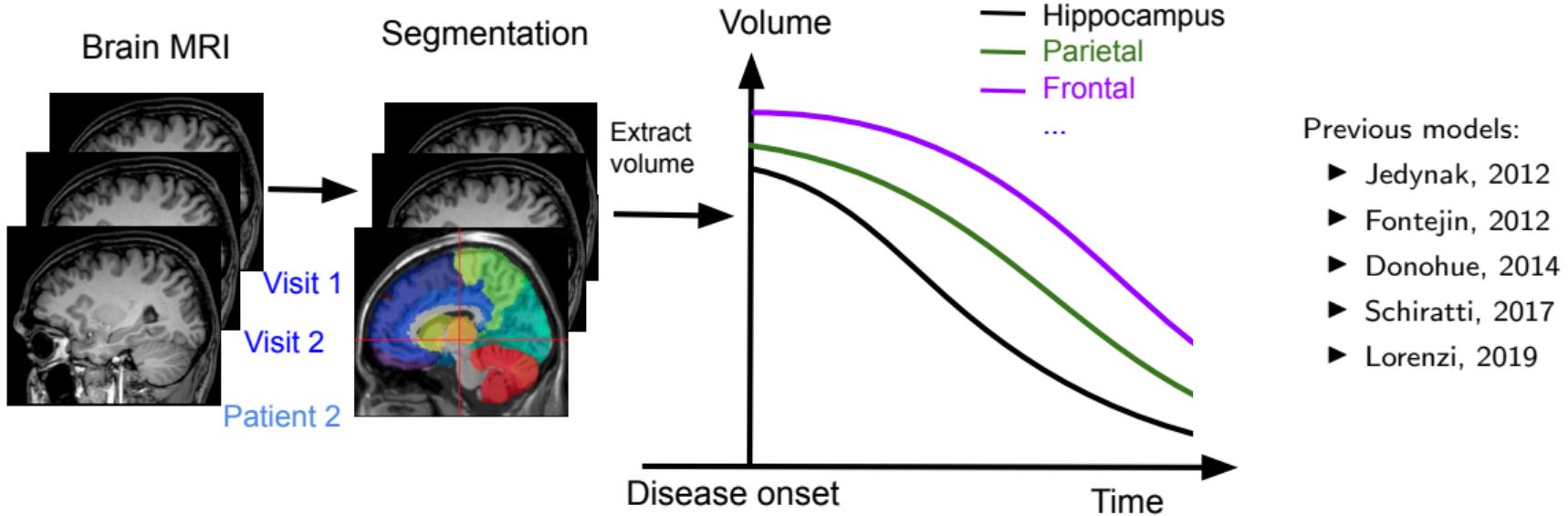
- ▶ Can now build population model
- ▶ Early diagnosis

# Building a Disease Progression Model



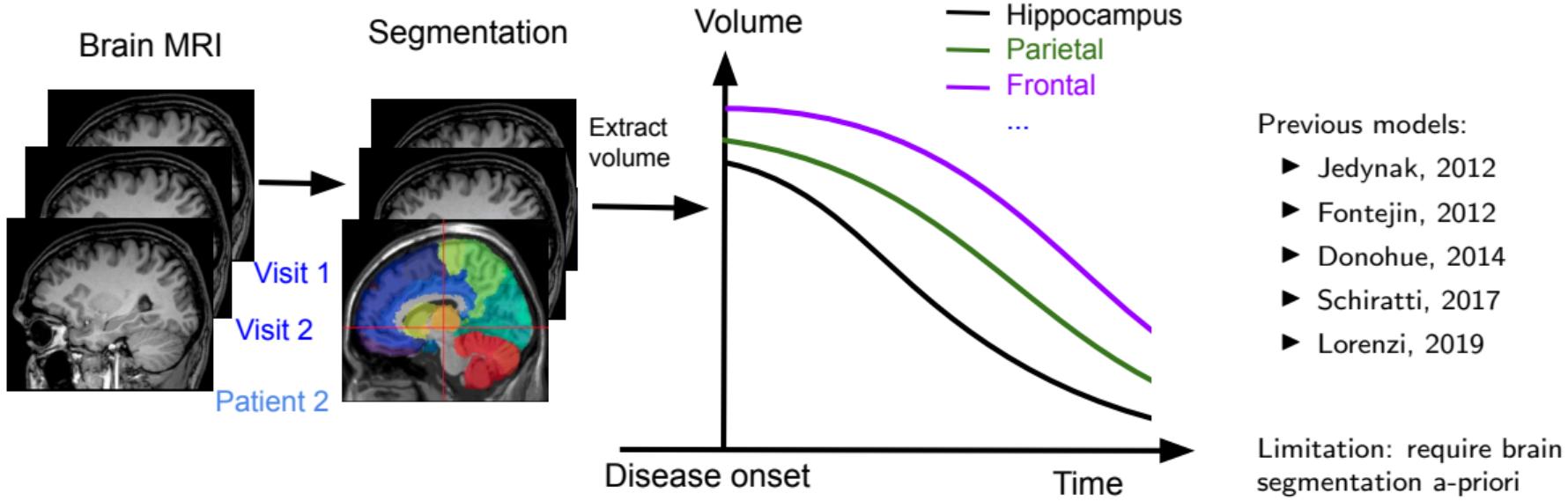
- ▶ Can now build population model
- ▶ Early diagnosis
- ▶ More accurate by analyzing all brain regions

# Building a Disease Progression Model



- ▶ Can now build population model
- ▶ Early diagnosis
- ▶ More accurate by analyzing all brain regions

# Building a Disease Progression Model

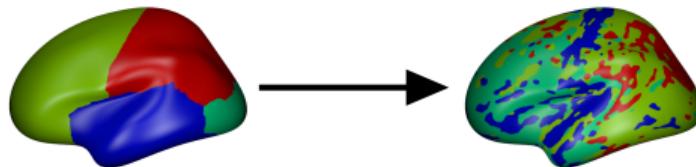


- ▶ Can now build population model
- ▶ Early diagnosis
- ▶ More accurate by analyzing all brain regions

## Aim: Build a disease progression model for vertexwise data

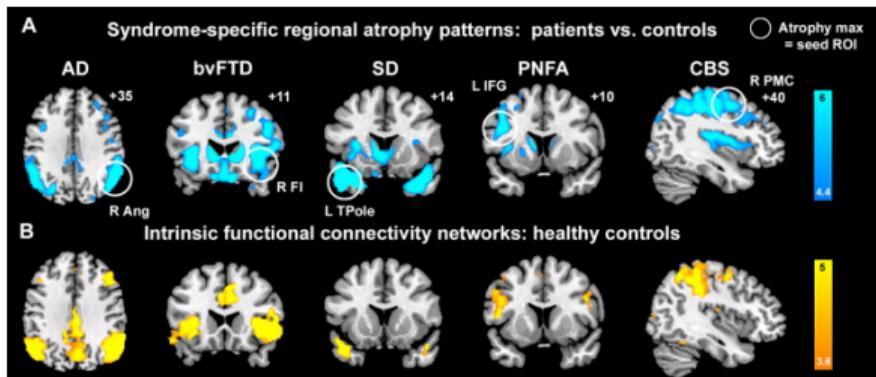
### Aim: Move from segmentation-based analysis to vertexwise

- vertex = point on the brain surface



### Why:

1. Atrophy correlates with functional networks, which are spatially disconnected (Seeley et al., 2009)
  - Atrophy = breakdown of neurons
  - Functional network = connections between neurons
2. Better prediction and disease staging



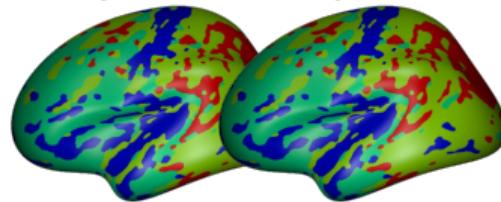
Seeley et al., Neuron, 2009

Our model clusters vertices with similar trajectories of pathology

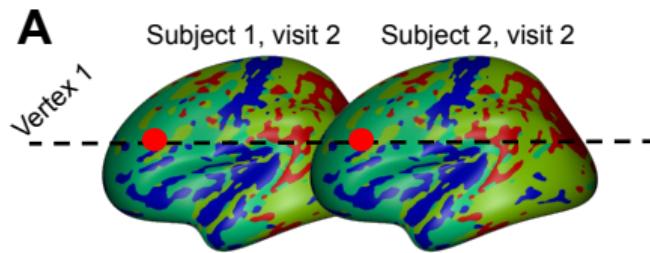
**A**

Subject 1, visit 2

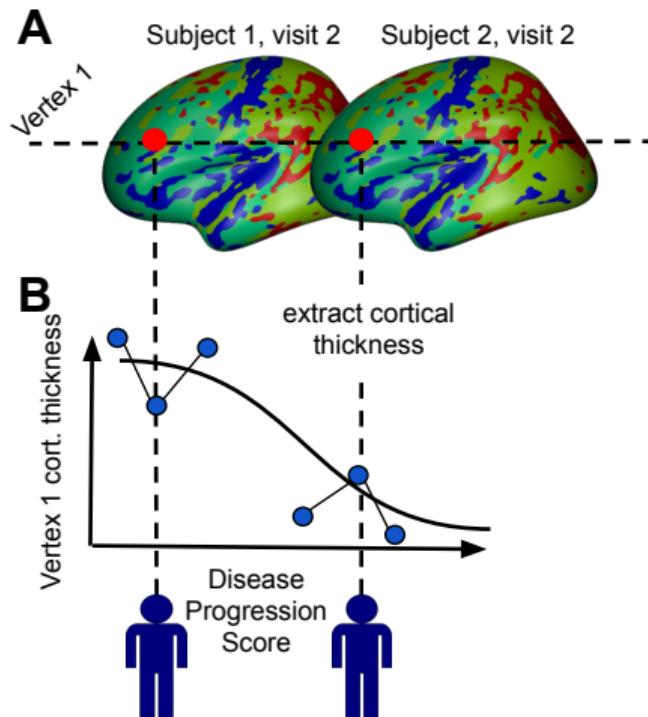
Subject 2, visit 2



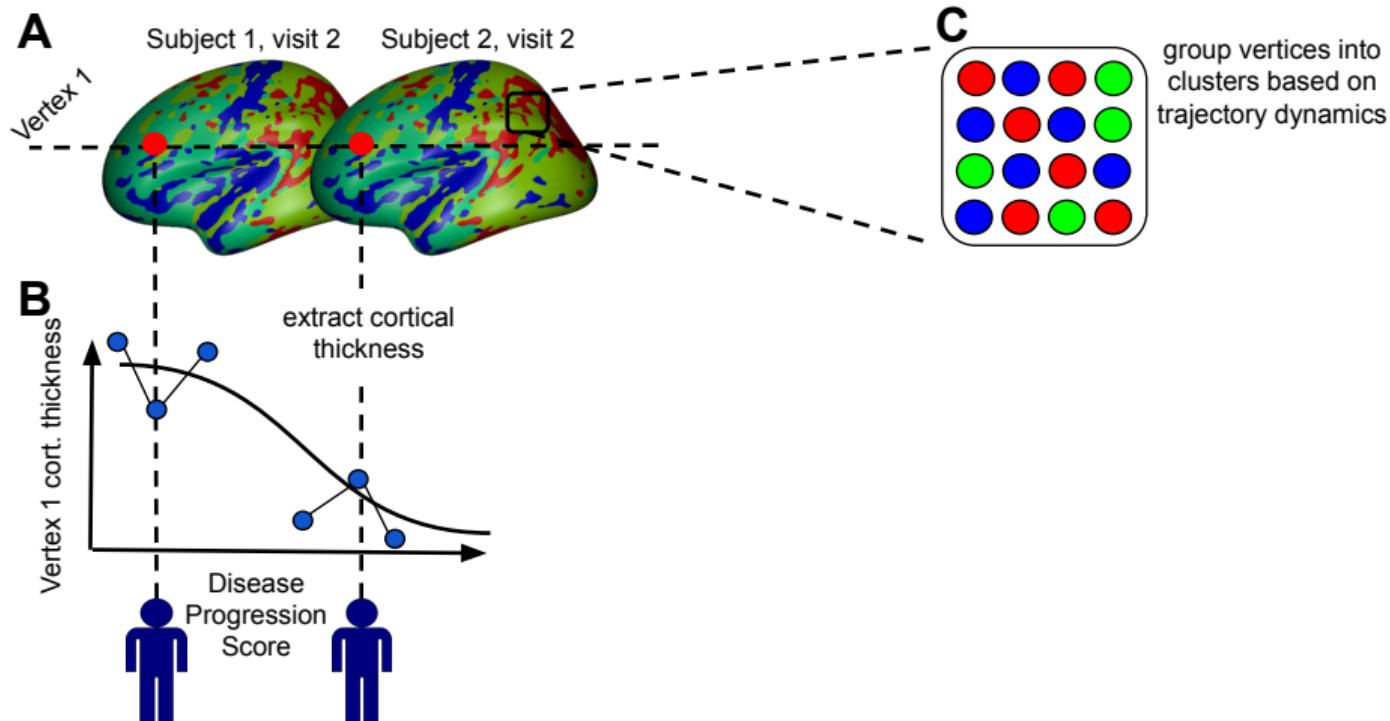
Our model clusters vertices with similar trajectories of pathology



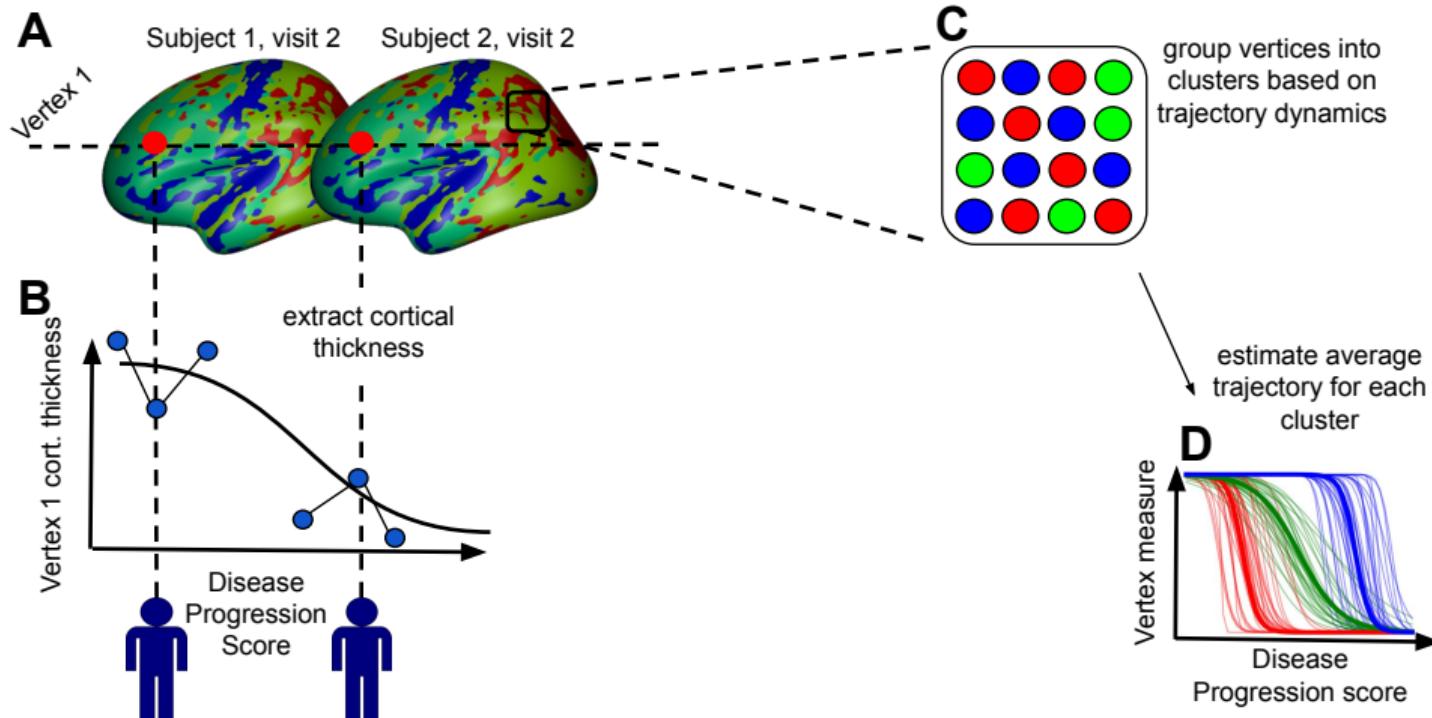
Our model clusters vertices with similar trajectories of pathology



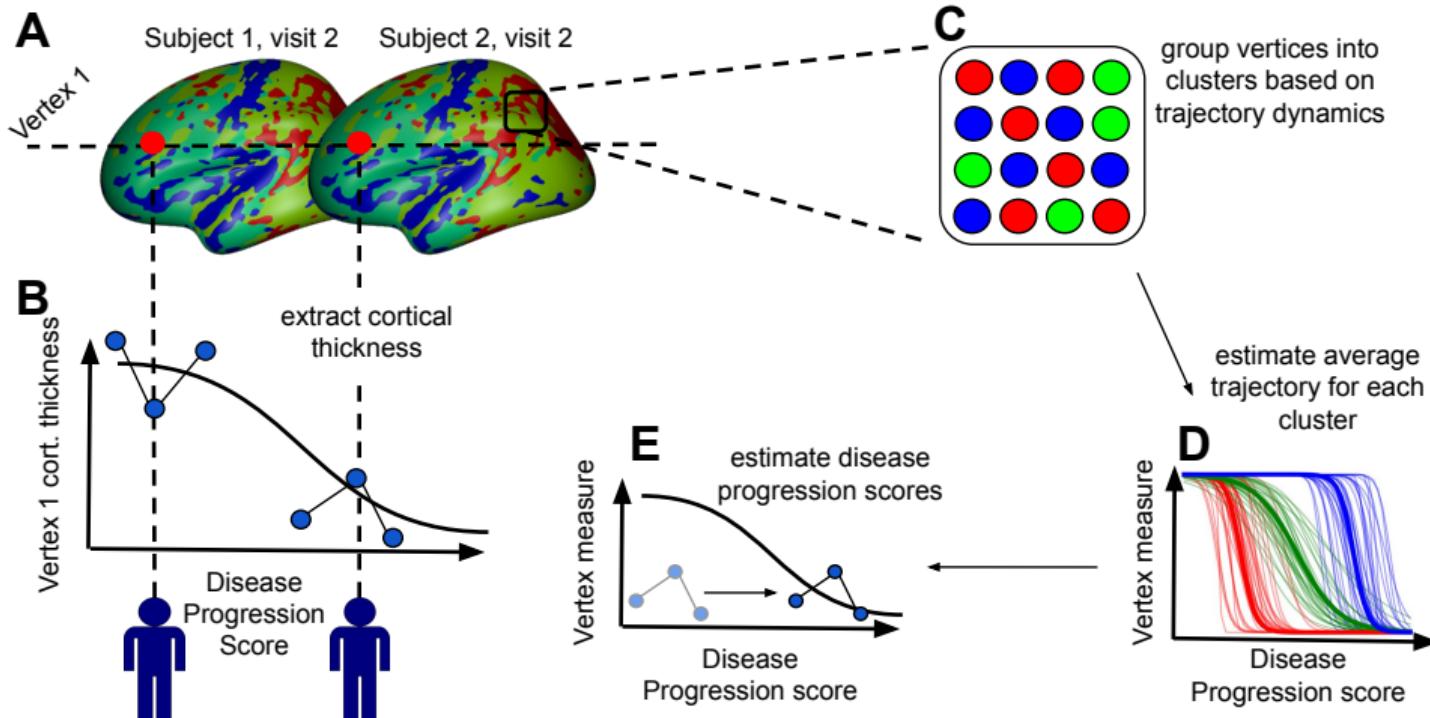
Our model clusters vertices with similar trajectories of pathology



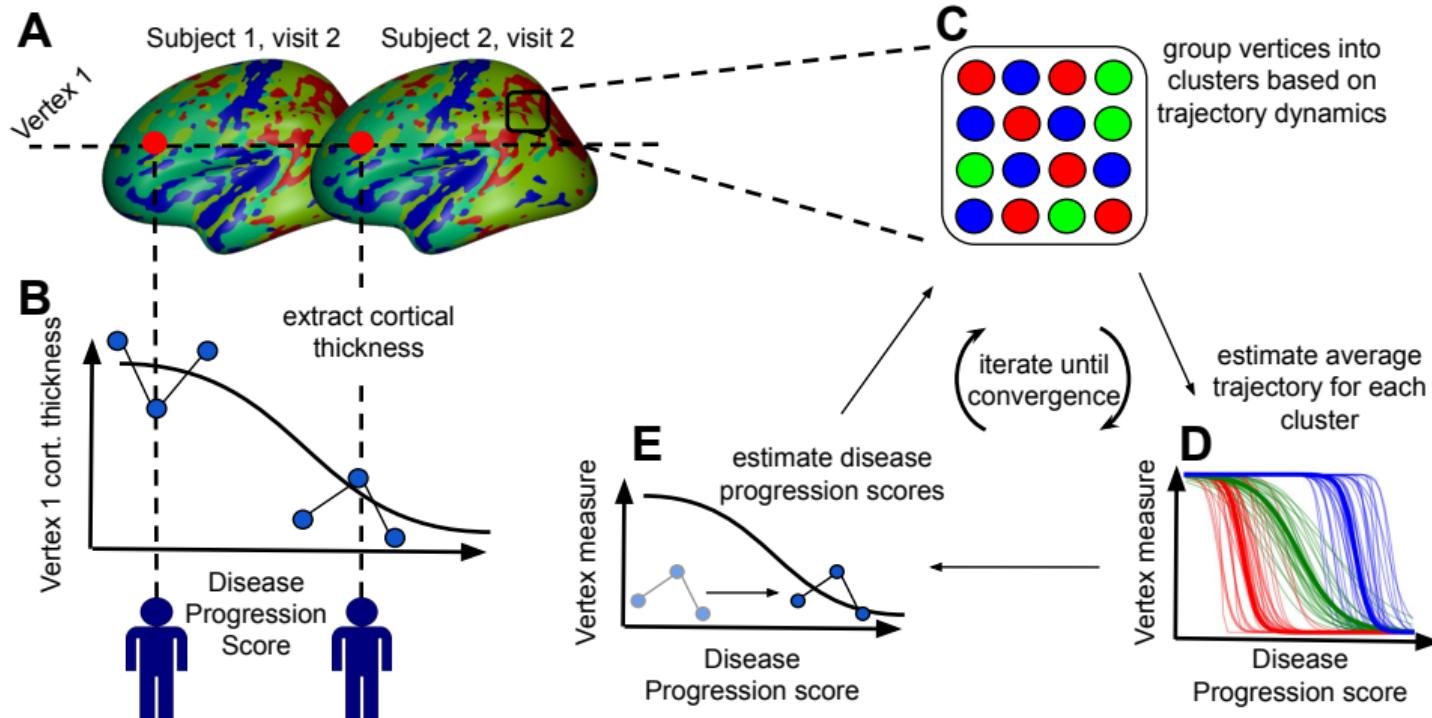
# Our model clusters vertices with similar trajectories of pathology



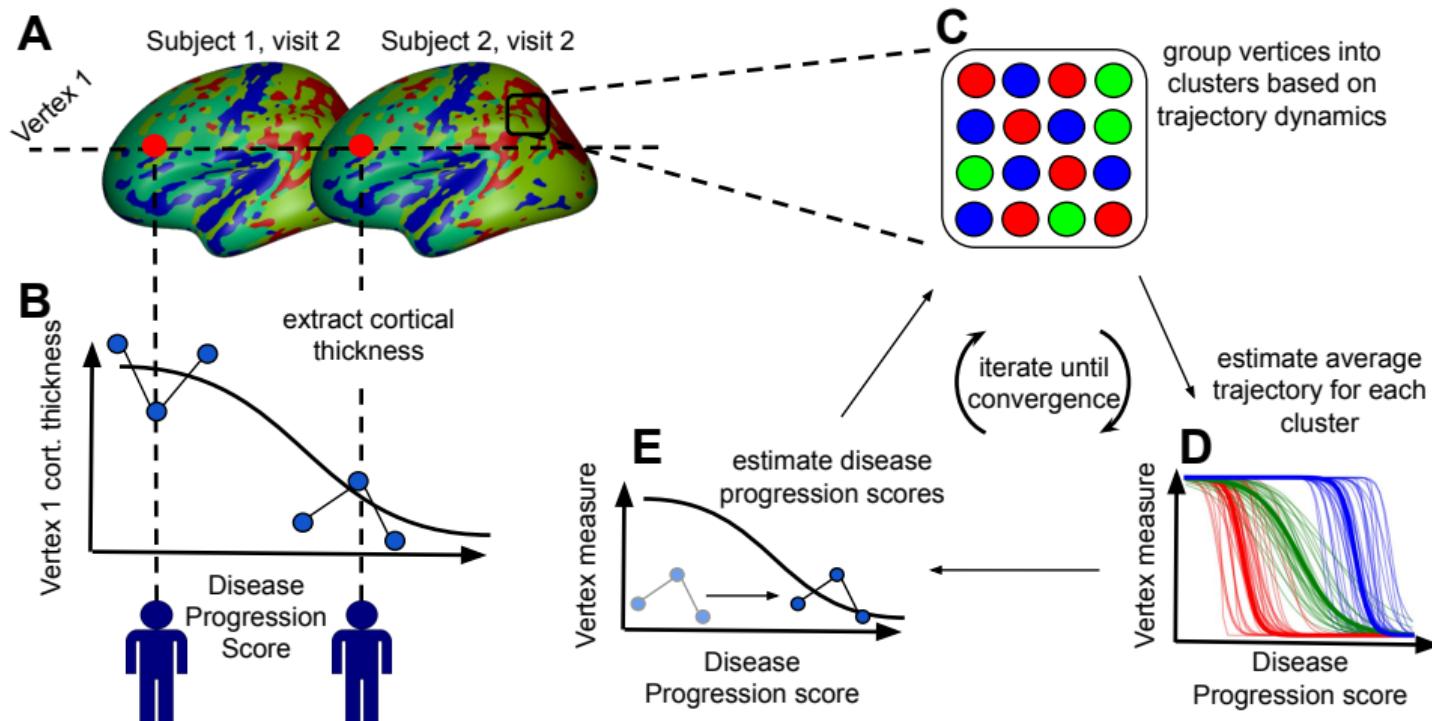
# Our model clusters vertices with similar trajectories of pathology



# Our model clusters vertices with similar trajectories of pathology



Our model clusters vertices with similar trajectories of pathology

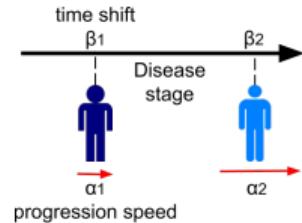


**Contribution:** Model can estimate pathology evolution **at each point** on the brain surface

## Building the model using a generative Bayesian framework

1. Model disease progression score for one subject  $i$  at visit  $j$ :

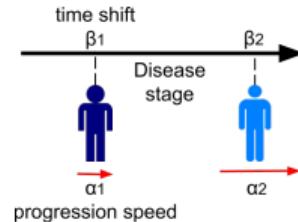
$$s_{ij} = \alpha_i t_{ij} + \beta_i$$



## Building the model using a generative Bayesian framework

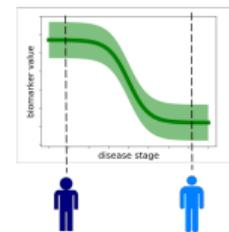
1. Model disease progression score for one subject  $i$  at visit  $j$ :

$$s_{ij} = \alpha_i t_{ij} + \beta_i$$



2. Model trajectory of cortical thickness at one location  $l$  on the brain:

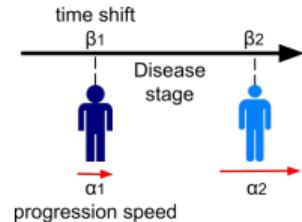
$$p(V_l^{ij} | \alpha_i, \beta_i, \theta_k, \sigma_k) \sim N(f(\alpha_i t_{ij} + \beta_i; \theta_k), \sigma_k)$$



## Building the model using a generative Bayesian framework

1. Model disease progression score for one subject  $i$  at visit  $j$ :

$$s_{ij} = \alpha_i t_{ij} + \beta_i$$

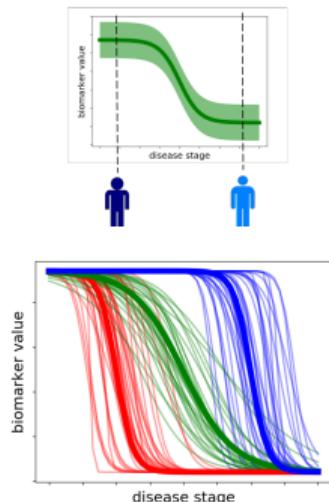


2. Model trajectory of cortical thickness at one location  $l$  on the brain:

$$p(V_l^{ij} | \alpha_i, \beta_i, \theta_k, \sigma_k) \sim N(f(\alpha_i t_{ij} + \beta_i; \theta_k), \sigma_k)$$

3. Extend to all locations and subjects:

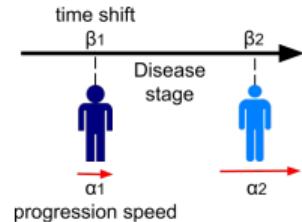
$$p(V, Z | \alpha, \beta, \theta, \sigma) = \prod_l^L \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i; \theta_{Z_l}), \sigma_{Z_l})$$



# Building the model using a generative Bayesian framework

1. Model disease progression score for one subject  $i$  at visit  $j$ :

$$s_{ij} = \alpha_i t_{ij} + \beta_i$$



2. Model trajectory of cortical thickness at one location  $l$  on the brain:

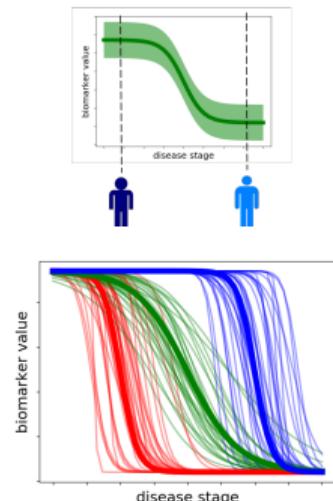
$$p(V_l^{ij} | \alpha_i, \beta_i, \theta_k, \sigma_k) \sim N(f(\alpha_i t_{ij} + \beta_i; \theta_k), \sigma_k)$$

3. Extend to all locations and subjects:

$$p(V, Z | \alpha, \beta, \theta, \sigma) = \prod_l^L \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i; \theta_{Z_l}), \sigma_{Z_l})$$

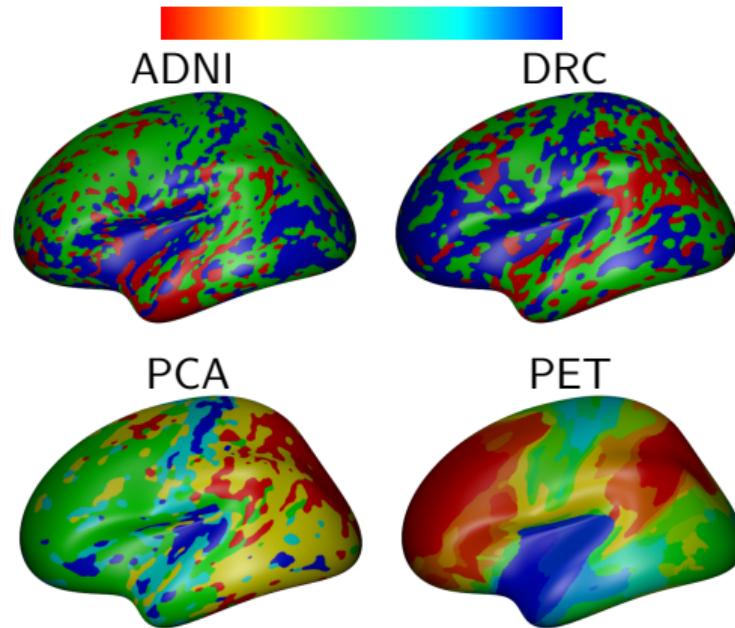
4. Marginalise over the hidden variables  $Z_l$  (cluster assignments):

$$p(V | \alpha, \beta, \theta, \sigma) = \prod_{l=1}^L \sum_{k=1}^K p(Z_l = k) \prod_{(i,j) \in I} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i ; \theta_k), \sigma_k)$$



## Our Model Finds Plausible Atrophy Patterns on Four Datasets

- ▶ Similar patterns of atrophy in independent Alzheimer's MRI datasets (ADNI vs DRC)
- ▶ Distinct patterns of atrophy in different diseases (Alzheimer's vs PCA) and modalities (MRI vs PET)  
severe pathology                      moderate pathology

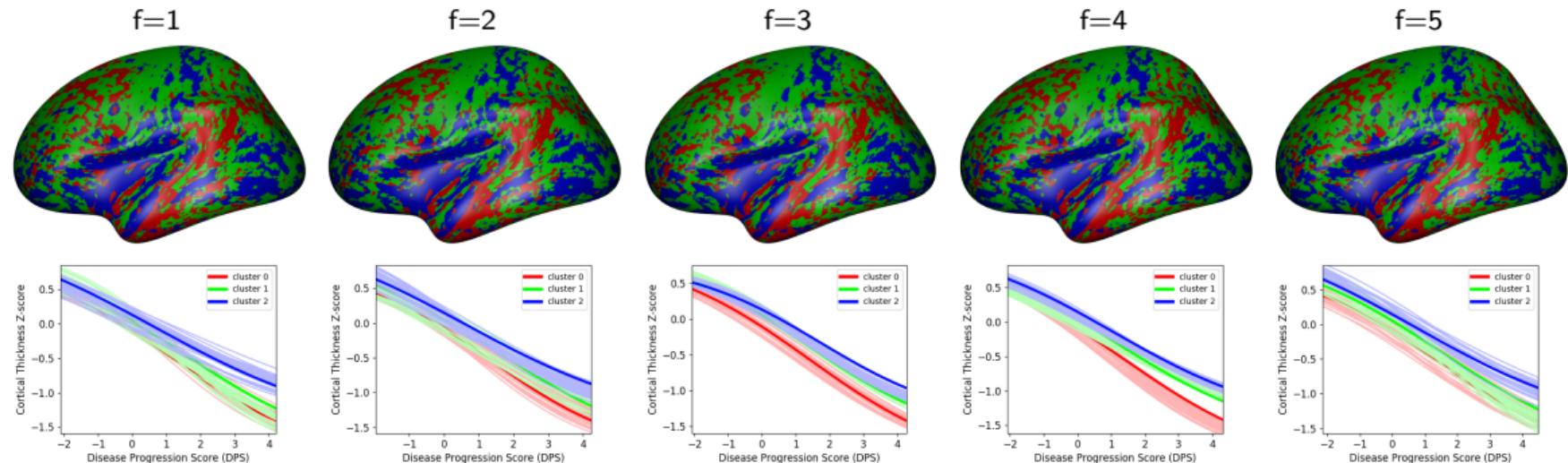


Marinescu et al., NeuroImage, 2019

# Validation - Model Robustly Estimates Atrophy Patterns

**Method:** Tested the consistency of the spatial clustering in ADNI using 10-fold CV

**Results:** Good agreement in terms of spatial distribution (dice score 0.89)



Marinescu et al., Neuroimage, 2019

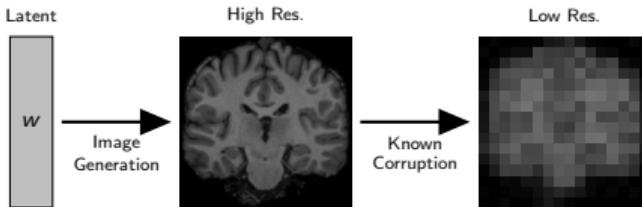
- ▶ We modelled the continuous progression of Alzheimer's disease and related dementias
- ▶ Used generative bayesian model that does not require labels (unsupervised)
- ▶ Plausible results on four different datasets
- ▶ However, such models require good quality data, to perform registration and extract disease markers
- ▶ How can we do such modelling for scans with limited resolution and contrast?

# Outline

1. Disease progression modelling of Alzheimer's disease
  - 1.1 Towards unsupervised clustering of biomarker trajectories



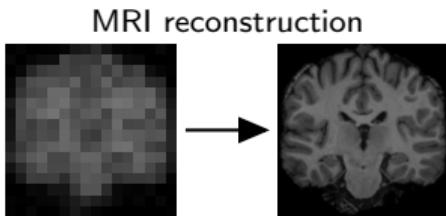
2. Image Reconstruction using Deep Generative Models



3. Future work towards brain anatomy simulators

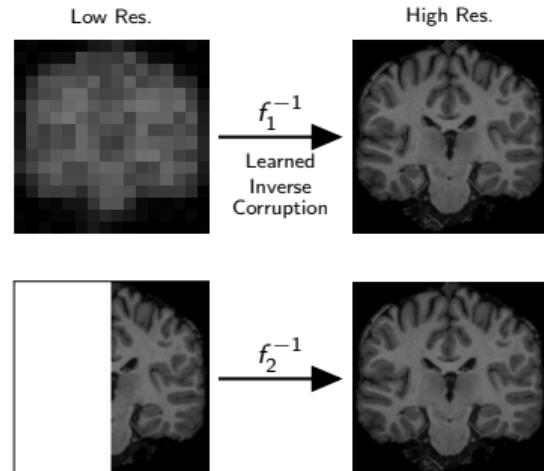
Aim: image reconstruction using **\*pre-trained\*** generator models

- Adapt the state-of-the-art StyleGAN2 for medical image reconstruction



## Current image reconstruction methods have several limitations

- ▶ Require large computational resources and data
- ▶ Are specific to particular corruption tasks
- ▶ Cannot deal with distribution shifts:
  - ▶ in inputs: e.g. older populations
  - ▶ in corruption type: e.g. change in blur kernel
- ▶ Are anti-causal, so they don't follow the data-generation process



## Limitation 1: State-of-the-art DL methods have large computational requirements

- Requirements = Computation Time + Advanced Hardware + Large Datasets
- Most computation now runs on clouds

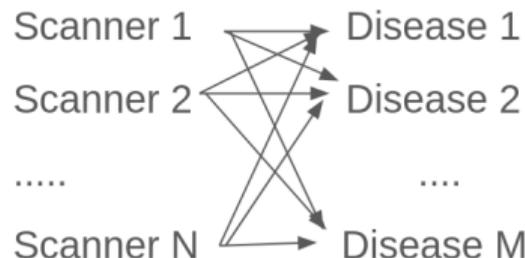


- Currently few labs/companies have the resources to train state-of-the-art models
  - StyleGAN2: 9 days on 4 GPUs
  - GPT-3: 355 years on single GPU
- Solutions moving forward:
  - Adapting previously-trained models
  - Combine smaller models into larger ones

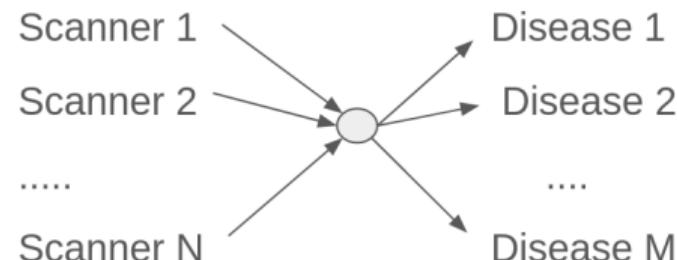
## Limitation 2: Distribution shifts require model re-training

- ▶ Distribution shifts happen all the time:
  - ▶ Changes in hospital scanners, protocols, software upgrades
  - ▶ Can be continuous: population getting older due to better healthcare
- ▶ Shifts can result in combinatorial effects in number of re-training instances!
- ▶ Compositionality is one potential solution

Without compositionality: **N x M**

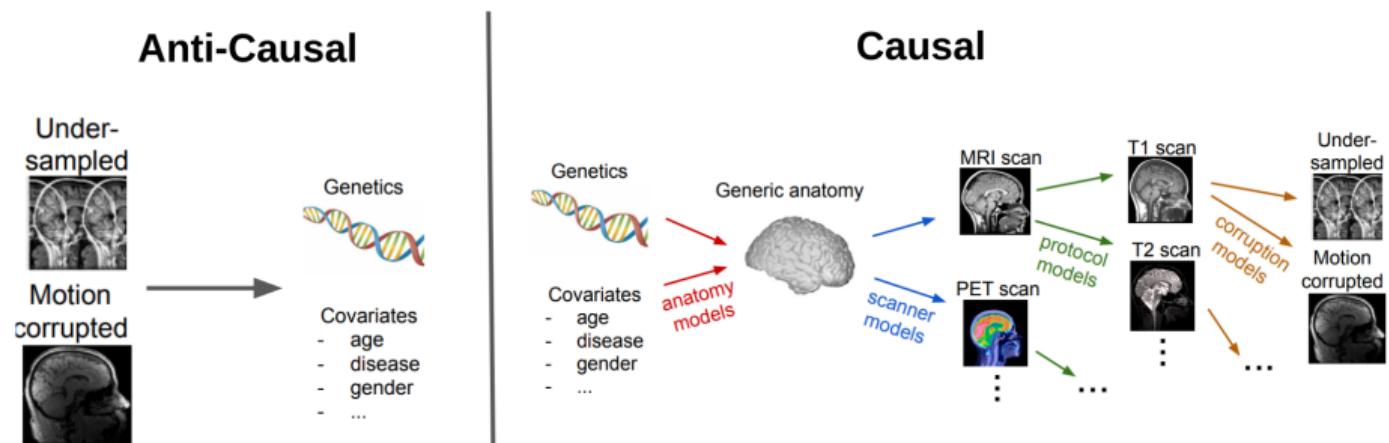


With compositionality: **N + M**



## Limitation 3: Models are anti-causal

- ▶ Existing model don't follow the data-generation process
  - ▶ Discriminative modelling easier than generative
- ▶ Causal modelling is the **right solution** to deal with distribution shifts



Recent models can perform image reconstruction using **pre-trained** generative models

Image2StyleGAN++ (Abdal et al, 2020)



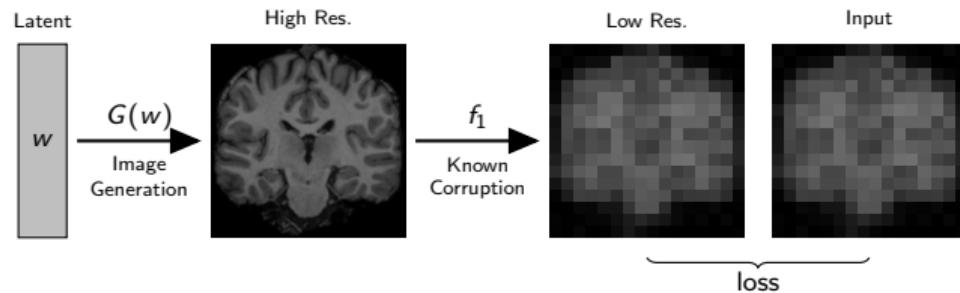
PULSE (Menon et al, 2020)



- ▶ These methods can generalise for any corruption process, because they don't use an embedder network
- ▶ **Cannot characterize uncertainty and recover multiple solutions**
- ▶ We will aim to construct a Bayesian formulation that can fully characterize the posterior over all potential solutions

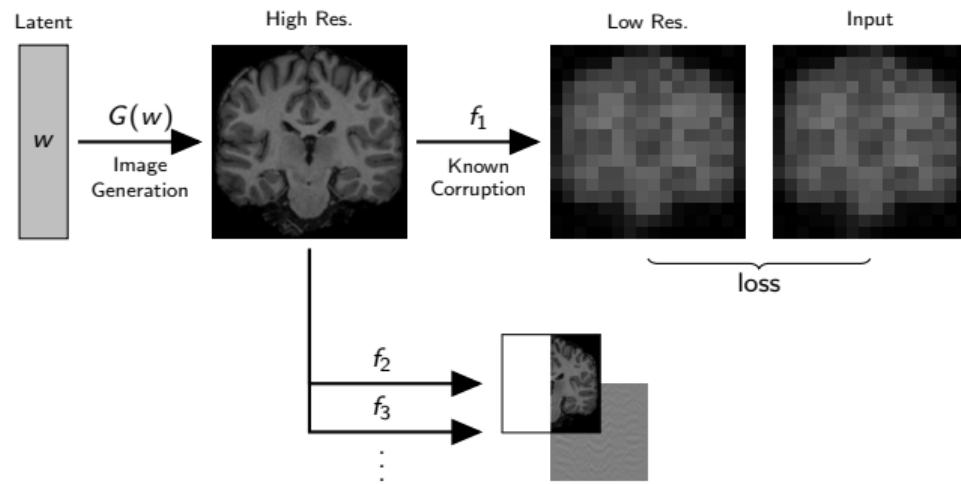
Method: We perform image reconstruction by combining two models

1. a pre-trained generator  $G$  (StyleGAN2)
2. a known forward corruption model  $f_1$



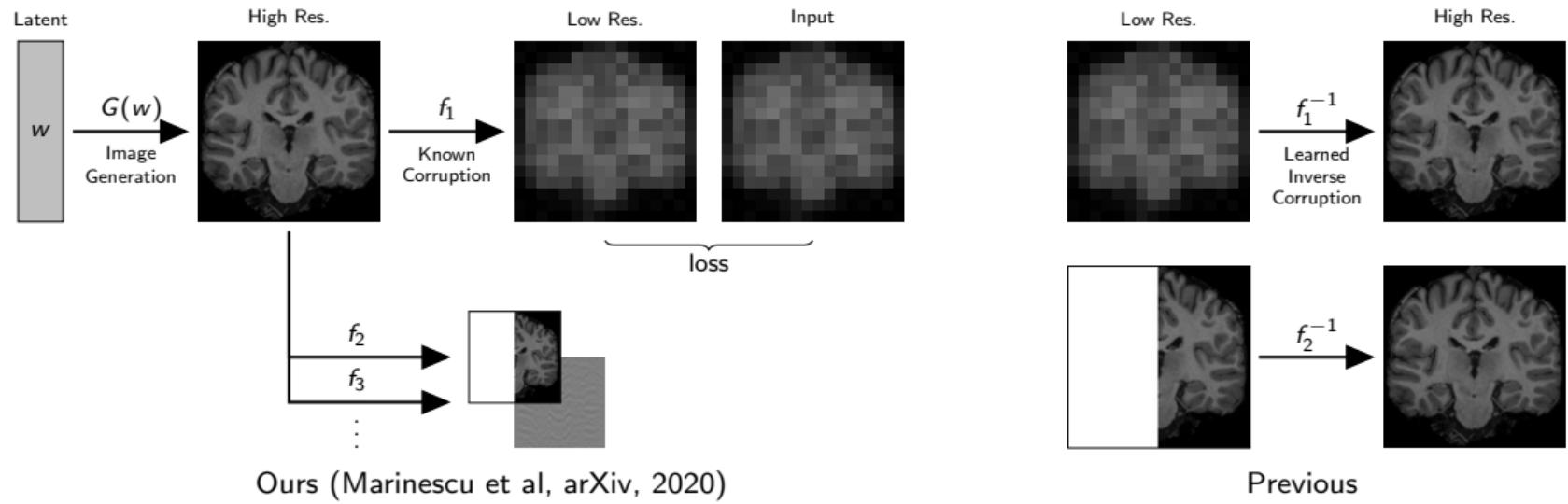
Method: We perform image reconstruction by combining two models

1. a pre-trained generator  $G$  (StyleGAN2)
2. a known forward corruption model  $f_1$



Method: We perform image reconstruction by combining two models

1. a pre-trained generator  $G$  (StyleGAN2)
2. a known forward corruption model  $f_1$



Reconstructed image is given by computing the Bayesian maximum a-posteriori (MAP) estimate

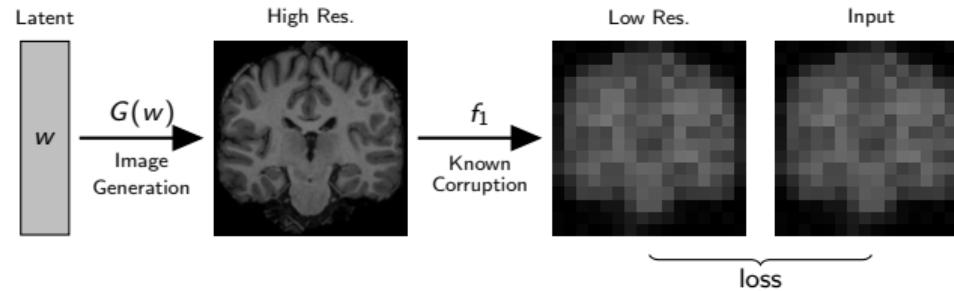
- We optimise:

$$w^* = \arg \max_w p(w)p(I|w)$$

- For uninformative prior  $p(w)$  and Gaussian noise model (pixelwise independent), we get:

$$w^* = \arg \min_w \|I - f \circ G(w)\|_2^2$$

- This can be optimised with SGD
- Once we get  $w^*$ , the the reconstructed image is  $G(w^*)$



## Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion

## Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes

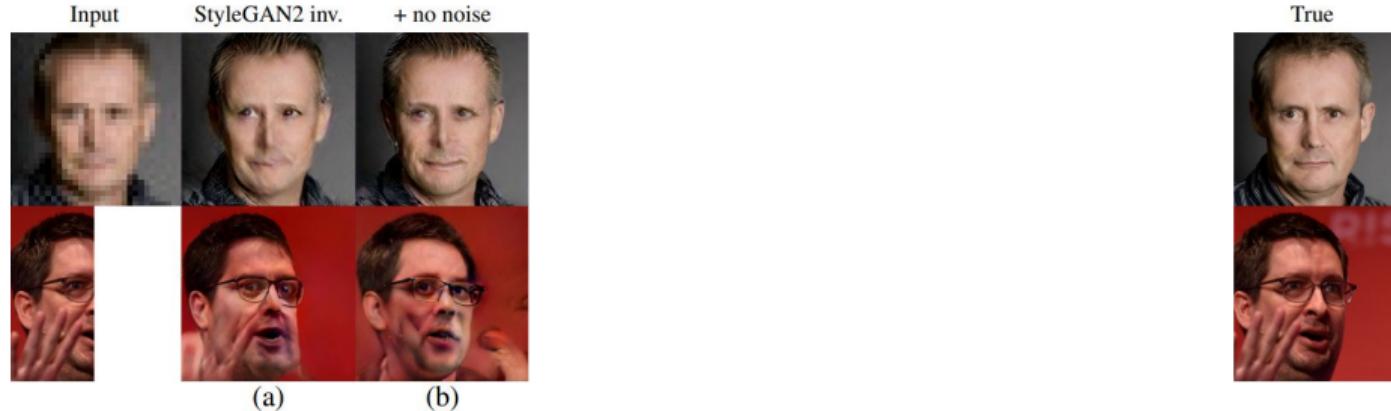
$$w^*, \eta^* = \arg \min_{w, \eta} \|\phi(I) - \phi \circ f \circ G(w, \eta)\|_2^2$$



## Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
  - remove noise layers

$$w^* = \arg \min_w \|\phi(I) - \phi \circ f \circ G(w)\|_2^2$$

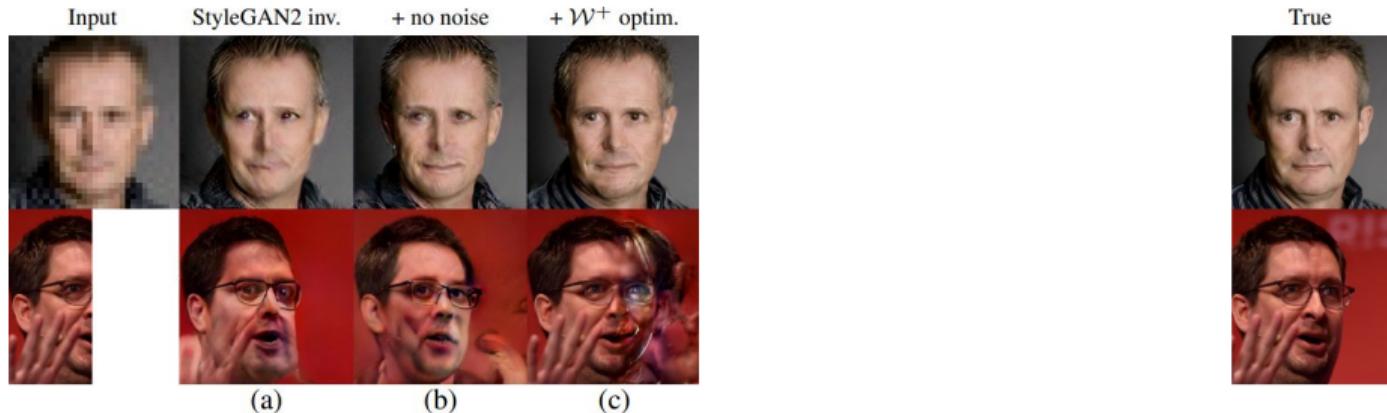


## Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
  - optimize latents at all resolutions

$$\mathbf{w} = \mathbf{w}_1, \dots, \mathbf{w}_L$$

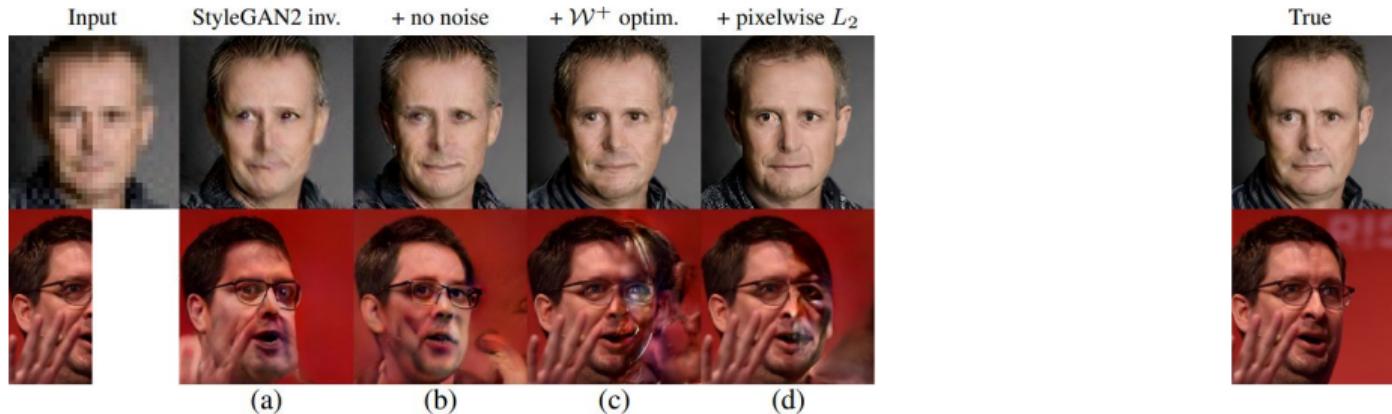
$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\phi(I) - \phi \circ f \circ G(\mathbf{w})\|_2^2$$



## Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
  - add pixelwise loss

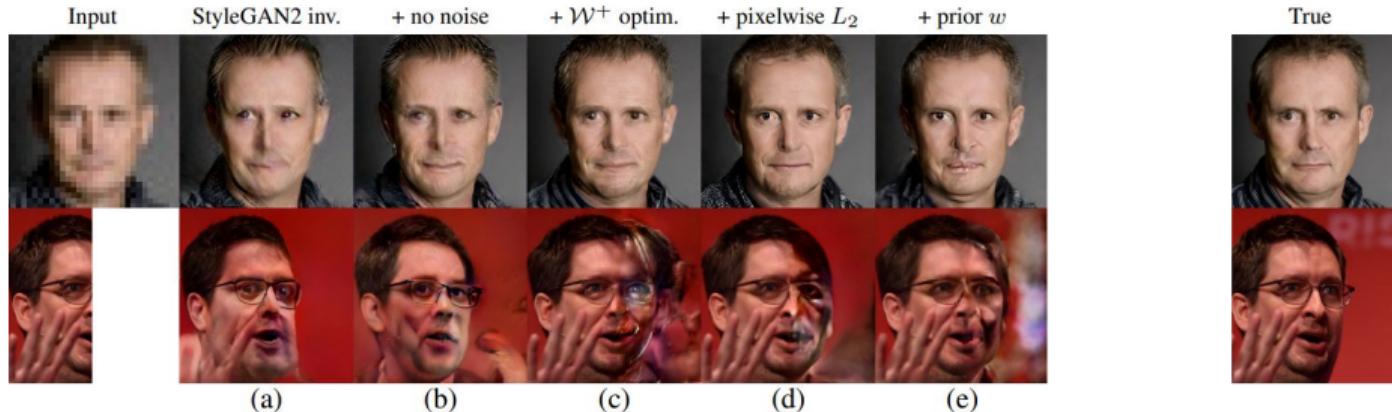
$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\phi(I) - \phi \circ f \circ G(\mathbf{w})\|_2^2 + \|I - f \circ G(\mathbf{w})\|_2^2$$



## Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
  - gaussian prior on latents

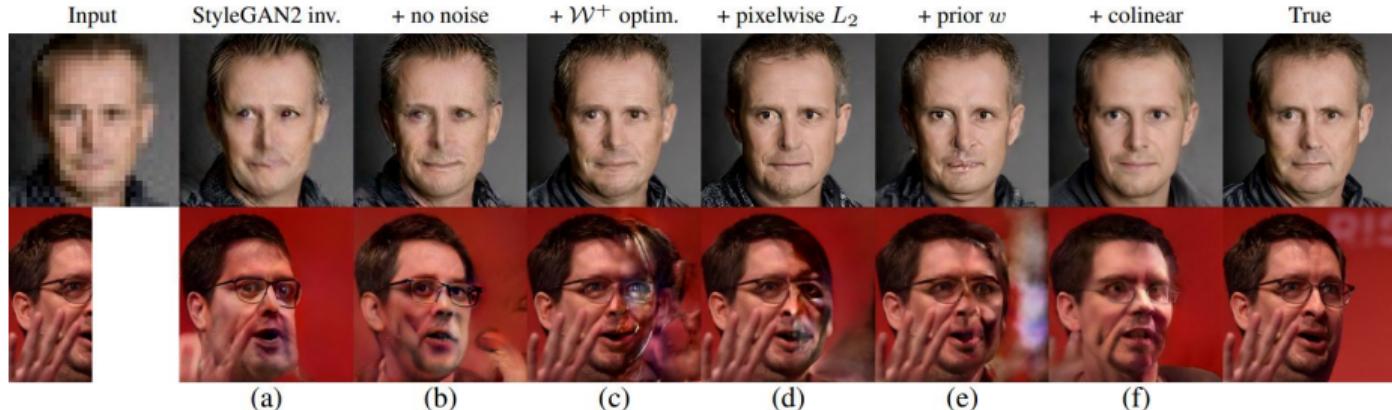
$$\begin{aligned}\mathbf{w}^* = \arg \min_{\mathbf{w}} & ||\phi(I) - \phi \circ f \circ G(\mathbf{w})||_2^2 + ||I - f \circ G(\mathbf{w})||_2^2 + \\ & + \sum_i \left( \frac{w_i - \mu}{\sigma_i} \right)^2\end{aligned}$$



## Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
  - force latents to be colinear

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\phi(I) - \phi \circ f \circ G(\mathbf{w})\|_2^2 + \|I - f \circ G(\mathbf{w})\|_2^2 + \\ + \sum_i \left( \frac{\mathbf{w}_i - \mu}{\sigma_i} \right)^2 - \sum_{i,j} \frac{\mathbf{w}_i \mathbf{w}_j^T}{\|\mathbf{w}_i\| \|\mathbf{w}_j\|}$$



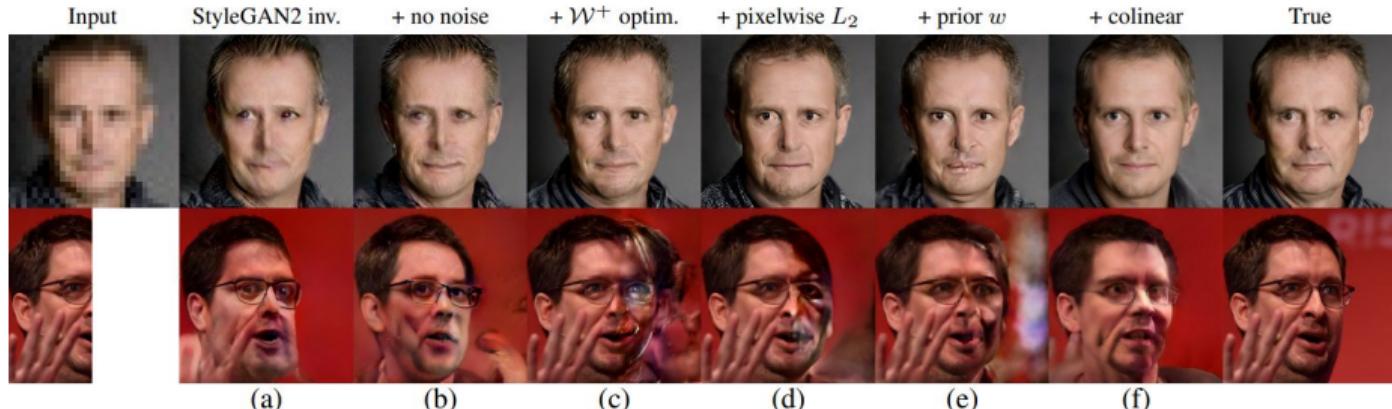
## Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion

- Yet the reconstruction was not good → required several changes

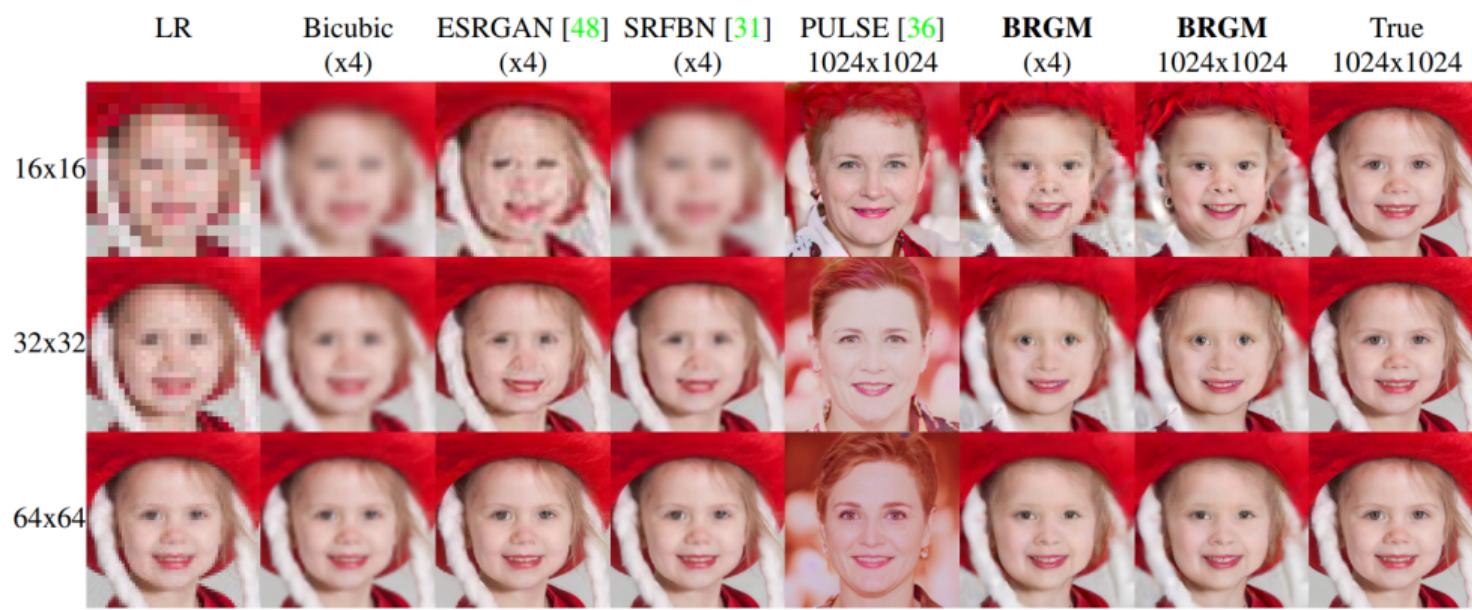
- Analytically expressed the full likelihood (Marinescu et al, 2021)

$$\begin{aligned} \mathbf{w}^* = \arg \min_{\mathbf{w}} & ||\phi(I) - \phi \circ f \circ G(\mathbf{w})||_2^2 + ||I - f \circ G(\mathbf{w})||_2^2 + \\ & + \sum_i \left( \frac{\mathbf{w}_i - \mu}{\sigma_i} \right)^2 - \sum_{i,j} \frac{\mathbf{w}_i \mathbf{w}_j^T}{|\mathbf{w}_i| |\mathbf{w}_j|} \end{aligned}$$



## Results on super-resolution using the FFHQ dataset

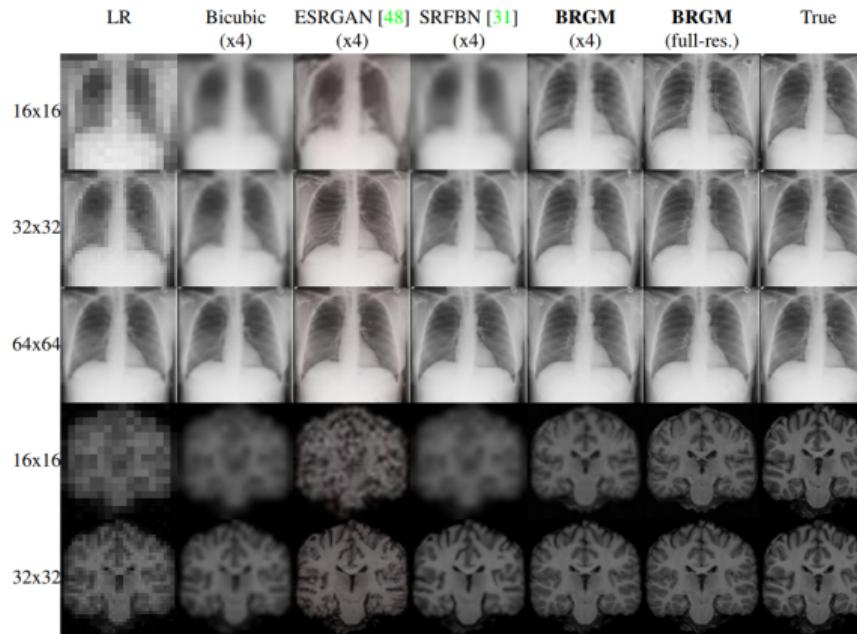
- We achieve state-of-the-art (SOTA) results on small inputs resolutions 16x16
- On larger resolutions (>32x32), we achieve very good results, albeit not SOTA



Marinescu et al, arXiv, 2020

## Similar results on super-resolution for medical datasets

- We achieve state-of-the-art (SOTA) results on small inputs resolutions 16x16
- On larger resolutions (>32x32), we achieve very good results, albeit not SOTA



Marinescu et al, arXiv, 2020

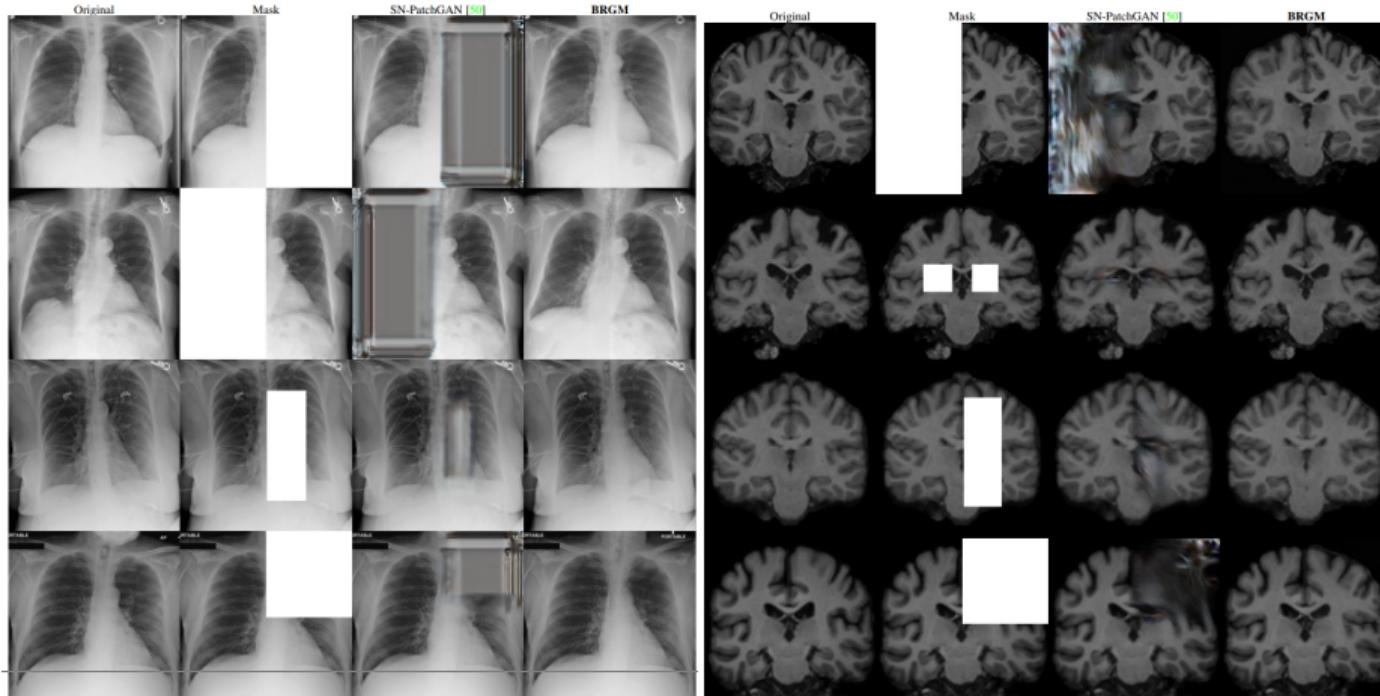
## Inpainting also achieves state-of-the-art results

- ▶ Best previous method (SN-PatchGAN, CVPR 2019) does not work for large masks
- ▶ Our method can “hypothesize” missing structure



## Inpainting also achieves state-of-the-art results

- Best previous method (SN-PatchGAN, CVPR 2019) does not work for large masks
- Our method can “hypothesize” missing structure



# Results confirmed through quantitative evaluation

- ▶ Three different datasets, at different resolutions
- ▶ Human study with 20 raters

## Super-resolution

Dataset	BRGM	PULSE [36]	ESRGAN [48]	SRFBN [31]
FFHQ 16 <sup>2</sup>	<b>0.24</b> /25.66	0.29/27.14	0.35/29.32	0.33/ <b>22.07</b>
FFHQ 32 <sup>2</sup>	0.30/18.93	0.48/42.97	0.29/23.02	<b>0.23</b> / <b>12.73</b>
FFHQ 64 <sup>2</sup>	0.36/16.07	0.53/41.31	0.26/18.37	<b>0.23</b> / <b>9.40</b>
FFHQ 128 <sup>2</sup>	0.34/15.84	0.57/34.89	0.15/15.84	<b>0.09</b> / <b>7.55</b>
X-ray 16 <sup>2</sup>	<b>0.18</b> / <b>11.61</b>	-	0.32/14.67	0.37/12.28
X-ray 32 <sup>2</sup>	0.23/10.47	-	0.32/12.56	<b>0.21</b> / <b>6.84</b>
X-ray 64 <sup>2</sup>	0.31/10.58	-	0.30/8.67	<b>0.22</b> / <b>5.32</b>
X-ray 128 <sup>2</sup>	0.27/10.53	-	0.20/7.19	<b>0.07</b> / <b>4.33</b>
Brains 16 <sup>2</sup>	<b>0.12</b> / <b>12.42</b>	-	0.34/22.81	0.33/12.57
Brains 32 <sup>2</sup>	<b>0.17</b> /11.08	-	0.31/14.16	0.18/ <b>6.80</b>

## Inpainting

Dataset	BRGM				SN-PatchGAN [50]			
	LPIPS	RMSE	PSNR	SSIM	LPIPS	RMSE	PSNR	SSIM
FFHQ	<b>0.19</b>	<b>24.28</b>	<b>21.33</b>	<b>0.84</b>	0.24	30.75	19.67	0.82
X-ray	<b>0.13</b>	<b>13.55</b>	<b>27.47</b>	<b>0.91</b>	0.20	27.80	22.02	0.86
Brains	<b>0.09</b>	<b>8.65</b>	<b>30.94</b>	<b>0.88</b>	0.22	24.74	21.47	0.75

## Human evaluation

Dataset	BRGM	PULSE [36]	ESRGAN [48]	SRFBN [31]
FFHQ 16 <sup>2</sup>	<b>0.42</b>	0.32	0.11	0.15
FFHQ 32 <sup>2</sup>	0.39	0.02	0.12	<b>0.47</b>
FFHQ 64 <sup>2</sup>	0.14	0.08	0.32	<b>0.45</b>
FFHQ 128 <sup>2</sup>	0.14	0.10	<b>0.39</b>	0.38

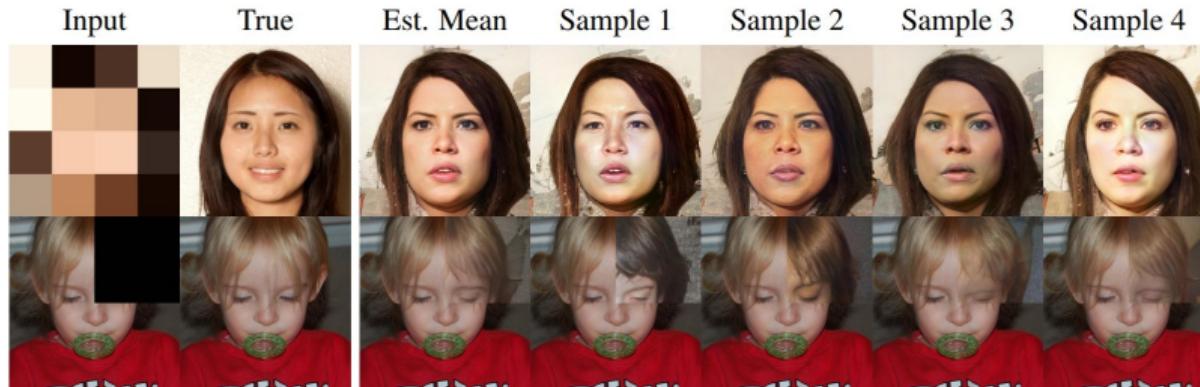
## Sampling multiple reconstructions through Variational Inference

- Variational inference can allow us to sample from the posterior distribution:

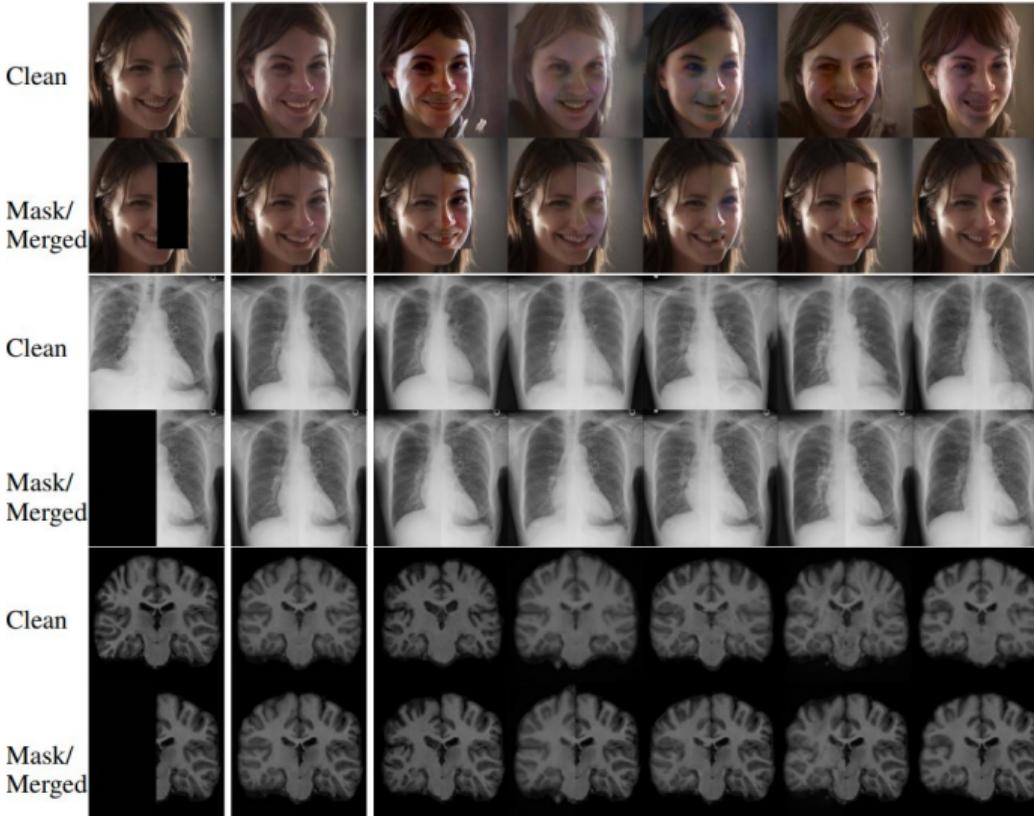
$$\theta^* = \arg \min_{\theta} KL[q(w|\theta)||p(w|I)] = \arg \min_{\theta} \int q(w|\theta) \log \frac{q(w|\theta)}{p(w)p(I|w)} dw$$

- We approximate the integral using Monte Carlo samples  $w^{(i)}$  taken from  $q(w|\theta)$

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^n \log q(w^{(i)}|\theta) - \log p(w^{(i)}) - \log p(I|w^{(i)})$$

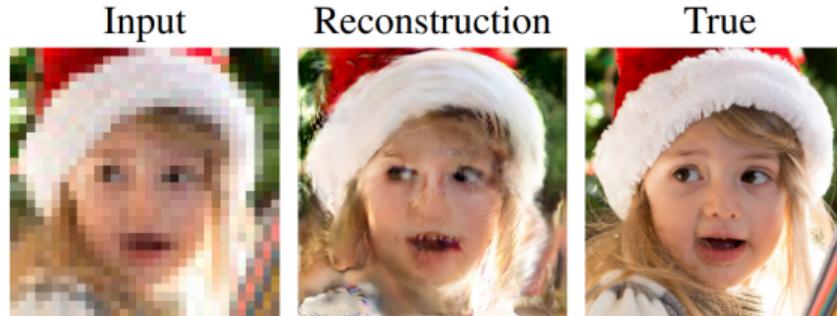


## More examples using Variational Inference

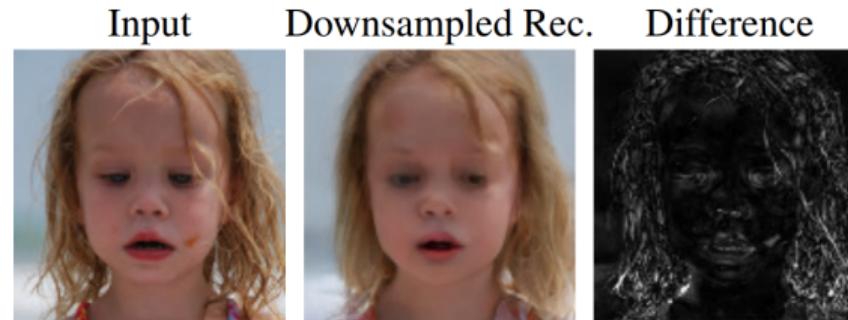


Our method also has limitations that we plan to address

- ▶ It can fail for images that are too dissimilar to the training ones
  - ▶ Because generator cannot extrapolate easily



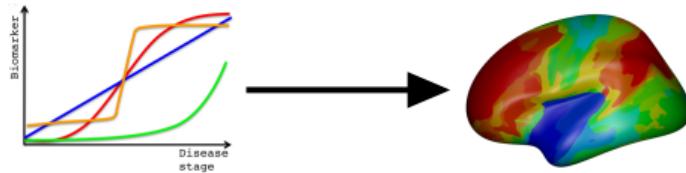
- ▶ Can be inconsistent with the input image



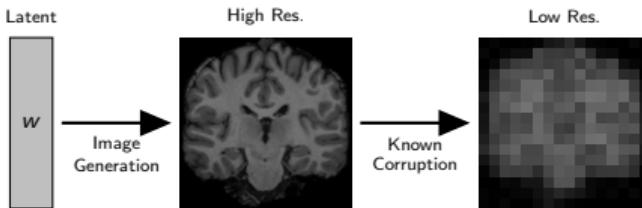
- ▶ Proposed a method for image reconstruction using pre-trained deep generative models
- ▶ Solutions given by both (i) Bayesian MAP estimates and (ii) Variational Inference
- ▶ State-of-the-art results on super-resolution and inpainting

# Outline

1. Disease progression modelling of Alzheimer's disease
  - 1.1 Towards unsupervised clustering of biomarker trajectories



2. Image Reconstruction using Deep Generative Models



3. Future work towards brain anatomy simulators

## Accurate diagnosis and prognosis through AI

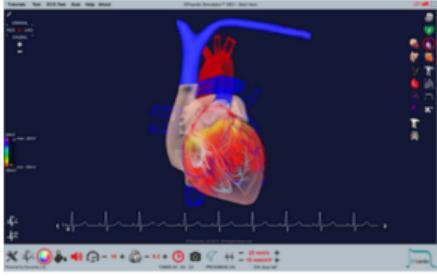


## AI to augment doctors

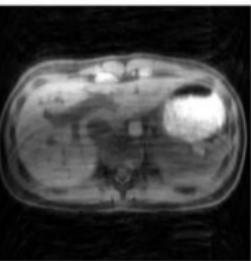
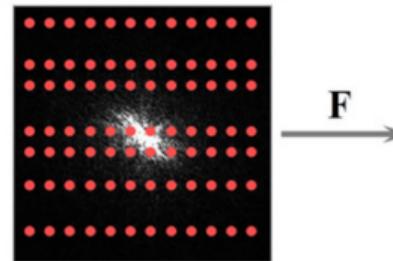


## Future work

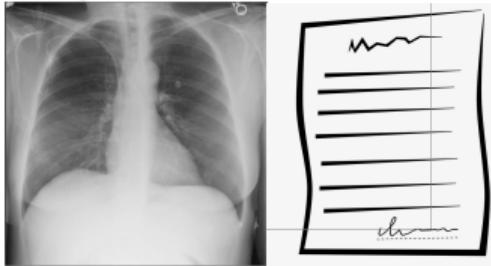
### Biological simulators



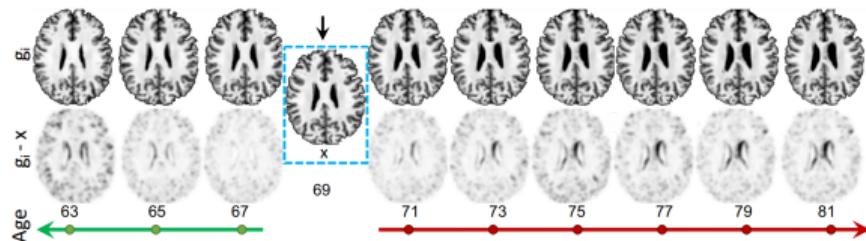
### Better and faster reconstruction of medical images Undersampled k-space



### Multimodal modelling images + text + structural data



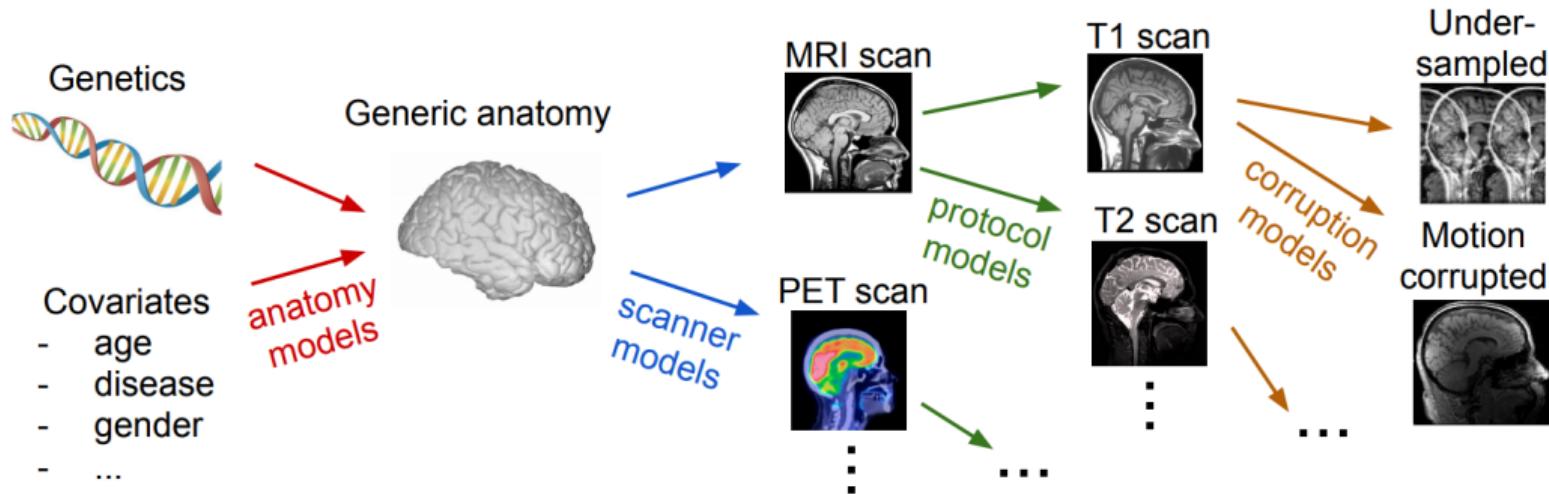
### Disease Progression Modelling



## Future work: Brain tissue and anatomy simulator

Simulator for brain anatomy from genetics:

- ▶ Using deep generative models
- ▶ Accounting for distributions shifts
- ▶ Following causal principles

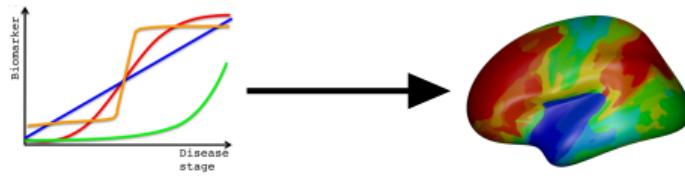


**Problem: Lack of good labels**

**Problem: Lack of good input data**

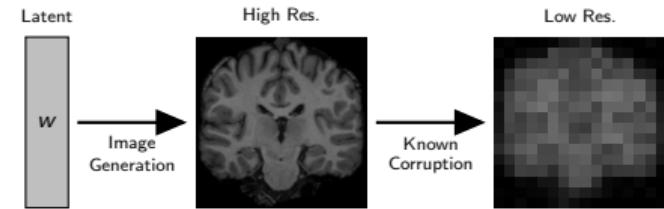
## Problem: Lack of good labels

Solution: Unsupervised Learning through  
Disease Progression Modelling



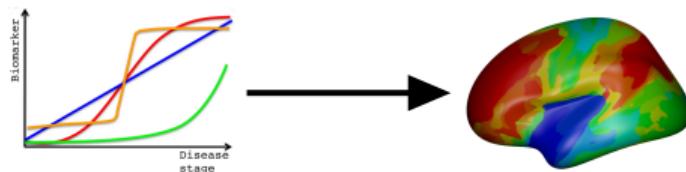
## Problem: Lack of good input data

Solution: Image Reconstruction  
using Deep Generative Models



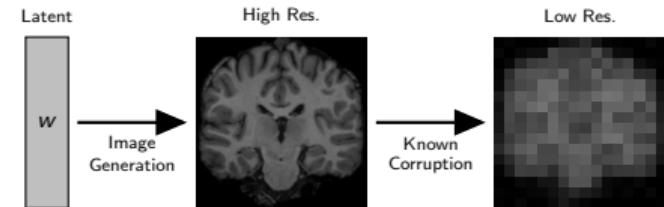
## Problem: Lack of good labels

Solution: Unsupervised Learning through  
Disease Progression Modelling



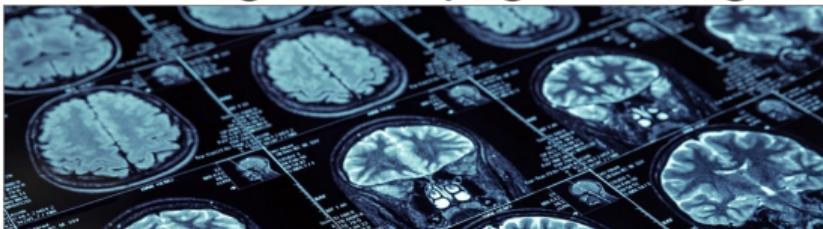
## Problem: Lack of good input data

Solution: Image Reconstruction  
using Deep Generative Models



Long-term vision

## Accurate diagnosis and prognosis through AI



## AI to augment doctors

