

Medical Image Generation and Analysis using Bayesian Generative Models

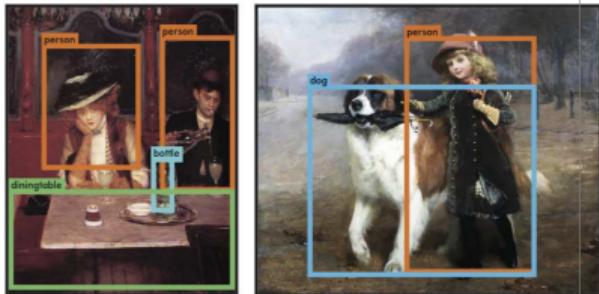
Răzvan V. Marinescu

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.

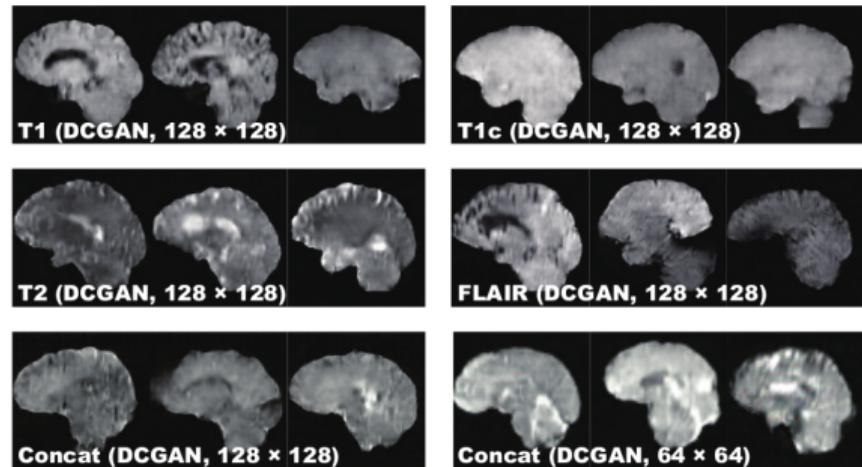
However, such milestones have not been translated to medical applications

► Prediction of clinical variables not always working:

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

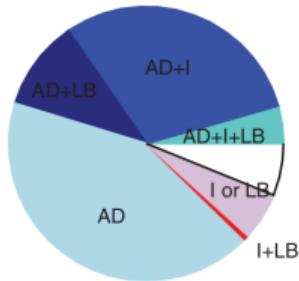
- Generated images are crude, not high-resolution, mostly 2D

Brain MRI generation (Han, 2018)



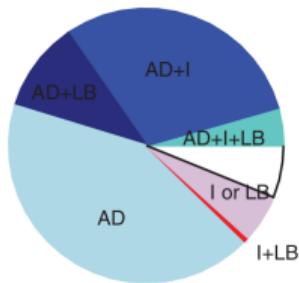
Lack of good labels

- Alzheimer's diagnosis accuracy just 42%

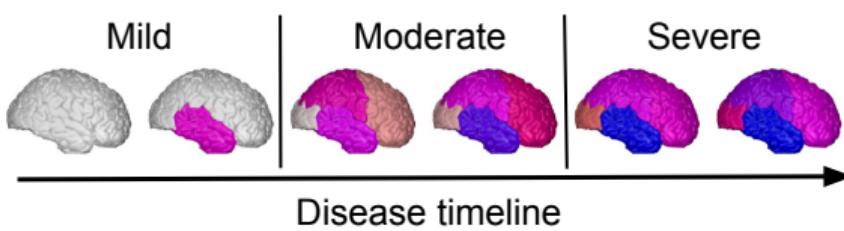


Lack of good labels

- ▶ Alzheimer's diagnosis accuracy just 42%



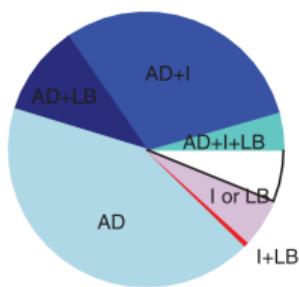
- ▶ Labels are categorical instead of continuous



Why are Machine Learning models not working on medical applications?

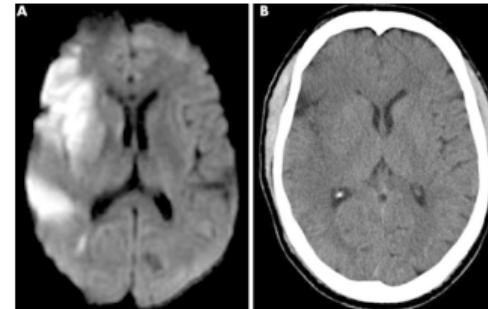
Lack of good labels

- Alzheimer's diagnosis accuracy just 42%

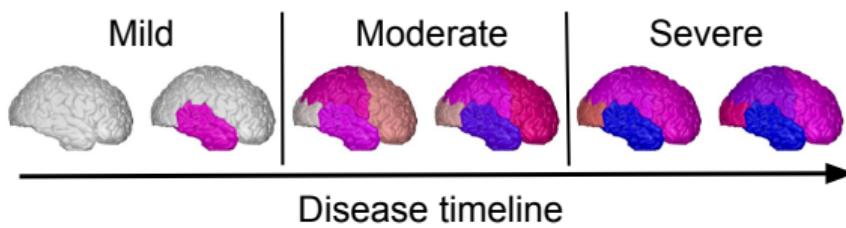


Lack of good input data/signal

- Limited contrast



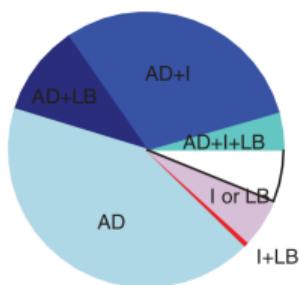
- Labels are categorical instead of continuous



Why are Machine Learning models not working on medical applications?

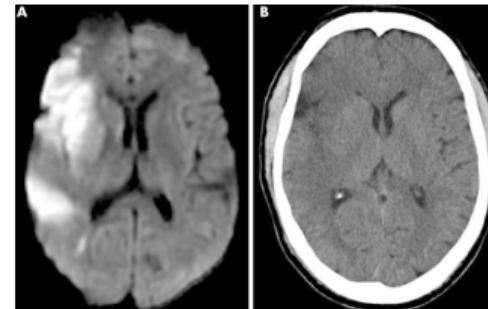
Lack of good labels

- Alzheimer's diagnosis accuracy just 42%



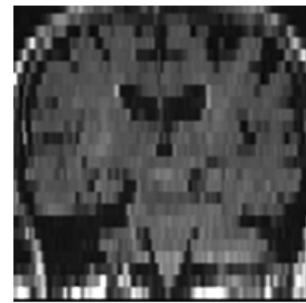
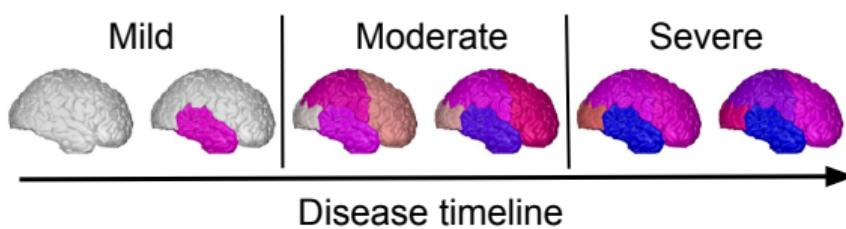
Lack of good input data/signal

- Limited contrast



- Low-resolution

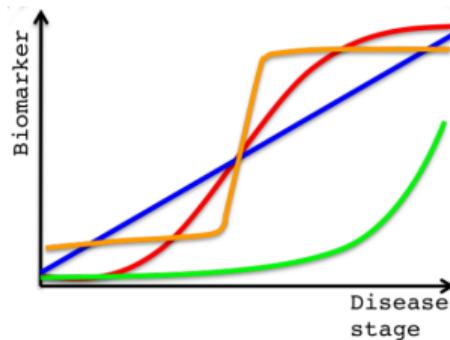
- Labels are categorical instead of continuous



What can we do?

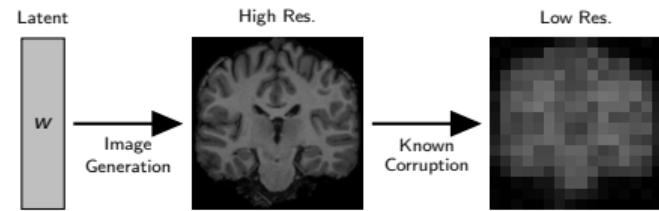
Lack of good labels

Solution: Unsupervised Learning of Continuous Dynamics
= Disease Progression Modelling



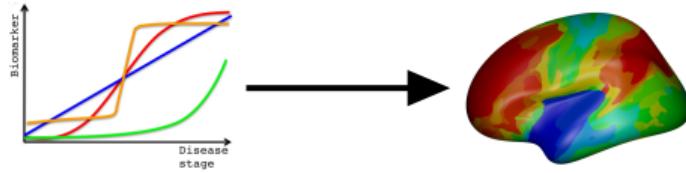
Lack of good input data/signal

Solution: Image Reconstruction using Deep Generative Models

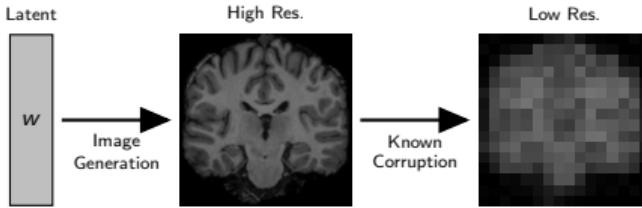


1. Disease progression modelling of Alzheimer's disease

1.1 Towards unsupervised clustering of biomarker trajectories



2. Image Reconstruction using Deep Generative Models



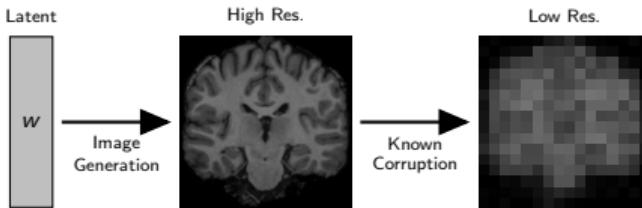
3. Future work

1. Disease progression modelling of Alzheimer's disease

1.1 Towards unsupervised clustering of biomarker trajectories



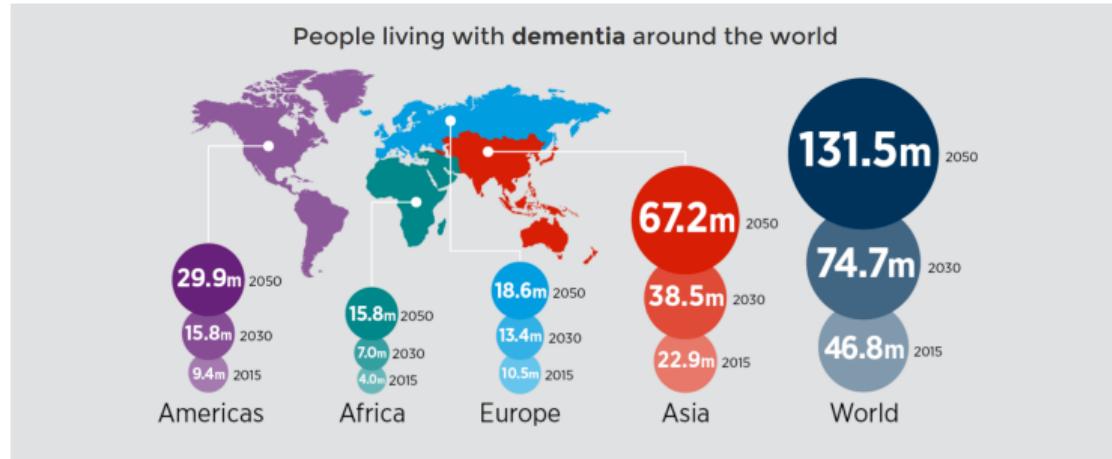
2. Image Reconstruction using Deep Generative Models



3. Future work

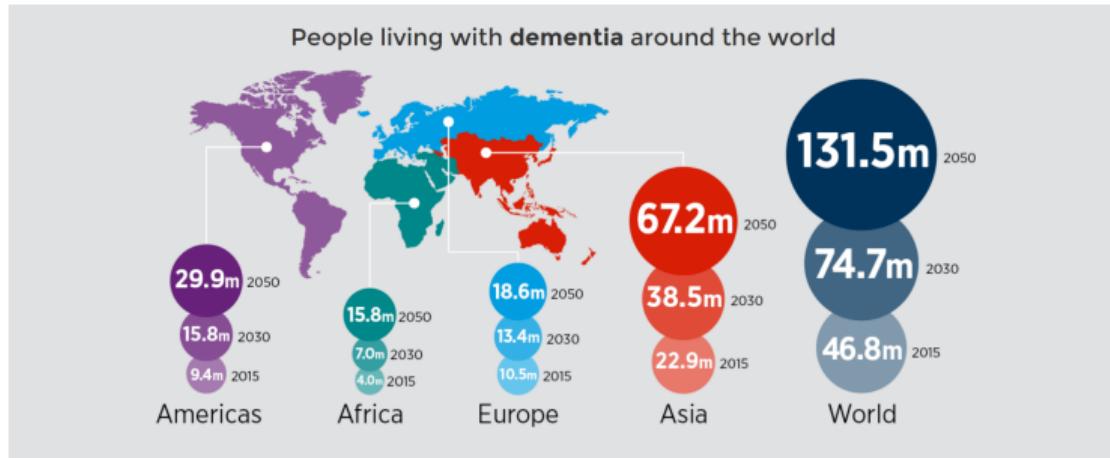
Alzheimer's Disease is a Devastating Disease

- 46 million people affected worldwide



Alzheimer's Disease is a Devastating Disease

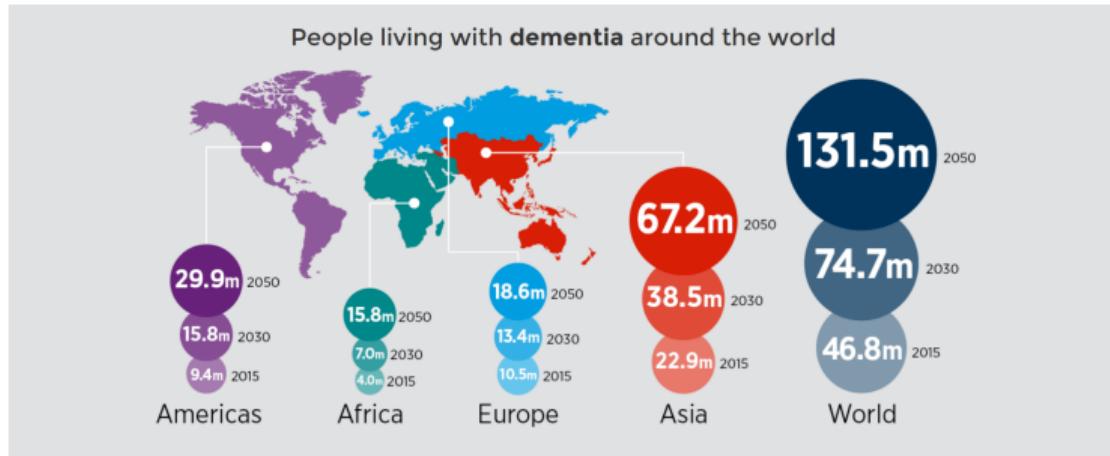
- ▶ 46 million people affected worldwide



- ▶ 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

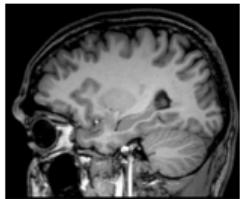
- ▶ 46 million people affected worldwide



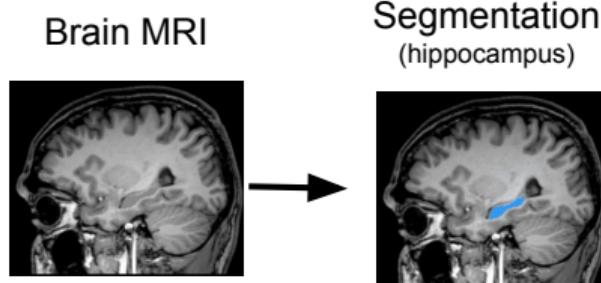
- ▶ No treatments available that stop or slow down cognitive decline
- ▶ Q: Why did clinical trials fail? A: Treatments were not administered early enough
- ▶ Q: How can we then identify subjects **early** in order to administer treatments?
- ▶ A: Disease progression model ...

Building a Disease Progression Model

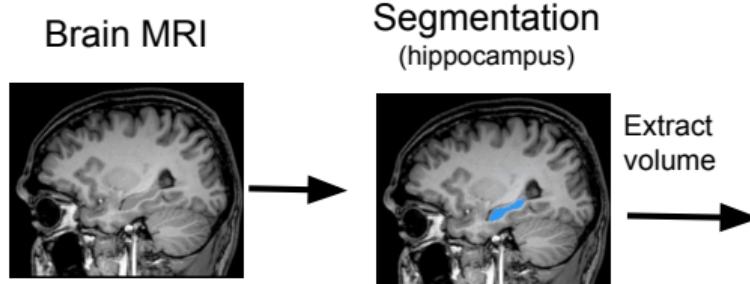
Brain MRI



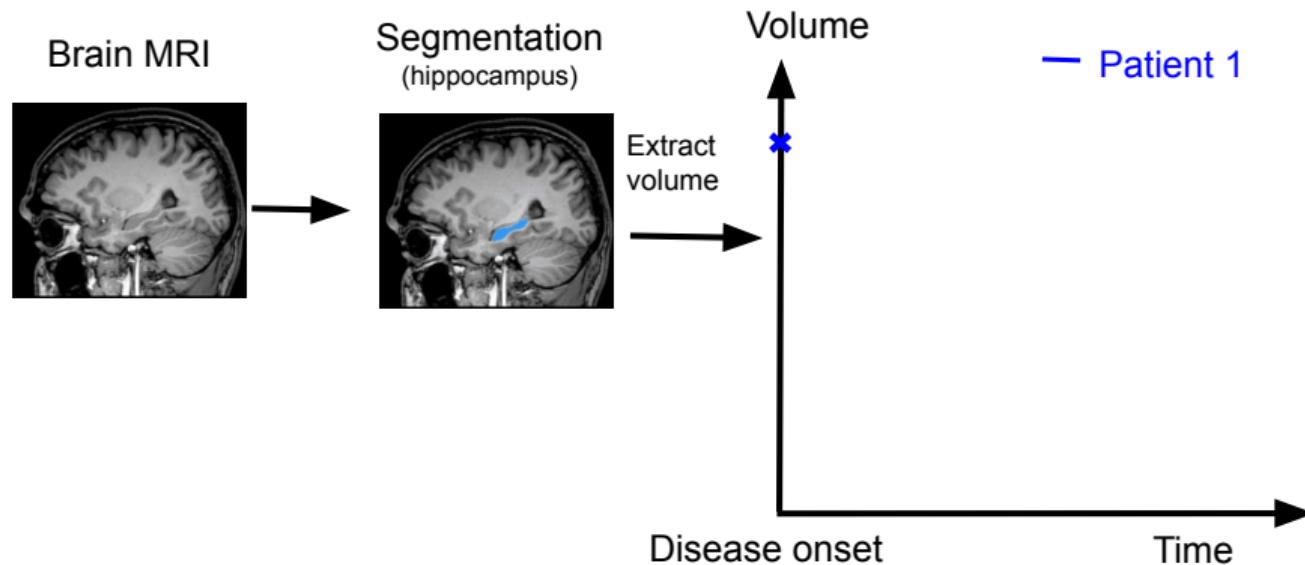
Building a Disease Progression Model



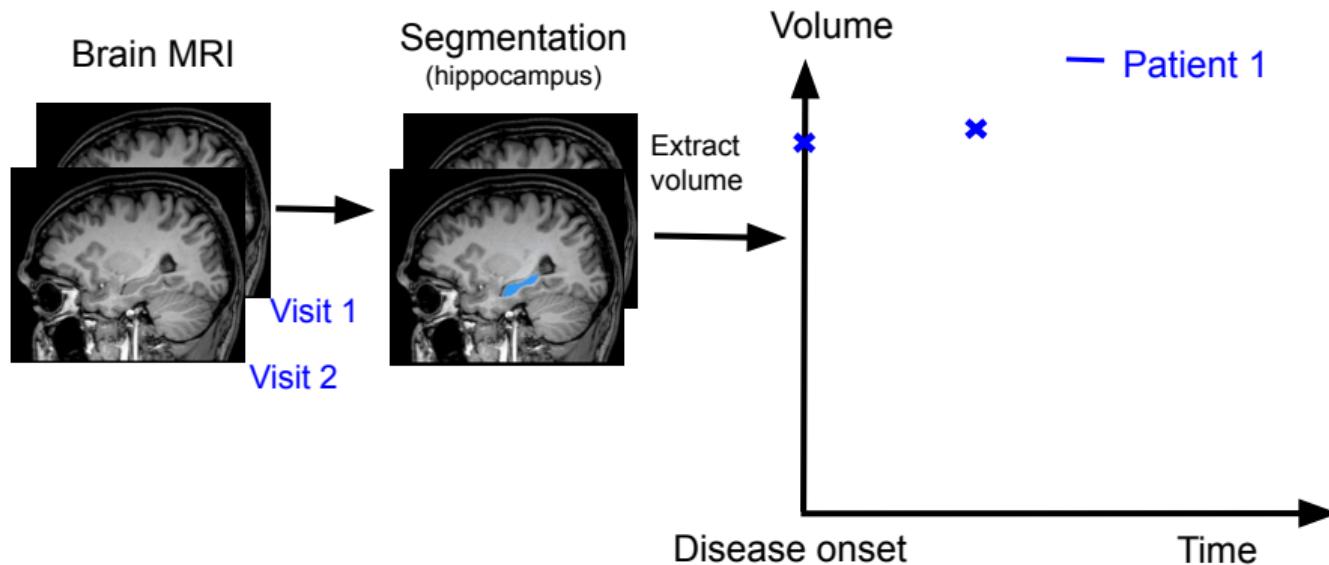
Building a Disease Progression Model



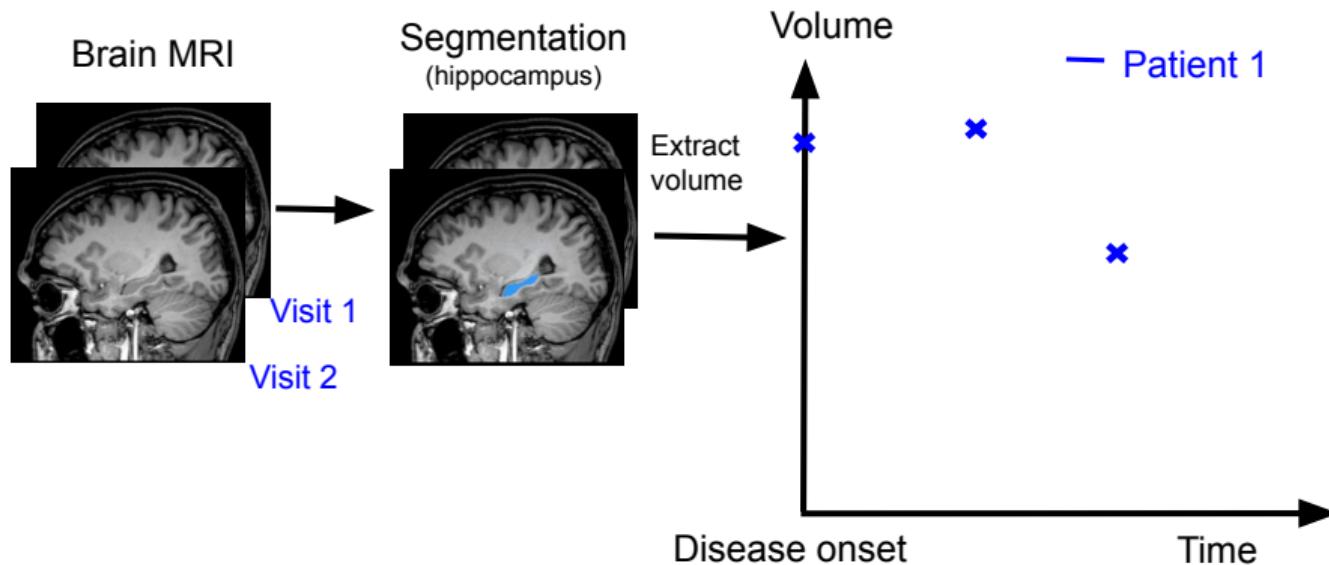
Building a Disease Progression Model



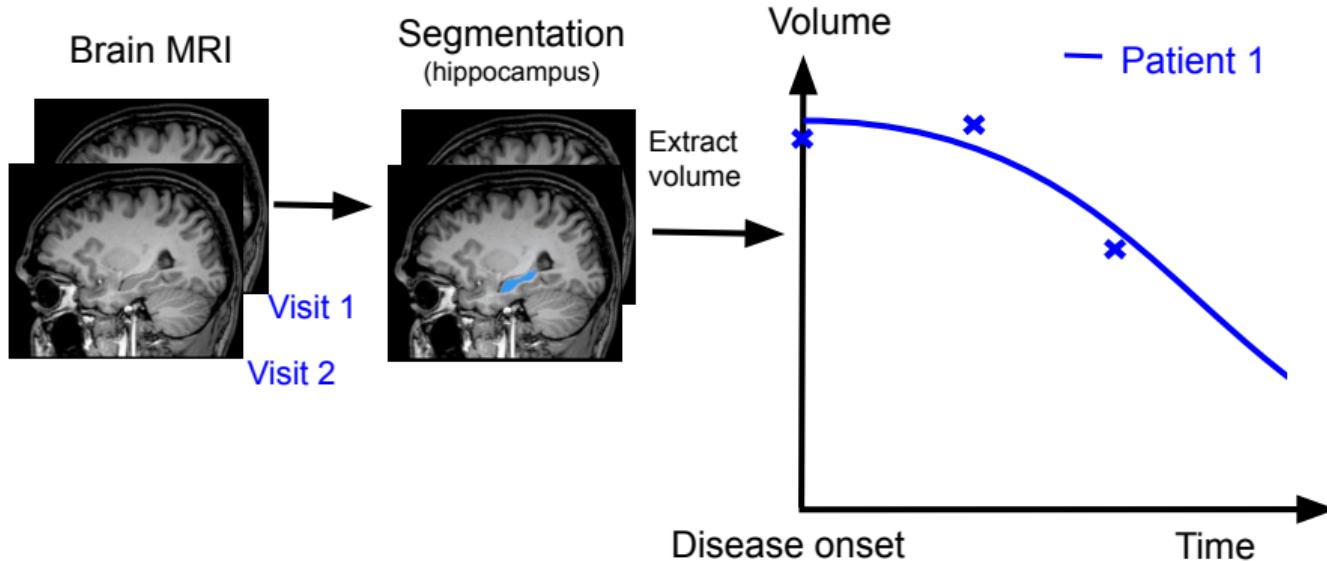
Building a Disease Progression Model



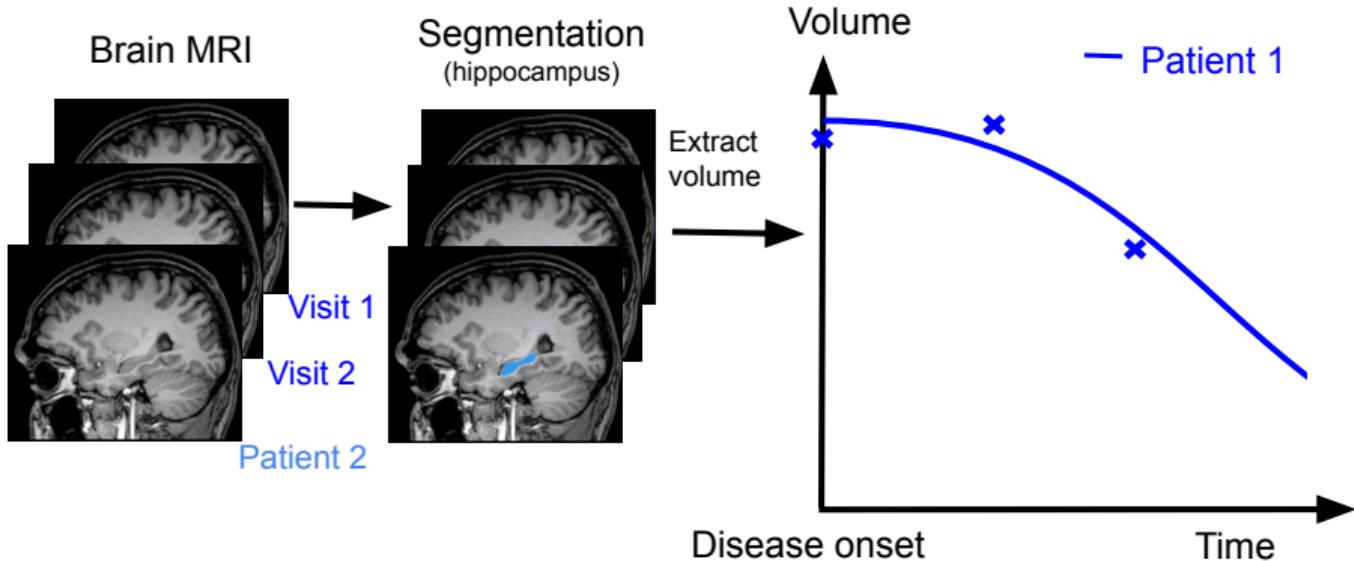
Building a Disease Progression Model



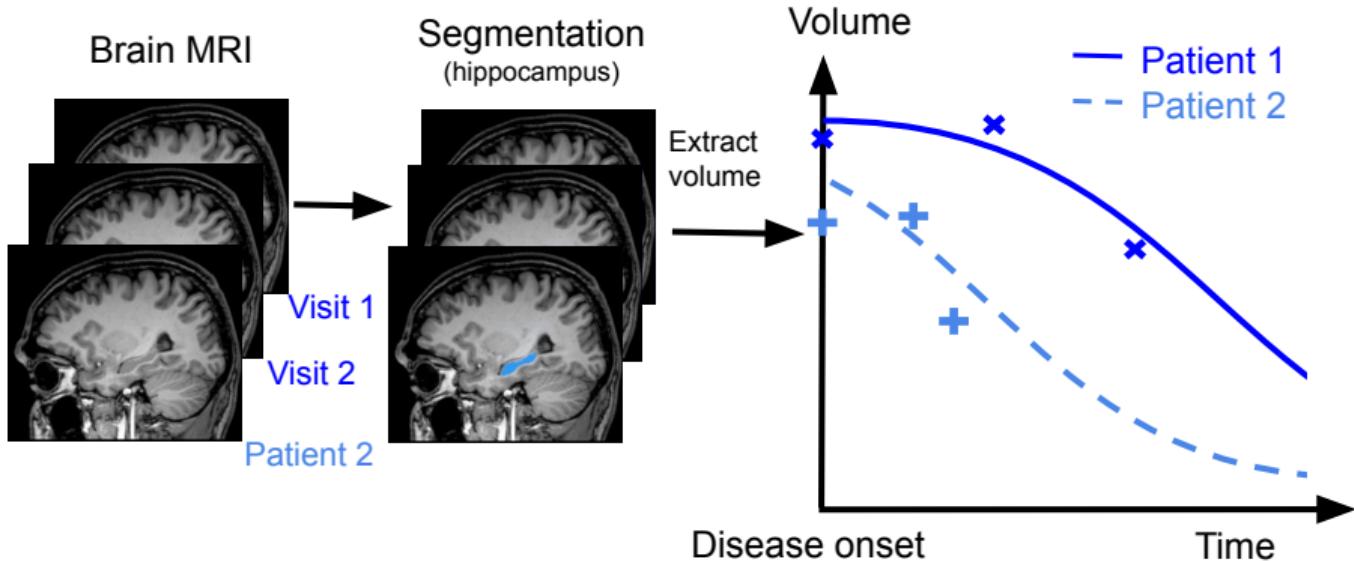
Building a Disease Progression Model



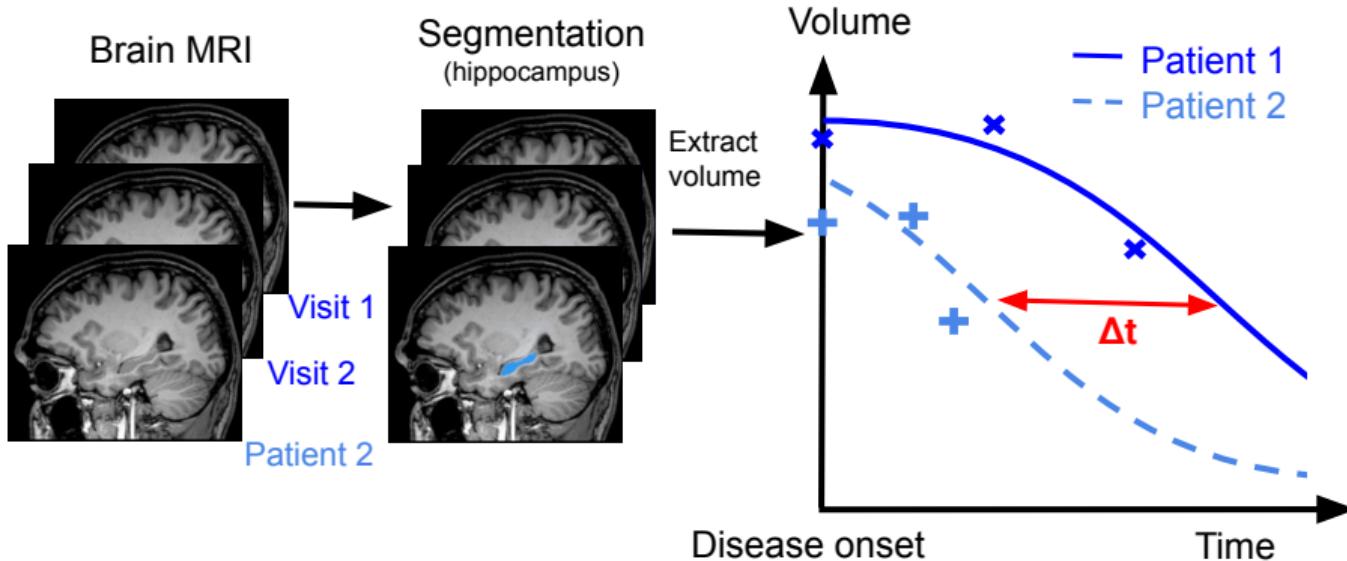
Building a Disease Progression Model



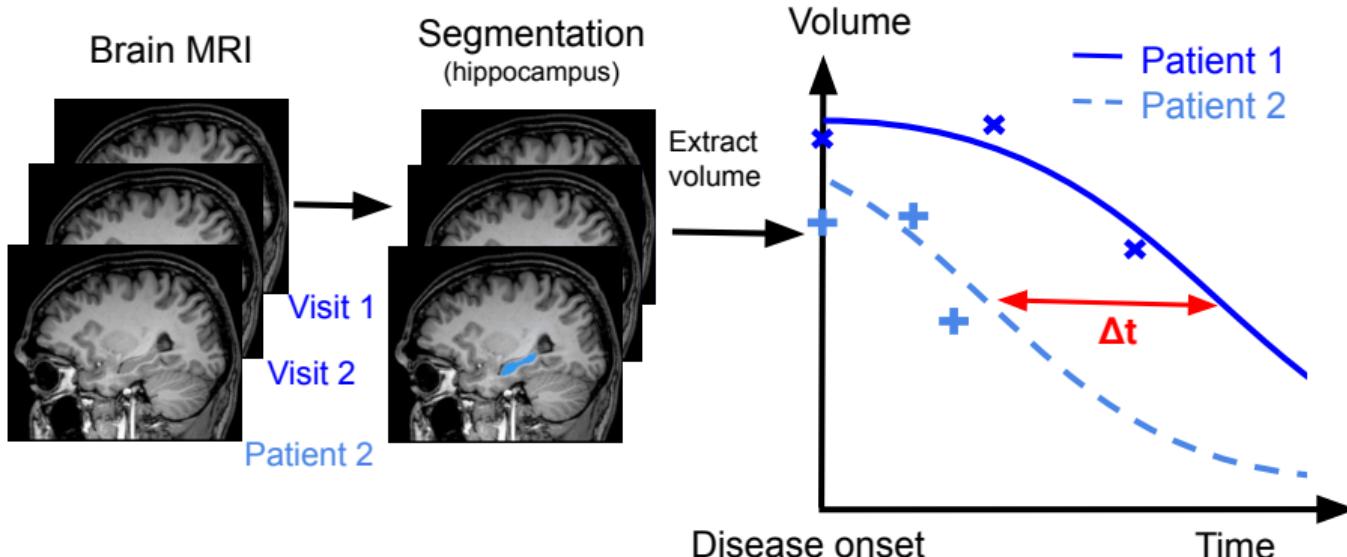
Building a Disease Progression Model



Building a Disease Progression Model

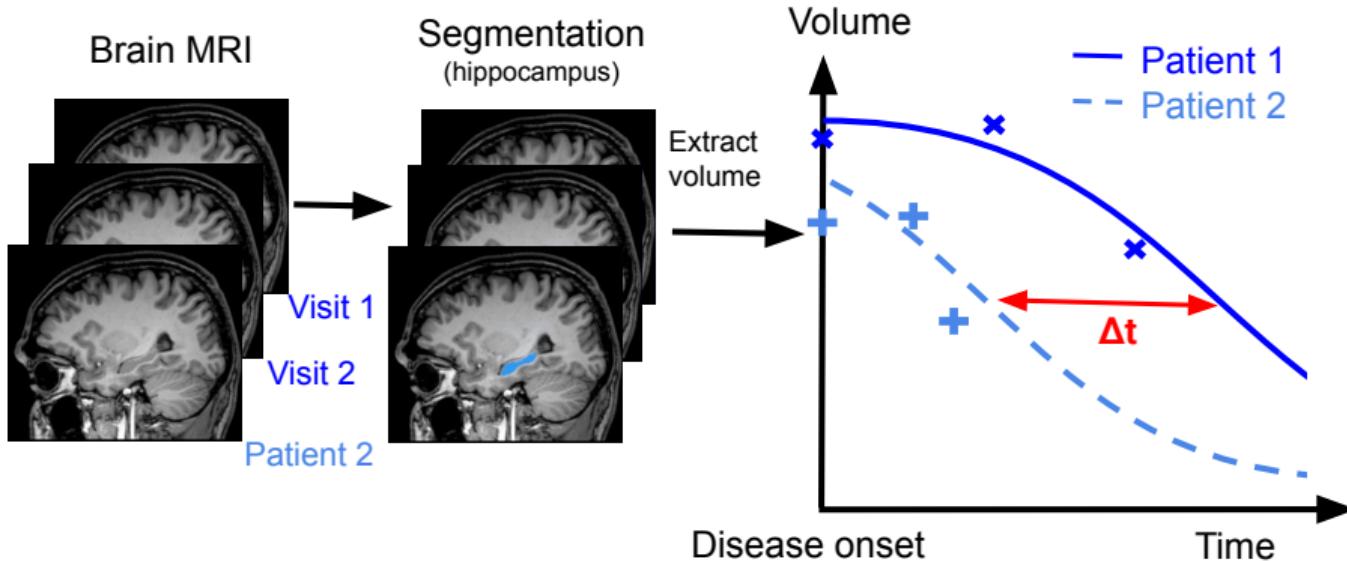


Building a Disease Progression Model



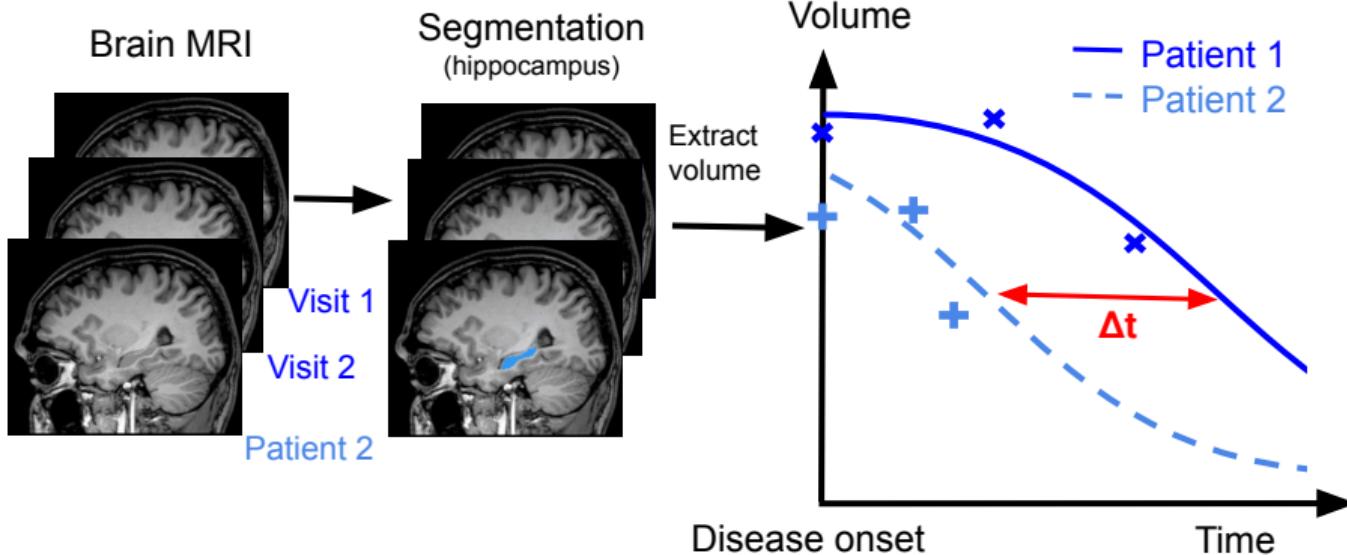
- ▶ Can now build population model

Building a Disease Progression Model



- ▶ Can now build population model
- ▶ Early diagnosis

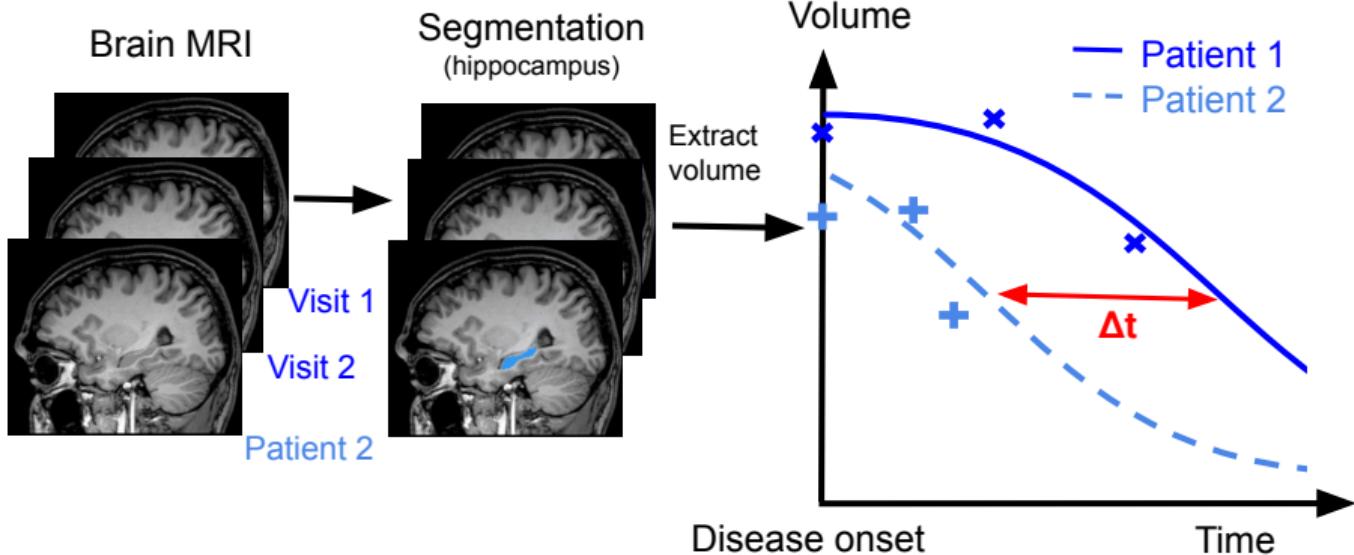
Building a Disease Progression Model



- Previous models:
- ▶ Jedynak, 2012
 - ▶ Fontejin, 2012
 - ▶ Donohue, 2014
 - ▶ Schiratti, 2017
 - ▶ Lorenzi, 2019

- ▶ Can now build population model
- ▶ Early diagnosis

Building a Disease Progression Model



Previous models:

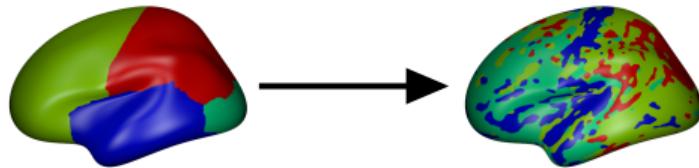
- ▶ Jedynak, 2012
- ▶ Fontejin, 2012
- ▶ Donohue, 2014
- ▶ Schiratti, 2017
- ▶ Lorenzi, 2019

Limitation: require brain segmentation a-priori

- ▶ Can now build population model
- ▶ Early diagnosis

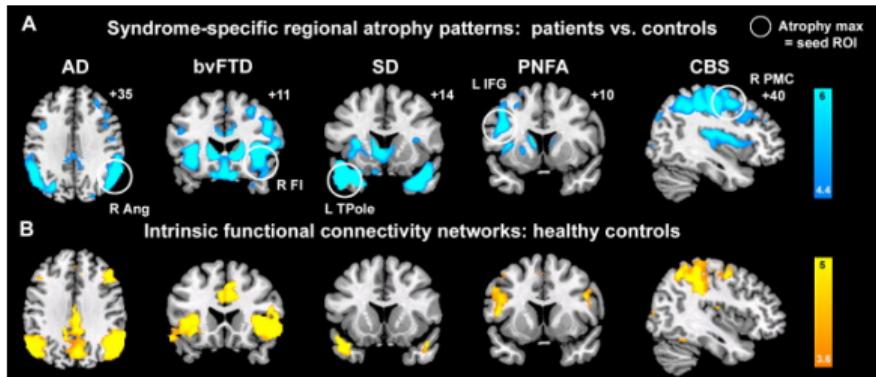
Aim: Build a disease progression model for voxelwise data

Aim: Move from segmentation-based analysis to voxelwise



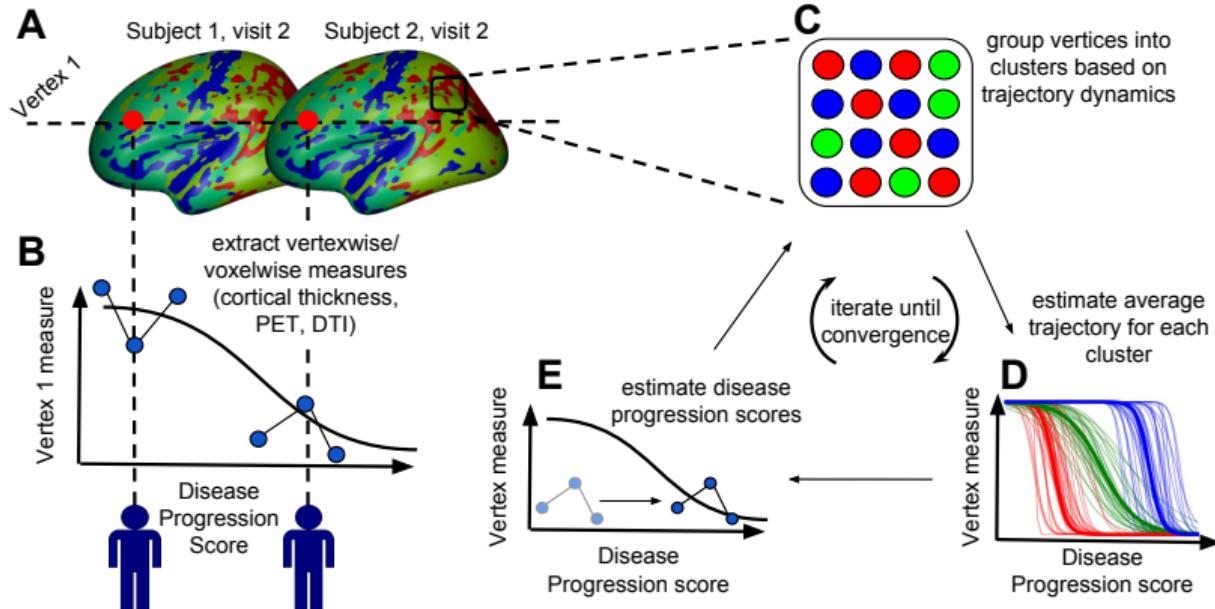
Why:

1. Atrophy correlates with functional networks, which are not spatially connected (Seeley et al., Neuron, 2009)
2. Better biomarker prediction and disease staging



Seeley et al., Neuron, 2009

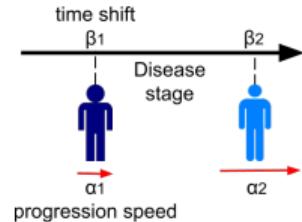
DIVE clusters vertices/voxels with similar trajectories of pathology



Building the model using a generative Bayesian framework

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

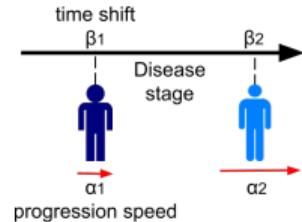
$$s_{ij} = \alpha_i t_{ij} + \beta_i$$



Building the model using a generative Bayesian framework

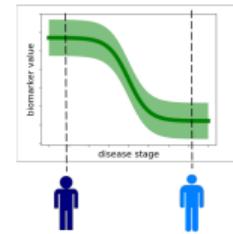
1. Model disease progression score for one subject i at visit j :

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



2. Model biomarker trajectory of one vertex (point) l on the brain:

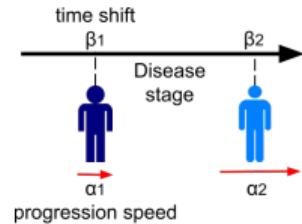
$$p(V_l^{ij} | \alpha_i, \beta_i, \theta_k, \sigma_k) \sim N(f(\alpha_i t_{ij} + \beta_i; \theta_k), \sigma_k)$$



Building the model using a generative Bayesian framework

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

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

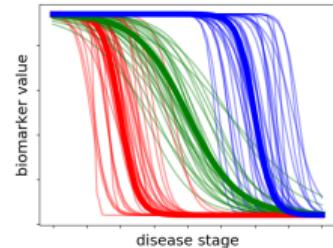
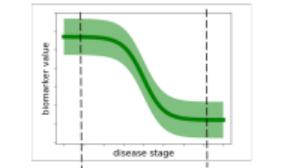


2. Model biomarker trajectory of one vertex (point) I on the brain:

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

3. Extend to all vertices and subjects:

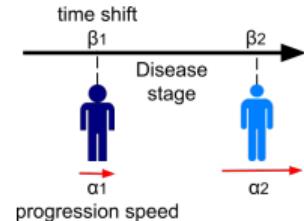
$$p(V, Z | \alpha, \beta, \theta, \sigma) = \prod_I^L \prod_{(i,j) \in I} N(V_I^{ij} | f(\alpha_i t_{ij} + \beta_i; \theta_Z), \sigma_{Z_I})$$



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 biomarker trajectory of one vertex (point) I on the brain:

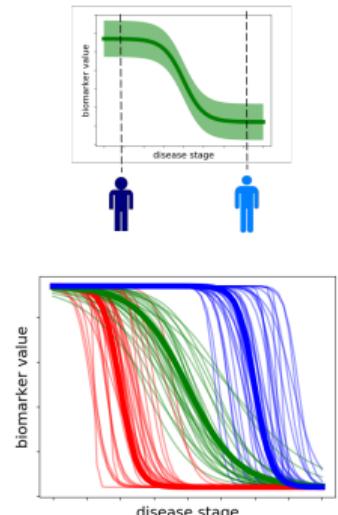
$$p(V_I^{ij} | \alpha_i, \beta_i, \theta_k, \sigma_k) \sim N(f(\alpha_i t_{ij} + \beta_i; \theta_k), \sigma_k)$$

3. Extend to all vertices and subjects:

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

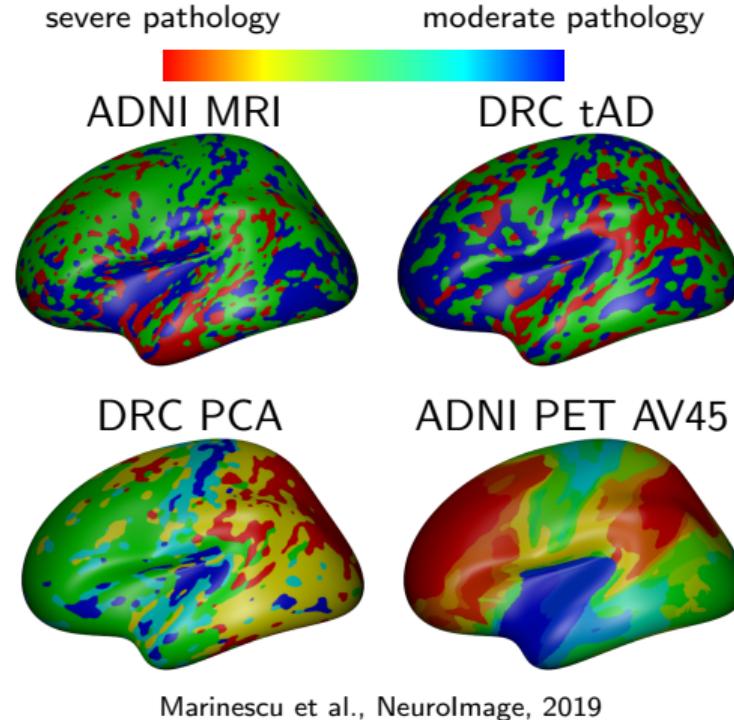
4. Marginalise over the hidden variables Z_I (cluster assignments):

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



DIVE Finds Plausible Atrophy Patterns on Four Datasets

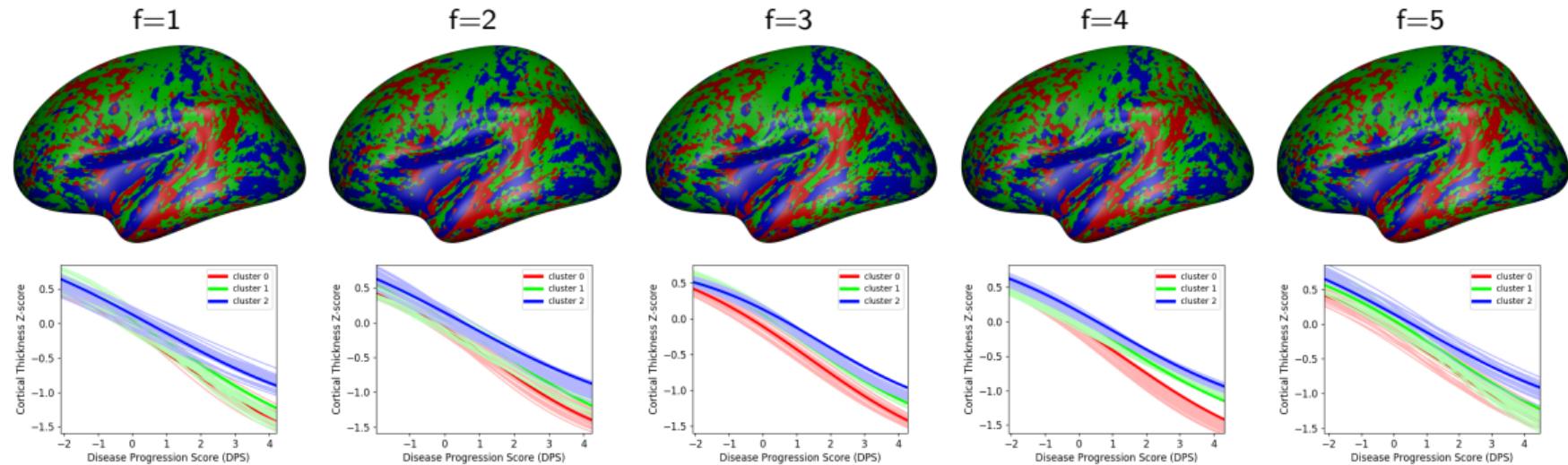
- ▶ Similar patterns of tAD atrophy in independent datasets: ADNI and UCL DRC
- ▶ Distinct patterns of atrophy in different diseases (tAD and PCA) and modalities (MRI vs PET)



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)

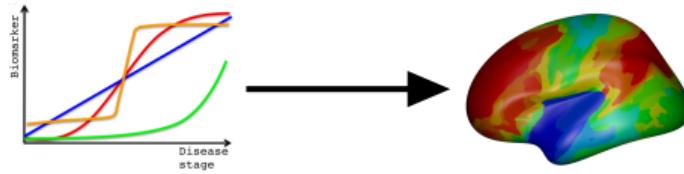


Marinescu et al., Neuroimage, 2019

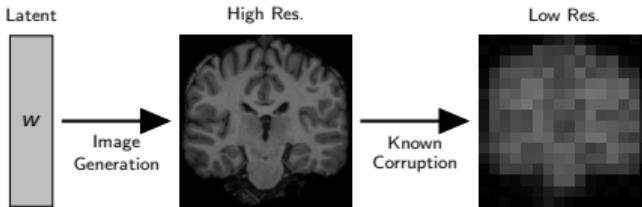
- ▶ We modelled the continuous progression of Alzheimer's disease and related dementias
- ▶ Used generative bayesian model that does not require labels (unsupervised)
- ▶ However, such models require good quality data, to perform segmentation and extract disease markers
- ▶ How can we do such modelling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

1. Disease progression modelling of Alzheimer's disease

1.1 Towards unsupervised clustering of biomarker trajectories



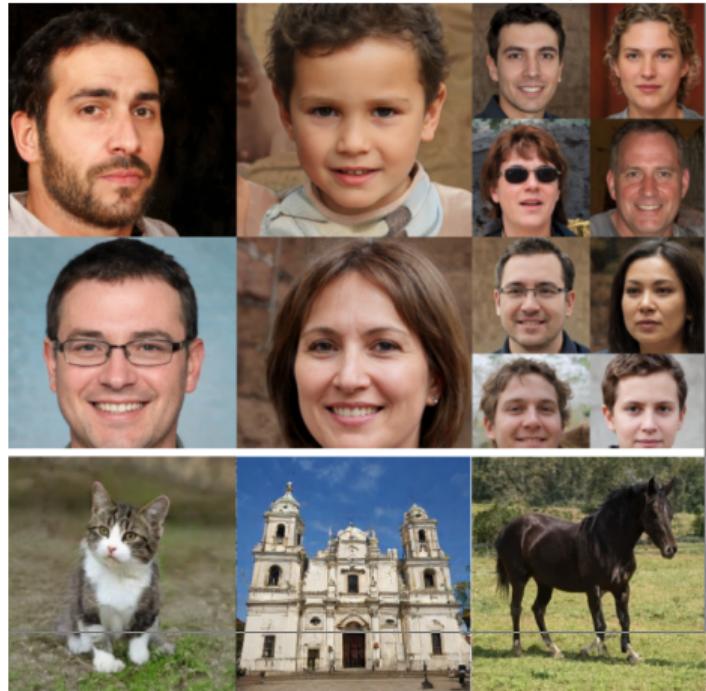
2. Image Reconstruction using Deep Generative Models



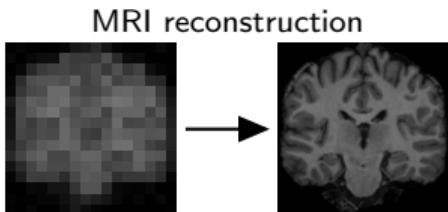
3. Future work

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

StyleGAN2 (Karras et al, 2019)

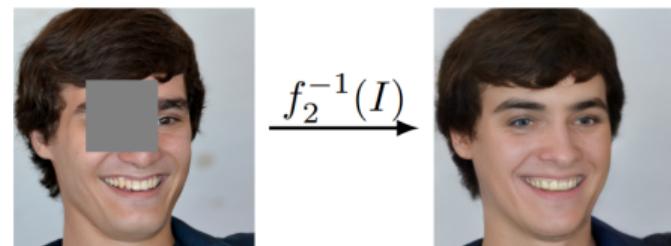
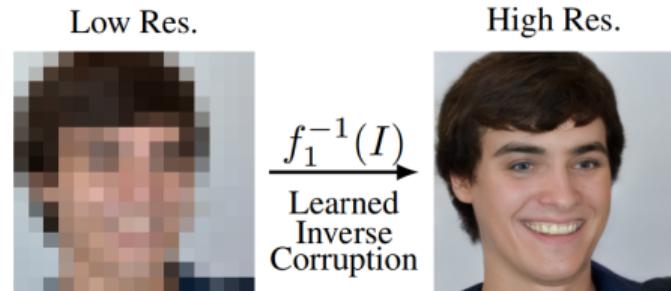


- Adapt the state-of-the-art StyleGAN2 for medical image reconstruction



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



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

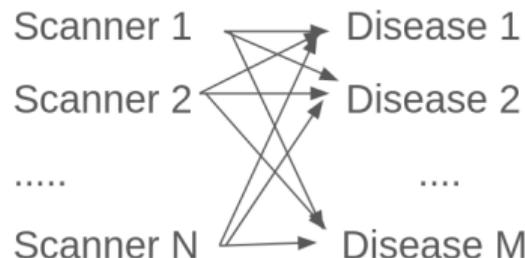


- Currently few labs/companies have the resources to train state-of-the-art models
 - StyleGAN2: 9 days on 4 GPUs
 - GPT-3: 355 years on single GPU
- Solutions moving forward:
 - Adapting previously-trained models
 - Combine smaller models into larger ones

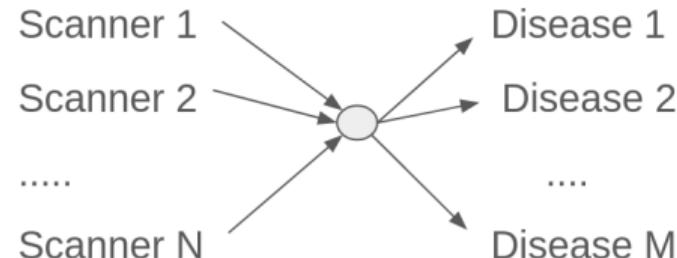
Limitation 2: Distribution shifts require model re-training

- ▶ 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**

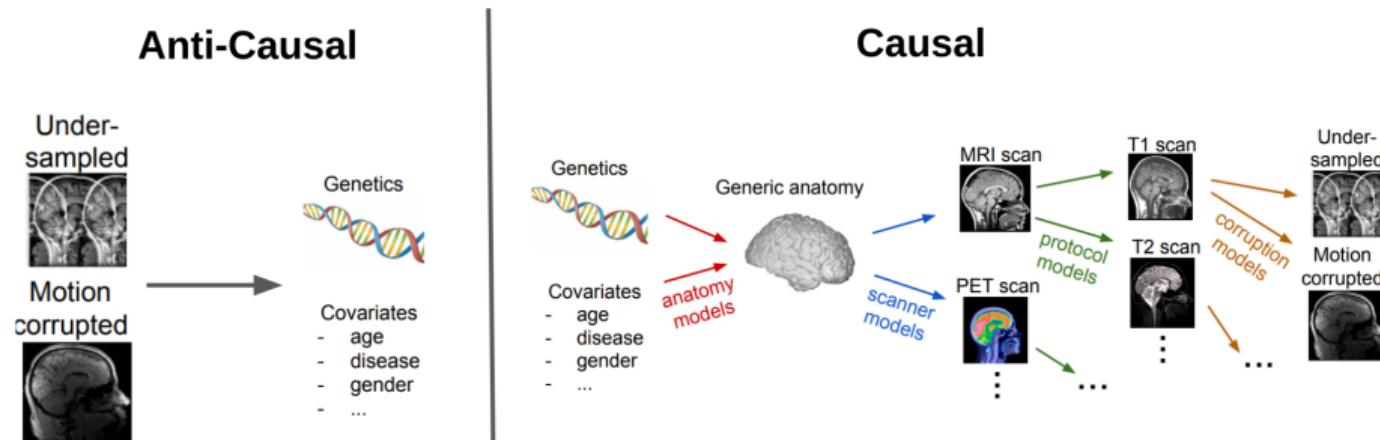


With compositionality: **N + M**



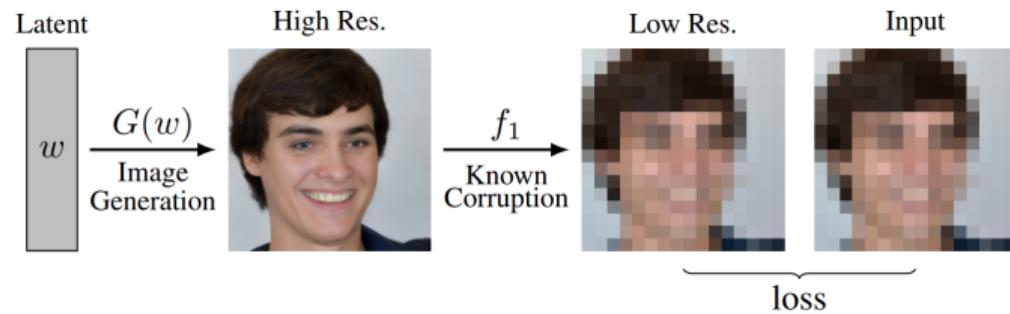
Limitation 3: Models are anti-causal

- ▶ 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



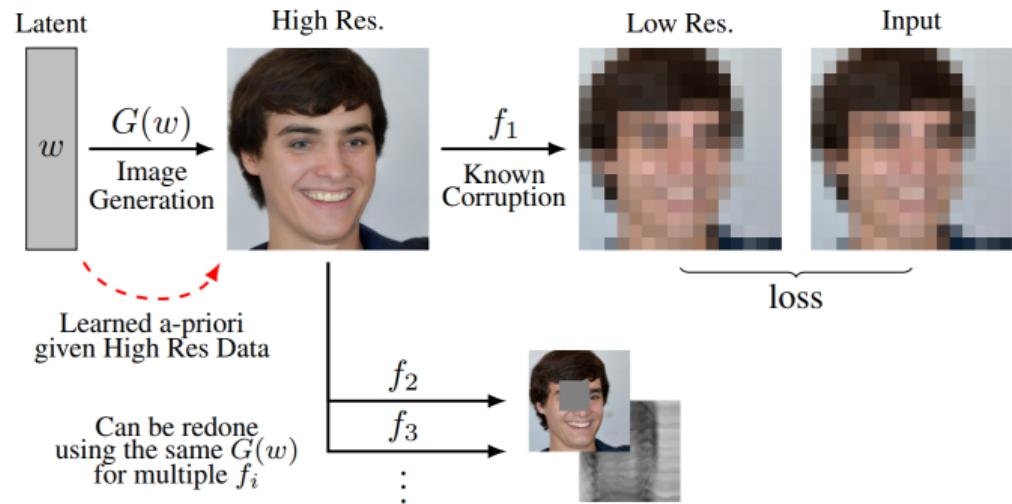
Method: We perform image reconstruction by combining two models

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



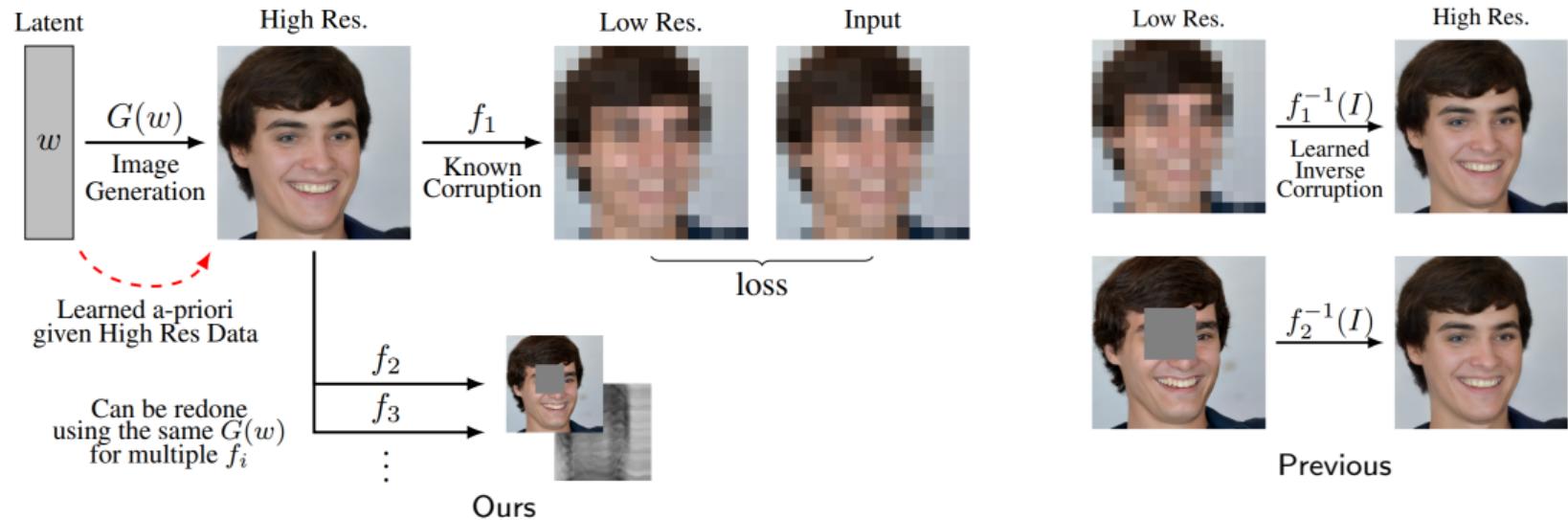
Method: We perform image reconstruction by combining two models

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

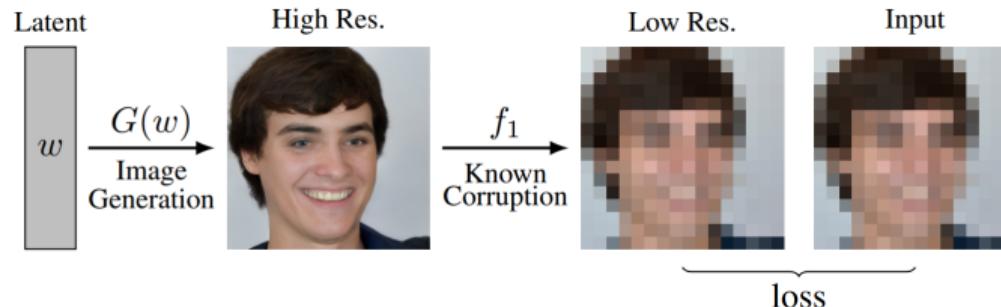
- We optimise:

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

- 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

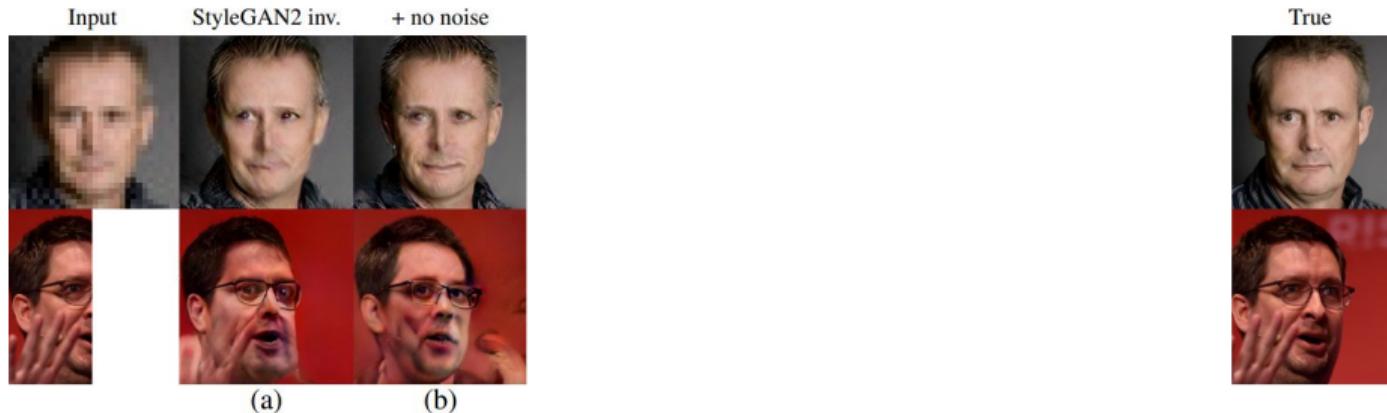
$$w^*, \eta^* = \arg \min_{w, \eta} \|\phi(I) - \phi \circ f \circ G(w, \eta)\|_2^2$$



Good reconstructions require further modifications

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

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

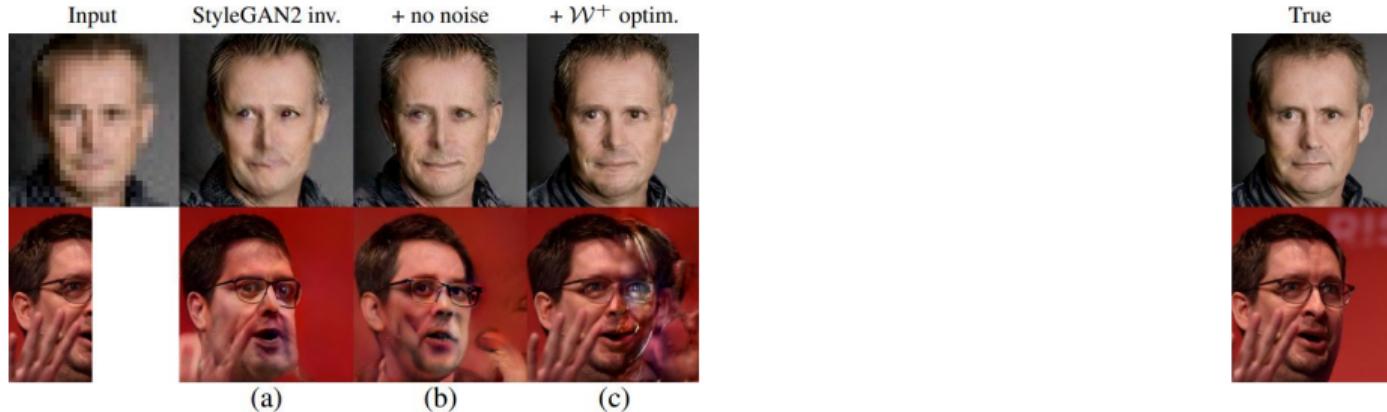


Good reconstructions require further modifications

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

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

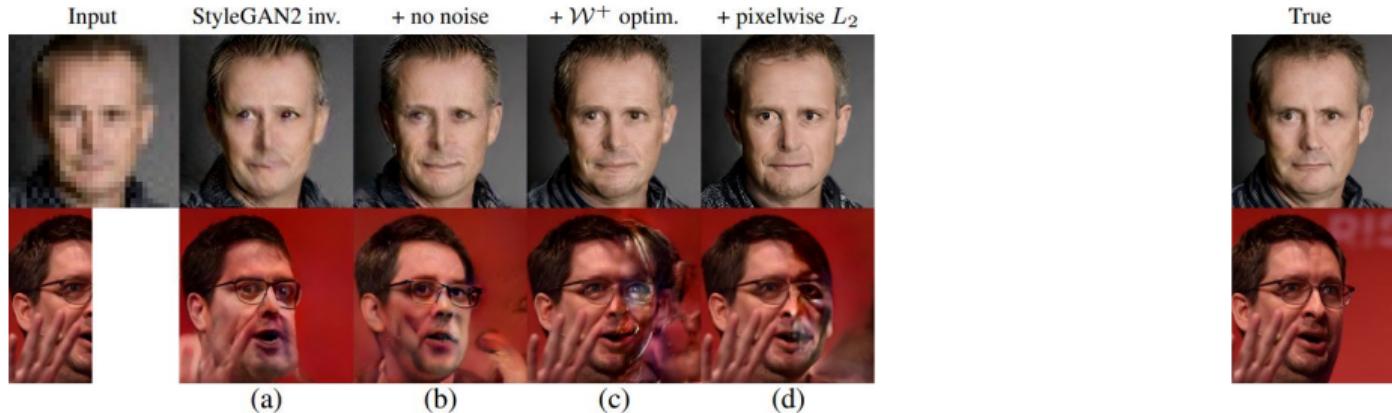
$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\phi(I) - \phi \circ f \circ G(\mathbf{w})\|_2^2$$



Good reconstructions require further modifications

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

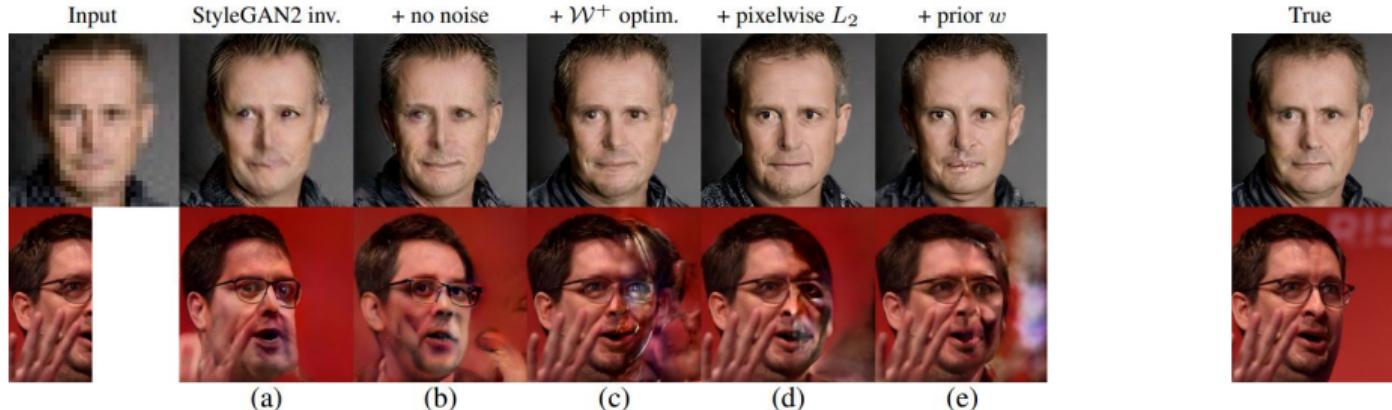
$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \|\phi(I) - \phi \circ f \circ G(\mathbf{w})\|_2^2 + \|I - f \circ G(\mathbf{w})\|_2^2$$



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

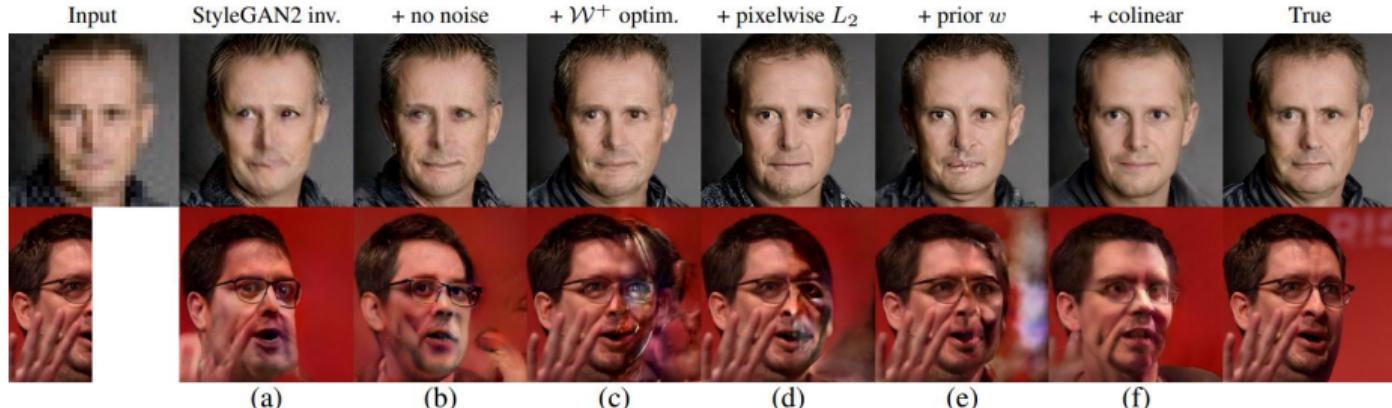
$$\begin{aligned}\mathbf{w}^* = \arg \min_{\mathbf{w}} & ||\phi(I) - \phi \circ f \circ G(\mathbf{w})||_2^2 + ||I - f \circ G(\mathbf{w})||_2^2 + \\ & + \sum_i \left(\frac{w_i - \mu}{\sigma_i} \right)^2\end{aligned}$$



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

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



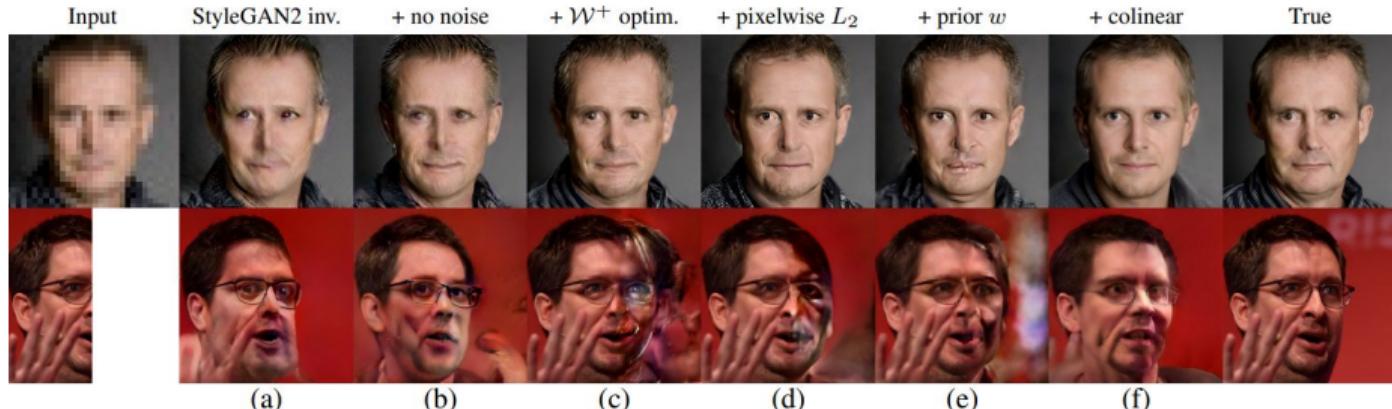
Good reconstructions require further modifications

- We started from the original StyleGAN2 inversion

- Yet the reconstruction was not good → required several changes

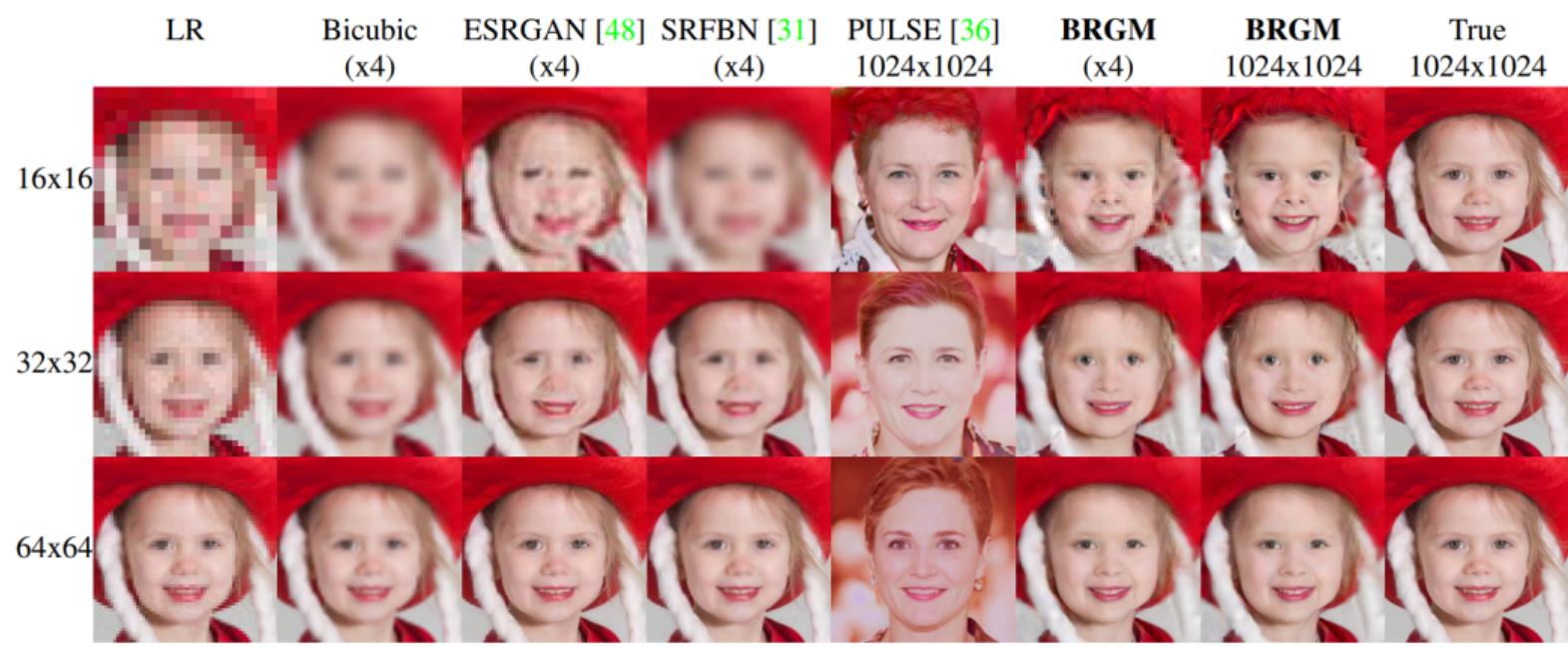
- Analytically expressed the full likelihood (Marinescu et al, 2021)

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



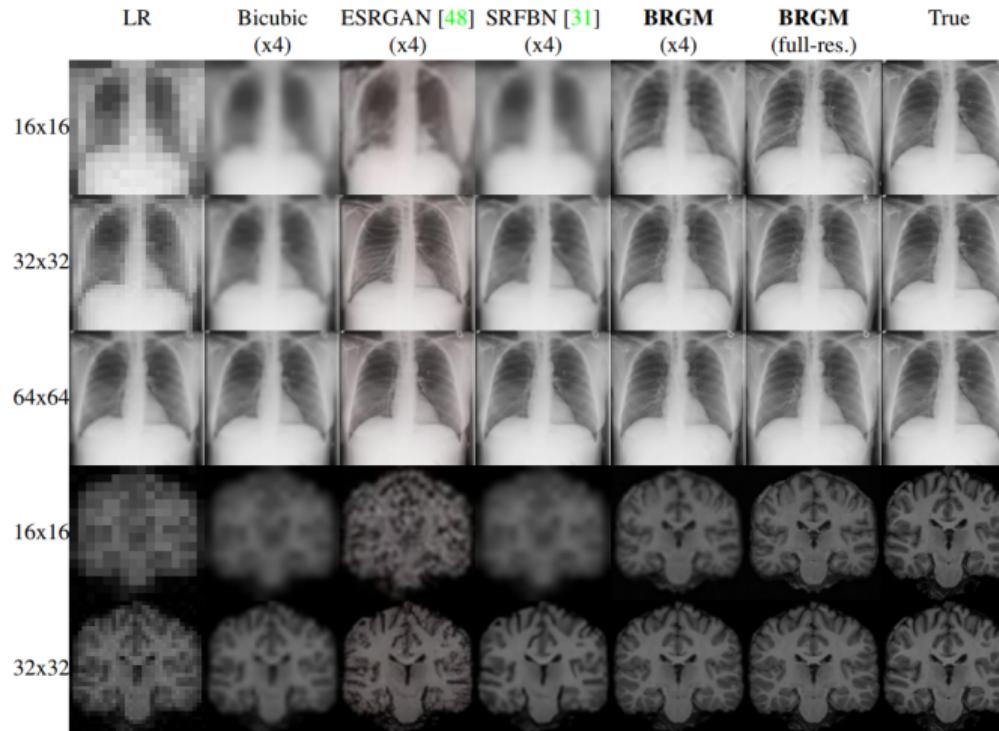
Results on super-resolution using the FFHQ dataset

- 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



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



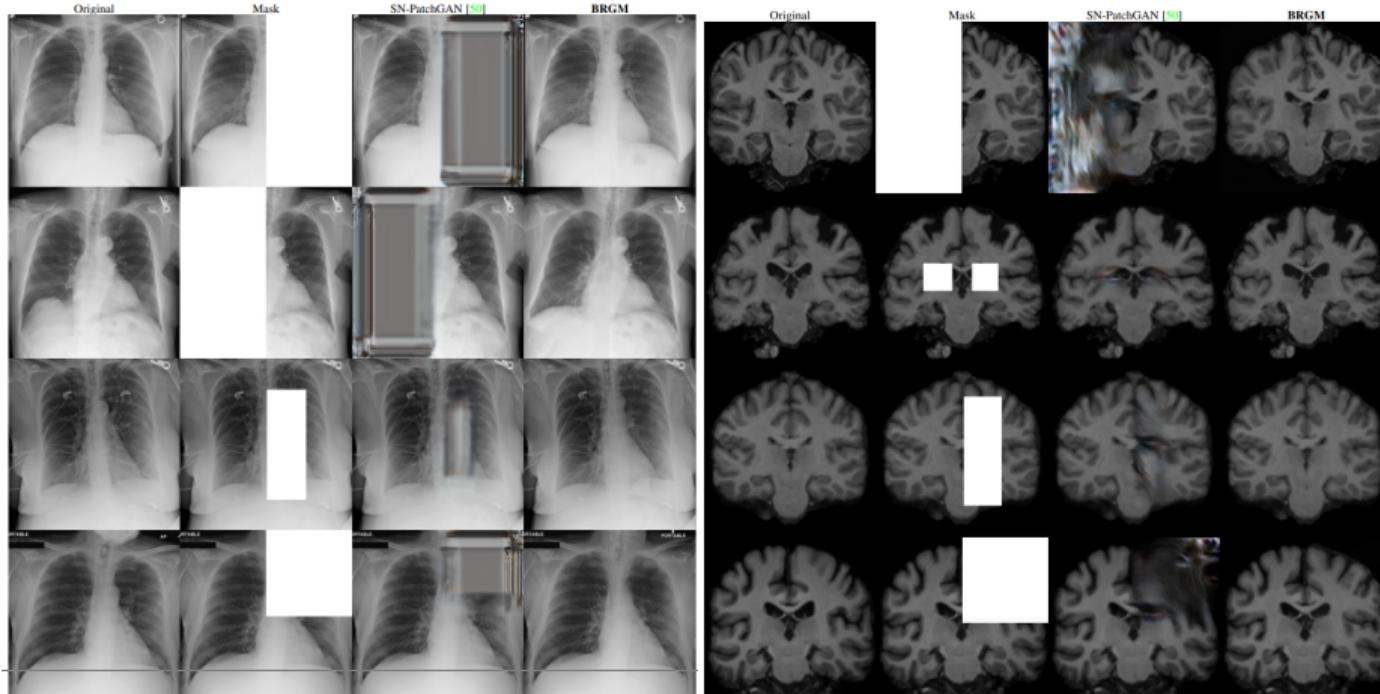
Inpainting also achieves state-of-the-art results

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

Inpainting

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

Human evaluation

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

Our method also has limitations

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



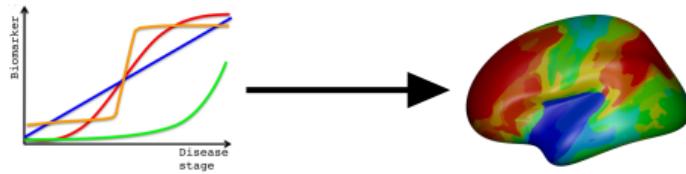
- ▶ Can be inconsistent with the input image



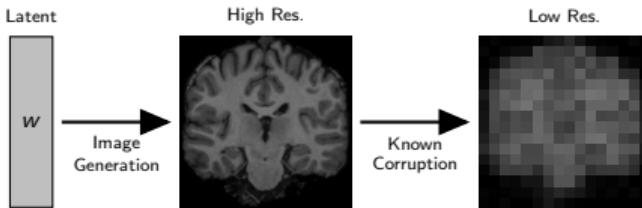
- ▶ Proposed a method for image reconstruction using pre-trained deep generative models
- ▶ Solution is given by the Bayesian MAP estimate
- ▶ State-of-the-art results on super-resolution and inpainting

1. Disease progression modelling of Alzheimer's disease

1.1 Towards unsupervised clustering of biomarker trajectories



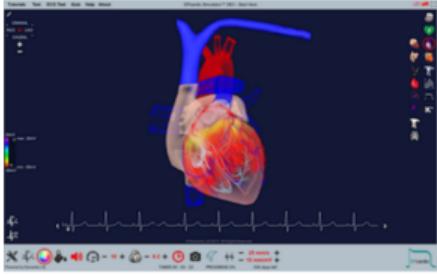
2. Image Reconstruction using Deep Generative Models



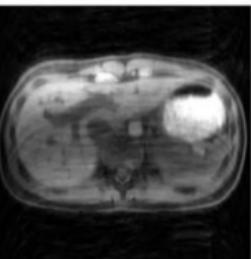
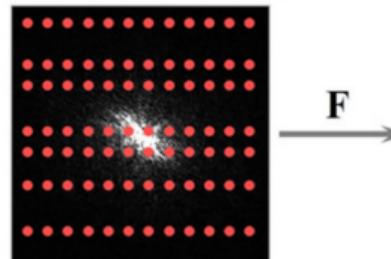
3. Future work

Future work

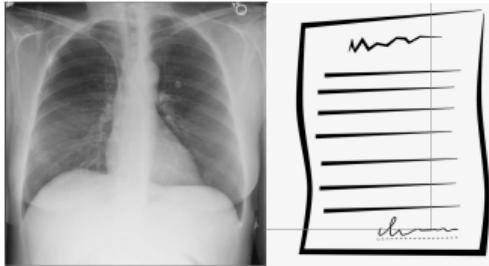
Biological simulators



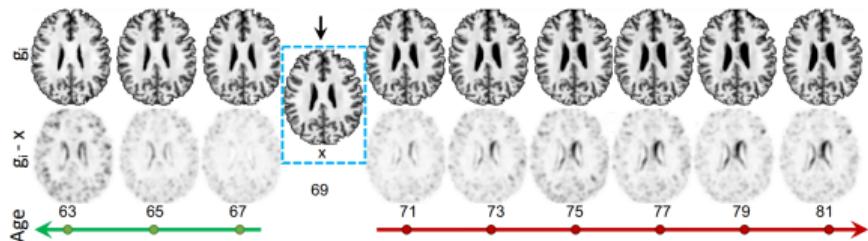
Better and faster reconstruction of medical images Undersampled k-space



Multimodal modelling images + text + structural data



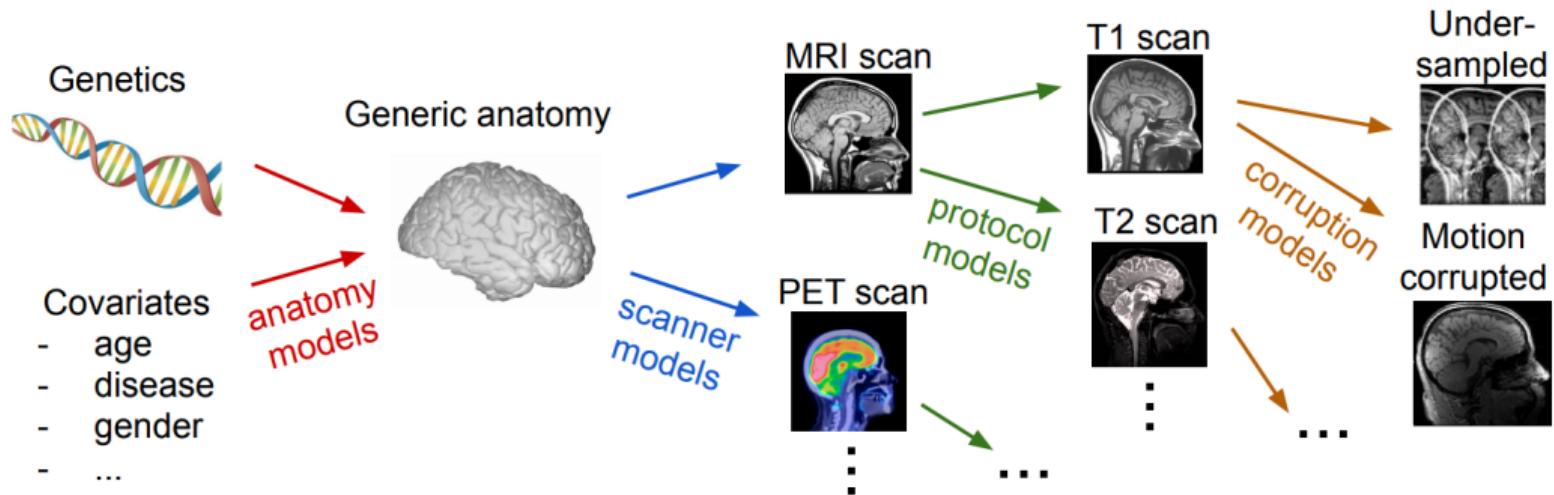
Disease Progression Modelling



Future work: Brain tissue and anatomy simulator

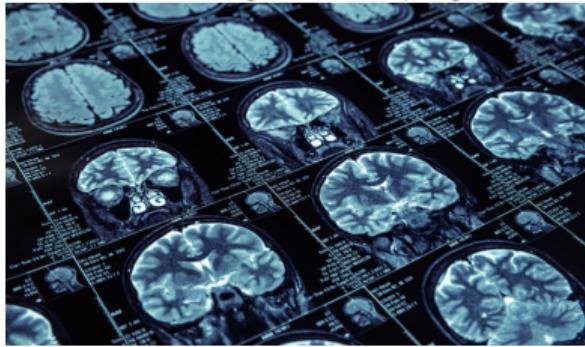
Simulator for brain anatomy from genetics:

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



Long-term vision: AI to make accurate diagnoses and predictions

Accurate diagnosis and prognosis



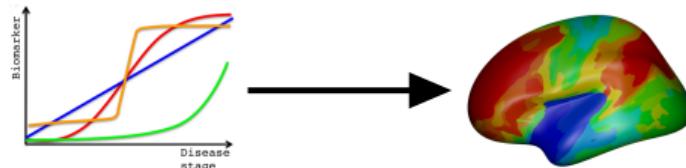
Augment humans



Key problems in Machine Learning in Medicine

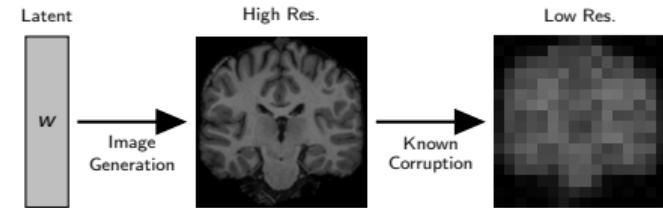
Lack of good labels

Solution: Unsupervised Learning through Disease Progression Modelling

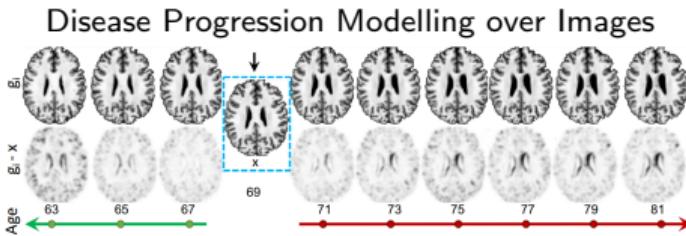


Lack of good input data

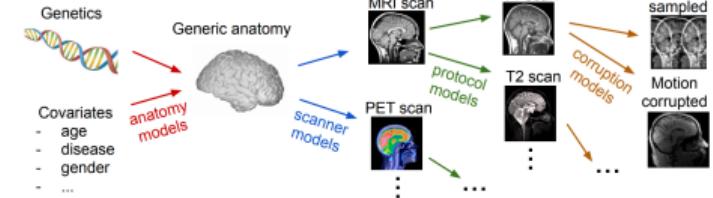
Solution: Image Reconstruction using Deep Generative Models



Future work



Brain Anatomy Simulator



Step 5: Modelling Spatial Correlation using Markov Random Fields

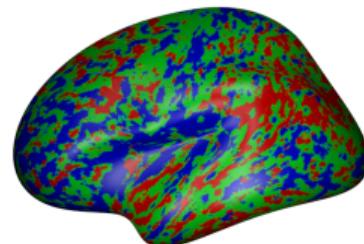
Motivation

- ▶ measurements from neighbouring vertices are inherently correlated
- ▶ can "fill-in holes", eliminate noisy cluster assignments due to noise

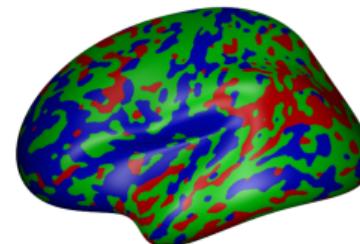
$$p(V, Z | \alpha, \beta, \theta, \sigma) = \prod_{l=1}^L \prod_{(i,j) \in I_l} N(V_l^{ij} | f(\alpha_i t_{ij} + \beta_i | \theta_{Z_l}), \sigma_{Z_l}) \prod_{l_1 \sim l_2} \Psi(Z_{l_1}, Z_{l_2})$$

where

- ▶ $\Psi(Z_{l_1} = k_1, Z_{l_2} = k_2) = \begin{cases} \exp(\lambda) & \text{if } k_1 = k_2 \\ \exp(-\lambda) & \text{otherwise} \end{cases}$
- ▶ λ - MRF parameter



(a) Without MRF



(b) With MRF, $\alpha = 5$.

Model Fitting with Expectation-Maximisation (EM)

► E-step:

- ▶ Estimate vertex assignment to clusters $z_{lk}^{(u)} = \zeta_{lk}(\lambda^{(u)})$:

$$\lambda^{(u)} = \arg \max_{\lambda} \sum_{l=1}^L \sum_{k=1}^K \zeta_{lk}(\lambda) \left[D_{lk} + \lambda \sum_{l_2 \in N_l} \zeta_{l_2 k}(\lambda) - \lambda^2 \sum_{l_2 \in N_l} (1 - \zeta_{l_2 k}(\lambda)) \right]$$
$$\zeta_{lk}(\lambda) \approx \exp \left(D_{lk} + \sum_{l_2 \in N_l} \log \left[\exp(-\lambda^2) + z_{l_2 k}^{(u-1)} (\exp(\lambda) - \exp(-\lambda^2)) \right] \right)$$

where:

$$D_{lk} = -\frac{1}{2} \log (2\pi (\sigma_k^{(u)})^2 |I|) - \frac{1}{2(\sigma_k^{(u)})^2} \sum_{i,j \in I} (V_i^{ij} - f(\alpha_i^{(u)} t_{ij} + \beta_i^{(u)} | \theta_k^{(u)}))^2$$

► M-step:

- ▶ Update trajectories:

$$\theta_k = \arg \min_{\theta_k} \left[\sum_{l=1}^L z_{lk} \sum_{(i,j) \in I} (V_i^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 \right] - \log p(\theta_k) \quad (1)$$

- ▶ Update subject progression scores:

$$\alpha_i, \beta_i = \arg \min_{\alpha_i, \beta_i} \left[\sum_{l=1}^L \sum_{k=1}^K z_{lk} \frac{1}{2\sigma_k^2} \sum_{j \in I_l} (V_i^{ij} - f(\alpha_i t_{ij} + \beta_i | \theta_k))^2 \right] - \log p(\alpha_i, \beta_i) \quad (2)$$