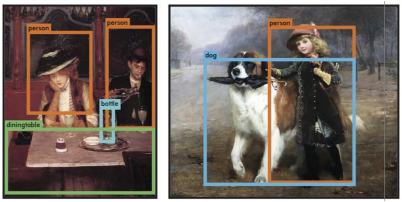
Medical Image Generation and Analysis using Bayesian Generative Models

Răzvan V. Marinescu

Massachusetts Institute of Technology

Machine Learning algorithms have achieved impressive milestones

Object detection (YOLO)



Text-to-Image Generation (DALL-E)

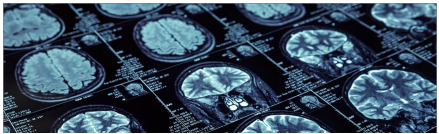
prompt: "an armchair in the shape of an avocado"

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



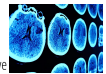
Top 12 Ways Artificial Intelligence Will Impact Healthcare

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

Augment doctors

How Artificial Intelligence Improves Medical Imaging in Hospitals

Deep learning software, such as artificial intelligence, can improve



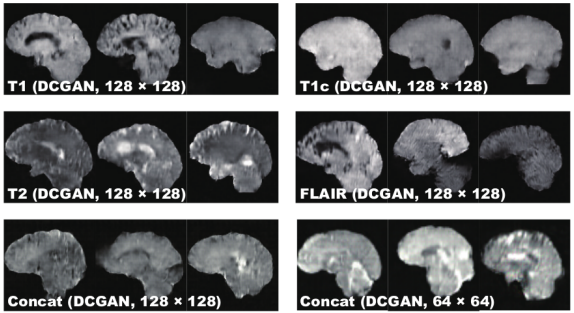
Prediction of clinical variables not always working

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 generate forecasts that were significantly better than random guessing

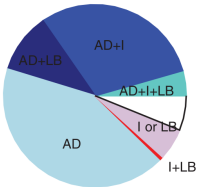
Generated images are crude, not high-resolution, mostly 2D

Brain MRI generation (Han, 2018)



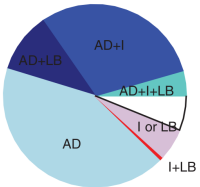
Lack of good labels

- Alzheimer's diagnosis accuracy just 42%

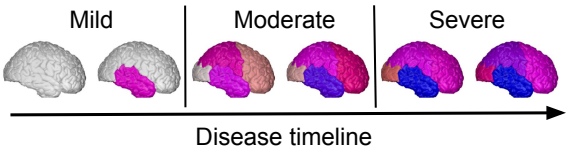


Lack of good labels

- Alzheimer's diagnosis accuracy just 42%

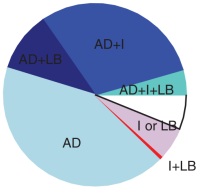


- Labels are categorical instead of continuous

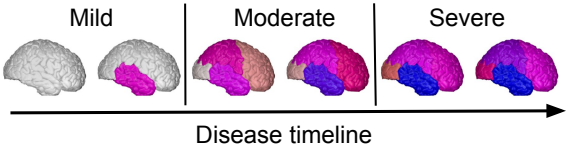


Lack of good labels

- Alzheimer's diagnosis accuracy just 42%

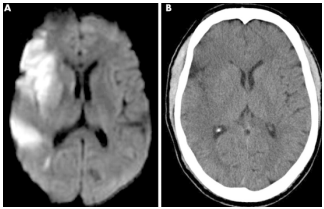


- Labels are categorical instead of continuous



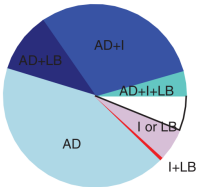
Lack of good input data/signal

- Limited contrast

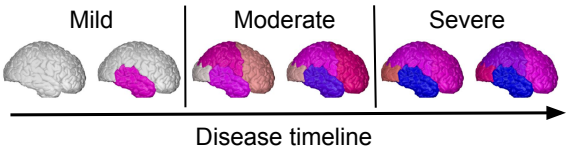


Lack of good labels

- Alzheimer's diagnosis accuracy just 42%

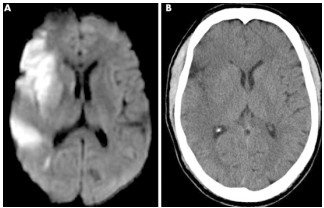


- Labels are categorical instead of continuous

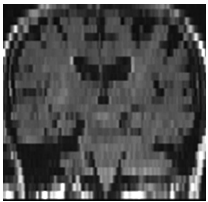


Lack of good input data/signal

- Limited contrast

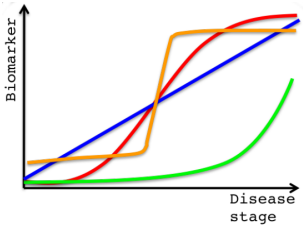


- Low-resolution



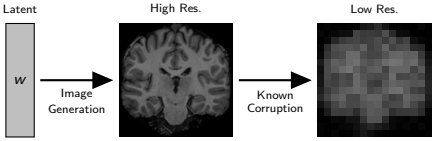
Lack of good labels

Solution: Unsupervised Learning of Continuous Dynamics
= Disease Progression Modelling



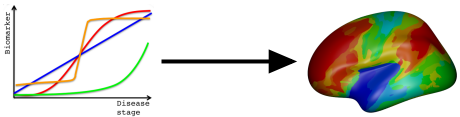
Lack of good input data/signal

Solution: Image Reconstruction
using Deep Generative Models

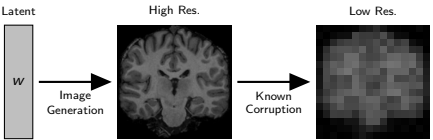


Outline

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



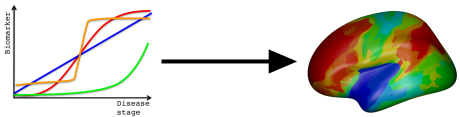
2. Image Reconstruction using Deep Generative Models



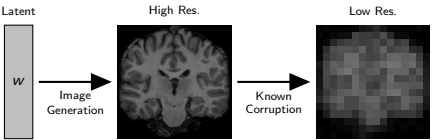
3. Future work towards brain anatomy simulators

Outline

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



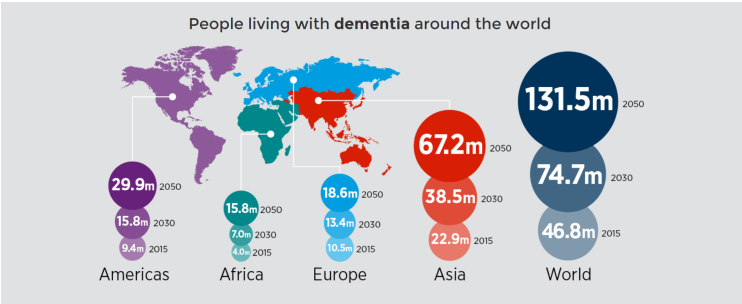
2. Image Reconstruction using Deep Generative Models



3. Future work towards brain anatomy simulators

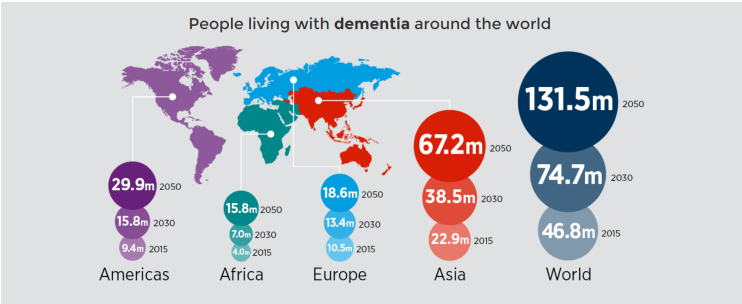
Alzheimer's Disease is a Devastating Disease

- 46 million people affected worldwide



Alzheimer's Disease is a Devastating Disease

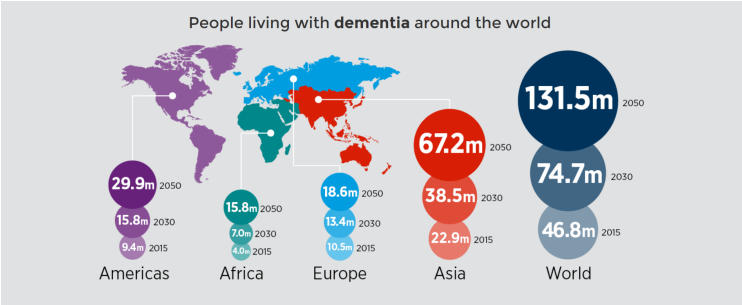
- 46 million people affected worldwide



- 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

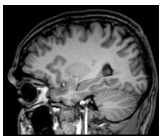
- 46 million people affected worldwide



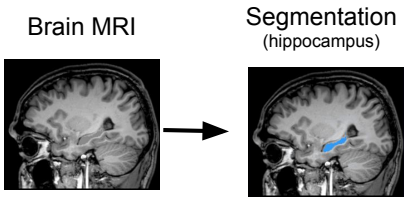
- No treatments available that stop or slow down cognitive decline
- Q: Why did clinical trials fail? A: Treatments were not administered early enough

- Q: How can we then identify subjects **early** in order to administer treatments?
- A: Disease progression model ...

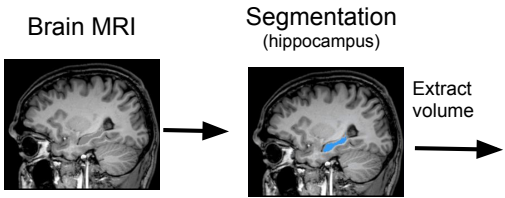
Brain MRI

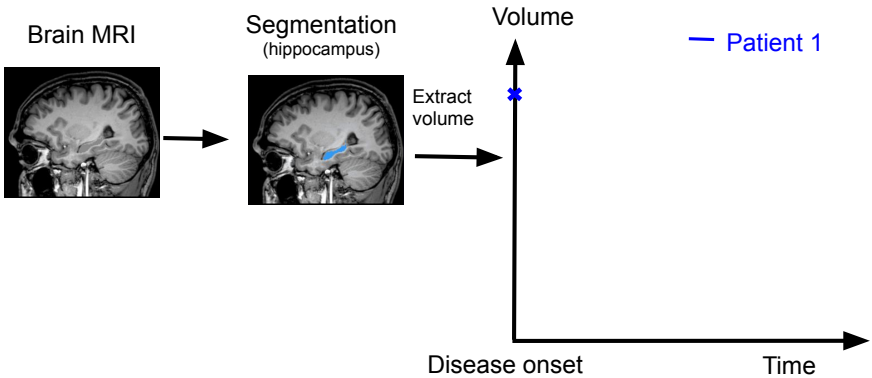


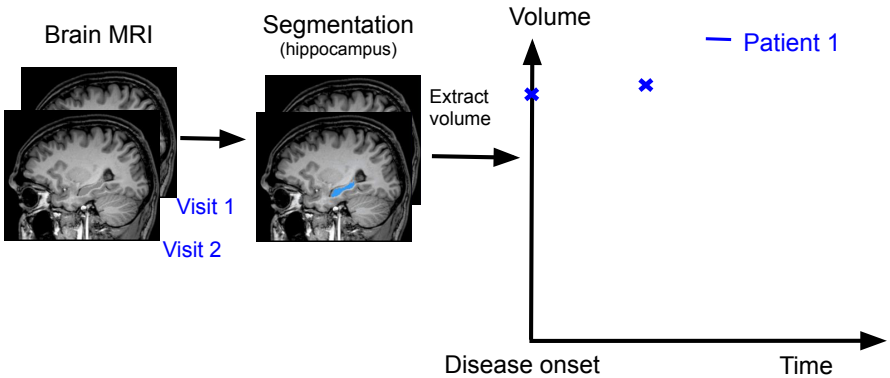
Building a Disease Progression Model

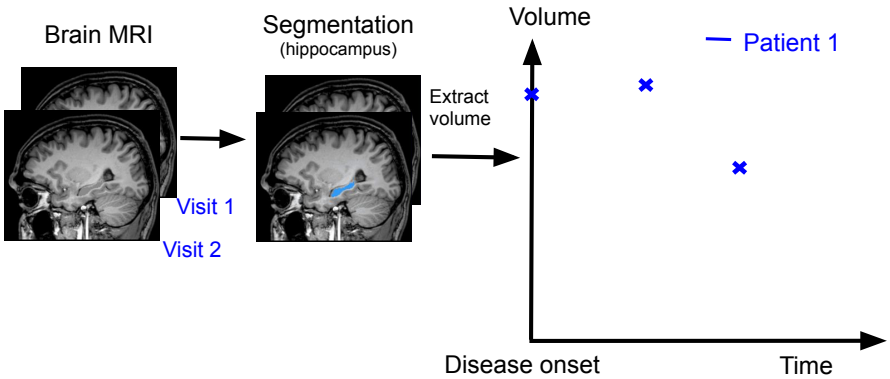


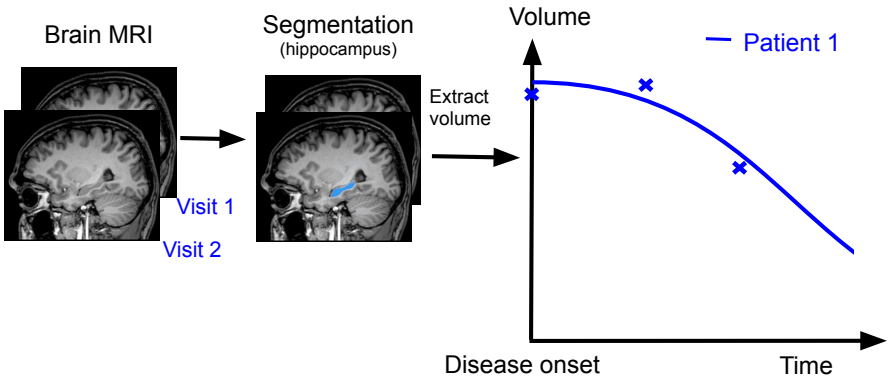
Building a Disease Progression Model



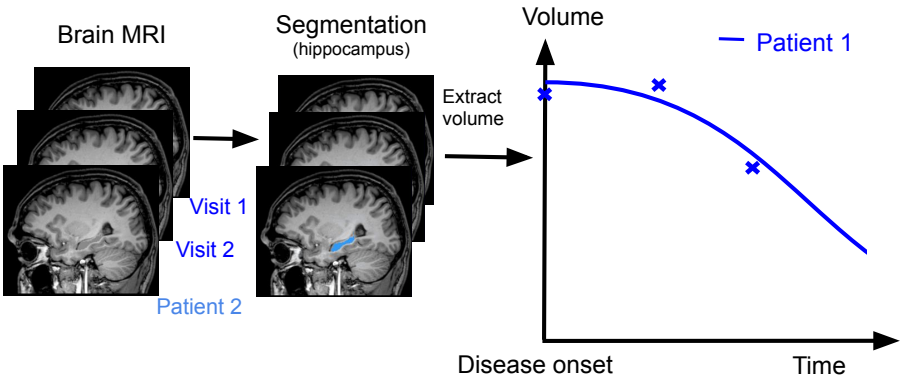

Brain MRI
Segmentation
(hippocampus)
Extract
volume
Volume
Patient 1
Disease onset
Time


Brain MRI
Segmentation
(hippocampus)
Visit 1
Visit 2
Extract
volume
Volume
Patient 1
Disease onset
Time

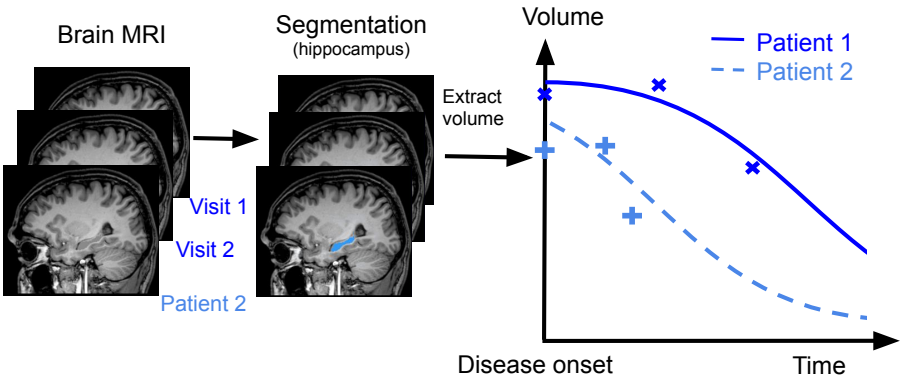

Brain MRI
Segmentation
(hippocampus)
Visit 1
Visit 2
Extract
volume
Volume
Patient 1
Disease onset
Time

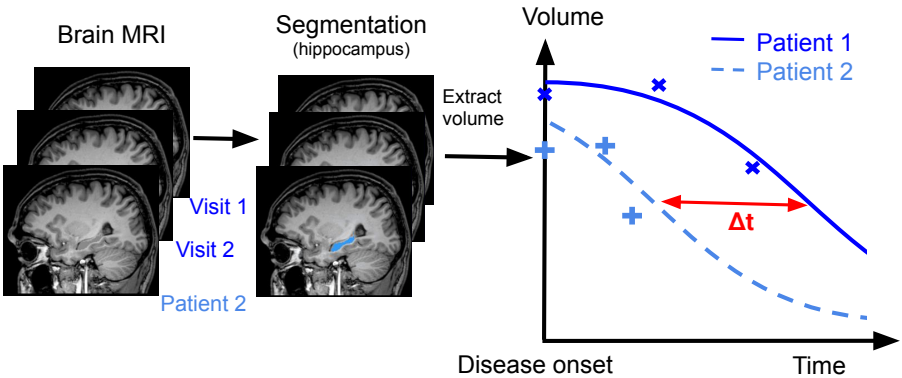

Brain MRI
Segmentation
(hippocampus)
Visit 1
Visit 2
Extract
volume
Volume
Patient 1
Disease onset
Time

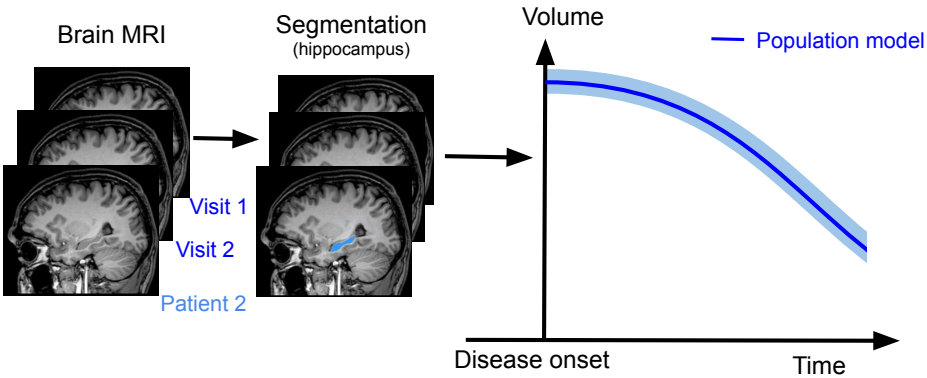
Building a Disease Progression Model



Building a Disease Progression Model

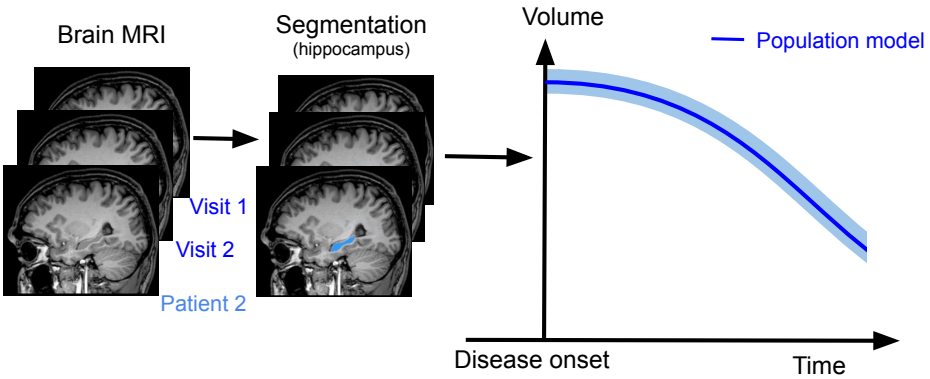



Brain MRI
Segmentation
(hippocampus)
Visit 1
Visit 2
Patient 2
Extract
volume
Volume
Patient 1
Patient 2
Δt
Disease onset
Time



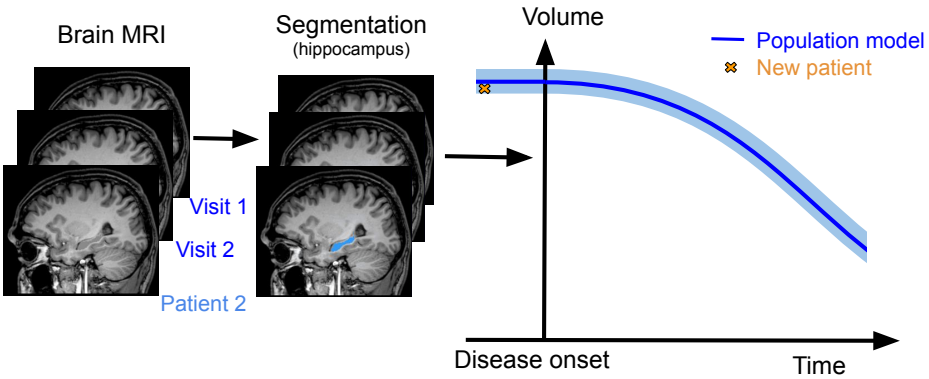
- Can now build population model

Building a Disease Progression Model



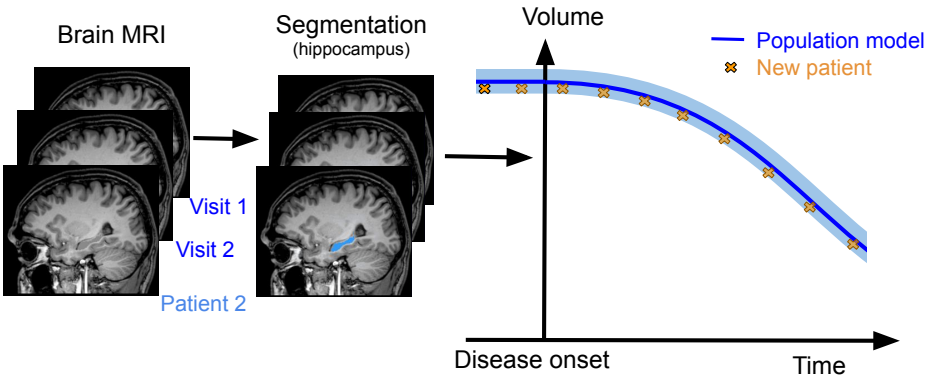
- Can now build population model
- Early diagnosis

Building a Disease Progression Model



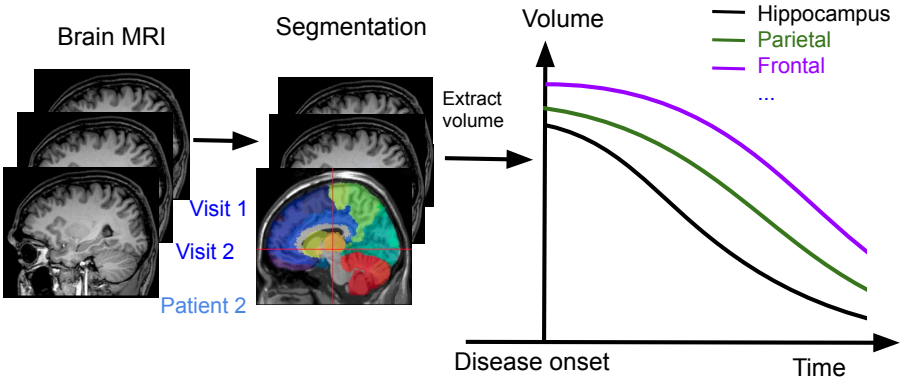
- Can now build population model
- Early diagnosis

Building a Disease Progression Model



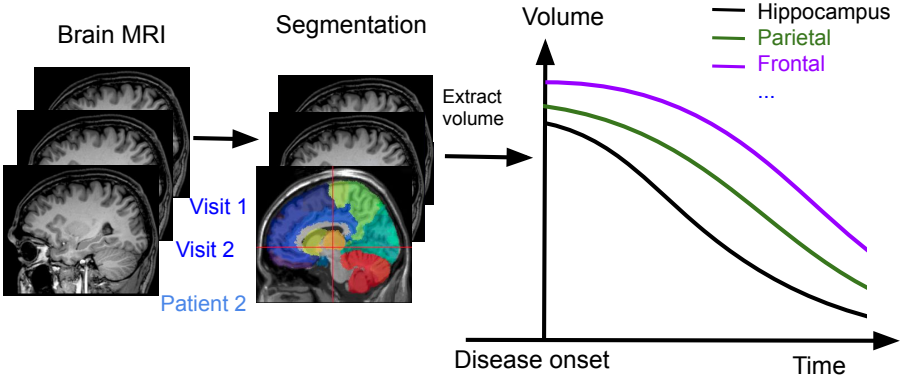
- Can now build population model
- Early diagnosis

Building a Disease Progression Model



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

Building a Disease Progression Model

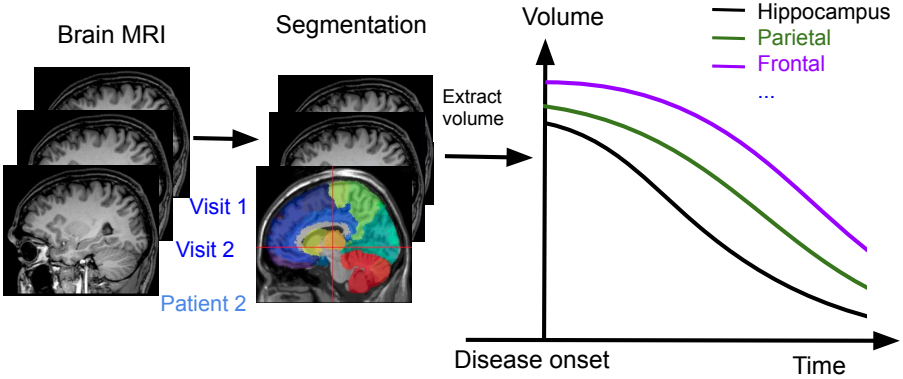


Previous models:

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

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

Building a Disease Progression Model



Previous models:

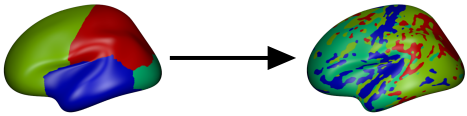
- Jedynak, 2012
- Fontejin, 2012
- Donohue, 2014
- Schiratti, 2017
- Lorenzi, 2019

Limitation: require brain segmentation a-priori

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

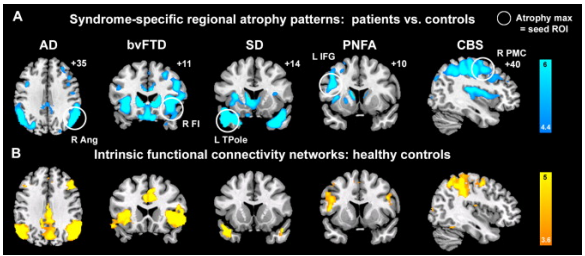
Aim: Move from segmentation-based analysis to vertexwise

- vertex = point on the brain surface



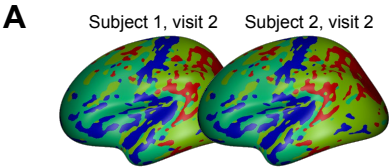
Why:

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

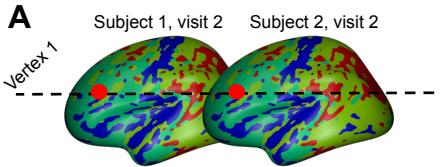


Seeley et al., Neuron, 2009

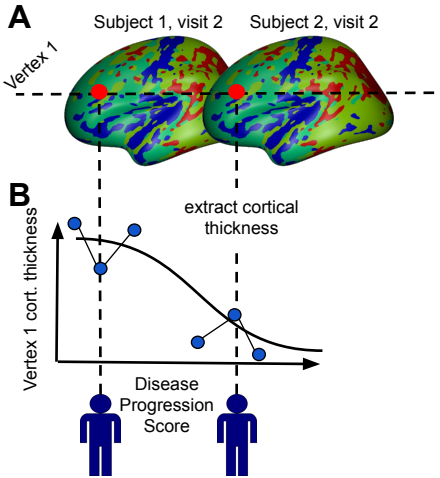
Our model clusters vertices with similar trajectories of pathology



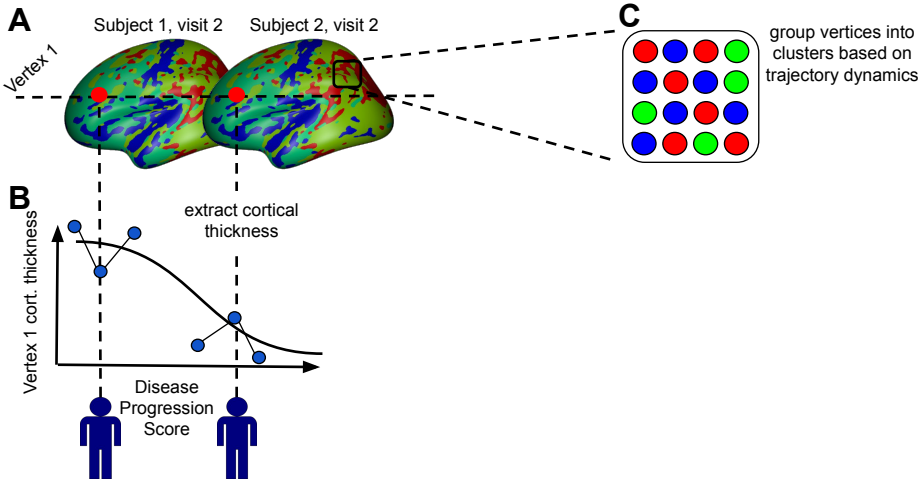
Our model clusters vertices with similar trajectories of pathology



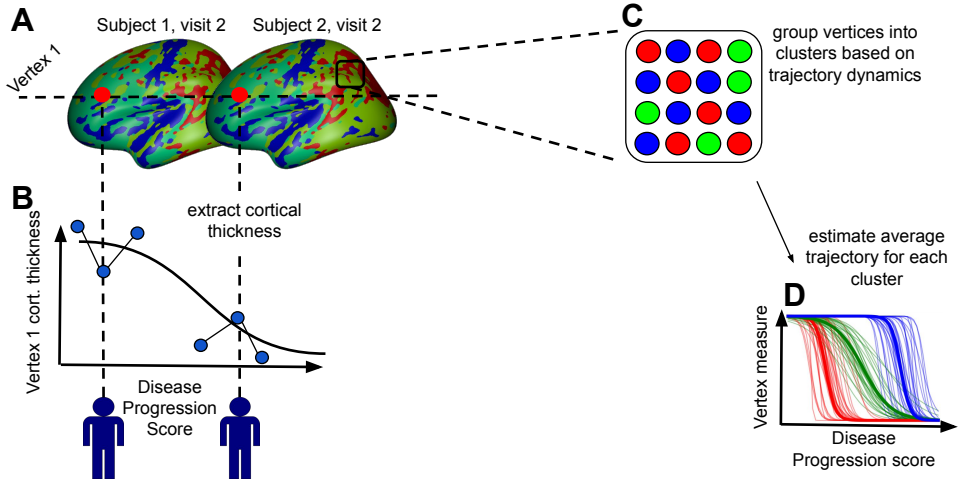
Our model clusters vertices with similar trajectories of pathology



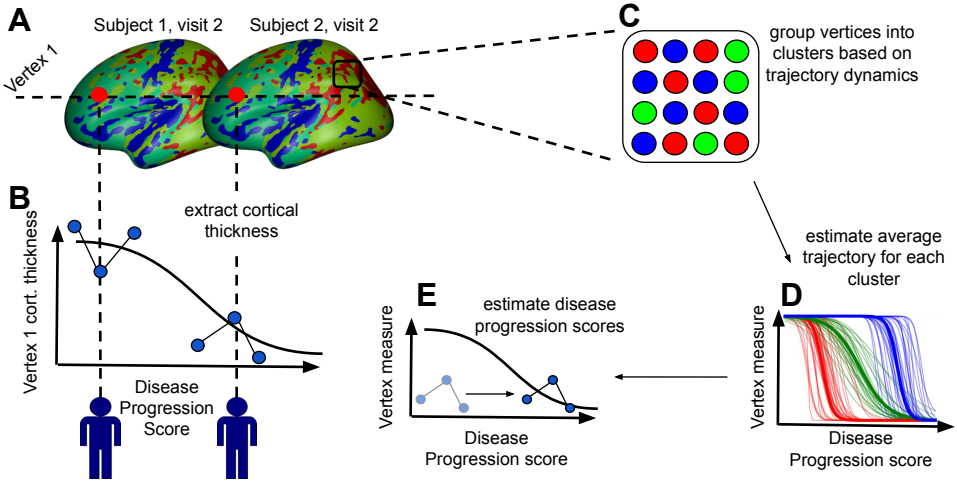
Our model clusters vertices with similar trajectories of pathology



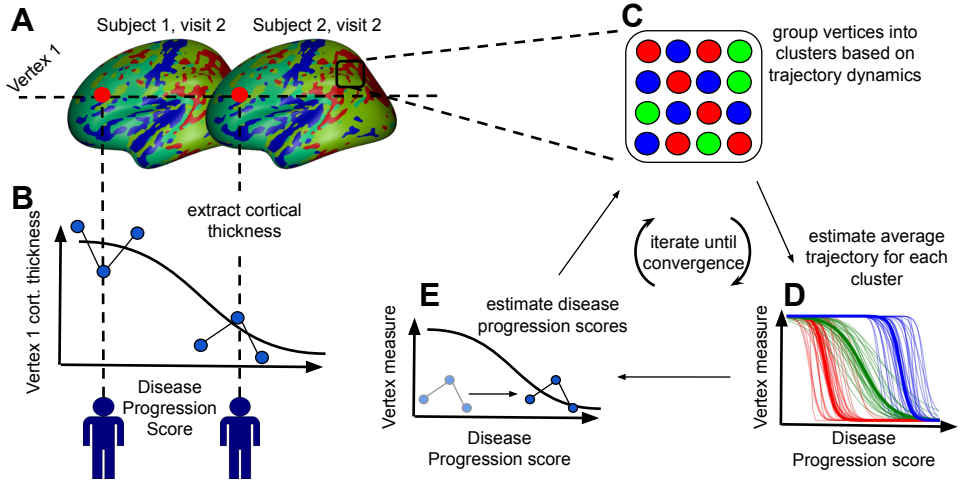
Our model clusters vertices with similar trajectories of pathology



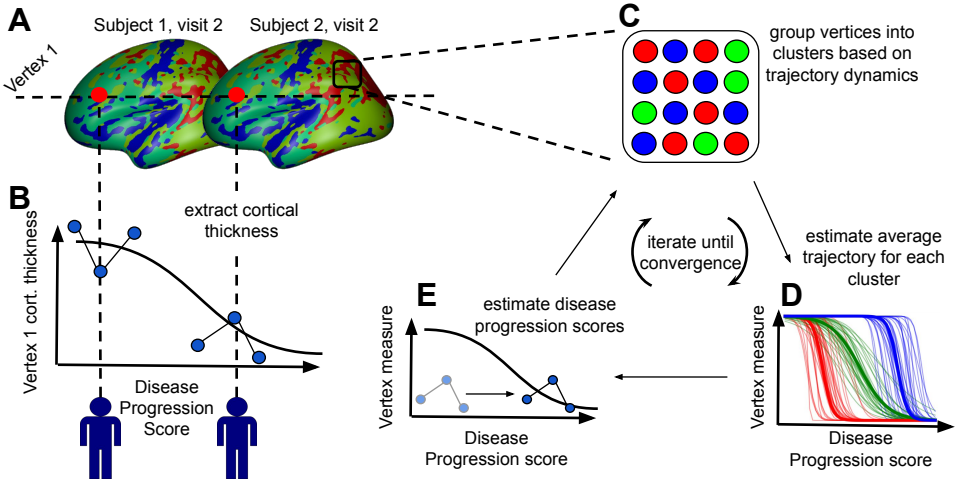
Our model clusters vertices with similar trajectories of pathology



Our model clusters vertices with similar trajectories of pathology



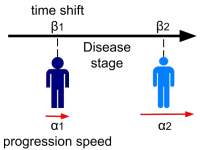
Our model clusters vertices with similar trajectories of pathology



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

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

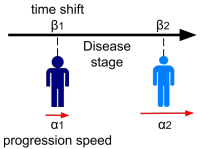
$$s_{ij} = \alpha_i t_{ij} + \beta_i$$



Building the model using a generative Bayesian framework

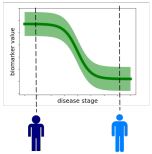
1. Model disease progression score for one subject i at visit j:

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



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

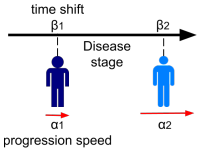
$$p(V_l^{ij} | \alpha_i, \beta_i, \theta_k, \sigma_k) \sim N(f(\alpha_i t_{ij} + \beta_i; \theta_k), \sigma_k)$$



Building the model using a generative Bayesian framework

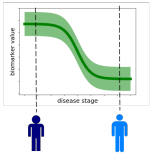
1. Model disease progression score for one subject i at visit j:

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



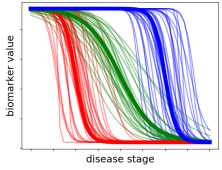
2. Model trajectory of cortical thickness at one location l on the brain:

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



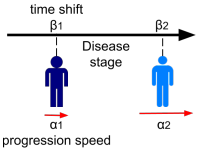
3. Extend to all locations and subjects:

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



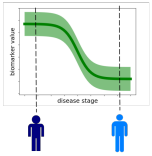
1. Model disease progression score for one subject i at visit j:

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



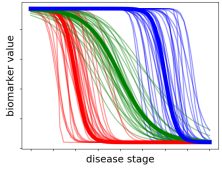
2. Model trajectory of cortical thickness at one location l on the brain:

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



3. Extend to all locations and subjects:

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

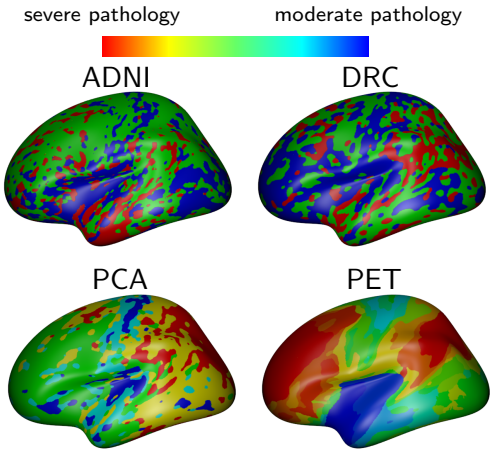


4. Marginalise over the hidden variables Z_l (cluster assignments):

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

Our Model Finds Plausible Atrophy Patterns on Four Datasets

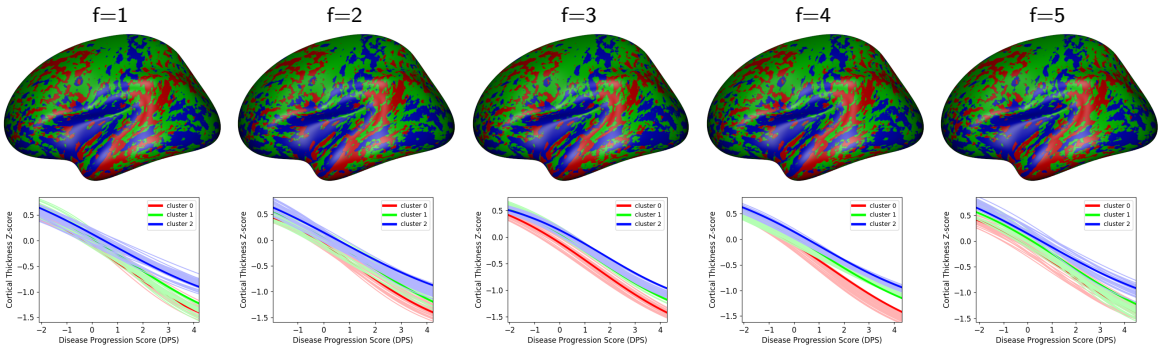
- Similar patterns of atrophy in independent Alzheimer's MRI datasets (ADNI vs DRC)
- Distinct patterns of atrophy in different diseases (Alzheimer's vs PCA) and modalities (MRI vs PET)



Marinescu et al., NeuroImage, 2019

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)



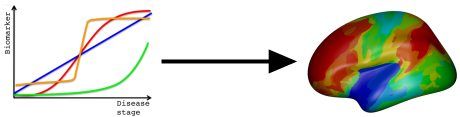
Marinescu et al., Neuroimage, 2019

Disease Progression Modelling – Summary

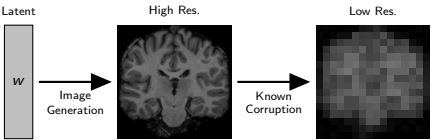
- We modelled the continuous progression of Alzheimer's disease and related dementias
- Used generative bayesian model that does not require labels (unsupervised)
- Plausible results on four different datasets
- However, such models require good quality data, to perform registration and extract disease 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**

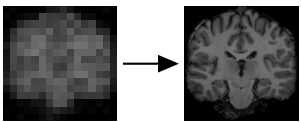


3. Future work towards brain anatomy simulators

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

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

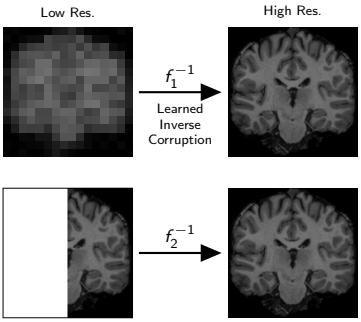
MRI reconstruction



StyleGAN2 (Karras et al, 2019)

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

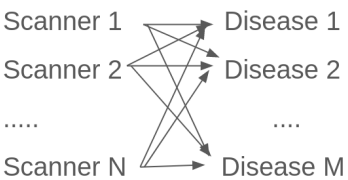
Limitation 2: Distribution shifts require model re-training

- Distribution shifts happen all the time:
  - Changes in hospital scanners, protocols, software upgrades
  - Can be continuous: population getting older due to better healthcare

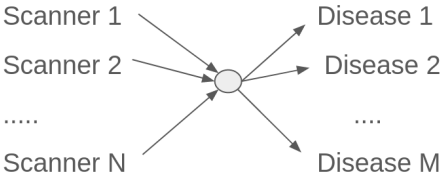
- Shifts can result in combinatorial effects in number of re-training instances!

- Compositionality is one potential solution

Without compositionality: **N x M**



With compositionality: **N + M**



- 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

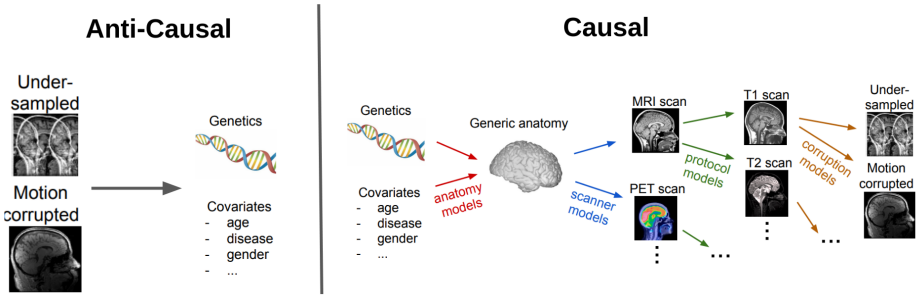
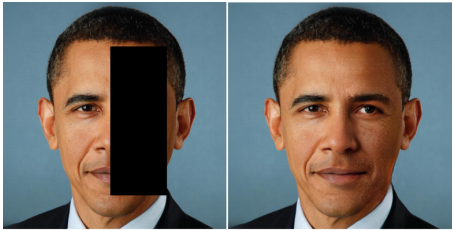


Image2StyleGAN++ (Abdal et al, 2020)

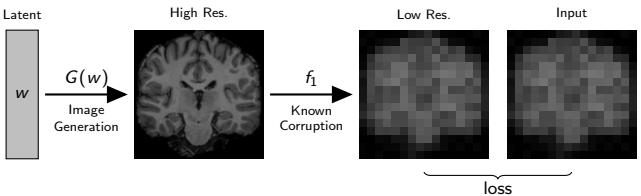


PULSE (Menon et al, 2020)

- 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

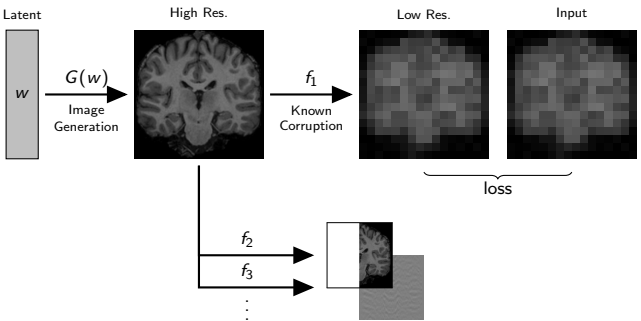
Method: We perform image reconstruction by combining two models

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



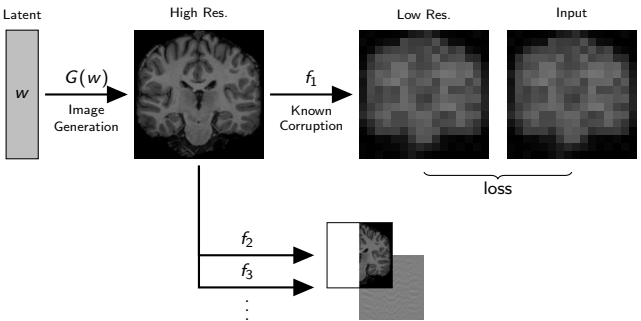
Method: We perform image reconstruction by combining two models

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

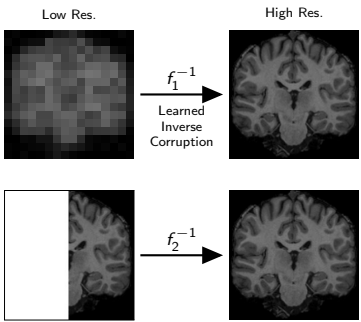


Method: We perform image reconstruction by combining two models

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



Ours (Marinescu et al, arXiv, 2020)



Previous

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

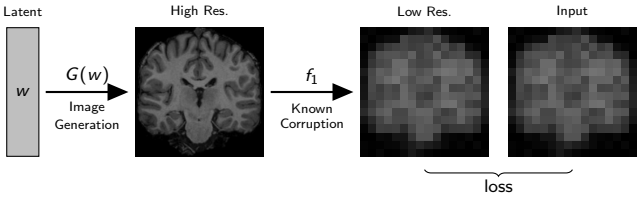
- We optimise:

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

- 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

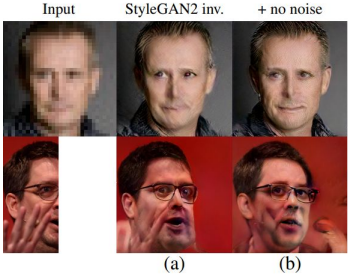
$$w^*, \eta^* = \arg\min_{w,\eta} ||\phi(I) - \phi \circ f \circ G(w, \eta)||_2^2$$

(a)

Good reconstructions require further modifications

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

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

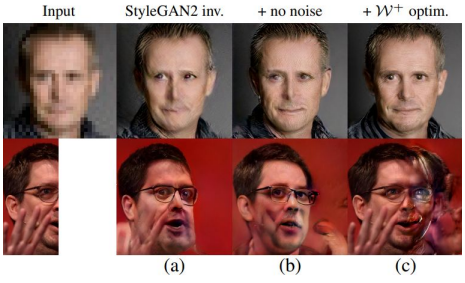


Good reconstructions require further modifications

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

$$\mathbf{w} = w_1, .., w_L$$

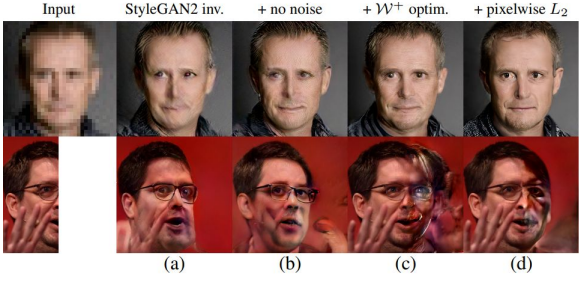
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} ||\phi(I) - \phi \circ f \circ G(\mathbf{w})||_2^2$$



Good reconstructions require further modifications

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

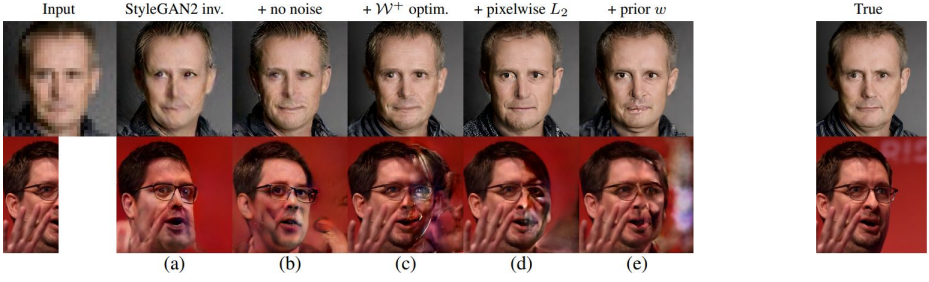
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} ||\phi(I) - \phi \circ f \circ G(\mathbf{w})||_2^2 + ||I - f \circ G(\mathbf{w})||_2^2$$



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

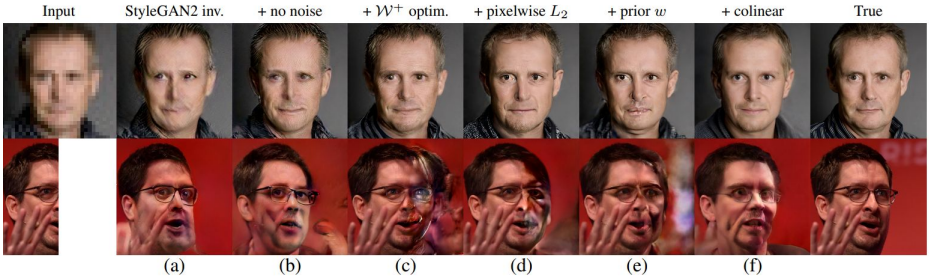
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} ||\phi(I) - \phi \circ f \circ G(\mathbf{w})||_2^2 + ||I - f \circ G(\mathbf{w})||_2^2 + \\ + \sum_i \left(\frac{w_i - \mu}{\sigma_i} \right)^2$$



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

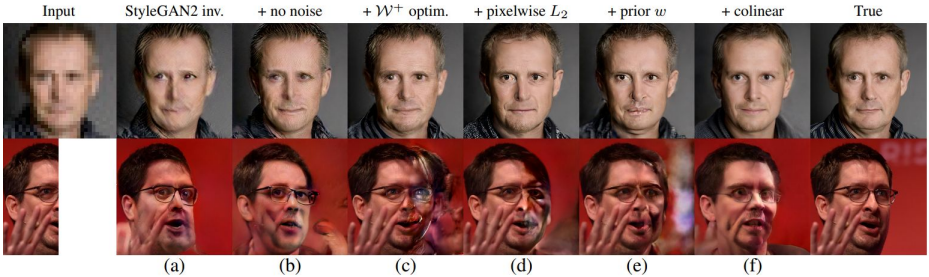
$$\mathbf{w}^* = \arg\min_{\mathbf{w}} ||\phi(I) - \phi \circ f \circ G(\mathbf{w})||_2^2 + ||I - f \circ G(\mathbf{w})||_2^2 +$$
$$+ \sum_i \left(\frac{w_i - \mu}{\sigma_i} \right)^2 - \sum_{i,j} \frac{w_i w_j^T}{|w_i||w_j|}$$



Good reconstructions require further modifications

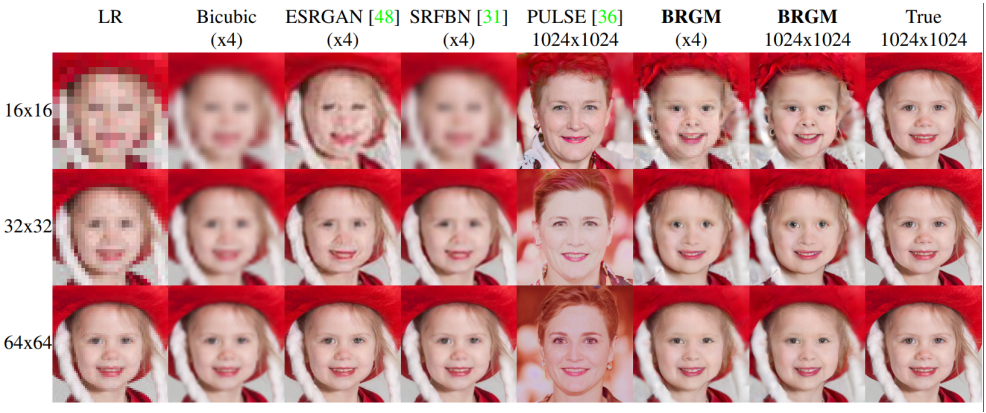
- We started from the original StyleGAN2 inversion
- Yet the reconstruction was not good → required several changes
  - Analytically expressed the full likelihood (Marinescu et al, 2021)

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



Results on super-resolution using the FFHQ dataset

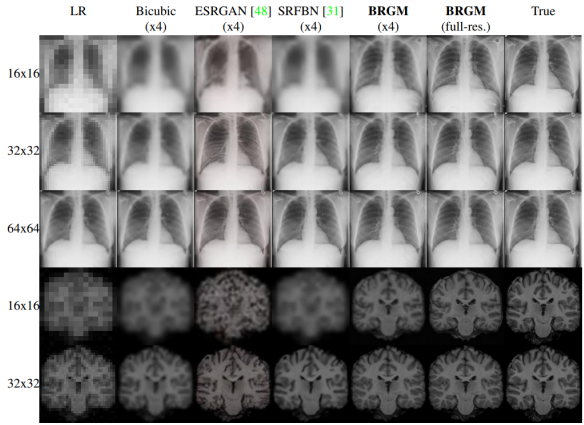
- We achieve state-of-the-art (SOTA) results on small inputs resolutions 16x16
- On larger resolutions (>32x32), we achieve very good results, albeit not SOTA



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



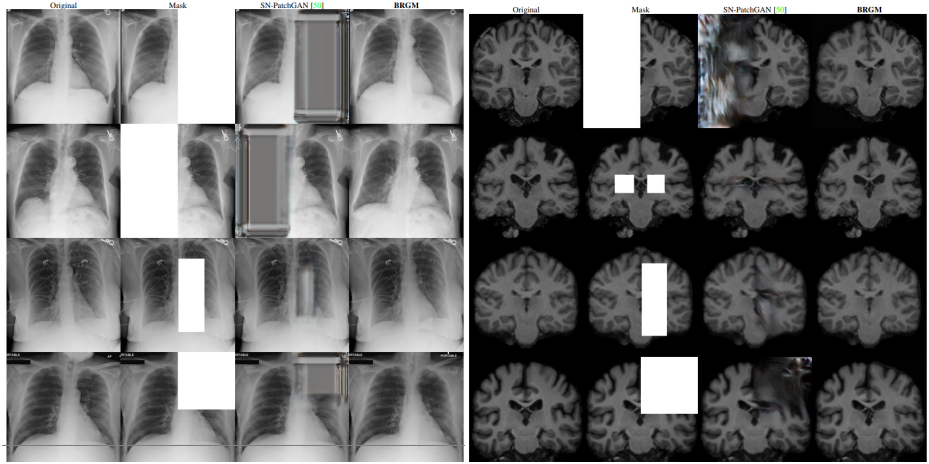
Marinescu et al, arXiv, 2020

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 16^2	**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

	BRGM				SN-PatchGAN [50]			
Dataset	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

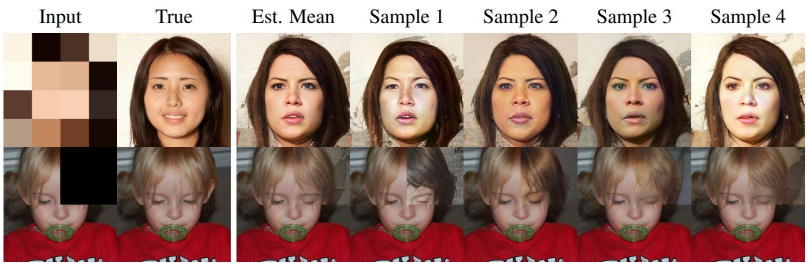
Dataset	**BRGM**	PULSE [36]	ESRGAN [48]	SRFBN [31]
FFHQ 16^2	**0.42**	0.32	0.11	0.15
FFHQ 32^2	0.39	0.02	0.12	**0.47**
FFHQ 64^2	0.14	0.08	0.32	**0.45**
FFHQ 128^2	0.14	0.10	**0.39**	0.38

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

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

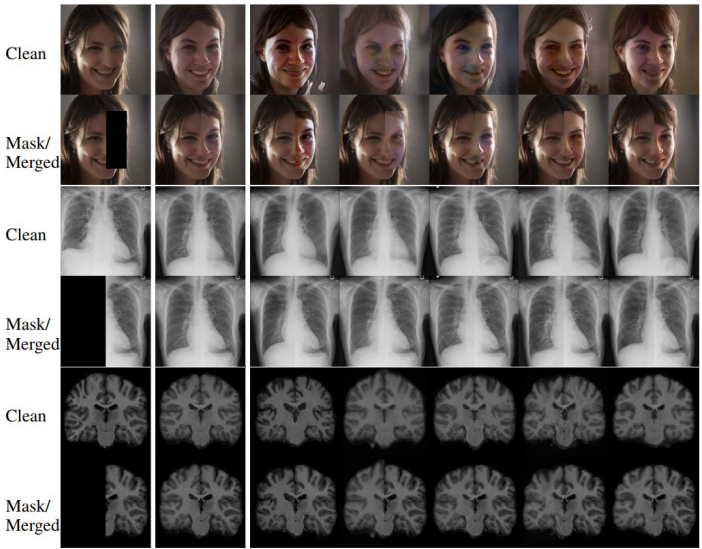
- We approximate the integral using Monte Carlo samples $w^{(i)}$ taken from $q(w|\theta)$

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



Marinescu et. al., 2020

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

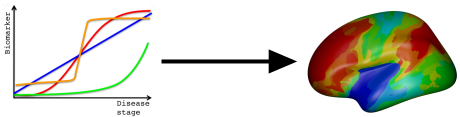
- Can be inconsistent with the input image

Bayesian Image Reconstruction – Summary

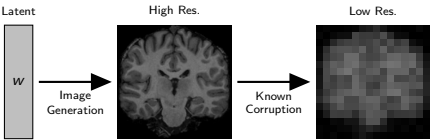
- Proposed a method for image reconstruction using pre-trained deep generative models
- Solution is given by the Bayesian MAP estimate
- State-of-the-art results on super-resolution and inpainting

Outline

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

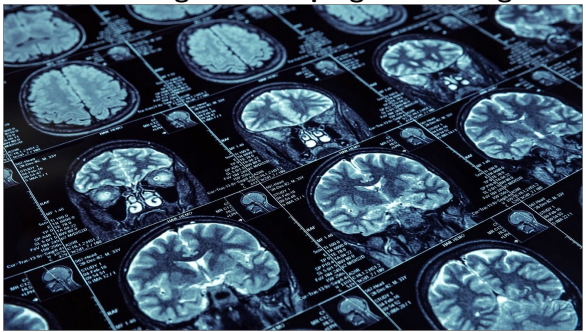


2. Image Reconstruction using Deep Generative Models



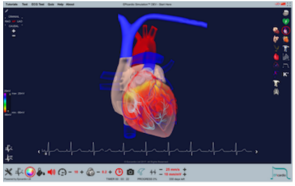
3. **Future work towards brain anatomy simulators**

Accurate diagnosis and prognosis through AI

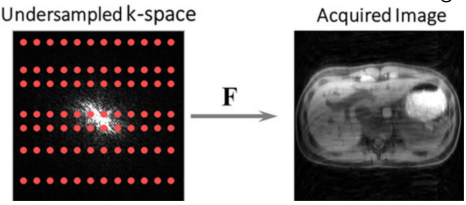


AI to augment doctors

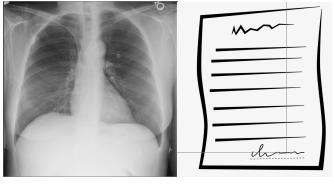
Biological simulators



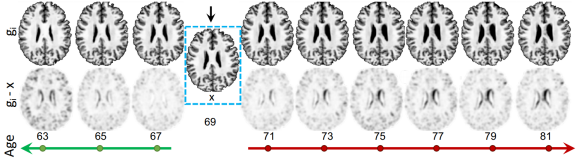
Better and faster reconstruction of medical images



Multimodal modelling
images + text + structural data



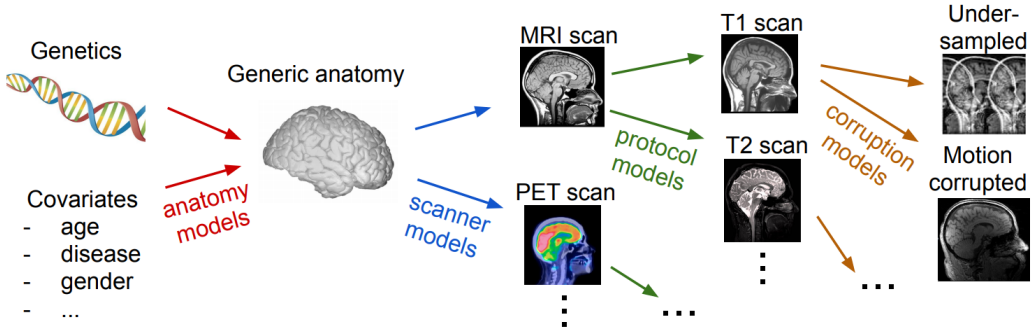
Disease Progression Modelling



Future work: Brain tissue and anatomy simulator

Simulator for brain anatomy from genetics:

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

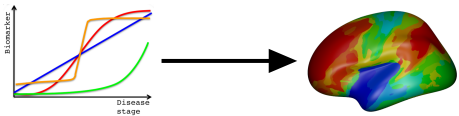


Problem: Lack of good labels

Problem: Lack of good input data

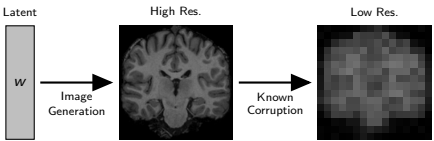
Problem: Lack of good labels

Solution: Unsupervised Learning through Disease Progression Modelling



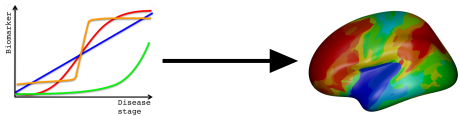
Problem: Lack of good input data

Solution: Image Reconstruction using Deep Generative Models



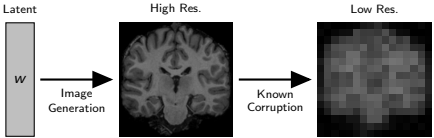
Problem: Lack of good labels

Solution: Unsupervised Learning through Disease Progression Modelling



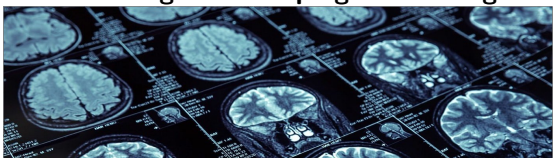
Problem: Lack of good input data

Solution: Image Reconstruction using Deep Generative Models



Long-term vision

Accurate diagnosis and prognosis through AI



AI to augment doctors