Learning to synthesise the ageing brain without longitudinal data

Medical Image Analysis (2021)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Learning to synthesise the ageing brain without longitudinal data

Tian Xia[a,*], Agisilaos Chartsias[a], Chengjia Wang[b], Sotirios A. Tsaftaris[a,c], for the Alzheimer's Disease Neuroimaging Initiative

[a] *Institute for Digital Communications, School of Engineering, University of Edinburgh, West Mains Rd, Edinburgh EH9 3FB, UK*
[b] *The BHF Centre for Cardiovascular Science, Edinburgh EH16 4TJ, UK*
[c] *The Alan Turing Institute, London NW1 2DB, UK*

ARTICLE INFO

Article history:

Keywords: Brain Ageing, Generative Adversarial Network, Neurodegenerative Disease, Magnetic Resonance Imaging (MRI)

ABSTRACT

How will my face look when I get older? Or, for a more challenging question: How will my brain look when I get older? To answer this question one must devise (and learn from data) a multivariate auto-regressive function which given an image and a desired target age generates an output image. While collecting data for faces may be easier, collecting longitudinal brain data is not trivial. We propose a deep learning-based method that learns to simulate subject-specific brain ageing trajectories *without* relying on longitudinal data. Our method synthesises images conditioned on two factors: age (a continuous variable), and status of Alzheimer's Disease (AD, an ordinal variable). With an adversarial formulation we learn the joint distribution of brain appearance, age and AD status, and define reconstruction losses to address the challenging problem of preserving subject identity. We compare with several benchmarks using two widely used datasets. We evaluate the quality and realism of synthesised images using ground-truth longitudinal data and a pre-trained age predictor. We show that, despite the use of cross-sectional data, our model learns patterns of gray matter atrophy in the middle temporal gyrus in patients with AD. To demonstrate generalisation ability, we train on one dataset and evaluate predictions on the other. In conclusion, our model shows an ability to separate age, disease influence and anatomy using only 2D cross-sectional data that should be useful in large studies into neurodegenerative disease, that aim to combine several data sources. To facilitate such future studies by the community at large our code is made available at `https://github.com/xiat0616/BrainAgeing`.

1. Introduction

The ability to predict the future state of an individual can be of great benefit for longitudinal studies (Ziegler et al., 2012). However, such learned phenomenological predictive models need to capture anatomical and physiological changes due to ageing and separate the factors that influence future state. Recently, deep generative models have been used to simulate and predict future degeneration of a human brain using existing scans (Ravi et al., 2019a; Rachmadi et al., 2019, 2020). However, current methods require considerable amount of longitudinal data to sufficiently approximate an auto-regressive model. Here, we propose a new conditional adversarial training procedure that does *not* require longitudinal data to train. Our approach (shown in Fig. 1) synthesises images of aged brains for a desired age and health state.

Brain ageing, accompanied by a series of functional and physiological changes, has been intensively investigated (Zecca et al., 2004; Mattson and Arumugam, 2018). However, the un-

*Corresponding author.
e-mail: `tian.xia@ed.ac.uk` (Tian Xia)

derlying mechanism has not been completely revealed (López-Otín et al., 2013; Cole et al., 2019). Prior studies have shown that a brain's chronic changes are related to different factors, e.g. the biological age (Fjell and Walhovd, 2010), degenerative diseases such as Alzheimer's Disease (AD) (Jack et al., 1998), binge drinking (Coleman Jr et al., 2014), and even education (Taubert et al., 2010). Accurate simulation of this process has great value for both neuroscience research and clinical applications to identify age-related pathologies (Cole et al., 2019; Fjell and Walhovd, 2010).

One particular challenge is inter-subject variation: every individual has a unique ageing trajectory. Previous approaches built a spatio-temporal atlas to predict average brain images at different ages (Davis et al., 2010; Huizinga et al., 2018). However, individuals with different health status follow different ageing trajectories. An atlas may not preserve subject-specific characteristics; thus, may preclude accurate modelling of individual trajectories and further investigation on the effect of different factors, e.g. age, gender, education, etc (Ravi et al., 2019b). Recent studies proposed subject-specific ageing progression with neural networks (Ravi et al., 2019a; Rachmadi et al., 2019), although they require longitudinal data to train. Ideally, the longitudinal data should cover a long time span with frequent sampling to ensure stable training. However, such data are difficult and expensive to acquire, particularly for longer time spans. Even in ADNI (Petersen et al., 2010), one of the most well-known large-scale datasets, longitudinal images are acquired at few time points and cover only a few years. Longitudinal data of sufficient time span remain an open challenge.

In this paper, we build the foundations of a model that can be trained without longitudinal data. A simplified schematic of our model is shown in Fig. 1 along with example results. Given a brain image, our model produces a brain of the same subject at target age. The input image is first encoded into a latent space, which is modulated by two vectors representing target age difference and health state (AD status in this paper), respectively. The conditioned latent space is finally decoded to an output image, i.e. the synthetically aged image.

Under the hood, what trains the generator, is a deep adversarial method that learns the joint distribution of brain appearance, age and health state. The quality of the output is encouraged by a discriminator that judges whether an output image is representative of the distribution of brain images of the desired age and health state. A typical problem in synthesis which is exacerbated with *cross-sectional* data (Ziegler et al., 2012) is loss of *subject identity*[1], i.e. the synthesis of an output that may not correspond to the input subject's identity. We propose, and motivate, two loss functions towards retaining *subject identity* by regularising the amount of change introduced by ageing. In addition, we motivate the design of our conditioning mechanisms and show that ordinal binary encoding for both discrete and continuous variables improves performance significantly.

We consider several metrics and evaluation approaches to verify the quality and biological plausibility of our results. We quantitatively evaluate our simulation results using longitudinal data from the ADNI dataset (Petersen et al., 2010) with classical metrics that estimate image fidelity. Since the longitudinal data only cover a limited time span, it is difficult to evaluate the quality of synthesized aged images across decades. For brain ageing synthesis, a good synthetic brain image should be accurate in terms of age, i.e. be close to the target age that we want it to be, and also preserve subject identity, i.e. should be from the same subject as the input. Thus, we pre-train a deep network to estimate the apparent age from output images. The estimated ages are used as a proxy metric for *age accuracy*. We also show qualitative results, including ageing simulation on different health states and long-term ageing synthesis. Both quantitative and qualitative results show that our method outperforms benchmarks with more accurate simulations that capture the characteristics specific to each individual on different health states. Furthermore, we train our model on Cam-CAN and evaluate it on ADNI to demonstrate the generalisation ability to unseen data. In addition, to demonstrate the realism of synthetic results, we perform volume synthesis and evaluate deformation. We also estimate gray matter atrophy in middle temporal gyrus and find that our model, even without longitudinal data, has learned that ageing and disease leads to atrophy. Ablation studies investigate the effect of loss components and different ways of embedding clinical variables into the networks.

Our contributions are summarised as follows:[2]

- Our main contribution is a deep learning model that learns to simulate the brain ageing process, and perform subject-specific brain ageing synthesis, trained on *cross-sectional* data overcoming the need for longitudinal data.
- For our model to be able to change output based on desired input (age and health state), we use an (ordinal) embedding mechanism that guides the network to learn the joint distribution of brain images, age and health state.
- Since we do not use longitudinal data that can constrain the learning process, we design losses that aim to preserve subject identity, while encouraging quality output.
- We provide an experimental framework to verify the quality and biological validity of the synthetic outputs.

[1] A classical computer vision example is generating a human face resembling another individual instead of the input subject. Even with faces, humans find it difficult to assess identity loss. It remains hard to define detailed structural changes during ageing, e.g. balding, nose shape change, eye colour change. There are some common patterns that we can expect, such as wrinkles and gray/white hair, but it is difficult to define other more detailed changes. Therefore, even in face ageing, 'subject identity' is defined as young and old images should be from the same person. In brain synthesis, it is even more difficult to define 'subject identity', as human eyes are less able to visually ascertain brain image identity particularly as modulated by age and pathology. In this paper, we followed a similar analogue of 'identity': a "synthetic image should be from the same subject as the input image".

[2] We advance our preliminary work (Xia et al., 2019) in the following aspects: 1) we extend our model to condition on age and AD status, which enables more accurate simulation of ageing progression of different health states; 2) we introduce additional regularisation to smooth the simulated progression; 3) we offer more experiments and a detailed analysis of performance, using longitudinal data, including new metrics and additional benchmark methods for comparison; 4) we introduce analysis based on measuring deformation and atrophy; and 5) several ablation studies.

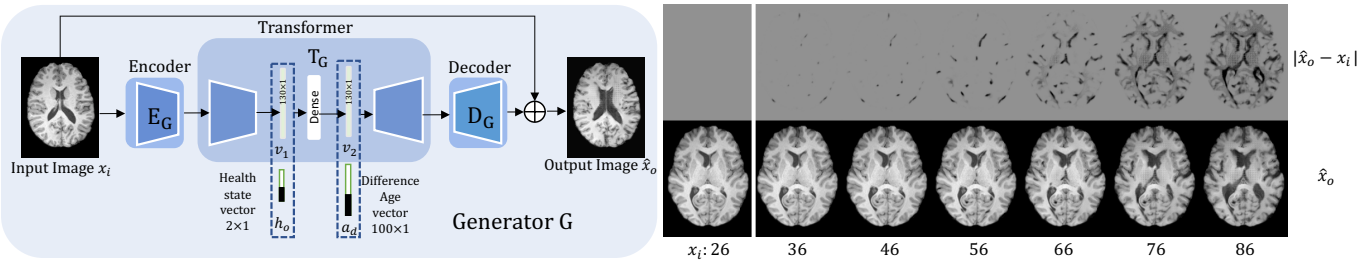


Fig. 1. Left: The input is a brain image x_i, and the network synthesises an aged brain image $\hat{x}_o$ from x_i, conditioned on the target health state vector h_o and target age difference $a_d = a_o - a_i$ between input a_i and target a_o ages, respectively. Right: For an image x_i of a 26 year old subject, bottom row shows outputs $\hat{x}_o$ given different target age. The top row shows the corresponding image differences $|\hat{x}_o - x_i|$ to highlight progressive changes.

While our first contribution is the most important one, it is the combination of our proposed losses and embedding mechanisms that lead to the method's robustness, as extensive experiments and ablation studies on two publicly available datasets, namely Cam-CAN (Taylor et al., 2017) and ADNI (Petersen et al., 2010) show.

The manuscript proceeds as follows: Section 2 reviews related work on brain ageing simulation and prediction. Section 3 details the proposed method. Section 4 describes the experimental setup and training details. Section 5 presents results and discussion. Finally, Section 6 offers conclusions.

2. Related Work

We first discuss *brain ageing simulation*, i.e. simulating the ageing process from data. For completeness, we also briefly discuss *brain age prediction*, i.e. estimating age from an image.

2.1. Brain ageing simulation

Given variables such as age, one can synthesise the corresponding brain image to enable visual observation of brain changes. For instance, patch-based dictionary learning (Zhang et al., 2016), kernel regression (Huizinga et al., 2018; Ziegler et al., 2012; Serag et al., 2012), linear mixed-effect modelling (Lorenzi et al., 2015; Sivera et al., 2019) and non-rigid registration (Sharma et al., 2010; Modat et al., 2014; Pieperhoff et al., 2008; Camara et al., 2006) have been used to build spatio-temporal atlases of brains at different ages. However, by relying on population averages as atlases subject-specific ageing trajectories are harder to capture. Recently, Khanal et al. (2017) build a biophysical model assuming brain atrophy, but without considering age or other clinical factors (e.g. AD status).

Deep generative methods have also been used for this task. While Rachmadi et al. (2019, 2020) and Wegmayr et al. (2019) used formulations of Generative Adversarial Networks (GAN) (Goodfellow et al., 2014) to simulate brain changes, others (Ravi et al., 2019a) used a conditional adversarial autoencoder as the generative model, following a recent face ageing approach (Zhang et al., 2017). Irrespective of the model, these methods need longitudinal data, which limits their applicability.

In Bowles et al. (2018), a GAN-based method is trained to add or remove atrophy patterns in the brain using image arithmetics. However, the atrophy patterns were modelled linearly, with morphological changes assumed to be the same for all subjects. Milana (2017) used a Variational Autoencoder (VAE) to synthesise aged brain images, but the target age is not controlled, and the quality of the synthesised image appears poor (blurry). Recently, Pawlowski et al. (2020) showed that a VAE-based structural causal model can generate brain images. However, they did not provide a quantitative evaluation of the generated images perhaps due to known issues of low quality outputs when using a VAE. Similarly, Zhao et al. (2019) used a VAE to disentangle the spatial information from temporal progression and then used the first few layers of the VAE as feature extractor to improve the age prediction task. As their focus is on age prediction, the synthetic brain images only contain the ventricular region and are population averages.

In summary, most previous methods either built average atlases (Zhang et al., 2016; Huizinga et al., 2018; Ziegler et al., 2012; Serag et al., 2012), or required longitudinal data (Rachmadi et al., 2019, 2020; Ravi et al., 2019a; Wegmayr et al., 2019) to simulate brain ageing. Other methods either did not consider subject identity (Bowles et al., 2018; Milana, 2017), or did not evaluate in detail morphological changes (Pawlowski et al., 2020; Zhao et al., 2019).

To address these shortcomings, we propose a conditional adversarial training procedure that learns to simulate the brain ageing process by being *specific* to the input subject, and by learning from *cross-sectional* data i.e. without requiring longitudinal observations.

2.2. Brain age prediction

These methods predict age from brain images learning a relationship between image and age; thus, for completeness we briefly mention two key directions. For example, Franke, Katja and Ziegler, Gabriel and Klöppel, Stefan and Gaser, Christian and Alzheimer's Disease Neuroimaging Initiative and others (2010) predicted age with hand-crafted features and kernel regression whereas Cole and Franke (2017) used Gaussian Processes. Naturally performance relies on the effectiveness of the hand-crafted features.

Recently, deep learning models have been used to estimate the brain age from imaging data. For example, Cole et al. (2017) used a VGG-based model (Simonyan and Zisserman, 2015) to predict age and detect degenerative diseases, while Jonsson et al. (2019) proposed to discover genetic associations with the brain degeneration using a ResNet-based network (He et al., 2016). Similarly, Peng et al. (2021) used a

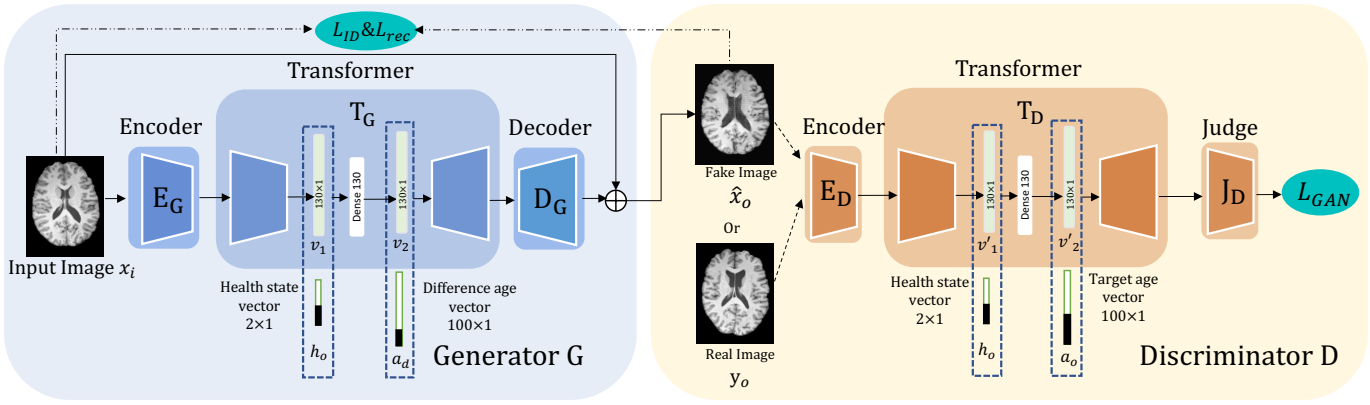


Fig. 2. An overview of the proposed method (training). $\mathbf{x}_i$ is the input image; $\mathbf{h}_o$ is the target health state; $\mathbf{a}_d$ is the difference between the starting age a_i and target age a_o: $a_d = a_o - a_i$; $\hat{\mathbf{x}}_o$ is the output (aged) image (supposedly belong to the same subject as x_i) of the target age a_o and health state h_o. The *Generator* takes as input $\mathbf{x}_i$, $\mathbf{h}_o$ and $\mathbf{a}_d$, and outputs $\hat{\mathbf{x}}_o$; the *Discriminator* takes as input a brain image and $\mathbf{h}_o$ and $\mathbf{a}_o$, and outputs a discrimination score.

CNN-based model to predict age. Cole et al. (2015) used the age predicted by a deep network to detect traumatic brain injury. While most previous works achieved mean absolute error (MAE) of 4-5 years, Peng et al. (2021) achieved state-of-the-art performance with MAE of 2.14 years. However, these methods did not consider the morphological changes of brain, which is potentially more informative (Costafreda et al., 2011).

3. Proposed approach

3.1. Problem statement, notation and overview

In the rest of the paper, we use **bold** notations for vectors/images, and *italics* notations for scalars. For instance, a represents an age while $\mathbf{a}$ is a vector that represents age a. We denote a brain image as $\mathbf{x}_s$ (and $\mathcal{X}_s$ their distribution such that $\mathbf{x}_s \sim \mathcal{X}_s$), where s are the subject's clinical variables including the corresponding age a and health state (AD status) h. Given a brain image $\mathbf{x}_i$ of age a_i and health state h_i, we want to synthesise a brain image $\hat{\mathbf{x}}_o$ of target age a_o and health state h_o. Critically, the synthetic brain image $\hat{\mathbf{x}}_o$ should retain the subject identity, i.e. belong to the same subject as the input $\mathbf{x}_i$, throughout the ageing process. The contributions of our approach, shown in Fig. 2, are the design of the conditioning mechanism; our model architecture that uses a Generator to synthesise images, and a Discriminator to help learn the joint distribution of clinical variables and brain appearance; and the losses we use to guide the training process. We detail all these below.

3.2. Conditioning on age and health state

In our previous work (Xia et al., 2019), we simulate the ageing brain with age as the single factor. Here, we improve our previous approach by involving the health state, i.e. AD status, as another factor to better simulate the ageing process. [3]

[3] Additional fine-grained information on AD effects on different, local, brain regions could be provided if clinical scores are used instead. As our work is the first to attempt to learn without longitudinal data, for simplicity we focused on variables capturing global effects. In the conclusion section, we note the addition of fine-grained information as an avenue for future improvement.

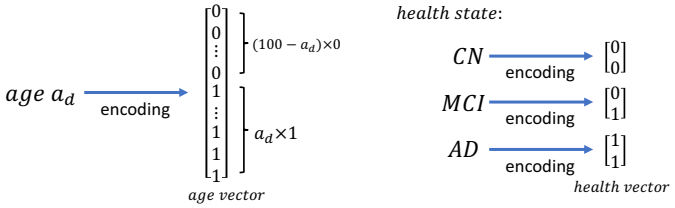


Fig. 3. Ordinal encoding of age and health state. Left shows how we represent age $\mathbf{a}_d$ using a binary vector with first $\mathbf{a}_d$ elements as 1 and the rest as 0; Right is the encoding of health state, where we use a 2×1 vector to represent three categories of AD status: control normal (CN), mildly cognitive impaired (MCI), and Alzheimer's Disease (AD).

We use ordinal binary vectors, instead of one-hot vectors as in Zhang et al. (2017), to encode both age and health state, which are embedded in the bottleneck layer of the Generator and Discriminator (detailed in Section 3.3). We assume a maximal age of 100 years and use a 100×1 vector to encode age a. Similarly, we use a 2×1 vector to encode health state. A simple illustration of this encoding is shown in Fig. 3. An ablation study presented in Section 5.4 illustrates the benefits of *ordinal* v.s. *one-hot* encoding.

3.3. Proposed model

The proposed method consists of a *Generator* and a *Discriminator*. The Generator synthesises aged brain images corresponding to a target age and a health state. The Discriminator has a dual role: firstly, it discriminates between ground-truth and synthetic brain images; secondly, it ensures that the synthetic brain images correspond to the target clinical variables. The Generator is adversarially trained to generate realistic brain images of the correct target age. The detailed network architectures are shown in Fig. 4.

3.3.1. Generator

The Generator G takes as input a 2D brain image $\mathbf{x}_i$, and ordinal binary vectors for target health state h_o and age difference a_d. Here, we condition on the age difference between input age a_i and target age a_o: $a_d = a_o - a_i$, such that when input and output ages are equal $a_d = 0$, the network is encouraged to recreate

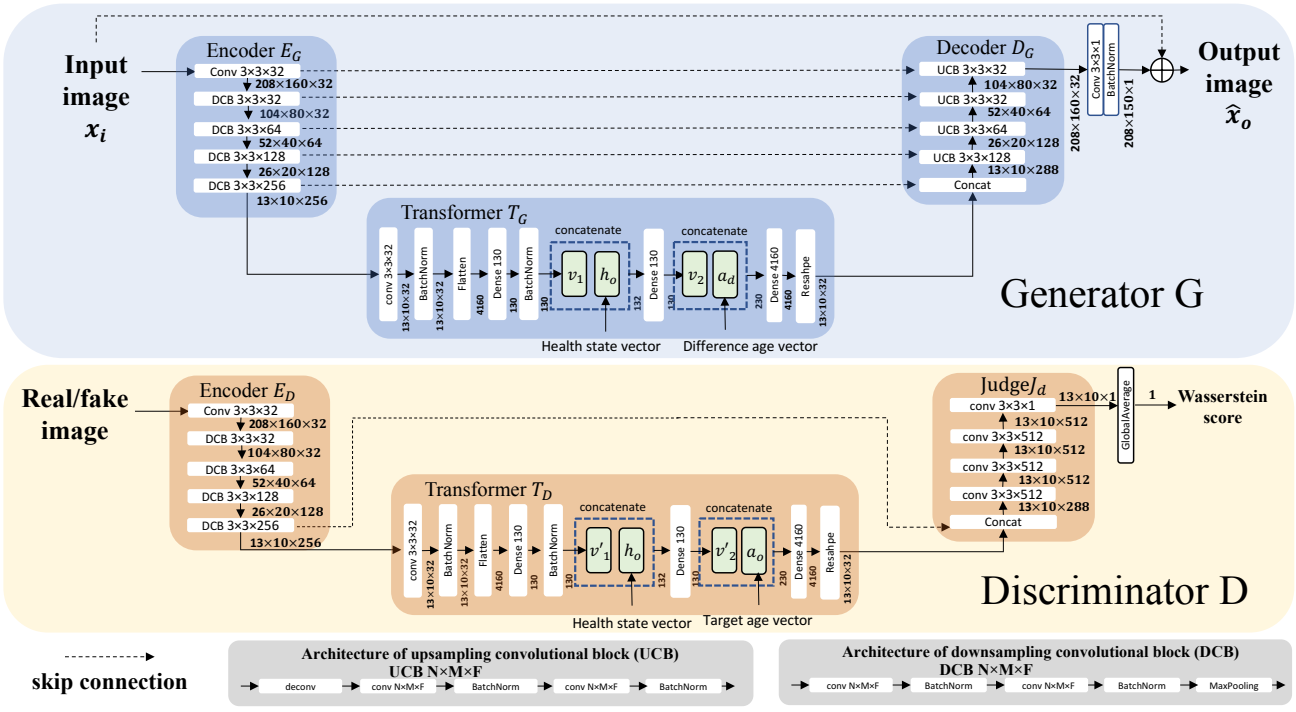


Fig. 4. Detailed architectures of *Generator* and *Discriminator*. The Generator contains three parts: an Encoder to extract latent features; a Transformer to involve target age and health state; and a Decoder to generate aged images. Similarly, we use the same conditioning mechanism for the Discriminator to inject the information of age and health state, and a long skip connection to better preserve features of input image.

the input. The output of G is a 2D brain image $\hat{\mathbf{x}}_o$ corresponding to the target age and health state.[4]

G has three components: the *Encoder* E_G, the *Transformer* T_G and the *Decoder* D_G. E_G first extracts latent features from the input image $\mathbf{x}_i$; T_G involves the target age and health state into the network. Finally, D_G generates the aged brain image from the bottleneck features. To embed age and health state into our model, we first concatenate the latent vector $\mathbf{v}_1$, obtained by E_G, with the health state vector h_o. The concatenated vector is then processed by a dense layer to output latent vector $\mathbf{v}_2$, which is then concatenated with the difference age vector $\mathbf{a}_d$. Finally, the resulting vector is used to generate the output image.[5] We adopt long-skip connections (Ronneberger et al., 2015) between layers of E_G and D_G to preserve details of the input image and improve the sharpness of the output images. Overall, the Generator's forward pass is: $\hat{\mathbf{x}}_o = G(\mathbf{x}_i, \mathbf{a}_d, \mathbf{h}_o)$.

3.3.2. Discriminator

Similar to the Generator, the Discriminator D contains three subnetworks: the *Encoder* E_D that extracts latent features, the *Transformer* T_D that involves the conditional variables, and the *Judge* J_D that outputs a discrimination score. For the discriminator to learn the joint distribution of brain image, age, and health state, we embed the age and health vectors into the discriminator with a similar mechanism as that of the Generator.

Note that D is conditioned on the target age $\mathbf{a}_o$ instead of age difference $\mathbf{a}_d$, to learn the joint distribution of brain appearance and age, such that it can discriminate between real and synthetic images of correct age. The forward pass for the Discriminator is $w_{fake} = D(\hat{\mathbf{x}}_o, \mathbf{a}_o, \mathbf{h}_o)$ and $w_{real} = D(\mathbf{y}_o, \mathbf{a}_o, \mathbf{h}_o)$.

3.4. Losses

We train with a multi-component loss function containing *adversarial*, *identity-preservation* and *self-reconstruction* losses. We detail these below.

3.4.1. Adversarial loss

We adopt the Wasserstein loss with gradient penalty (Gulrajani et al., 2017) to predict a realistic aged brain image $\hat{\mathbf{x}}_o$ and force $\hat{\mathbf{x}}_o$ to correspond to the target age a_o and health state h_o:

$$L_{GAN} = \mathbb{E}_{\mathbf{y}_o \sim \mathcal{X}_o, \hat{\mathbf{x}}_o \sim \hat{\mathcal{X}}_o}[D(\mathbf{y}_o, \mathbf{a}_o, \mathbf{h}_o) - D(\hat{\mathbf{x}}_o, \mathbf{a}_o, \mathbf{h}_o) + \lambda_{GP}(\|\nabla_{\tilde{\mathbf{z}}} D(\tilde{\mathbf{z}}, \mathbf{a}_o, \mathbf{h}_o)\|_2 - 1)_2], \quad (1)$$

where $\hat{\mathbf{x}}_o$ is the output image: $\hat{\mathbf{x}}_o = G(\mathbf{x}_i, \mathbf{a}_d, \mathbf{h}_o)$ (and $\mathbf{a}_d = \mathbf{a}_o - \mathbf{a}_i$); $\mathbf{y}_o$ is a ground truth image from another subject of target age a_o and health state h_o; and $\tilde{\mathbf{z}}$ is the average sample defined by $\tilde{\mathbf{z}} = \epsilon\hat{\mathbf{x}}_o + (1-\epsilon)\mathbf{y}_o, \epsilon \sim U[0,1]$. The first two terms measure the Wasserstein distance between ground-truth and synthetic samples; the last term is the gradient penalty involved to stabilise training. As in Gulrajani et al. (2017) and Baumgartner et al. (2018) we set $\lambda_{GP} = 10$.

[4] Note that the target health state can be different from the corresponding input state. This encourages learning a joint distribution between brain images and clinical variables.

[5] We tested the ordering of $\mathbf{h}_o$ and $\mathbf{a}_d$, and it did not affect the results. We also tried to concatenate $\mathbf{h}_o$, $\mathbf{a}_d$ and $\mathbf{v}_1$ together into one vector, and use the resulting vector to generate the output. However, we found that the model tended to ignore the information of $\mathbf{h}_o$. This might be caused by the dimensional imbalance between h_o (2×1) and a_d (100×1).

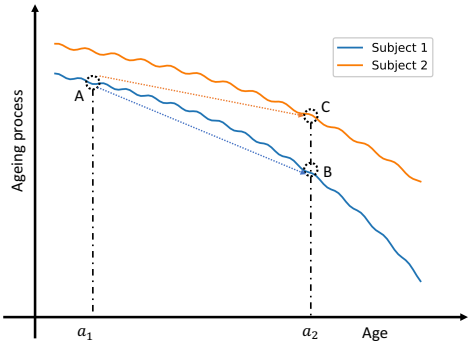


Fig. 5. Illustration of ageing trajectories for two subjects. For a subject of age a_1 (A), the network can learn a mapping from A to C, which could still fool the Discriminator, but loses the identity of Subject 1 (orange line).

3.4.2. Identity-preservation loss

While L_{GAN} encourages the network to synthesise realistic brain images, these images may lose subject identity. For example, it is easy for the network to learn a mapping to an image that corresponds to the target age and health state, but belongs to a different subject. An illustration is presented in Fig. 5, where ageing trajectories of two subjects are shown. The task is to predict the brain image of subject 1 at age a_2 starting at age a_1, by learning a mapping from point A to point B. But there are no ground-truth data to ensure that we stay on the trajectory of subject 1. Instead, the training data contain brain images of age a_2 belonging to subject 2 (and other subjects). Using only L_{GAN}, the Generator may learn a mapping from A to C to fool the Discriminator, which will lose the identity of subject 1. To alleviate this and encourage the network to learn mappings along the trajectory (i.e. from A to B), we adopt:

$$L_{ID} = \mathbb{E}_{\mathbf{x}_i \sim \mathcal{X}_i, \hat{\mathbf{x}}_o \sim \mathcal{X}_o} \|\mathbf{x}_i - \hat{\mathbf{x}}_o\|_1 \cdot e^{-\frac{|a_o - a_i|}{|a_{max} - a_{min}|}}, \tag{2}$$

where $\mathbf{x}_i$ is the input image of age a_i and $\hat{\mathbf{x}}_o$ is the output image of age a_o ($a_o > a_i$). The term $e^{-\frac{|a_o - a_i|}{|a_{max} - a_{min}|}}$ encourages $\|\mathbf{x}_i - \hat{\mathbf{x}}_o\|_1$ to positively correlate with the difference $|a_o - a_i|$. The health state is not involved in L_{ID} as we do not aim to precisely model the ageing trajectory. Instead, L_{ID} is used to encourage identity preservation by penalising major changes between images close in age, and to stabilise training. A more accurate ageing prediction, which is also correlated with health state, is achieved by the adversarial loss. An ablation study illustrating the critical role of L_{ID} is included in Section 5.4.

3.4.3. Self-reconstruction loss

We use a self-reconstruction loss,

$$L_{rec} = \mathbb{E}_{\mathbf{x}_i \sim \mathcal{X}_i, \hat{\mathbf{x}}_o \sim \mathcal{X}_i} \|\mathbf{x}_i - \hat{\mathbf{x}}_o\|_1, \tag{3}$$

to explicitly encourage that the output $\hat{\mathbf{x}}_o$ is a faithful reconstruction of the input $\mathbf{x}_i$ for the same age and health state. Although L_{rec} is similar to L_{ID}, their roles are different: L_{ID} helps to preserve subject identity when generating aged images, while L_{rec} encourages smooth progression via self-reconstruction. An ablation study on L_{rec} in Section 5.4 shows the importance of stronger regularisation.[6]

4. Experimental setup

Datasets: We use two datasets, as detailed below.

Cambridge Centre for Ageing and Neuroscience (Cam-CAN) (Taylor et al., 2017) is a cross-sectional dataset containing normal subjects aged 18 to 88. We split subjects into different age groups spanning 5 years. We randomly selected 38 volumes from each age group and used 30 for training and 8 for testing. To prevent data imbalance, we discarded subjects under 25 or over 85 years old, because there are underrepresented in the dataset. We use Cam-CAN to demonstrate consistent brain age synthesis across the whole lifespan. *Alzheimer's Disease Neuroimaging Initiative (ADNI)* (Petersen et al., 2010) is a longitudinal dataset of subjects being cognitively normal (CN), mildly cognitive impaired (MCI) and with AD. We use ADNI to demonstrate brain image synthesis, conditioned on different health states. Since ADNI has longitudinal data, we used these data to quantitatively evaluate the quality of synthetically aged images. We chose 786 subjects as training (279 CN, 260 MCI, 247 AD), and 136 subjects as testing data (49 CN, 46 MCI, 41 AD). The age difference between baseline and followup images in the testing set is 2.93 ± 1.35 years.

Pre-processing: All volumetric data are skull-stripped using DeepBrain[7], and linearly registered to MNI 152 space using FSL-FLIRT (Woolrich et al., 2009). We normalise brain volumes by clipping the intensities to $[0, V_{99.5}]$, where $V_{99.5}$ is the 99.5% largest intensity value within each volume, and then rescale the resulting intensities to the range $[-1, +1]$. Such intensity pre-processing also helps alleviate potential intensity harmonisation issues between datasets in a manner that creates no leakage (see footnote on section 5.2.3 why this is important). We select the middle 60 axial slices from each volume, and crop each slice to the size of [208, 160]. During training, we only use *cross-sectional* data, i.e. one subject only has one volume of a certain age. During testing, we use the longitudinal ADNI data covering more than 2 years, and discard data where images are severely misaligned due to registration errors.

Benchmarks: We compare with the following benchmarks[8]:
Conditional GAN: We use a conditional image-to-image translation approach (Mirza and Osindero, 2014) and train different Conditional GANs for transforming young images to different older age groups. Therefore, a single model of ours is compared with age-group specific Conditional GANs.
CycleGAN: We use CycleGAN (Zhu et al., 2017), with two translation paths: from 'young' to 'old' to 'young', and from 'old' to 'young' to 'old'. Similarly to Conditional GAN, we train several CycleGANs for different target age groups.

[6] In our previous work (Xia et al., 2019), Eq. 2 did not have the $a_o > a_i$ constraint and would randomly include the case of $a_o = a_i$ to encourage self-reconstruction. However, as shown in Section 5.4, we found that stronger regularisation is necessary.

[7] https://github.com/iitzco/deepbrain

[8] We also used the official implementation of Milana (2017); however, our experiments confirmed the poor image quality reported by the author.

CAAE: We compare with Zhang et al. (2017), a recent paper for face ageing synthesis. We use the official implementation[9], modified to fit our input image shape. This method used a Conditional Adversarial Autoencoder (CAAE) to perform face ageing synthesis by concatenating a one-hot age vector with the bottleneck vector. They divided age into discrete age groups.

Implementation details: The optimization function is:

$$L = \min_{G} \max_{D} \lambda_1 L_{GAN} + \lambda_2 L_{ID} + \lambda_3 L_{rec}, \quad (4)$$

where $\lambda_1 = 1$, $\lambda_2 = 100$ and $\lambda_3 = 10$ are hyper-parameters used to balance each loss. The λ parameters are chosen experimentally. We chose λ_2 as 100 following Baumgartner et al. (2018) and Xia et al. (2019), and λ_3 as a smaller value to put emphasis on synthesis rather than self-reconstruction.

To train our model, we divide subjects into a young group and an old group, and randomly draw a image $\mathbf{x}_i$ the young group and an image $\mathbf{y}_o$ from the old group to synthesise the aged image $\hat{\mathbf{x}}_o$ (of $\mathbf{x}_i$) with target age a_o and health state h_o (of those corresponding to $\mathbf{y}_o$). Here $\hat{\mathbf{x}}_o$ is the synthetically aged version of $\mathbf{x}_i$, and the target age a_o and health state h_o are the same as those of the selected old sample $\mathbf{y}_o$. Afterwards, $\mathbf{y}_o$ and $\hat{\mathbf{x}}_o$ are fed into the discriminator as real and fake samples, respectively. Note that for all samples $a_o > a_i$, and h_o could be different than h_i. Since Alzheimer's Disease is an irreversible neurodegenerative disease, we select samples where the input health status is not worse than the output health status. We train all methods for 600 epochs. We update the generator and discriminator iteratively (Arjovsky et al., 2017; Goodfellow et al., 2014). Since the discriminator of a Wasserstein GAN needs to be close to optimal during training, we update the discriminator for 5 iterations per generator update. Initially, for the first 20 epochs, we update the discriminator for 50 iterations per generator update. We use Keras (Chollet et al., 2015) and train with Adam (Kingma and Ba, 2015) with a learning rate of 0.0001 and decay of 0.0001. Code is available at `https://github.com/xiat0616/BrainAgeing`.

Evaluation metrics: To evaluate the quality of synthetically aged images, we first use the longitudinal data from ADNI dataset. We select follow-up studies covering >2 years to allow observable neurodegenerative changes to happen. We used standard definitions of *mean squared error* (MSE), *peak signal-to-noise ratio* (PSNR) and *structural similarity* (SSIM) of window length of 11 (Wang et al., 2003) to evaluate the closeness of the predicted images to the ground-truth.

Predicted age difference (PAD) as a metric: Longitudinal data in ADNI only cover a short time span, i.e. the age difference between baseline and followup images is only several years. To assess output even when we do not have corresponding follow-up ground truth, we use a proxy metric of apparent age to evaluate image output. To develop our proxy metric, we first train a learning based age predictor to assess apparent brain age. We pre-train a VGG-like (Simonyan and Zisserman, 2015) network to predict age from brain images, then use this age predictor, f_{pred}, to estimate the apparent age of output images. To train this age predictor f_{pred} we combine Cam-CAN and healthy (CN) ADNI training data to ensure good age coverage. On a held out testing set it achieves a MAE of 5.1 ± 3.1 years. When the held out dataset is restricted to ADNI healthy subjects alone, MAE is 3.9 ± 2.8 years.

We use the difference between the predicted and desired target age to assess how close the generated images are to the (desired) target age. Formally, our proxy metric *predicted age difference* (PAD) is:

$$PAD = \mathbb{E}_{\hat{\mathbf{x}}_o \sim \mathcal{X}_o} |f_{pred}(\hat{\mathbf{x}}_o) - a_o|, \quad (5)$$

where f_{pred} is the trained age predictor, $\hat{\mathbf{x}}_o$ is the synthetically aged image, and a_o is the target age. Here we choose to measure the mean absolute error as we want to avoid the neutralization of positive and negative errors. By adopting PAD, we have a quantitative metric to measure the quality of synthetic results in terms of age accuracy. Observe that PAD does not compare baseline and follow-up scans. Given that the age predictor is only trained on healthy data it will estimate age on how normal brains will look like. Thus, it should capture when brain ageing acceleration occurs in AD, as others have demonstrated before us (Cole et al., 2019). This will increase PAD error when we synthesise with AD or MCI target health state, but given that we use PAD to compare between different methods this error should affect all methods. With advances in brain ageing estimation Peng et al. (2021) the fidelity of PAD will also increase. Here since we use PAD to compare across methods even a biased estimator is still a useful method of comparison.

Statistics: All results are obtained on testing sets, and we show average and standard deviation (std, as subscript on all tables), estimated by sample mean and variance on the testing set. We use **bold** font to denote the best performing method (for each metric) and an asterisk (*) to denote statistical significance. We use a paired t-test (at 5% level assessed via permutations) to test the null hypothesis that there is no difference between our methods and the best performing benchmark.

5. Results and discussion

We start by showing quantitative and qualitative results on ADNI with detailed evaluation demonstrating quality of the generated images. We then train our model on Cam-CAN to show long-term brain ageing synthesis. We conclude with ablation studies to illustrate the effect of design choices.

5.1. Brain ageing synthesis on different health states (ADNI)

In this section, we train and evaluate our model on ADNI dataset, which contains CN, MCI and AD subjects. Our model is trained only on *cross-sectional* data. The results and discussions are detailed below.

5.1.1. Quantitative results

The quantitative results are shown in Table 1, employing the metrics defined in Section 4. For ADNI we also obtained a non-learned naïve baseline that simply calculates performance comparing ground-truth baseline and follow-up images. The naïve

[9] https://zzutk.github.io/Face-ageing-CAAE/

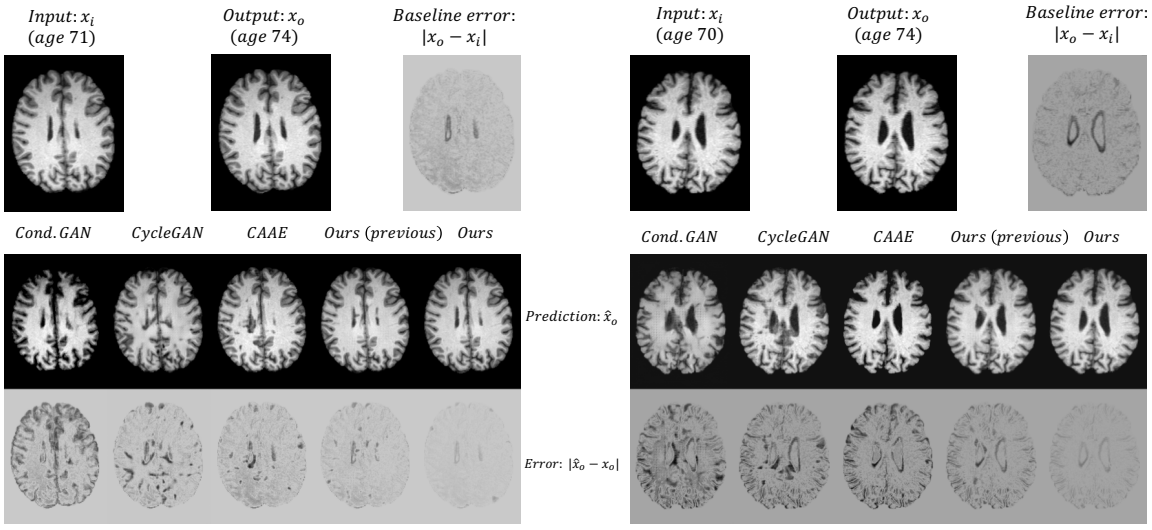


Fig. 6. Example results of subjects with ground-truth follow-up studies. We predict output $\hat{x}_o$ from input x_i using benchmarks and our method. We also show errors between the outputs and the ground-truths as $|\hat{x}_o - x_o|$. We can observe that our method achieves the most accurate results outperforming our previous method (Xia et al., 2019) and benchmarks. As a comparison, we also visualized the difference between inputs and ground-truth outputs as $|x_o - x_i|$. For more details see text.

Table 1. Quantitative evaluation on ADNI dataset (testing set) for several metrics. We report average and std (as subscript) with BOLD, * indicating best performance and statistical significance, respectively (see Section 4).

	SSIM	PSNR	MSE	PAD	PAD (CN)	PAD (MCI)	PAD (AD)
Naïve baseline	$0.71_{\pm 0.09}$	$22.1_{\pm 3.3}$	$0.097_{\pm 0.013}$	$7.2_{\pm 3.9}$	$6.3_{\pm 3.8}$	$6.8_{\pm 3.9}$	$8.7_{\pm 4.0}$
Cond. GAN	$0.39_{\pm 0.08}$	$14.2_{\pm 3.5}$	$0.202_{\pm 0.012}$	$9.5_{\pm 4.7}$	$8.7_{\pm 4.8}$	$9.1_{\pm 4.7}$	$10.9_{\pm 4.7}$
CycleGAN	$0.46_{\pm 0.07}$	$16.3_{\pm 3.3}$	$0.193_{\pm 0.008}$	$9.7_{\pm 5.1}$	$8.9_{\pm 4.9}$	$9.4_{\pm 5.2}$	$11.0_{\pm 5.2}$
CAAE	$0.64_{\pm 0.07}$	$20.3_{\pm 2.9}$	$0.114_{\pm 0.011}$	$5.4_{\pm 4.5}$	$4.4_{\pm 4.3}$	$5.1_{\pm 4.4}$	$6.9_{\pm 4.7}$
Ours-previous	$0.73_{\pm 0.06}$	$23.3_{\pm 2.2}$	$0.081_{\pm 0.009}$	$5.0_{\pm 3.7}$	$4.0_{\pm 3.5}$	$4.6_{\pm 3.6}$	$6.6_{\pm 4.0}$
Ours	$\mathbf{0.79^{*}_{\pm 0.06}}$	$\mathbf{26.1^{*}_{\pm 2.6}}$	$\mathbf{0.042^{*}_{\pm 0.006}}$	$\mathbf{4.2^{*}_{\pm 3.9}}$	$\mathbf{3.1^{*}_{\pm 3.6}}$	$\mathbf{3.9^{*}_{\pm 3.8}}$	$\mathbf{5.9^{*}_{\pm 4.2}}$

baseline result is obtained by subtracting from the followup the baseline (input) image. We involve this non-learned baseline as a lower bound to check if the proposed algorithm synthesises images that are closer to the follow-up than the baseline images or not. As reported in Section 4, the average age prediction error (MAE) of the age predictor on the ADNI testing data is 3.9 years. Estimating PAD separately for CN, MCI and AD testing subjects (see Table 1) shows that the best PAD results are obtained on healthy (CN) data. This is expected as the age predictor used to estimate PAD it is trained on healthy data only. However, this bias affects all methods, and thus still allows comparisons between them. Indeed, we can observe that our method achieves the best results in all metrics, with second best being the previous (more simple incarnation) (Xia et al., 2019) of the proposed model. Embedding health state improves performance, because it permits the method to learn an ageing function specific for each state as opposed to the one learned by the method in Xia et al. (2019). The other benchmarks achieve a lower performance compared to the baseline. The next best results are achieved by CAAE (Zhang et al., 2017), where age is divided into 10 age groups and represented by a one-hot vector. To generate the aged images at the target age (the age of the follow-up studies), we use the age group to which the target age belongs, i.e. if the target age is 76, then we choose the age group of age 75-78. We see the benefits of encoding age into ordinal vectors, where the difference between two vectors positively correlates with the difference between two ages in a finely-grained fashion. CycleGAN and Conditional GAN achieve the poorest results unsurprisingly, since conditioning here happens explicitly by training separate models according to different age groups.

5.1.2. Qualitative results

Visual examples on two images from ADNI, are shown in Fig. 6. For both examples, our method generates most accurate predictions, followed by our previous method Xia et al. (2019), offering visual evidence to the observations above. The third best results are achieved by CAAE, where we can see more errors between prediction $\hat{x_o}$ and ground-truth x_o. CycleGAN and Conditional GAN produced the poorest output images, with observable structural differences from ground-truth, indicating loss of subject identity. We can also observe that the brain ventricle is enlarged in our results and the difference between x_i and x_o is reduced, which is consistent with known knowledge that ventricle increases during ageing.

Furthermore, we show visual results of the same subject at different target health states h_o, in Fig. 7. We observe that for all h_o, the brain changes gradually as age (a_o) increases. However, the ageing rate varies based on health state (h_o). Specifically, when h_o is CN, ageing is slower than that of MCI and AD,

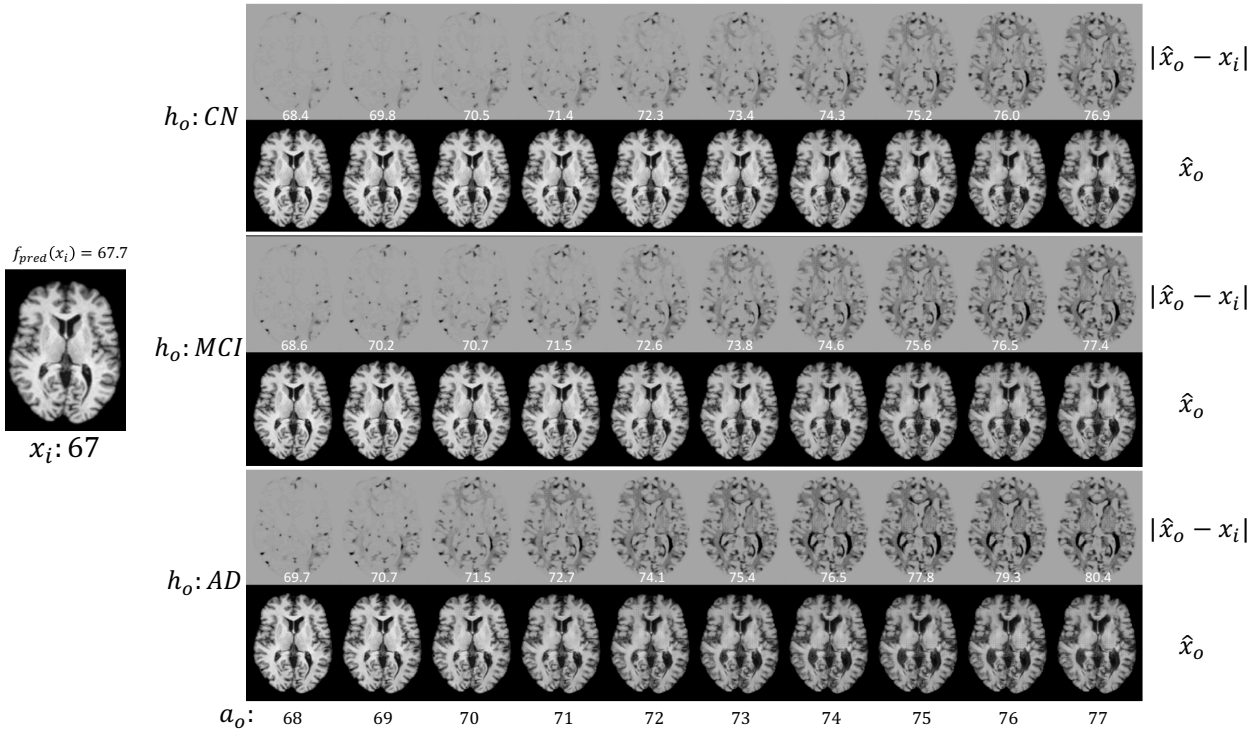


Fig. 7. Brain ageing progression for a healthy (CN) subject $\mathbf{x}_i$ (at age 67) from ADNI dataset. We synthesise the aged images $\hat{\mathbf{x}}_o$ at different target ages a_o on different health states h_o: CN, MCI and AD, respectively. We also visualise the difference between $\mathbf{x}_i$ and $\hat{\mathbf{x}}_o$, $|\hat{\mathbf{x}}_o - \mathbf{x}_i|$, and show the predicted (apparent) ages of $\hat{\mathbf{x}}_o$ as obtained by our pre-trained age predictor (white text overlaid on each difference image). For more details see text.

as one would expect; when h_o is AD, ageing changes accelerate. We also report the estimated ages of these synthetic images as predicted by f_{pred}. While these results show one instance, we synthesised aged images of different health status from 49 ADNI test set CN subjects, with target ages 10 years older than the original age. We then used f_{pred} to estimate the age of these synthetic images. We find that on average, synthetic AD images are 4.9 ± 2.3 years older than the target age, whereas synthetic MCI and CN images are 1.8 ± 2.0 and 1.5 ± 2.1 years older than the target age, respectively. These observations are consistent with prior findings that AD accelerates brain ageing (Petersen et al., 2010). We also observe that the gray/white matter contrast decreases as age increases, which is consistent with existing findings (Westlye et al., 2009; Farokhian et al., 2017).

5.2. Does our model capture realistic morphological changes of ageing and disease?

Here we want to assess whether our model captures known ageing-related brain degeneration. It is known that brain ageing is related to gray matter reduction in middle temporal gyrus (MTG) (Guo et al., 2014; Sullivan et al., 1995). We wanted to assess whether synthetic volumes could act as drop-in replacements of ground-truth follow-up in assessing MTG gray matter volume change. We focus here on the MTG as this is well covered by the range of slices we use to train our synthesis method. Before we proceed we first illustrate that we can synthesise 3D volumes slice-by-slice, and then show that our model can capture realistic morphological changes.

5.2.1. Volume synthesis by stacking 2D slices

We show that, even with our 2D model, we can still produce 3D volumes that show consistency. We applied our model on 2D axial slices and obtained a 3D volume by stacking the synthetic slices. An example result of a stacked synthetic 3D volume in sagittal and coronal views is shown in Figure 8. Compared to the respective ground-truth from the same subject, we observe that both sagittal and coronal views of the synthetic volume look realistic and are close to the follow-up images. Note here that our model is trained only on 2D axial slices, for which we chose middle 60 slices from each volume. Our model uses a residual connection and thus makes minimal changes to the regions affected by age instead of synthesising the whole brain image. This helps preserve details and continuity across slices. These results illustrate that we can produce 3D volumes that maintain consistency in different views.

5.2.2. Do we capture morphological changes?

We use an ℓ_1 loss to restrict (in pixel space) the amount of change between input and output images. This is computationally efficient, but to show that it also restricts deformations, we measure the deformation between input (baseline) and synthetic or ground-truth follow-up images in ADNI. We obtain for each subject the baseline image $\mathbf{x}_i$, the follow-up image $\mathbf{x}_o$ and the synthetic image $\hat{\mathbf{x}}_o$, respectively. We first rigidly register $\mathbf{x}_o$ to $\mathbf{x}_i$ using Advanced Normalization Tools (ANTs) (Avants et al., 2008) rigid transformation. Then we non-rigidly register $\mathbf{x}_o$ to $\mathbf{x}_i$ and obtain the Jacobian determinant map $\mathbf{J}_{\mathbf{x}_o \to \mathbf{x}_i}$ that describes the transformation from $\mathbf{x}_o$ to $\mathbf{x}_i$, using ANTs "SyN" transformation (Avants et al., 2008). Similarly, we obtain $\mathbf{J}_{\hat{\mathbf{x}}_o \to \mathbf{x}_i}$

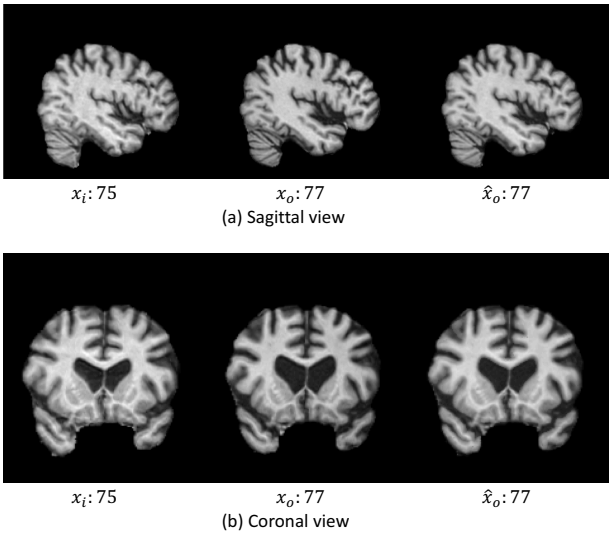


Fig. 8. Example results of a synthetic 3D volume $\hat{\mathbf{x}}_o$ in sagittal view (top) and coronal view (bottom) from ADNI dataset. Here we construct the 3D volume by stacking the 2D synthetic axial slices of our model. From left to right are slices from a baseline volume $\mathbf{x}_i$, the corresponding follow-up volume $\mathbf{x}_o$, and the stacked synthetic volume $\hat{\mathbf{x}}_o$.

that describes the non-linear transformation from $\hat{\mathbf{x}}_o$ to $\mathbf{x}_i$. Fig. 9 shows an example of the Jacobian maps for one subject.

From Fig. 9, we observe that $\mathbf{J}_{\hat{\mathbf{x}}_o \to \mathbf{x}_i}$ is close to $\mathbf{J}_{\mathbf{x}_o \to \mathbf{x}_i}$. To quantify their difference, we calculate the mean relative error between the Jacobian determinant maps, defined as:

$$E = \mathbb{E}_{\mathbf{x}_i \sim \mathcal{X}_i, \mathbf{x}_o \sim \mathcal{X}_o, \hat{\mathbf{x}}_o \sim \hat{\mathcal{X}}_o} \frac{\|\mathbf{J}_{\mathbf{x}_o \to \mathbf{x}_i} - \mathbf{J}_{\hat{\mathbf{x}}_o \to \mathbf{x}_i}\|_1}{\|\mathbf{J}_{\mathbf{x}_o \to \mathbf{x}_i}\|_1}, \tag{6}$$

where $\|.\|_1$ is 1-norm of matrices. We find the mean relative error to be 3.49% on the testing set of 136 images. Similarly, we perform the same evaluations for the results of Conditional GAN, CycleGAN, CAAE and our previous method, and find the mean relative errors to be 9.87%, 8.76%, 5.91% and 4.43%, respectively. Both qualitative and quantitative results suggest that synthetically aged images capture realistic morphological changes of the brain ageing process.

5.2.3. *Measuring middle temporal gyrus (MTG) gray matter atrophy.*

We further evaluate the quality of the synthetic results by assessing if they can act as a drop-in replacement to real data in a

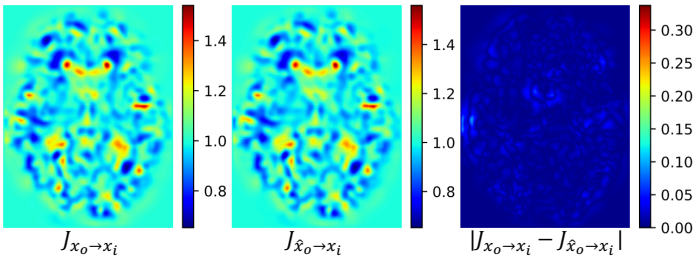


Fig. 9. An example of Jacobian determinant maps for a subject. From left to right are the Jacobian determinant maps $\mathbf{J}_{\mathbf{x}_o \to \mathbf{x}_i}$, $\mathbf{J}_{\hat{\mathbf{x}}_o \to \mathbf{x}_i}$, and the error map between them: $|\mathbf{J}_{\mathbf{x}_o \to \mathbf{x}_i} - \mathbf{J}_{\hat{\mathbf{x}}_o \to \mathbf{x}_i}|$.

Table 2. Analysis of MTG gray matter relative change between baseline and follow-up real or synthetic. Mean and std are reported as well as the corresponding F-statistic of a one-way ANOVA test (between relative change and patient type), with asterisk indicating significance ($p < 0.05$).

	Relative change	F-statistic
real (RC_{real})	$-0.071_{\pm 0.0096}$	4.008^*
synthetic (RC_{syn})	$-0.083_{\pm 0.0099}$	4.539^*

simple study of brain atrophy. We performed ageing synthesis with our model on 136 ADNI testing subjects, such that for each subject we have: a baseline image $\mathbf{x}_i$; a real follow-up image $\mathbf{x}_o$; and a synthetic image $\hat{\mathbf{x}}_o$ of the same target age and health state as of $\mathbf{x}_o$. We then assembled volumes by stacking 2D images. Then we affinely registered both $\mathbf{x}_o$ and $\hat{\mathbf{x}}_o$ and the Human-Brainnetome based on Connectivity Profiles (HCP) atlas (Fan et al., 2016) to $\mathbf{x}_i$. After that, we obtained the MTG segmentation of $\mathbf{x}_i$, $\mathbf{x}_o$ and $\hat{\mathbf{x}}_o$ by means of label propagation from HCP using the deformation fields. Then we obtained the gray matter segmentation of $\mathbf{x}_i$ using FSL-FAST (Zhang et al., 2001). The gray matter segmentation of x_o and $\hat{\mathbf{x}}_o$ was subsequently obtained by non-linearly registering $\mathbf{x}_o$ and $\hat{\mathbf{x}}_o$ to $\mathbf{x}_i$ and propagating anatomical labels using ANTs (Avants et al., 2008). These steps yield the MTG gray matter volume of $\mathbf{x}_i$, $\mathbf{x}_o$ and $\hat{\mathbf{x}}_o$, termed as $\mathbf{V}_{base}$, $\mathbf{V}_{fol}$, and $\mathbf{V}_{syn}$, respectively. Then, we calculate the relative change between $\mathbf{V}_{base}$ and $\mathbf{V}_{fol}$ as $RC_{real} = \frac{\mathbf{V}_{fol} - \mathbf{V}_{base}}{\mathbf{V}_{base}}$, and the relative change for synthetic data as $RC_{syn} = \frac{\mathbf{V}_{syn} - \mathbf{V}_{base}}{\mathbf{V}_{base}}$. We repeat this for several subjects in three patient type groups, i.e. CN (49), MCI (46) and AD (41).

We expect, following Guo et al. (2014) and Sullivan et al. (1995), to see a statistical relationship between patient type and RC_{real} when assessed with a one-way analysis of variance (ANOVA). If a similar relationship is shown also with synthetic data RC_{syn}, it will demonstrate that for this statistical test, our synthetic data can act as a drop-in replacement to real data, and as such have high quality and fidelity.

The results are summarised in Table 2, where we report also the F-statistic of the omnibus one-way ANOVA test. We observe that MTG gray matter volume reduces in both real and synthetic volumes. This indicates that our synthetic results achieve good quality and similar statistical conclusions can be drawn using real or synthetic data in this simple atrophy study.

5.3. *Long term brain ageing synthesis*

In this section, we want to see how our model performs in long term brain ageing synthesis. As ADNI dataset only covers old subjects, we use Cam-CAN dataset which contains subjects of all ages. We train our model with Cam-CAN dataset where no longitudinal data are available, but evaluate it on the longitudinal part of ADNI to assess the generalisation performance of our model when trained on one dataset and tested on another.

5.3.1. *Qualitative results*

In Fig. 10, we demonstrate the simulated brain ageing process throughout the whole lifespan, where the input images are two young subjects from Cam-CAN. We observe that the output gradually changes as a_o increases, with ventricular enlargement

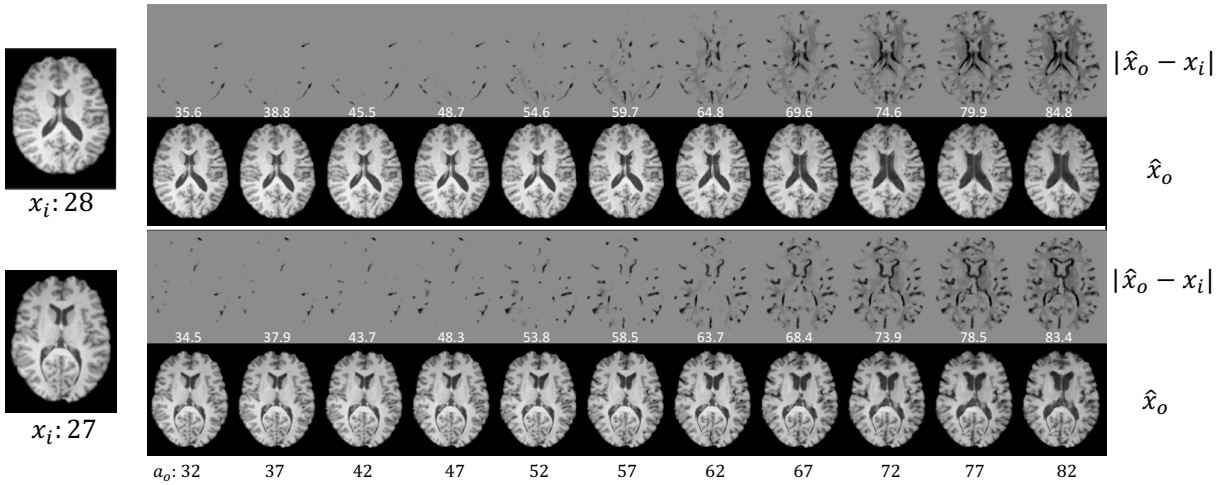
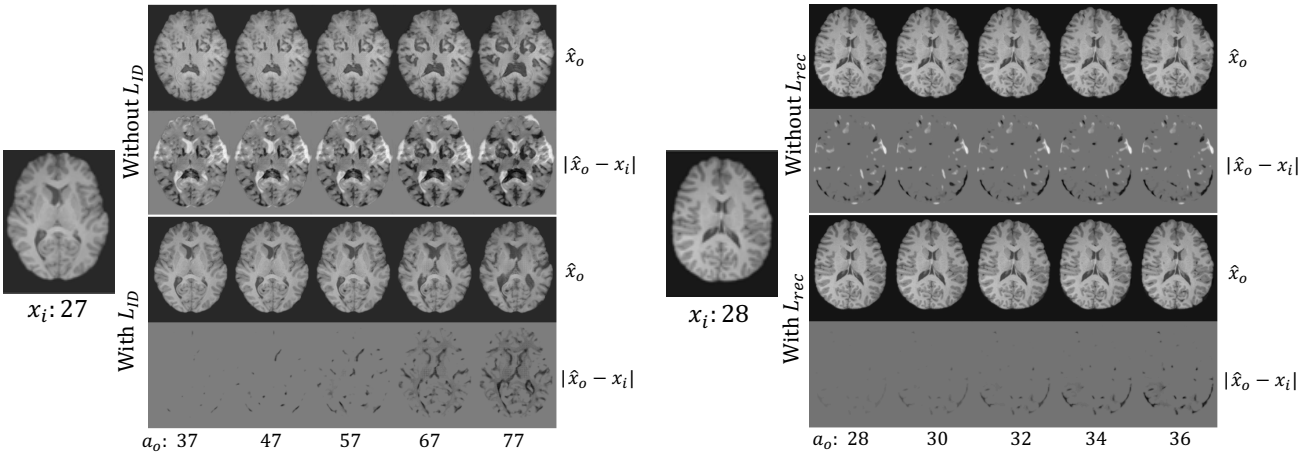


Fig. 10. Long-term brain ageing synthesis on Cam-CAN dataset. We synthesise the aged images $\hat{\mathbf{x}}_o$ at different target ages a_o and show the difference between input images $\mathbf{x}_i$ and $\hat{\mathbf{x}}_o$, $|\hat{\mathbf{x}}_o - \mathbf{x}_i|$, and show the predicted (apparent) ages of $\hat{\mathbf{x}}_o$ as obtained by our pre-trained age predictor (white text overlaid on each difference image). Note here $\mathbf{x}_i$: N means an input image at age N. For more details see text.



(a) Ablation study on *identity loss* L_{id} (b) Ablation study on *self-reconstruction loss* L_{rec}

Fig. 11. Ablation studies for loss components. Left: ablation study of L_{ID}. Top row shows that without L_{ID}, the network can lose the subject identity. Bottom row shows that the use of L_{ID} can enforce the preservation of subject identity, such that the changes as ages are smooth and consistent. Right: ablation study on L_{rec}. When L_{rec} is not used (top two rows), there are sudden changes at the beginning of ageing progression simulation (even at the original age), which hinders the preservation of subject identity. In contrast, when L_{rec} is used (bottom two rows), the ageing progression is smoother, which demonstrates better identity preservation. Note here x_i: N means an input image at age N.

and brain tissue reduction. This pattern is consistent with previous studies (Good et al., 2001; Mietchen and Gaser, 2009), showing that our method learns to synthesise the ageing brain throughout the lifespan even trained on cross-sectional data.[10] Fig. 10 offers only only a qualitative visualization to show the potential of life-time simulation. We cannot quantitatively evaluate the quality of these synthetic images due to the lack of longitudinal data in Cam-CAN. However, both the previous section on ADNI where we train and test on ADNI, and the next section, where we use longitudinal ADNI as testing set we but train on Cam-CAN data, offer considerable quantitative experiments.

5.3.2. Quantitative results (generalisation performance on ADNI)

To evaluate how accurate our longitudinal estimation is, even when training with cross sectional data from *another* dataset, we train a model on Cam-CAN and evaluate it on ADNI. We use the longitudinal portion of ADNI data, and specifically only the CN cohort, to demonstrate generalisation performance.[11] Given an image of ADNI we use our Cam-CAN trained model to predict an output at the same age as the real follow up image. We compare our prediction with the ground truth follow up im-

[10] We observe checkerboard artefacts near the ventricles after target age 67. Such artefacts are a known problem in computer vision and mostly likely due to the use of deconvolutional layers in the decoder (Oramas et al., 2018).

[11] We purposely do not use any intensity harmonisation that uses both datasets, e.g. histogram matching. Such methods will leak information from ADNI to Cam-CAN. Any leakage would skew (to our favour) the generalisation ability which we want to avoid. Thus, our experiments also indirectly evaluate how design choices (e.g. using a residual connection in the generator) help with differences in intensities between datasets.

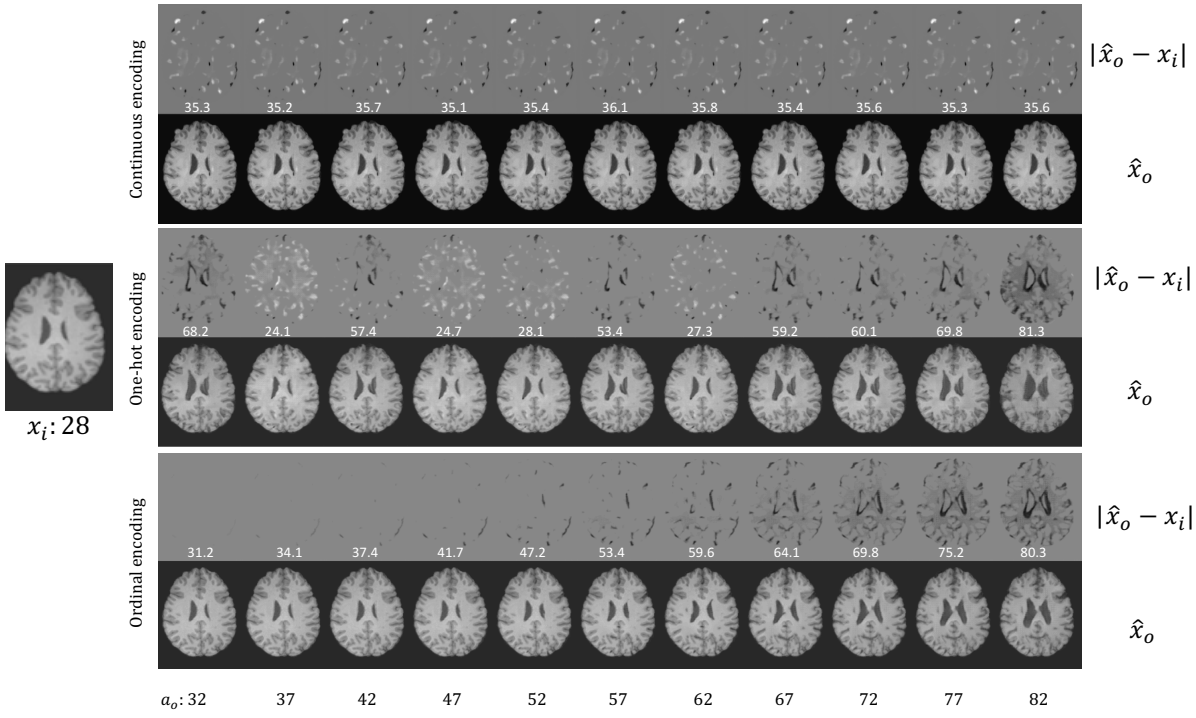


Fig. 12. Example results for *continuous*, *one-hot* and *ordinal* encoding on the Cam-CAN dataset for an image ($\mathbf{x}_i$) of a 28 year old subject. We synthesise aged images $\hat{\mathbf{x}}_o$ at different target ages a_o. We also show the difference between $\mathbf{x}_i$ and $\hat{\mathbf{x}}_o$, $|\hat{\mathbf{x}}_o - \mathbf{x}_i|$, and report estimated age (white text overlaid at the bottom of each difference image). The proposed ordinal encoding shows consistent and progressive changes.

Table 3. Quantitative evaluation of methods trained on Cam-CAN and evaluated on ADNI.

	SSIM	PSNR	MSE	PAD
Cond. GAN	$0.38_{\pm 0.12}$	$13.9_{\pm 4.2}$	$0.221_{\pm 0.021}$	$11.3_{\pm 5.6}$
CycleGAN	$0.42_{\pm 0.09}$	$14.4_{\pm 3.8}$	$0.212_{\pm 0.016}$	$10.2_{\pm 5.5}$
CAAE	$0.59_{\pm 0.10}$	$19.3_{\pm 3.9}$	$0.121_{\pm 0.012}$	$5.9_{\pm 4.7}$
Ours-previous	$0.68_{\pm 0.08}$	$22.7_{\pm 2.8}$	$0.095_{\pm 0.014}$	$5.3_{\pm 3.8}$
Ours	$\mathbf{0.74}^{*}_{\pm \mathbf{0.08}}$	$\mathbf{24.2}^{*}_{\pm \mathbf{2.7}}$	$\mathbf{0.043}^{*}_{\pm \mathbf{0.009}}$	$\mathbf{5.0}_{\pm \mathbf{3.6}}$

age (in the ADNI dataset). The results are shown in Table 3. We observe that though our model is trained and evaluated on different datasets, it still achieves comparable results with those of Table 1 and outperforms benchmarks.

5.4. Ablation studies

We ablate loss components, explore different conditioning mechanisms, and explore latent space dimensions.

5.4.1. Effect of loss components

We demonstrate the effect of L_{ID} and L_{rec} by assessing the model performance when each component is removed. In Table 4 we show quantitative results on ADNI dataset. In Fig. 11 we illustrate qualitative results on Cam-CAN dataset to visualise the effect. We can observe that the best results are achieved when all loss components are used. Specifically, without L_{ID}, the synthetic images lost subject identity severely throughout the whole progression, i.e. the output image appears to come from a different subject; without L_{rec}, output images suffer from sudden changes at the beginning of progression, even when $a_o = a_i$. Both quantitative and qualitative results show that the design of L_{ID} and L_{rec} improves preservation of subject identity and enables more accurate brain ageing simulation.

5.4.2. Effect of different embedding mechanisms

We investigate the effect of different embedding mechanisms. Our embedding mechanism is described in Section 3. We considered to encode age as a normalized *continuous* value (between 0 and 1) or using a *one-hot* vector, which was then concatenated with the latent vector at the bottleneck. The qualitative results are shown in Fig. 12. We can see that when age is represented as a normalized *continuous* value, this is ignored by the network, thus generating similar images regardless of changes in target age a_o. When we use *one-hot* vectors to encode age, the network still generates realistic images, but the ageing progression is not consistent, i.e. synthetic brains appear to have ventricle enlarging or shrinking in random fashion across age. In contrast, with *ordinal encoding*, the model simulates the ageing process consistently. This observation is confirmed by the estimated ages of the output images by f_{pred}.

We also compare with an embedding strategy where we concatenate $\mathbf{h}_o$, $\mathbf{a}_d$ and the bottleneck latent vector $\mathbf{v}_1$ together, and

Table 4. Ablations on using different combinations of cost functions.

	SSIM	PSNR	MSE
L_{GAN}	$0.55_{\pm 0.14}$	$18.4_{\pm 3.7}$	$0.132_{\pm 0.013}$
$L_{GAN} + L_{rec}$	$0.62_{\pm 0.12}$	$19.6_{\pm 3.2}$	$0.089_{\pm 0.014}$
$L_{GAN} + L_{ID}$	$0.74_{\pm 0.07}$	$24.3_{\pm 2.5}$	$0.074_{\pm 0.010}$
$L_{GAN} + L_{ID} + L_{rec}$	$\mathbf{0.79}^{*}_{\pm \mathbf{0.08}}$	$\mathbf{26.1}^{*}_{\pm \mathbf{2.6}}$	$\mathbf{0.042}^{*}_{\pm \mathbf{0.006}}$

Table 5. Quantitative results of different embedding mechanisms.

	SSIM	PSNR	MSE	PAD
One-hot	$0.54_{\pm 0.14}$	$17.3_{\pm 3.8}$	$0.177_{\pm 0.014}$	$9.7_{\pm 4.9}$
$\text{concat}_{\text{all}}$	$0.74_{\pm 0.09}$	$23.9_{\pm 2.9}$	$0.065_{\pm 0.011}$	$5.2_{\pm 3.9}$
Ours	$\mathbf{0.79}^{*}_{\pm \mathbf{0.08}}$	$\mathbf{26.1}^{*}_{\pm \mathbf{2.6}}$	$\mathbf{0.042}^{*}_{\pm \mathbf{0.006}}$	$\mathbf{5.0}_{\pm \mathbf{3.6}}$

Table 6. Quantitative results of different choices of the v_2 dimension.

	SSIM	PSNR	MSE	PAD
65×1	$0.73_{\pm 0.09}$	$23.6_{\pm 3.1}$	$0.065_{\pm 0.012}$	$5.6_{\pm 4.1}$
260×1	$0.76_{\pm 0.10}$	$24.9_{\pm 2.9}$	$0.055_{\pm 0.012}$	$5.3_{\pm 3.8}$
130×1 (ours)	$\mathbf{0.79}^{*}_{\pm \mathbf{0.08}}$	$\mathbf{26.1}^{*}_{\pm \mathbf{2.6}}$	$\mathbf{0.042}^{*}_{\pm \mathbf{0.006}}$	$\mathbf{5.0}_{\pm \mathbf{3.6}}$

the concatenated vector is processed by the Decoder to generate the output image. We refer to this embedding strategy as $concat_{all}$. Results on ADNI are shown in Table 5. We found with $concat_{all}$, the network tends to ignore the health state vector $\mathbf{h}_o$ and only use $\mathbf{a}_d$. This can be caused by the dimensional imbalance between $\mathbf{h}_o$ (2×1) and $\mathbf{a}_d$ (100×1). When *one-hot encoding* is used, performance deteriorates even more.

5.4.3. Effect of latent space dimension

We explored whether latent dimension affects performance. We altered the length of the latent vector ($\mathbf{v}_2$) from 130×1 to twice smaller/larger and compared the corresponding models on ADNI. Our findings are shown in Table 6. We find that our choice (130×1) achieved the best results. It appears that too small is not enough to represent image information well, and too large can cause dimension imbalance.

5.4.4. Comparison with longitudinal model

To compare our method with models that use longitudinal data[12], we created a new benchmark where we train a fully supervised generator using only longitudinal ADNI data. The results are shown in Table 7. We see that our method has slightly better performance than the longitudinal model. This is because the proposed model is trained on 786 subjects (cross-sectional data), while the longitudinal model is trained on a longitudinal cohort of ADNI of 98 subjects. This illustrates the benefit of using cross-sectional data. Note that our SSIM results are similar to those presented in Ravi et al. (2019a).

5.4.5. Data augmentation for AD classification

We explore whether we can use our model to generate synthetic data used to augment training sets for training an Alzheimer's disease classifier. We select 200 ADNI subjects as training data (100 AD, 100 CN), 40 subjects as validation data (20 AD, 20 CN), and 80 (40 AD, 40 CN) subjects as testing data. For each subject, there are 60 2D slices. Next, we train classifiers of the same VGG architecture to classify AD and CN subjects varying the composition of the training data combining real and synthetic data obtained by our generator.

[12]Ravi et al. (2019a) also synthesise subject-specific aged brain images based on longitudinal data relyin on complex biological constraints. Our attempts to replicate their work (an official code is not available) have been fruitless.

Table 7. Quantitative results of a longitudinal benchmark and our method.

	SSIM	PSNR	MSE
Longitudinal	$0.72_{\pm 0.09}$	$24.2_{\pm 3.0}$	$0.076_{\pm 0.013}$
Ours	$0.79_{\pm 0.08}$	$26.1_{\pm 2.6}$	$0.042_{\pm 0.006}$

We always evaluate the classifiers on the same testing set. The synthetic data are generated from the training set using our proposed method by randomly selecting target ages larger than the original ages. As shown in Table 8, we first train classifiers only on real data varying the size of the training data (1st and 2nd rows). Then we compose mixed sets of the same size of 200 subjects varying the ratio of real vs. synthetic data (3rd and 4th rows), e.g. 10%+90% means this set is composed of 10% real data and 90% synthetic data. Note here the 90% synthetic data are not generated from the whole training set, but from the 10% real data.

We can observe that when training on 10% of real training data, the accuracy reduces by almost 40% compared to when using the full training data. However, the performances improve when synthetic data are involved. The results demonstrate that our method can be used as data augmentation to improve AD classification especially when the training data are not sufficient.

Furthermore, we perform another experiment to demonstrate our model's potential to improve the classification accuracy for specific age groups and thus target augmentation to treat data imbalance. We evaluate the classification model trained with 100% real data on test set subjects of age 65 to 70 years old. We find an accuracy of 67.2%, which is much lower than the overall accuracy (89.5%, Table 8). This may be likely due to training data imbalance: we have only 5 training subjects with age between 65 and 70 yrs. To alleviate this data imbalance, we use our model to generate 25 synthetic subjects with target ages between 65 and 70 yrs from younger subjects in the training set. Then we train a new AD classifier on 100% real data and the 25 synthetic subjects, and evaluate its performance on the same testing and age group. Accuracy now increases to 80.1% a substantial change from 67.2%.

6. Conclusion

We present a method that learns to simulate subject-specific aged images *without* longitudinal data. It relies on a Generator to generate the images and a Discriminator that captures the joint distribution of brain images and clinical variables, i.e. age and health state (AD status). We propose an embedding mechanism to encode the information of age and health state into our network, and age-modulated and self-reconstruction losses to preserve *subject identity*. We present qualitative results showing that our method is able to generate consistent and realistic images conditioned on the target age and health state. We evaluate with longitudinal data from ADNI for image quality and *age accuracy*. We demonstrate on ADNI and Cam-CAN datasets that our model outperforms benchmarks both qualitatively and quantitatively and, via a series of ablations, illustrate the importance of each design decision.

Table 8. Quantitative results of VGG-based AD/CN classifiers trained on different datasets. The first two rows show results when trained on varying size of real training data, e.g. 10% means this model is trained on 10% of the real training data; the last two rows show results when trained on mixed datasets with different ratios of real and synthetic data, e.g. 10%+90% means this model is trained on 10% real training data and 90% synthetic data.

Real data	10%	30%	50%	70%	100%
Accuracy (%)	51.3	55.7	64.6	74.0	
Real data + synthetic data	10%+90%	30%+70%	50%+50%	70%+30%	89.5
Accuracy (%)	58.7	64.0	72.6	81.7	

Potential applications. The proposed method has several potential applications. For example, a common problem in longitudinal studies is missing data due to patient dropout or poor-quality scans. The proposed method offers an opportunity to impute missing data at any time point. Furthermore, when there is insufficient longitudinal training data, the proposed method can be used to include cross-sectional data within a study. The simple experiment in 5.2.3 shows a glimpse of this potential.

This in turn will make further clinical analysis of ageing patterns, e.g. to evaluate the incidence of white matter hyperintensities (Wardlaw et al., 2015), and large studies into neurodegenerative diseases, possible. Finally, from an AI perspective we advocated earlier on the paper about the importance of capturing and understanding current state from a machine learning perspective. In fact, recently this has been cast in a causal inference and counterfactual setting Pawlowski et al. (2020). While our work didn't explicitly use a causal inference framework, our generated outputs can be seen as counterfactuals.

Limitations and avenues for improvement. The notion of subject identity is context specific and we do note that others in the literature also follow the same simple assumptions we make. We do agree though that identity should be defined as what remains invariant under ageing and neurodegenerative disease. Although we used several losses to help preserve subject identity of synthetic aged images, there is no guarantee that a subject's identity will be preserved, and new losses or mechanisms that could further improve identity preservation will be of high value. Unfortunately, without access to large data where we exhaustively explore all possible combinations of variables that we want to be equivalent (to identity) or invariant (to age, pathology) preservation of identity can only be proxied. The proposed model only considers predicting older brain images from young ones. However, performing both brain ageing and rejuvenation will provide more utility but will require more advanced design of encoding and network architecture. The proposed method allows for change of health status between input and output images. However, it does not model change of health state in between input and output. This is a common limitation of current works in this area (Ravi et al., 2019a; Pawlowski et al., 2020; Rachmadi et al., 2019). A potential solution is recursive image synthesis: generating a suitable intermediate image before generating the desired target output of an older age and state. Advances in architectures that improve image quality will enable such recursive image generation in the future. Conditioning mechanisms that reliably embed prior information into neural networks enabling finer control over the outputs of models are of considerable interest in deep learning. In this paper we design a simple yet effective way to encode both *age* (continuous) and *AD status* (ordinal) factors into the image generation network. However, as classification of MCI is challenging, use of further (fine-grained) clinical information (e.g. clinical score) to reflect health status can be of benefit. Incorporating additional clinical variables, e.g. gender, genotypes, etc., can become inefficient with our current approach as it may involve more dense layers. While new techniques are available (Huang and Belongie, 2017; Perez et al., 2018; Park et al., 2019; Lee et al., 2019) and some prior examples on few conditioning variables (Jacenkow et al., 2019) or disentanglement (Chartsias et al., 2019) are promising, their utility in integrating clinical variables, and replacing the need for ordinal pre-encoding of continuous or ordinal variables, with imaging data is under investigation. Although we used brain data, the approach could be extended to other organs. Furthermore, here we focus on the use of cross-sectional data to train a model to predict aged brain images. If longitudinal data are also available, e.g. within a large study aggregating several data sources, model performance could be further improved by introducing supervised losses; however, adding more losses requires that they are well balanced –a known problem in semi-supervised learning (Sener and Koltun, 2018). We showed that our synthetic volumes (composed by stacking 2D images) can achieve good quality. In all our attempts with 3D architectures, the parameter space exploded due to their size. We expect that advances in network compression (Han et al., 2016) will eventually permit us to adapt a 3D design which should further improve visual quality and consistency of the approach and allow us to repeat our atrophy analysis not only in the MTG.

7. Acknowledgements

This work was supported by the University of Edinburgh by PhD studentships to T. Xia and A. Chartsias. This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1. We thank Nvidia for donating a Titan-X GPU. S.A. Tsaftaris acknowledges the support of Canon Medical and the Royal Academy of Engineering and the Research Chairs and Senior Research Fellowships scheme (grant RCSRF1819/8/25). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (NIH grant U01 AG024904) and Department of Defense ADNI (award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks, in: ICML.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis 12, 26–41.

Baumgartner, C.F., Koch, L.M., Can Tezcan, K., Xi Ang, J., Konukoglu, E., 2018. Visual feature attribution using wasserstein GANs, in: CVPR, pp. 8309–19.

Bowles, C., Gunn, R., Hammers, A., Rueckert, D., 2018. Modelling the progression of Alzheimer's disease in MRI using generative adversarial networks, in: Medical Imaging 2018: Image Processing.

Camara, O., Schweiger, M., Scahill, R.I., Crum, W.R., Sneller, B.I., Schnabel, J.A., Ridgway, G.R., Cash, D.M., Hill, D.L., Fox, N.C., 2006. Phenomenological model of diffuse global and regional atrophy using finite-element methods. TMI 25, 1417–1430.

Chartsias, A., Joyce, T., Papanastasiou, G., Semple, S., Williams, M., Newby, D.E., Dharmakumar, R., Tsaftaris, S.A., 2019. Disentangled representation learning in cardiac image analysis. Medical Image Analysis 58, 101535.

Chollet, F., et al., 2015. Keras. `https://keras.io`.

Cole, J.H., Franke, K., 2017. Predicting age using neuroimaging: innovative brain ageing biomarkers. Trends in Neurosciences 40, 681–690.

Cole, J.H., Leech, R., Sharp, D.J., Initiative, A.D.N., 2015. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. Annals of neurology 77, 571–581.

Cole, J.H., Marioni, R.E., Harris, S.E., Deary, I.J., 2019. Brain age and other bodily 'ages': implications for neuropsychiatry. Molecular Psychiatry 24, 266.

Cole, J.H., Ritchie, S.J., Bastin, M.E., Hernández, M.V., Maniega, S.M., Royle, N., Corley, J., Pattie, A., Harris, S.E., Zhang, Q., et al., 2017. Brain age predicts mortality. Molecular Psychiatry .

Coleman Jr, L.G., Liu, W., Oguz, I., Styner, M., Crews, F.T., 2014. Adolescent binge ethanol treatment alters adult brain regional volumes, cortical extracellular matrix protein and behavioral flexibility. Pharmacology Biochemistry and Behavior 116, 142–151.

Costafreda, S.G., Dinov, I.D., Tu, Z., Shi, Y., Liu, C.Y., Kloszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., et al., 2011. Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment. Neuroimage 56, 212–219.

Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S., 2010. Population shape regression from random design data. IJCV .

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., et al., 2016. The human brainnetome atlas: a new brain atlas based on connectional architecture. Cerebral cortex 26, 3508–3526.

Farokhian, F., Yang, C., Beheshti, I., Matsuda, H., Wu, S., 2017. Age-related gray and white matter changes in normal adult brains. Aging and disease 8, 899.

Fjell, A.M., Walhovd, K.B., 2010. Structural brain changes in aging: courses, causes and cognitive consequences. Reviews in the Neurosciences 21, 187–222.

Franke, Katja and Ziegler, Gabriel and Klöppel, Stefan and Gaser, Christian and Alzheimer's Disease Neuroimaging Initiative and others, 2010. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. Neuroimage 50, 883–892.

Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S., 2001. A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage 14, 21–36.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: NeurIPS, pp. 2672–2680.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C., 2017. Improved training of Wasserstein GANs, in: NeurIPS, pp. 5767–5777.

Guo, Y., Zhang, Z., Zhou, B., Wang, P., Yao, H., Yuan, M., An, N., Dai, H., Wang, L., Zhang, X., et al., 2014. Grey-matter volume as a potential feature for the classification of alzheimer's disease and mild cognitive impairment: an exploratory study. Neuroscience bulletin 30, 477–489.

Han, S., Mao, H., Dally, W.J., 2016. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding, in: International Conference on Learning Representations (ICLR).

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: CVPR.

Huang, X., Belongie, S., 2017. Arbitrary style transfer in real-time with adaptive instance normalization, in: CVPR, pp. 1501–1510.

Huizinga, W., Poot, D.H., Vernooij, M.W., Roshchupkin, G., Bron, E., Ikram, M.A., Rueckert, D., Niessen, W.J., Klein, S., Initiative, A.D.N., et al., 2018. A spatio-temporal reference model of the aging brain. NeuroImage 169, 11–22.

Jacenkow, G., Chartsias, A., Mohr, B., Tsaftaris, S.A., 2019. Conditioning convolutional segmentation architectures with non-imaging data, in: MIDL.

Jack, C., Petersen, R.C., Xu, Y., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1998. Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. Neurology 51, 993–999.

Jonsson, B., Bjornsdottir, G., Thorgeirsson, T., Ellingsen, L., Walters, G.B., Gudbjartsson, D., Stefansson, H., Stefansson, K., Ulfarsson, M., 2019. Deep learning based brain age prediction uncovers associated sequence variants. bioRxiv , 595801.

Khanal, B., Ayache, N., Pennec, X., 2017. Simulating longitudinal brain MRIs with known volume changes and realistic variations in image intensity. Frontiers in Neuroscience 11, 132.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. International Conference on Learning Representations .

Lee, M.C.H., Petersen, K., Pawlowski, N., Glocker, B., Schaap, M., 2019. Tetris: Template transformer networks for image segmentation with shape priors. TMI 38, 2596–2606.

López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., Kroemer, G., 2013. The hallmarks of aging. Cell 153, 1194–1217.

Lorenzi, M., Pennec, X., Frisoni, G.B., Ayache, N., Initiative, A.D.N., et al., 2015. Disentangling normal aging from alzheimer's disease in structural magnetic resonance images. Neurobiology of aging 36, S42–S52.

Mattson, M.P., Arumugam, T.V., 2018. Hallmarks of brain aging: adaptive and pathological modification by metabolic states. Cell Metabolism 27, 1176–1199.

Mietchen, D., Gaser, C., 2009. Computational morphometry for detecting changes in brain structure due to development, aging, learning, disease and evolution. Frontiers in Neuroinformatics 3, 25.

Milana, D., 2017. Deep generative models for predicting Alzheimer's disease progression from MR data. Master's thesis. Politecnico Di Milano.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 .

Modat, M., Simpson, I.J., Cardoso, M.J., Cash, D.M., Toussaint, N., Fox, N.C., Ourselin, S., 2014. Simulating neurodegeneration through longitudinal population analysis of structural and diffusion weighted MRI data, in: MICCAI, Springer. pp. 57–64.

Oramas, J., Wang, K., Tuytelaars, T., 2018. Visual explanation by interpretation: Improving visual feedback capabilities of deep neural networks, in: International Conference on Learning Representations.

Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y., 2019. Semantic image synthesis with spatially-adaptive normalization, in: CVPR, pp. 2337–2346.

Pawlowski, N., Coelho de Castro, D., Glocker, B., 2020. Deep structural causal models for tractable counterfactual inference. Advances in Neural Information Processing Systems 33.

Peng, H., Gong, W., Beckmann, C.F., Vedaldi, A., Smith, S.M., 2021. Accurate brain age prediction with lightweight deep neural networks. Medical Image Analysis 68, 101871.

Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A., 2018. FILM: Visual reasoning with a general conditioning layer, in: AAAI.

Petersen, R.C., Aisen, P., Beckett, L.A., Donohue, M., Gamst, A., Harvey, D.J., Jack, C., Jagust, W., Shaw, L., Toga, A., et al., 2010. Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. Neurology 74, 201–209.

Pieperhoff, P., Südmeyer, M., Hömke, L., Zilles, K., Schnitzler, A., Amunts, K., 2008. Detection of structural changes of the human brain in longitudinally acquired mr images by deformation field morphometry: methodological analysis, validation and application. NeuroImage 43, 269–287.

Rachmadi, M.F., Valdés-Hernández, M.d.C., Makin, S., Wardlaw, J., Komura, T., 2020. Automatic spatial estimation of white matter hyperintensities evolution in brain mri using disease evolution predictor deep neural networks. Medical Image Analysis , 101712.

Rachmadi, M.F., Valdés-Hernández, M.d.C., Makin, S., Wardlaw, J.M., Komura, T., 2019. Predicting the Evolution of White Matter Hyperintensities in Brain MRI using Generative Adversarial Networks and Irregularity Map.

MICCAI .

Ravi, D., Alexander, D.C., Oxtoby, N.P., 2019a. Degenerative Adversarial NeuroImage Nets: Generating Images that Mimic Disease Progression. MICCAI .

Ravi, D., Blumberg, S.B., Mengoudi, K., Xu, M., Alexander, D.C., Oxtoby, N.P., 2019b. Degenerative adversarial neuroimage nets for 4d simulations: Application in longitudinal mri. arXiv preprint arXiv:1912.01526 .

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: MICCAI, Springer. pp. 234–241.

Sener, O., Koltun, V., 2018. Multi-task learning as multi-objective optimization, in: NeurIPS, pp. 527–538.

Serag, A., Aljabar, P., Ball, G., Counsell, S.J., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D., 2012. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. NeuroImage 59, 2255–2265.

Sharma, S., Noblet, V., Rousseau, F., Heitz, F., Rumbach, L., Armspach, J.P., 2010. Evaluation of brain atrophy estimation algorithms using simulated ground-truth data. Medical image analysis 14, 373–389.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, in: ICLR.

Sivera, R., Delingette, H., Lorenzi, M., Pennec, X., Ayache, N., Initiative, A.D.N., et al., 2019. A model of brain morphological changes related to aging and alzheimer's disease from cross-sectional assessments. NeuroImage 198, 255–270.

Sullivan, E.V., Marsh, L., Mathalon, D.H., Lim, K.O., Pfefferbaum, A., 1995. Age-related decline in mri volumes of temporal lobe gray matter but not hippocampus. Neurobiology of aging 16, 591–606.

Taubert, M., Draganski, B., Anwander, A., Müller, K., Horstmann, A., Villringer, A., Ragert, P., 2010. Dynamic properties of human brain structure: learning-related changes in cortical areas and associated fiber connections. Journal of Neuroscience 30, 11670–11677.

Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., Dixon, M., Tyler, L.K., Henson, R.N., 2017. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. NeuroImage 144, 262–9.

Wang, Z., Simoncelli, E.P., Bovik, A.C., 2003. Multiscale structural similarity for image quality assessment, in: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, IEEE. pp. 1398–1402.

Wardlaw, J.M., Valdés Hernández, M.C., Muñoz-Maniega, S., 2015. What are white matter hyperintensities made of? relevance to vascular cognitive impairment. Journal of the American Heart Association 4, e001140.

Wegmayr, V., Hörold, M., Buhmann, J.M., 2019. Generative Aging of Brain MR-Images and Prediction of Alzheimer Progression, in: German Conference on Pattern Recognition, Springer. pp. 247–260.

Westlye, L.T., Walhovd, K.B., Dale, A.M., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Greve, D.N., Fischl, B., Fjell, A.M., 2009. Increased sensitivity to effects of normal aging and alzheimer's disease on cortical thickness by adjustment for local variability in gray/white contrast: a multi-sample mri study. Neuroimage 47, 1545–1557.

Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M., 2009. Bayesian analysis of neuroimaging data in FSL. Neuroimage 45, S173–S186.

Xia, T., Chartsias, A., Tsaftaris, S.A., Initiative, A.D.N., 2019. Consistent brain ageing synthesis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 750–758.

Zecca, L., Youdim, M.B., Riederer, P., Connor, J.R., Crichton, R.R., 2004. Iron, brain ageing and neurodegenerative disorders. Nature Reviews Neuroscience 5, 863.

Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. TMI 20, 45–57.

Zhang, Y., Shi, F., Wu, G., Wang, L., Yap, P.T., Shen, D., 2016. Consistent spatial-temporal longitudinal atlas construction for developing infant brains. TMI 35, 2568–2577.

Zhang, Z., Song, Y., Qi, H., 2017. Age progression/regression by conditional adversarial autoencoder, in: CVPR, pp. 5810–5818.

Zhao, Q., Adeli, E., Honnorat, N., Leng, T., Pohl, K.M., 2019. Variational autoencoder for regression: Application to brain aging analysis. MICCAI .

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: ICCV, pp. 2223–2232.

Ziegler, G., Dahnke, R., Gaser, C., 2012. Models of the aging brain structure and individual decline. Frontiers in Neuroinformatics 6, 3.