A Holistic Vision: Modeling Patient Trajectories in Longitudinal Medical Imaging

Nico Disch[1,2,3*] **Saikat Roy**[1,3] **Constantin Ulrich**[1,7,8] **Robin Peretzke**[1,8]
David Zimmerer[1,2] **Jens Kleesiek**[5] **Rainer Stiefelhagen**[6] **Klaus Maier-Hein** [1,2,4]
[1]Division of Medical Image Computing, German Cancer Research Center
[2]HIDSS4Health - Helmholtz Information and Data Science School for Health,
Karlsruhe/Heidelberg, Germany [3]Faculty of Mathematics and Computer Science,
University of Heidelberg, Heidelberg, Germany
[4] Pattern Analysis and Learning Group, Department of Radiation Oncology,
Heidelberg University Hospital, Heidelberg, Germany
[5]Institute for AI in Medicine (IKIM), University Hospital Essen
[6] Karlsruhe Institute of Technology, Karlsruhe, Germany
[7] National Center for Tumor Diseases (NCT), NCT Heidelberg,
A partnership between DKFZ and University Medical Center Heidelberg, Germany
[8] Medical Faculty Heidelberg, University of Heidelberg, Heidelberg, Germany
[*]nico.disch@dkfz.de

Abstract

In medical data analysis, human practitioners have long excelled at adopting a holistic approach, considering a wide range of patient information, including multiple imaging sources and evolving medical histories. A prime example can be found in tumor boards, where, among others, radiologists evaluate a multitude of images while taking into account dynamic patient narratives. However, within the domain of medical image analysis, the current focus often narrows down to individual images, or if longitudinal data is used, the task is to infer non-dense predictions, like classification. However, if we can leverage multiple time points and model a patient trajectory, we can predict patient status on an image level at an arbitrary time point in the future. It could not only lead to better predictions for the current time point, i.e. for image segmentation, but it can also lead to a more holistic approach in medical image analysis, more akin to the human approach. In response to this disparity, our work motivates the need for longitudinal medical image analysis and we present a model that can deal with sparse and irregular longitudinal series, and without sacrificing generality, generate images.

Keywords: longitudinal medical imaging, generative AI, neural ODEs

1 Introduction

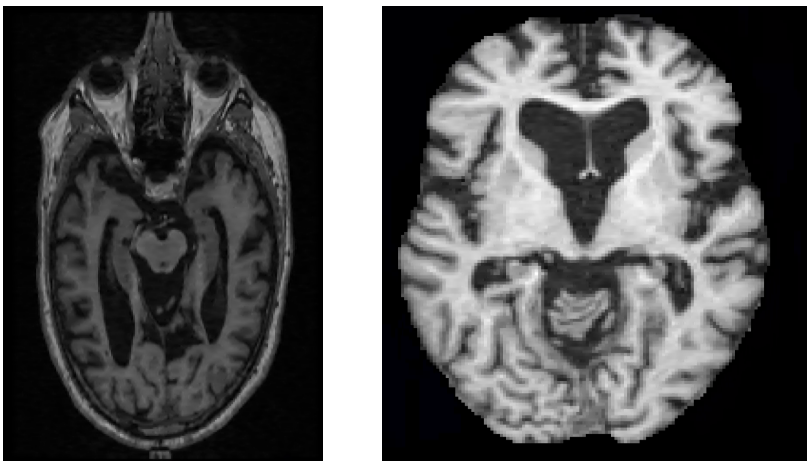
In recent years, the realm of artificial intelligence has witnessed a meteoric rise, such as large language models (LLMs) Brown et al. [4] or image generation models like in Rombach et al. [13]. The medical field, too, has experienced profound transformations owing to the integration of AI, as explored in Anaya-Isaza et al. [2]. Yet, while these successes are prominent, they are still far away from human analysis of medical problems. One such example are tumor boards, where the experts consider patient histories and multiple modalities, as described in as described in Okasako and Bernstein [11] and Mano et al. [9]. It is important to note that longitudinal medical images, in particular, exhibit more sparsity compared to natural videos due to e.g. factors such as the high cost of individual MRI acquisitions, which can reach up to $1430 in the USA as of 2017 [1]. Furthermore, the acquisition

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

of medical images is often irregular, as scans are conducted based on resource availability or in medical emergencies to diagnose underlying issues, and not during regular time intervals. Yet, the majority of deep learning models primarily analyze individual time points, or if longitudinal data is used, it is typically used for downstream tasks like as classification. So in this study, we take initial steps towards a more holistic patient analysis, by modelling patient trajectories to generate images of patient images at arbitrary time steps.

2 Data and pre-processing steps

Our research benefits from the extensive dataset offered by the Alzheimer's Disease Neuroimaging Initiative (ADNI), which is accessible at `https://adni.loni.usc.edu/`. While our research is not exclusively reliant on the ADNI dataset, including Alzheimer patients, its longitudinal aspects make it particularly suitable for our investigation of modeling longitudinal trajectories in the context of medical imaging. It's worth noting that we do not presume our data originates from an Alzheimer's dataset; instead, we utilize it for methodological validation. For repeatability of experiments, we describe data pre-processing, and we will later publish the model and the pre-processing here [1]. For the pre-processing, we convert the DICOM images using plastimatch, if applicable, registration to the MNI space using flirt Jenkinson et al. [7], brain extraction using HD-BET from Isensee et al. [6]. An illustrative pre-processed example is depicted in 1, and these pre-processing steps facilitate our model's training process.



(a) raw MRI (b) pre-processed MRI

Figure 1: This figure illustrates brain MRI data processing from the ADNI database. (a) displays a raw image obtained from a random patient. (b) shows a pre-processed MRI brain, where we've registered the image to the MNI space and extracted the brain.

3 Related work

We draw inspiration from the following approaches that serve as the foundation for our work. One similar work is found in Benoît and Durrleman [3], which leverages the ADNI database as its foundation. Their model is based on a Variational Autoencoder (VAE), augmented with a pre-defined geodesic function in the bottleneck whose parameters are learned during training. The use of defining a geodesic simplifies the latent trajectory, which yields among others, values like disease onset for Alzheimer disease, but it also needs prior knowledge of the specific disease. In Petersen et al. [12], their work focuses on predicting segmentation masks for glioblastoma, with a model that is independent from disease-specific predictions. However, since their data is non-public, their model cannot be used directly as a baseline out of the box. Also, apart from the constraints imposed by data limitations, their architecture choice has an added drawback; it is limited to inputs of $2D + t$ due to

[1] `https://github.com/NicoDisch/LongitudinalMedicalImaging`

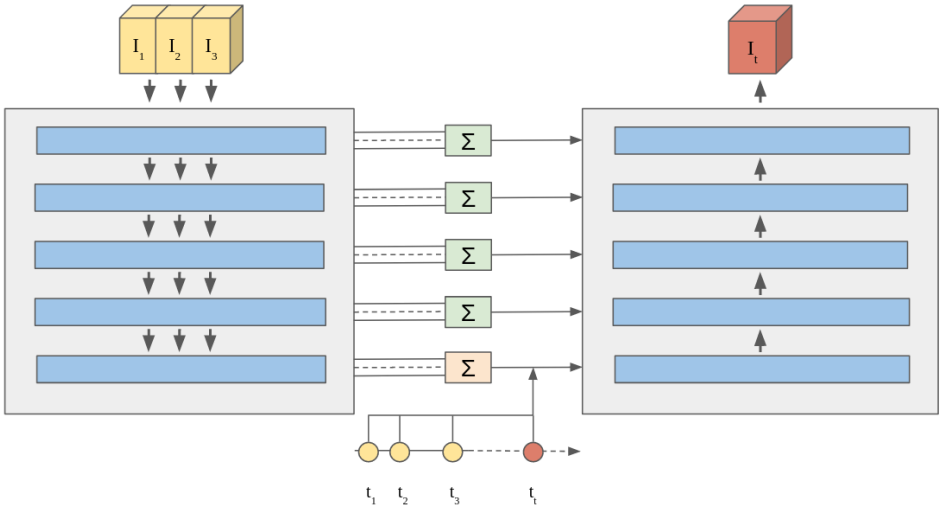


Figure 2: Basic network architecture overview: Yellow squares are the input images. Red is the target image. t_i with $i \in \{1, \ldots, n\}$ are the input times, and t_t is the target time. Green boxes are the summation of image representations. The gray boxes are encoder and decoder, with the blue boxes are convolutional blocks. The downward and upward arrows represent max-pooling and bilinear upsampling respectively. The orange box is summing up the representations of each image representation and concatenation of time representations. Note that we are free to chose a latent function instead of the summation, e.g. a neural ODE or a predefined trajectory.

memory constraints. Lastly, we mention the Neural ODEs, which are capable of modeling temporal irregular data, as described in Chen et al. [5]. Another noteworthy mention is research of Yıldız et al. [14], where they model trajectories of sparse image time series, with a VAE as a backbone for their model. Furthermore, we note the existence of other methods for longitudinal data, that solve a different task, for example they predict e.g. disease onset, as in Lu et al. [8]. And in the work of Zhang et al. [15], they implicitly take time as regularly sampled, which is not always applicable in the medical field, but it simplifies network architecture and it can be used in tasks that have regularly sampled time steps.

4 Methods

Our model architecture takes inspiration from the attentive segmentation process (ASP) introduced in Petersen et al. [12]. An illustration of the adapted model can be seen in 2. These adaption can be seen in the skip-connections and in the bottleneck. For the bottleneck, we employ a neural ODE after the summation, similar to the work of Norcliffe et al. [10]. Furthermore, we are not restricted to an neural ODE, but in principle we can include the bottleneck function as in the work of Benoît and Durrleman [3], to leverage prior knowledge of disease progression. We note here, that for the ASP model we use attention over whole image representations, so we are restricted to $2D$ in our experiments, due to memory constraints. In their original work, the ASP consists of an attention mechanism over the image representations, a linearly upscaled context and target time representations. Instead, we employ a naive summation of image representations, all else being equal to the ASP model.

The ASP model employs an attention across complete image representations over the lower resolutions. However, due to memory limitations, we are confined to $2D$ in our experiments. Petersen et al. [12] original model utilizes an attention mechanism over image representations, with context and target times linearly upscaled. This constrains the number of context times and restricts us to $2D$ images. To mitigate this, we have opted for a straightforward summation of image representations, while maintaining other aspects of the ASP model.

5 Results and further work

The results can be seen in table 1. While preliminary, these results indicate that a simple addition of image representation aren't enough to entirely replace the attention mechanism, but that the performance loss is marginal. We also note that the additional neural ODE does not significantly improve the results of either the ASP or our model, which can be attributed that the trajectory of Alzheimer's disease is roughly a sigmoid, akin to natural aging. To isolate the performance of the ODE within the network, a comparison with alternative trajectory models such as neural ODEs, Controlled Differential Equations (CDEs), or neural flows within a VAE is warranted. Furthermore, exploring different state representations for the neural ODE, such as an RNN in place of the current summation approach for image representations, may yield better insights. Most importantly, expanding our datasets beyond the ADNI cohort is a crucial priority. As our experimentation progresses, we aim to scale our model to $3D$, since the simple skip connection in our model scales better than the attention. The scalability is important for medical applications, since MRI and CT images are $3D$, and considering medical images in their whole is superior in their performance, as explored in Zhang et al. [16]. Even though there are more and more methods going from $2D$ to $3D$, called $2.5D$, it remains challenging and still less performant.

Table 1: Results comparing different model architectures. The task is predicting a $2D$ slice (randomly chosen from transverse, coronal, or sagittal axis) of a MRI patient with a randomly chosen time-point. The loss is comparing the reconstruction error between the network prediction and the ground truth. Trained using the $SSIM$ loss.

Model description	MSE	$SSIM$
ASP (adapted)[12]	0.071	0.691
ASP (adapted) + neural ODE	0.070	0.691
Summed representations 2	0.072	0.695
Summed representations + neural ODE	0.066	0.692

We recognize that our current method, which simply sums image representations, is basic. The temporal encoding is confined to the bottleneck, representing significantly less information than the pixels, which are encoded in every level of the network. Future work will aim to balance the encoding of pixel and temporal information, possibly drawing inspiration from the ASP model, but with a focus on efficiency. Ultimately, we hope our work will not only spur new model innovations but also encourage the collection of longitudinal data, providing accessible baselines for its utilization.

Acknowledgments

The present contribution is supported by the Helmholtz Association under the joint research school "HIDSS4Health – Helmholtz Information and Data Science School for Health."

References

[1] Average price MRI in selected countries 2017 | Statista. URL https://www.statista.com/statistics/312020/price-of-mri-diagnostics-by-country/.

[2] A. Anaya-Isaza, L. Mera-Jiménez, and M. Zequera-Diaz. An overview of deep learning in medical imaging. *Informatics in Medicine Unlocked*, 26:100723, 1 2021. ISSN 2352-9148. doi: 10.1016/J.IMU.2021.100723.

[3] S. Benoît and S. Durrleman. Progression Models for Imaging Data with Longitudinal Variational Auto Encoders. In Q. Wang Linwei
}and Dou, F. P. Thomas, S. Stefanie, and L. Shuo, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 3–13, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-16431-6.

[4] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language Models are Few-Shot Learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

[5] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural Ordinary Differential Equations. 6 2018. URL http://arxiv.org/abs/1806.07366.

[6] F. Isensee, M. Schell, I. Pflueger, G. Brugnara, . David Bonekamp, . Ulf Neuberger, A. Wick, H.-P. Schlemmer, S. Heiland, W. Wick, M. Bendszus, K. H. Maier-Hein, and . P. Kickingereder. Automated brain extraction of multisequence MRI using artificial neural networks. 2019. doi: 10.1002/hbm.24750. URL https://onlinelibrary.wiley.com/doi/10.1002/hbm.24750.

[7] M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2): 825–841, 2002. ISSN 10538119. doi: 10.1016/S1053-8119(02)91132-8. URL https://pubmed.ncbi.nlm.nih.gov/12377157/.

[8] D. Lu, K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg. Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images. *Scientific Reports 2018 8:1*, 8(1):1–13, 4 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-22871-z. URL https://www.nature.com/articles/s41598-018-22871-z.

[9] M. S. Mano, F. T. Çitaku, and P. Barach. Implementing multidisciplinary tumor boards in oncology: a narrative review. *Future oncology (London, England)*, 18(3):375–384, 1 2022. ISSN 1744-8301. doi: 10.2217/FON-2021-0471. URL https://pubmed.ncbi.nlm.nih.gov/34787482/.

[10] A. Norcliffe, C. Bodnar, B. Day, J. Moss, and P. Liò. Neural ODE Processes. 3 2021. URL http://arxiv.org/abs/2103.12413.

[11] J. Okasako and C. Bernstein. Multidisciplinary Tumor Boards and Guiding Patient Care: The AP Role. *Journal of the Advanced Practitioner in Oncology*, 13(3):227–230, 4 2022. ISSN 21500878. doi: 10.6004/JADPRO.2022.13.3.9.

[12] J. Petersen, F. Isensee, G. Köhler, P. F. Jäger, D. Zimmerer, U. Neuberger, W. Wick, J. Debus, S. Heiland, M. Bendszus, P. Vollmuth, and K. H. Maier-Hein. Continuous-Time Deep Glioma Growth Models. 6 2021. URL http://arxiv.org/abs/2106.12917.

[13] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. Technical report. URL https://github.com/CompVis/latent-diffusion.

[14] Yıldız, M. Heinonen, and H. Lähdesmäki. ODE$ˆ2$VAE: Deep generative second order ODEs with Bayesian neural networks. 5 2019. URL http://arxiv.org/abs/1905.10994.

[15] L. Zhang, L. Lu, X. Wang, R. M. Zhu, M. Bagheri, R. M. Summers, and J. Yao. Spatio-Temporal Convolutional LSTMs for Tumor Growth Prediction by Learning 4D Longitudinal Patient Data. *IEEE transactions on medical imaging*, 39(4):1114–1126, 4 2020. ISSN 1558-254X. doi: 10.1109/TMI.2019.2943841. URL https://pubmed.ncbi.nlm.nih.gov/31562074/.

[16] Y. Zhang, Q. Liao, L. Ding, and J. Zhang. Bridging 2D and 3D segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of 2.5D solutions. *Computerized Medical Imaging and Graphics*, 99:102088, 7 2022. ISSN 0895-6111. doi: 10.1016/J.COMPMEDIMAG.2022.102088.