Disentangling Structure from Detail for Dementia Prediction

Daniela Layer[*] **and Jacob Schäfer**[*]

[*]these authors contributed equally to this work

Introduction

Variational autoencoders (VAEs) are a tool for unsupervised representation learning and have applications in various fields, including biomedical research [1, 2]. Due to their nature, they focus primarily on the detection of low-frequency features [3], which poses a challenge for the detection of subtle pathological changes [4, 5].

In this technical report, the impact of integrating brain segmentation masks into VAEs is investigated. Using the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [6], our research aims to improve the model's ability to capture detailed features relevant to the assessment of cognitive decline.

We propose three VAE architectures, compare them and evaluate their performance in terms of hyperparameter optimisation and downstream task evaluation.

Context and Related Work

VAEs, introduced in [1], are a form of unsupervised representation learning and consist of two main parts: The encoder and the decoder. These models work together, with the encoder linking the input data x to a latent representation z and the decoder converting z back to the original input space. Furthermore, the decoder serves as a guide or constraint for the encoder to capture meaningful representations of the data [7].

Current research on biomedical information using VAEs focuses primarily in two directions: Data augmentation and representation learning in the fields molecular design, sequence data set analyses and medical imaging and image analyses. VAE use on medical imaging dataset includes image classification, segmentation, restoration and reconstruction [2].

With VAEs, the focus of training is on the detection of low-frequency features, primarily reconstructing the largest number of pixels. This phenomenon is referred to as spectral bias [4]. However, this predominant focus on low-frequency features poses a challenge, especially in the context of disease detection, where high-frequency variations indicate an early stage of pathology [5].

Disentanglement is a process that involves breaking down complex data into features, which are independently meaningful [8]. In raw data, however, factors often appear to be closely intertwined. The challenge here is to construct the representations of the data in such a way that they provide features that are suitable for a variety of possible tasks and can deal with the reality of the intertwined variation factors [9].

Research Question

Our research focuses on investigating whether integrating brain segmentation masks as additional input to VAEs can enhance feature representation and enable the model to learn more detailed features in medical imaging. We are not aware of any publications that report to use segmentation mask as additional input to the VAE.

To investigate this we use the ADNI database which was established in 2004. The goal of ADNI is to investigate the use of biological markers and imaging to determine the decline of cognition in Cognitively Normal (CN), Mild Cognitive Impairment (MCI) and Alzheimer Disease (AD) [6, 10].

Alzheimer disease is the most common form of dementia. The International Classification of Diseases 11 (ICD-11) defines dementia as the presence of significant impairment in one or more cognitive domains. The impairment is related to age and the expected level of cognitive function and represents a deterioration from the previous functional level. Depending on the extent of neurocognitive and functional impairment, the severity of dementia can be categorized as mild, moderate or severe [11].

To create the used dataset, the imaging data available in ADNI was filtered for T1-MPRAGE scans and skull-stripping was performed using the HD-BET [12] tool. With a 1mm resolution a non-linear registration was conducted on the MNI152 templates using the FLIRT and FNIRT commands of the FSL tool [13]. Finally, the segmentation masks of brain regions were extracted using the Synthseg software [14].

Solution and Implementation[1]

To answer the research question we compared three different model architectures. All of them are based on the VAEs provided by Pytorch Lightning Bolts. The first architecture serves as a baseline and does not make use of additional segmentation masks. The second one adds the segmentation masks as additional layers to the encoder. Thus, the images and segmentation masks are entangled in this case. Lastly, the third architecture adds a second encoder following the same architecture which now separates the images and segmentation masks. For all those architectures we perform the following steps:

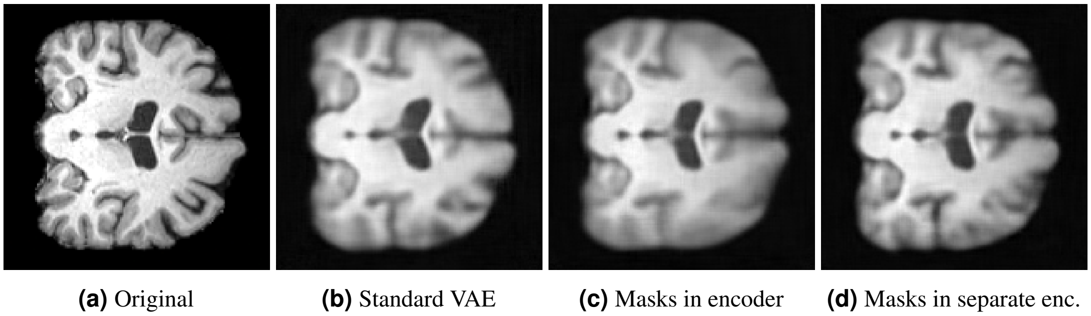
Data Preparation We use the ADNI dataset which contains 12791 3D brain scans and the corresponding segmentation masks of 1820 distinct patients. The dataset splitting is done based on the patients to avoid data leakage from training to validation and test sets. We use a split of 0.7, 0.15, 0.15 for the three datasets respectively. Since the number of scans per patient is variable, this results in dataset sizes 8847 (training), 1876 (validation), and 2014 (test).
Also, we perform several preprocessing steps per scan:

- Select the middle 2D slice on the coronal axis
- Crop the image to 160x160
- Clamp the values to the 0.98 percentile
- Normalize the image to mean zero and standard deviation one

The segmentation masks are stored as one hot encodings.

Hyperparameter Sweeps For each architecture, we optimize hyperparameters individually. We perform a bayesian sweep optimizing learning rate, batch size, latent dimensions, Kullback Leibler divergence coefficient, and whether to use an additional convolutional layer at the beginning and include a max pool layer. The sweep optimizes the validation reconstruction loss and includes 64 runs trained for 50 epochs each. With the best parameters, we train a final model for 400 epochs while saving the checkpoint with the lowest validation reconstruction loss.

Evaluation based on downstream task Finally, we evaluate for each architecture if a simple Logistic Regression Model can differentiate between cognitively normal and dementia based on the encodings. To do so, we keep the same dataset split but remove all entries which are marked as mild cognitive impairment or don't include a diagnosis. Note that this reduces the number of scans to 4642 (training), 998 (validation), and 1005 (test). As people age, their risk of developing dementia increases and on average, women tend to live longer than men [15, 16]. Therefore, we analyzed whether there could be a correlation in the data between dementia, age, and gender in predicting the onset of dementia. However, demographic factors like sex and age did not turn out to be predictive and we did not take additional measures for that (see Table 2). We use balanced accuracy as a metric for the classification task.



(a) Original **(b)** Standard VAE **(c)** Masks in encoder **(d)** Masks in separate enc.

Figure 1. Reconstructions of one of the validation images for our three different architectures.

[1] Link to GitHub repository: https://github.com/jacob271/Disentangling-Structure-from-Detail-for-Dementia-Prediction

Evaluation

As a result from the hyperparameter sweeps, we obtained one final model per architecture. See Table 1 for an overview of the final parameters as well as the performance of the models. The performance metrics are measured for the step where validation reconstruction loss was lowest.

Parameter/Metric	Standard VAE	Masks in encoder	Masks in separate encoder
learning rate	0.0001	0.0002	0.0003
batch size	64	64	32
latent dimensions	234	232	242
kl coefficient	0.048	0.049	0.042
Use first conv. layer	true	true	true
Use max pool layer	true	false	true
train recon. loss	0.062	0.072	0.056
val. recon. loss	0.0847	0.0864	0.0844
test recon. loss	0.1053	0.108	0.098

Table 1. This table shows the final configuration of our three VAEs as well as their performance at the step with minimal validation reconstruction loss.

In addition to the performance metrics, we created a number of reconstructions for images from the validation set which are shown in Figure 1.
As preparation for the downstream task, we first evaluated if sex and age are predictive for dementia. The logistic regression model achieved a balanced accuracy of 0.49 in this case, so we did not take additional measures.
With the best checkpoints from above, we obtained the following results for the downstream task of dementia diagnosis:

	Standard VAE	Masks in encoder	Masks in separate encoder	Sex and age as input
train balanced accuracy	0.798	0.784	0.821	0.502
val. balanced accuracy	0.706	0.709	0.702	0.494
test balanced accuracy	0.749	0.737	0.782	0.492

Table 2. Performance of the three different architectures for dementia prediction. The last column shows the predicitve performance based on demographic factors sex and age.

Conclusion

From the results of the evaluation we conclude that providing segmentation masks in a separate encoder improves dementia prediction. However, we also observed issues with overfitting which should be addressed in future work.
Despite evaluating the model checkpoints with lowest validation reconstruction loss, we still observe overfitting for the final VAE models. In an attempt to mitigate this, we ran an additional grid search hyperparameter sweep. Choosing 16 as number of latent dimensions and 0.1 for the Kullback-Leibler divergence coefficient did mitigate overfitting. However, the previous validation reconstruction loss was still better, so we did stick with the final VAEs from the first sweeps.
Interestingly, the test validation reconstruction loss is even worse compared to the validation reconstruction loss. This indicates, that our final models also overfit on the validation set.
The impact of this can be seen in the results of the downstream task. While the validation balanced accuracy is similar across all architectures and up to 0.119 worse than the training balanced accuracy, the models perform significantly better on the test set. Also, the test balanced accuracy differs between the three architectures. The version with a separate encoder for the segmentation masks has a 0.033 higher test balanced accuracy than the baseline architecture. We suspect that the difference between validation results and test results for the downstream task is due to the previously mentioned overfitting on the validation set.
Overall, adding information about the structure through segmentation masks in a separate encoder does improve the predictive capabilities for dementia. However, future work should further investigate issues regarding overfitting.

References

1. D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*.
2. R. Wei and A. Mahmood, "Recent advances in variational autoencoders with representation learning for biomedical informatics: A survey," *Ieee Access*, vol. 9, pp. 4939–4956, 2020.
3. G. Bredell, K. Flouris, K. Chaitanya, E. Erdil, and E. Konukoglu, "Explicitly minimizing the blur error of variational autoencoders," *arXiv preprint arXiv:2304.05939*.
4. N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, "On the spectral bias of neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 5301–5310.
5. D. Crosby, S. Bhatia, K. M. Brindle, L. M. Coussens, C. Dive, M. Emberton, S. Esener, R. C. Fitzgerald, S. S. Gambhir, P. Kuhn *et al.*, "Early detection of cancer," *Science*, vol. 375, no. 6586, p. eaay9040, 2022.
6. S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, "The alzheimer's disease neuroimaging initiative," *Neuroimaging Clinics*, vol. 15, no. 4, pp. 869–877, 2005.
7. D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *CoRR*, vol. abs/1906.02691, 2019. [Online]. Available: http://arxiv.org/abs/1906.02691
8. X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Advances in neural information processing systems*, vol. 29, 2016.
9. G. Desjardins, A. Courville, and Y. Bengio, "Disentangling factors of variation via generative entangling," *arXiv preprint arXiv:1210.5474*.
10. M. Weiner, R. Petersen, P. Aisen, M. Rafii, L. Shaw, J. Morris, and W. Jagust, "Alzheimer's disease neuroimaging initiative 3 (adni3) protocol," *Retrieved May*, vol. 24, p. 2016, 2016.
11. W. H. Organization, *ICD-11: International Classification of Diseases 11th Revision : the Global Standard for Diagnostic Health Information*. World Health Organization. [Online]. Available: https://books.google.de/books?id=H8WFzgEACAAJ
12. F. Isensee, M. Schell, I. Pflueger, G. Brugnara, D. Bonekamp, U. Neuberger, A. Wick, H.-P. Schlemmer, S. Heiland, W. Wick *et al.*, "Automated brain extraction of multisequence mri using artificial neural networks," *Human brain mapping*, vol. 40, no. 17, pp. 4952–4964, 2019.
13. M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.
14. B. Billot, D. N. Greve, O. Puonti, A. Thielscher, K. Van Leemput, B. Fischl, A. V. Dalca, and J. E. Iglesias, "Synthseg: Segmentation of brain MRI scans of any contrast and resolution without retraining," *Medical Image Analysis*, vol. 86, p. 102789, 2023.
15. J. E. Seifarth, C. L. McGowan, and K. J. Milne, "Sex and life expectancy," *Gender medicine*, vol. 9, no. 6, pp. 390–401, 2012.
16. M. J. Katz, R. B. Lipton, C. B. Hall, M. E. Zimmerman, A. E. Sanders, J. Verghese, D. W. Dickson, and C. A. Derby, "Age-specific and sex-specific prevalence and incidence of mild cognitive impairment, dementia, and alzheimer dementia in blacks and whites: a report from the einstein aging study," *Alzheimer Disease & Associated Disorders*, vol. 26, no. 4, pp. 335–343, 2012.