Generative neural networks for the sciences

Ex 05: Final-Project Proposal
Group: Lena Stelter, Kevin Klein , Erik Bartelme

Paper: DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation: [1]
Paper: Curriculum DeepSDF: [2]

Which scientifc questions do you want to answer, and why are they interesting?
We want to figure out how to represent 3D shapes in a way that is smooth and compact, using NNs to describe the surface instead of storing the entire shape as a mesh or blocks. So the input data could be a point cloud or is incomplete, some parts are missing and our Model should predict what is the most likely shape the input data could represent. We want that the data looks somooth so the final result (output of the model) is a Signed distance function (SDF) [3] that can be rendered with a technique called ray marching [4].

Which relevant papers did you fnd? What are their pros and cons? Which among these papers do you propose to build upon, and why?

Pros: We have provided the papers at the top. The papers introduces a continuous representation of 3D shapes using SDFs, which allows for resolution-independent and memory-efficient shape storage. It supports smooth interpolation between shapes and can reconstruct shapes even from incomplete data.

Cons: The method depends on a fixed latent vector for each shape, that limits the generalization to unseen shapes without retraining or optimization. Additionally, the approach will likely struggle with very complex or highly detailed geometries that require extremely precise sampling, for example 3D Fractals like the Mandelbulb [5].

What methods will you try, and why do you consider them promising to answer the questions
Like descibed in the paper we will use a concept similar to an autoencoder, where latent vectors encode 3D shapes compactly, and a neural network acts as a decoder to generate signed distance functions (SDFs). Generative methods include sampling and interpolating these latent vectors to create new shapes or blend existing ones. This approach is promising because it also compresses the data which is more memory effcient compared to data meshes, also we get a smooth shape reconstruction.

What will your data sources be? How large are the data sets and how high do you think the quality is? Can you use simulated data when real data is scarce?
The data sources for the paper include 3D shape datasets like ShapeNet, which provide a wide variety of objects in CAD or mesh formats. When real data is scarce, simulated data can be generated by creating procedural 3D models or synthesizing signed distance values for sampled points around virtual shapes.

What computational resources do you need (e.g. GPUs)? How will you get access to these? How much time will the computations need?
The method requires computational resources like GPUs. The training time depends on the dataset size and network complexity, typically taking several hours. We can ask if we get access to Helix.

What diffculties do you anticipate in the project?
Compute ressources, generating high-quality signed distance values, generalizing the model to unseen shapes, ensuring smooth interpolation between shapes in the latent space without artifacts.

Investigating XAI

The use of generative AI in medicine is often looked on with suspicion due to its black box nature. This lead in recent years to an increased exploration of explainable AI (XAI) methods in order to increase trust of AI methods. One of these new methods was proposed by Mauri et. al. in there paper “A lightweight generative model for interpretable subject-level prediction”[1].
We aim to investigate the ability of this model to accurately explain confounders present in training and test data and compare its performance with other XAI methods. We plan to investigate that using a similar approach as used in the paper “Right for the Wrong Reason: Can Interpretable ML Techniques Detect Spurious Correlations?”[2] by Sun et al .
We plan to use the OASIS-1[3] dataset as a basis. We will prepare with confouders modified versions of the data set. On these we then will train different Res-net, Attri-net and Mauri et al. models.
The result of post hoc explanation methods applied on Res-net, Attri-net and Mauri et. al. models is then compared using metrics introduced by Sun et al.

We plan to use google colab in order to train the around 15 models we need to train. One advantage we have is that the OASIS-1 dataset is relatively small with only 416 MRI scans. Because we are dependent on accurate diagnosies of the scans we can’t use generated data to augment the dataset.

The main difficultys of our project will be deciding on an realistic cofounder with which to modifie our dataset as well as the transfer of pixel based models (Res-net, Attri-net) to voxel based data(MRI-scans).

[1] https://www.sciencedirect.com/science/article/pii/S136184152400361X#fn6
[2] https://arxiv.org/abs/2307.12344
[3] https://sites.wustl.edu/oasisbrains/

Project proposal: Generating MRI modalities to enhance dataset completion

Scientific question and methods to try:
Main challenges in the field of Artificial Intelligence (AI) in medical imaging is the lack of large-scale, well-documented and balanced datasets, due to challenges like data heterogeneity and privacy concerns [3, 2]. If we take a look into the medical imagine technique Magnetic Resonance Imaging (MRI), biological tissue information, used to make clinical decisions, are oftentimes based on multiple MRI modalities. For brain MRI these modalities are T1-weighting, T2-weighting, FLAIR or T1 contrast-enhanced imaging.
A combination of these modalities result in better segmentation of brain tumors[1]. To produce the images of the different modalities, it can be very time consuming and expensive[4].

This exercise proposes to replicate and use a Generative Adversarial Network (GAN) that addresses the synthesis of missing MR imaging modalities proposed by Huang *et al.* in [4]. After successful replication of their results it would be interesting to use the generated images to evaluate the data on a nnUNet.
The article by Huang *et al.* in [4] proposes the synthesis of missing image modalities to address missing data in the clinical setting, but did not discuss the usage of the proposed models, namely the early-fusion CoCa-GAN (eCoCa-GAN) and the intermediate-fusion CoCa-GAN (iCoCa-GAN), in the medical image AI research.
As already mentioned, large, well-documented and balanced data sets - that are also publicly available - are scarce and also time and resource costly. Therefore, this exercise proposes that, after (hopefully) successfully replicating the results in the paper, to use a nnUNet to evaluate the data by:

- Use real T1-weighted image with synthesized T2-weighted, FLAIR and T1 contrast-enhanced images.
- Use real T2-weighted image with synthesized T1-weighted, FLAIR and T1 contrast-enhanced images.
- Use all four real imaging modalities.

Therefore, that all four data sources are combined as one input and evaluated with the nnUNet. The idea is, if there is a reasonably good performance of the nnUNet when having synthesized three of the four modalities, then it might be more time, cost and resource efficient to get new, promising datasets. The article states that T1 images can be synthesized well from all image modalities, while the FLAIR and T2 images can be better synthesized from a combination of FLAIR/T2 images - would an nnUNet be significantly affected by it?
The performances of the nnUNet for the different input settings can be evaluated by using the dice score.
Huang *et al.* compared their segmentation results with and without synthesized inputs to a segmentation model robust to missing MR image modalities to evaluate the segmentations, namely the *RobustMultiSeg* [1]. However, in the case of deep learning research, it would be interesting to see the general effects of synthesized images on training and testing performance of, as here proposed, the nnUNet.
Why not generating a medical imaging dataset? There are many proposals of different models to generate CT, MRI or also PET images by using a segmentation mask or one of these techniques to

generate another one of these techniques. There are 2D and 3D GANs, Variational Autoencoders (VAEs) as well as Denoising Diffusion Probabilistic Models (DDPMs) [3] which are built to generate the data. However, these models cannot be better than the data they are trained on. Therefore, by having larger datasets with high quality, these models performance might be improve too.
As the article of Huang et al. [4] only studies the effect of multiple modalities in brain MRI research, it could also be possible to extend the models to work for MRI datasets of different body parts. For example, for the abdominal area the T1-weighted images do not play as a crucial role as for brain images, however, here there might be made use of imaging modalities like Diffusion Weighted Imaging (DWI) or the Adherent Diffusion Coefficient (ADC). It can be experimented if it is possible to synthesize these imaging modalities too, by using different datasets.

Data sources:
Huang *et al.* worked with the BraTS19 data set, a data set with brain MRI images of the MRI modalities T1-weighted, T2-weighted, FLAIR and T1 contrast-enhanced imaging for around 330 cases. It is a data set that contains segmentation masks and class labels. It is possible to download it from Kaggle (here linked is BraTS 2023), Synapse.ord (here, for example BraTS 2024) or within the datasets published for the MICCAI conference (every year). This data set is updated every year, therefore by using an updated BraTS version the data set size will be bigger. Another brain MRI data set to try this for is the IXI data set witch contains T1-weighted, T2-weighted, Proton Density (PD), MR Angiogram (MRA) and Diffusion weighted images and contains around 600 MR images. MRI data to try the model for other body parts could be the HNS-MRG 2023 (data set of the MICCAI conference) that used the MR modalities T2-weighting, DWI and an image if the Adherent Diffusion Coefficient (ADC). It s a dataset for head and neck tumor segmentation.

About computational resources, access to resources and computation time:
To run the iCOCA-GAN and the eCOCA-GAN, GPUs are necessary. The article states that it took around 100 hours to train the model for 150 epochs. Due to memory issues, the output size of the generated images needed to be reduced. For testing, each modality was possible to be generated in around two seconds. Therefore, once the model is trained, fast usage is possible.
By using publicly accessible GPU, it might be possible that there will be an issue with the limited usage time of the GPUs as for example by using the GPUs from Google Colab or Kaggle. However, it could be possible to save checkpoints within training and restart the training multiple times starting from the last saved checkpoints.

Possible difficulties in this project:
This paper does not provide the code for both models, therefore this GAN needs to be created from scratch. In the article, they described what how they built the model, but it might be very time consuming to built the model - with the model built at the end not having similar performance to the proposed model.

References

[1] Cheng Chen et al. “Robust multimodal brain tumor segmentation via feature disentanglement and gated fusion”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*. Springer. 2019, pp. 447–456.

[2] Zolnamar Dorjsembe et al. “Conditional diffusion models for semantic 3D brain MRI synthesis”. In: *IEEE Journal of Biomedical and Health Informatics* (2024).

[3] Paul Friedrich et al. “Wdm: 3d wavelet diffusion models for high-resolution medical image synthesis”. In: *MICCAI Workshop on Deep Generative Models*. Springer. 2024, pp. 11–21.

[4] Pu Huang et al. “Common feature learning for brain tumor MRI synthesis by context-aware generative adversarial network”. In: *Medical Image Analysis* 79 (2022), p. 102472.