Towards Fair and Accurate Medical Image Embeddings

Florian Hunecke, Robert Graf, Prof. Dr. Daniel Rückert

Motivation

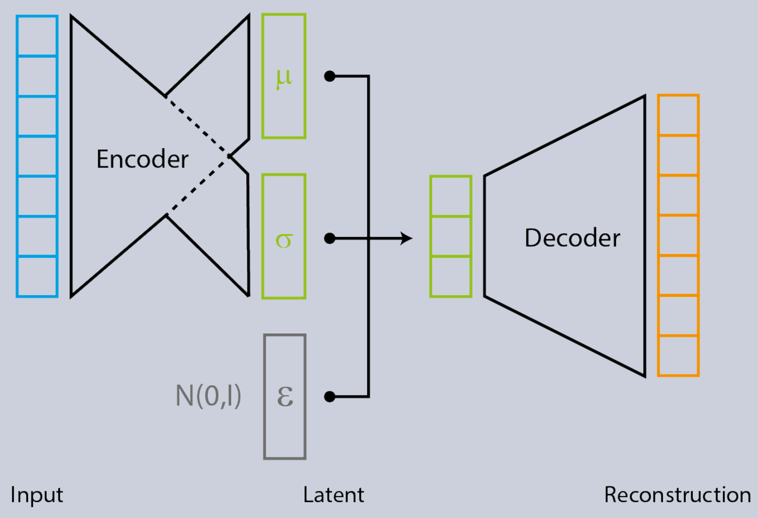
- Deep models require and learn from large datasets, which need thorough investigation to prevent learning of incorrect behavior
- This thesis leverages the nature of autoencoders (AE) to obtain new representations of image datasets making them more insightful
- In addition to visual analysis, common metrics for prediction accuracy and fairness quantification get applied

Methods

1. Unsupervised Training of Autoencoders

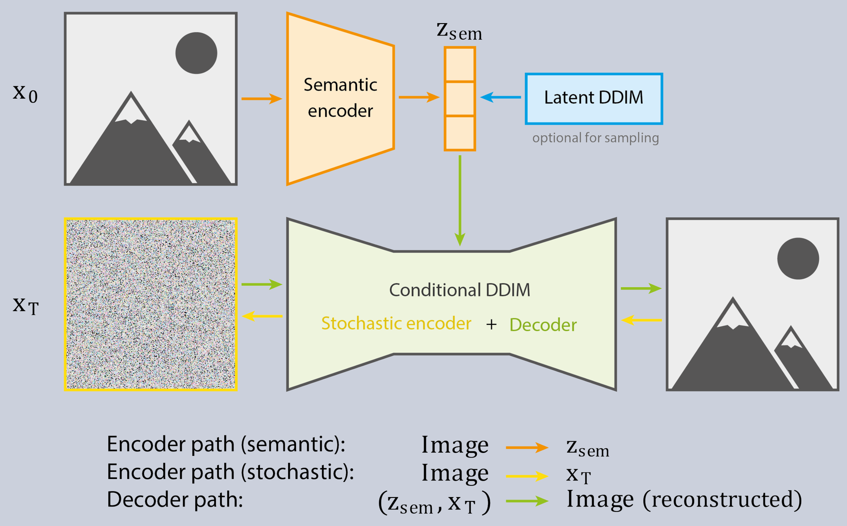
Variational Autoencoder (VAE)

- Regularized AE, encoding into a latent normal distribution, sampling the latent vector
- Continuous latent space



Diffusion Autoencoder (DAE)

- Iteratively add noise to input, train decoder for reversion
- Two latent spaces: semantic and detailed



Hyperparameter	GNC		CheXpert	
	VAE	DAE	VAE	DAE
Learning Rate	5×10^{-3}	1×10^{-3}	25×10^{-4}	5×10^{-4}
Hess. Pen. Weight	5×10^{-6}	1	5×10^{-7}	1
Input Shape (px)	256	256	256	256
Batch Size	64	5	128	64
Latent Dim.	128	512	128	512

2. Embedding creation and t-SNE

- Forward samples through encoder
- Reduce dimensionality using t-SNE to obtain 2D representations

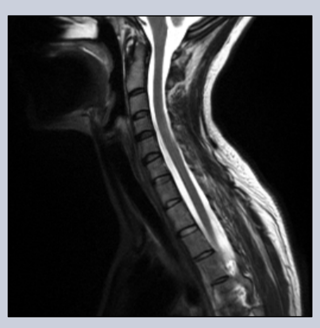
3. Prediction and Fairness Evaluation

- Embedding accuracy assessment with SVMs
  Scores include balanced accuracy, mean absolute error and AUC
- Various fairness metrics performed on embeddings and predictions

Datasets

German National Cohort (GNC) [1]

- Extensive study over multiple German institutes
- > 30,000 magnetic resonance images (MRI)
- Assessment: 2014 – 2018



CheXpert [2]

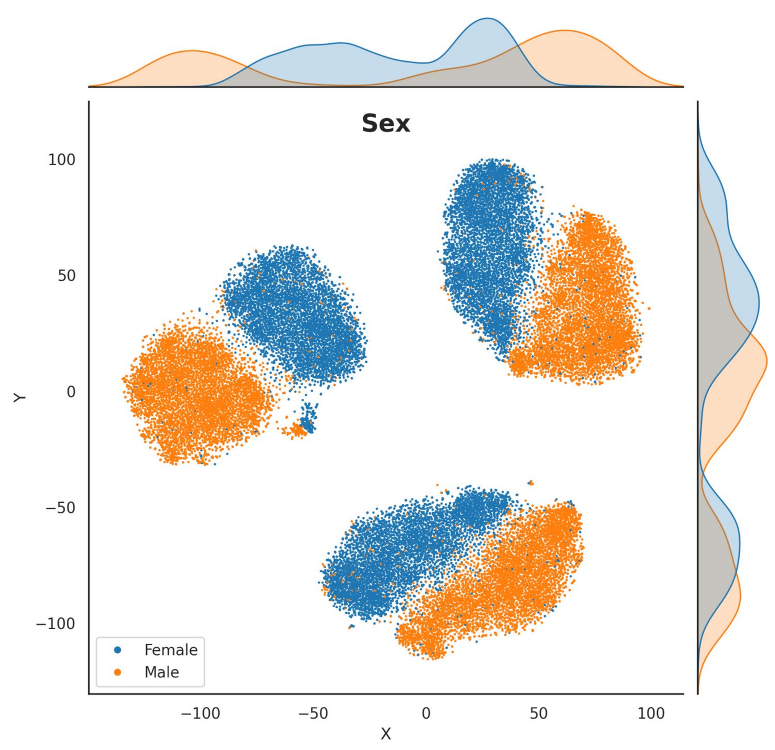
- Stanford University Hospital
- > 220,000 chest x-Rays
- Labeler for findings from free text reports
- Assessment: 2004 – 2017

References

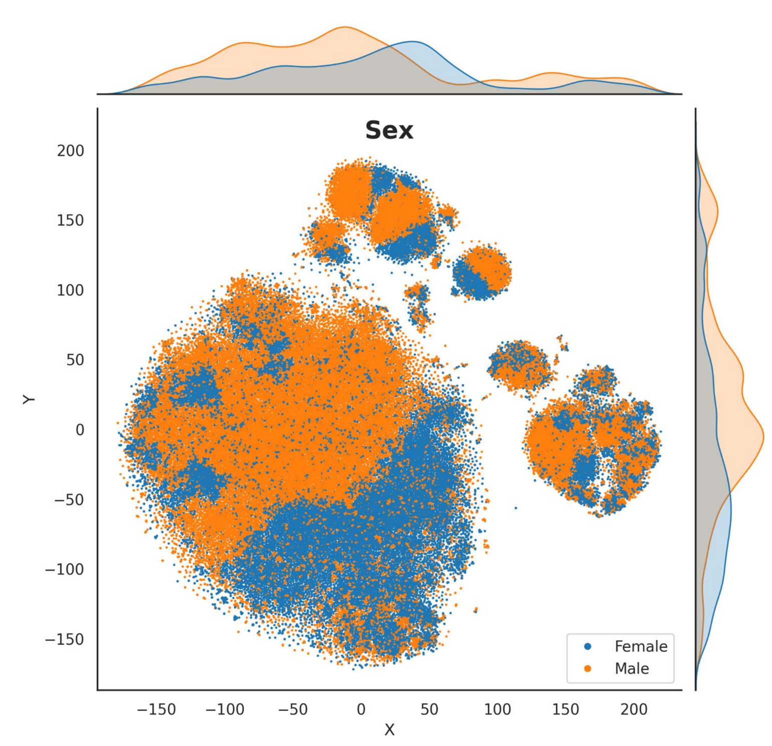
[1] German National Cohort (GNC) Consortium geschaeftsstelle@ nationale-kohorte. de. "The German National Cohort: aims, study design and organization." *European journal of epidemiology* 29.5 (2014): 371-382.
[2] Irvin, Jeremy, et al. "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. No. 01. 2019.

Embeddings (DAE)

GNC



CheXpert



Findings

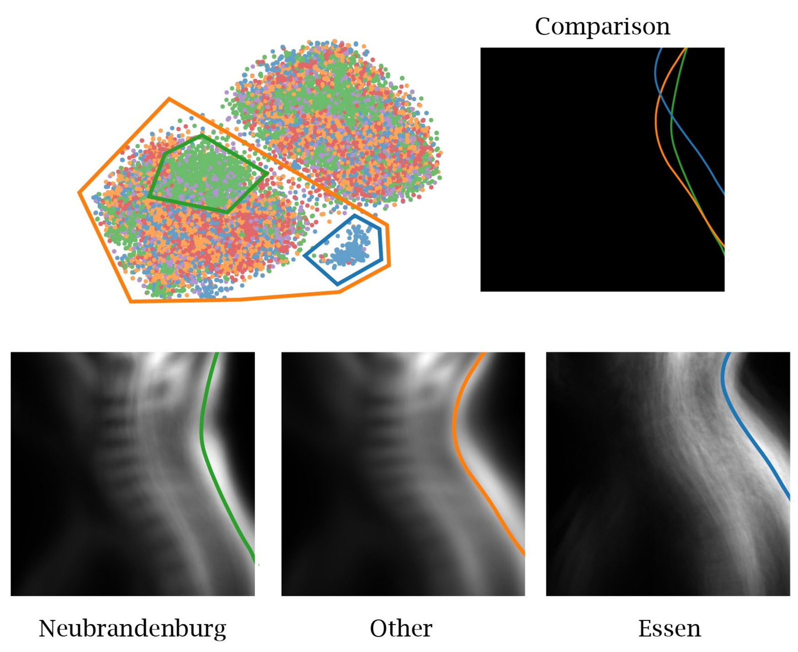
- AEs are highly capable of encoding patient information
- The embeddings show a distinct separation for discrete and a clear distribution for continuous attributes
- DAEs generally yield more reliable results

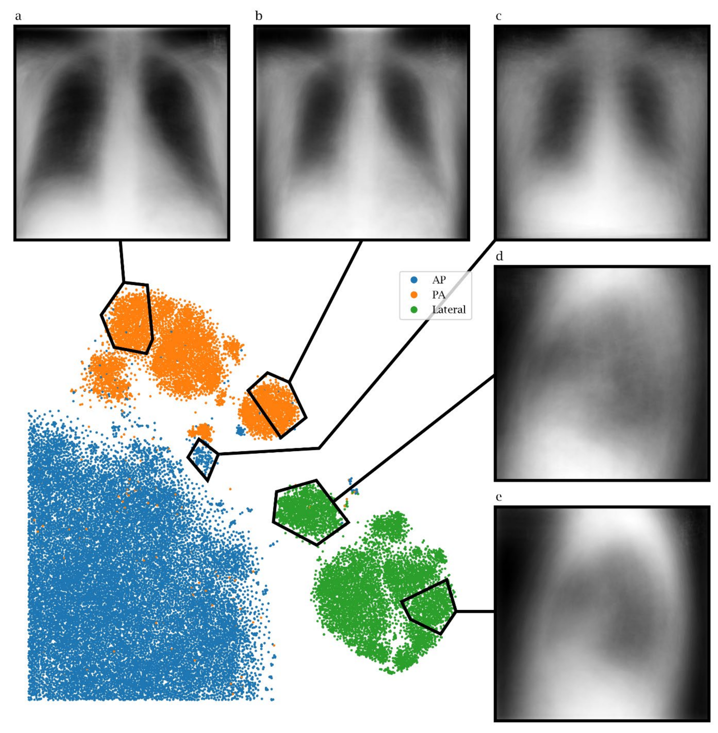
GNC

- Systematic shift in head position for certain institutions
- Likely caused by systematic errors in imaging process

CheXpert

- Cluster separation in lateral and PA images indicate discrete differences in the images
- Not explicable by any labels provided by the dataset





Accuracy

GNC	Region ACC	Sex ACC	Age MAE	Weight MAE	Height MAE
VAE	0.999	0.921	7.20	6.38	0.049
DAE	1.000	0.986	3.94	4.28	0.036
CheXpert	**View** ACC	**Sex** ACC	**Age** MAE	**Disease**[1] MAE	**Mean AUC**[2]
VAE	0.942	0.736	13.06	0.719	0.690
DAE	0.986	0.951	7.90	0.778	0.765

Balanced accuracy, mean absolute error and AUC scores for attribute prediction on the embeddings

1. Newly created binary attribute: finding/no finding
2. Mean of AUC scores of five selected findings used for calculating the CheXpert competition score

Conclusion

- Image embeddings grant valuable insights into inscrutable datasets
- Discovered systematic variances in imaging processes
- Found different types of bias with respect to patient attributes
- Similar accuracy compared with common classifiers

Klinikum rechts der Isar
Technische Universität München