

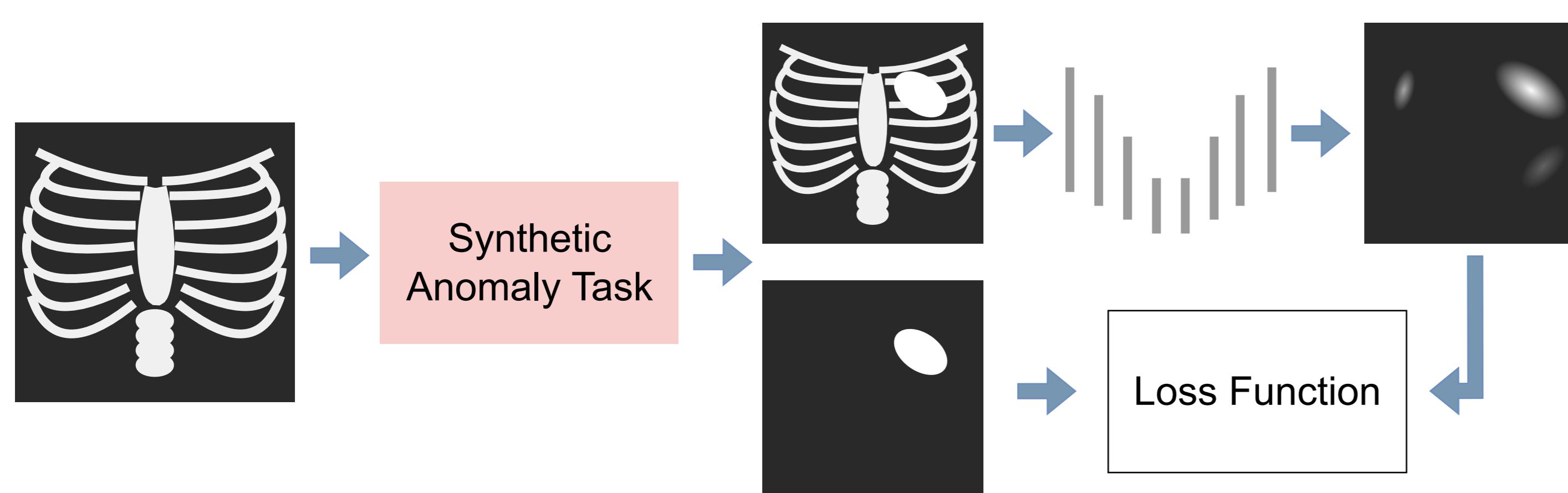
Many tasks make light work: Learning to localise medical anomalies from multiple synthetic tasks

Matthew Baugh, Supervisor: Dr. Bernhard Kainz

1

Self-supervised anomaly detection involves training models to identify synthetic anomalies introduced into otherwise normal data.

- Avoids need to collect data for every type of pathology
- Able to identify anomalies with extreme textures, which reconstruction-based methods struggle with



But existing methods lack a programmatic way to determine training time:

- Overtraining causes models to overfit to synthetic anomalies
- Manually choosing training duration risks tuning to test set

3

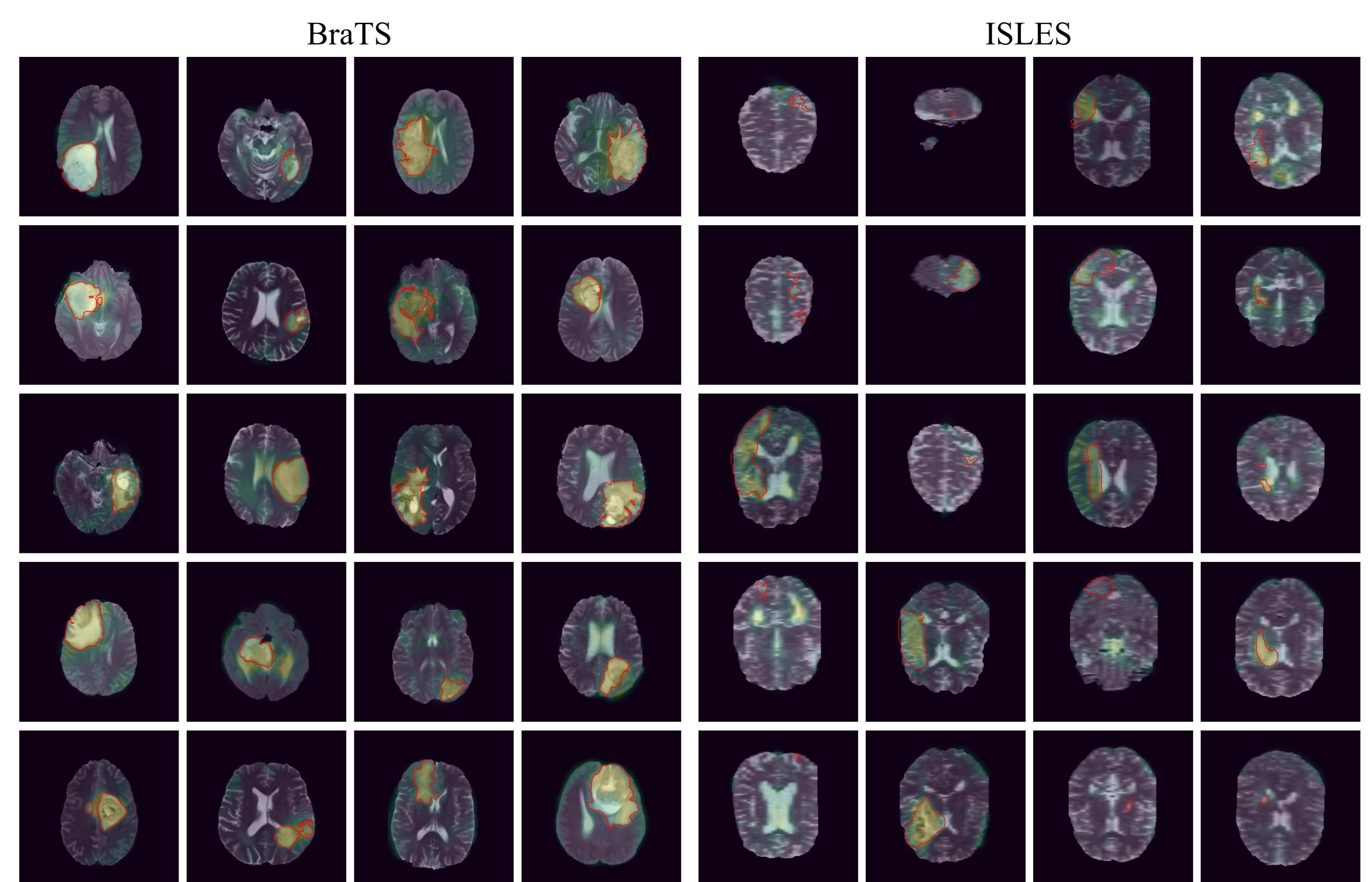
We readily outperform state-of-the-art methods on both brain MRI and chest X-rays.

- Brain MRI: Train on Human Connectome Project dataset [1], test on BraTS 2017 [2] and ISLES 2015 [3] (domain shift of research to clinical).
- Chest X-ray: Use VinDr-CXR dataset [4], following DDAD's train-test split [5], but also testing on the dedicated test set.

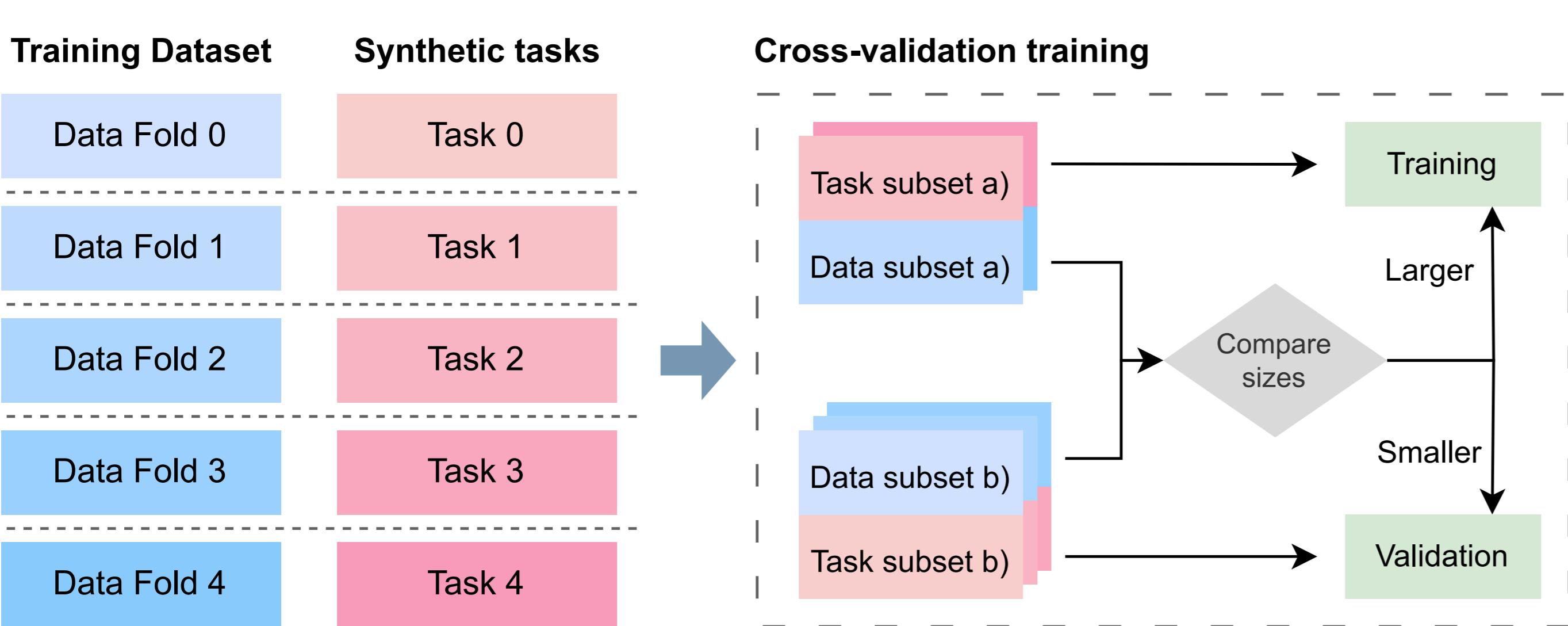
Methods	MRI	Brain MRI				Chest X-Ray (CXR)			
		Slice-wise		Pixel-wise		Sample-wise		Pixel-wise	
		BraTS17	ISLES	BraTS	ISLES	DDAD _{ts}	VinDr _{ts}	DDAD _{ts}	VinDr _{ts}
VAE		80.7/83.3	51.9/71.7	29.8/92.5	7.7/87.5	59.8/55.8	74.8/76.3	MemAE	15.3±3.7
ceVAE		85.6/86.5	54.1/72.7	48.3/94.8	14.5/87.9	72.8/73.8	49.9/48.2	f-AnoGAN	21.4
CRADL		81.9/82.6	54.9/69.3	38.0/94.2	18.6/89.8	49.9/48.2	65.8/65.9	AE-U	19.2±1.7
Ours		87.6/89.4	61.3/80.2	76.2/98.7	46.5/97.1	65.8/64.4	78.4/76.6	FPI	19.5±1.8
Random		49.0/50.0	36.6/50.0	2.4/50.0	1.1/50.0	50.0/50.0	31.6/50.0	PII	24.0
Ours		40.3/50.0	29.4/50.0	1.7/50.0	0.8/50.0	78.4/76.6	71.2/81.1	NSA	24.7
Random		87.6/92.2	62.0/84.6	76.2/99.1	45.9/97.9	78.4/76.6	71.2/81.1	CXR	2.7/50.0
Ours						74.7±4.9	66.3±4.4	21.1	21.4
Train/val. split abd.	1/4	all ens.	83.4±4.4 87.6	59.3±2.2 62.0	46.9±14.9 76.2	23.7±7.7 45.9	78.4	19.2±1.7	15.2±4.5
	2/3	all ens.	82.5±3.3 85.7	55.9±8.5 58.4	42.8±12.8 72.2	21.2±9.3 41.0	78.6±1.4 80.7	19.2±1.7	19.5±1.8
	3/2	all ens.	81.1±4.3 84.0	52.5±4.7 55.0	37.9±11.1 63.7	15.4±3.3 26.6	78.7±1.8 80.4	20.3±1.4	24.0
	4/1	all ens.	81.5±2.7 83.1	53.1±2.3 54.7	36.1±9.0 52.5	16.5±5.0 23.7	79.2±1.3 80.5	21.1±1.3	24.7
						78.4±0.9 73.6	20.4±0.9 23.5	21.1±0.9 24.5	

We investigate how model performance varies as different numbers of task are allocated to training versus validation.

- A more balanced allocation gives best results for chest X-ray data, as the greater variety of anomalies used for training allows the models to generalise to more types of variation.
- But the domain shift in the brain MRI data means that models trained with a greater emphasis on validation perform best, as the difference in data distribution (primarily the acquisition resolution) causes the more sensitive models to predict more false positives.

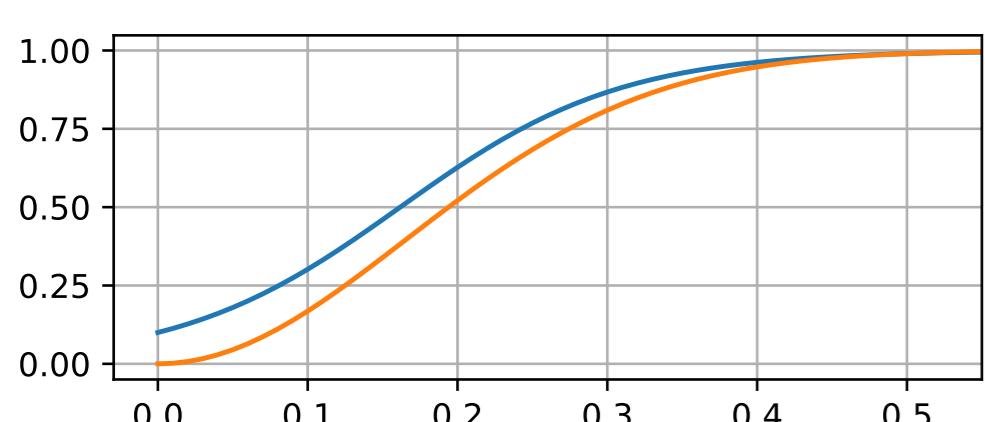


- Using unseen synthetic tasks for validation provides an estimate for the models performance on unknown real-world anomalies as each task has a distinct feature set.
- All tasks are seamlessly integrated to prevent the model learning trivial solutions, such as discontinuities at anomaly boundaries.
- Cross-validation is performed over pairs of data partitions and tasks, as cross-validating over both independently would be computationally expensive.



We also propose a C1 continuous labelling function

- Allows for natural image variation
- Stabilises training



List of publications:

- Baugh M et al.: Many tasks make light work: Learning to localise medical anomalies from multiple synthetic tasks. Early accepted to MICCAI 2023.
- Baugh M et al.: Zero-Shot Anomaly Detection with Pre-trained Segmentation Models. Won 3rd place in zero-shot track of Visual Anomaly and Novelty Detection 2023 Challenge (CVPR Workshop).
- Baugh M et al.: nnOOD: A Framework for Benchmarking Self-supervised Anomaly Localisation Methods. In UNSURE 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings 2022 Sep 14 (pp. 103-112).

References:

1. Van Essen, D. et al.: The human connectome project: A data acquisition perspective. *NeuroImage* 62(4), 2222–2231 (2012).
2. Bakas, S et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *ArXiv:1811.02629* (2018)
3. Maier, O. et al.: ISLES 2015 – a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical Image Analysis* 35, 250–269 (2017).
4. Nguyen, H.Q. et al.: Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations. *Scientific Data* 9(1), 429 (2022). *arXiv:2210.04227* (2022)
5. Cai, Y. et al.: Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. *arXiv:2210.04227* (2022)