Long Tail Image Generation Through Feature Space Augmentation and Iterated Learning

Rafael Elberg[1,2,3], Denis Parra[1,2,3], Mircea Petrache[1,2]

{rafael.elberg, denis.parra, mpetrache}@uc.cl

[1] Pontificia Universidad Católica de Chile, [2] CENIA, [3] iHealth
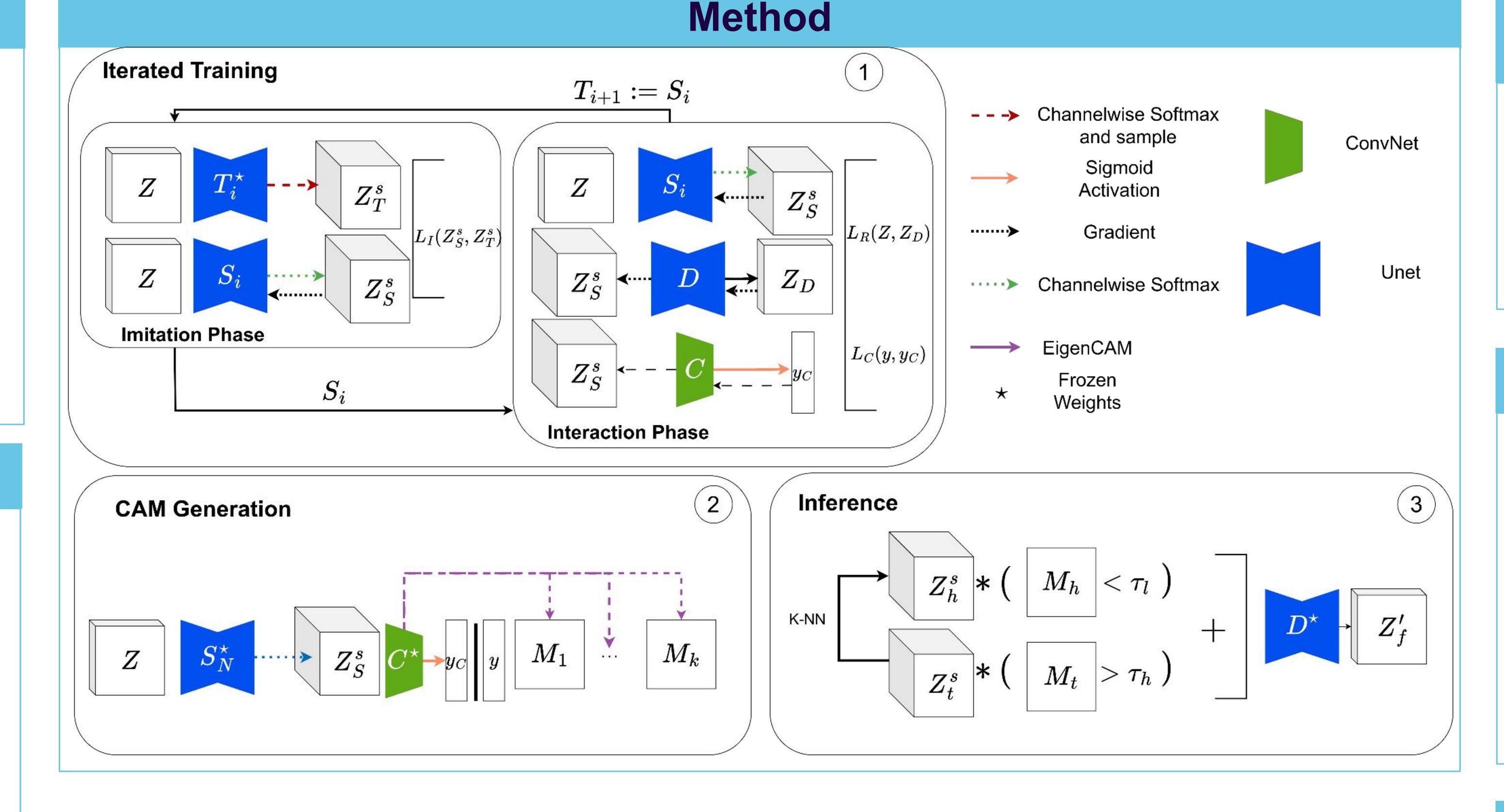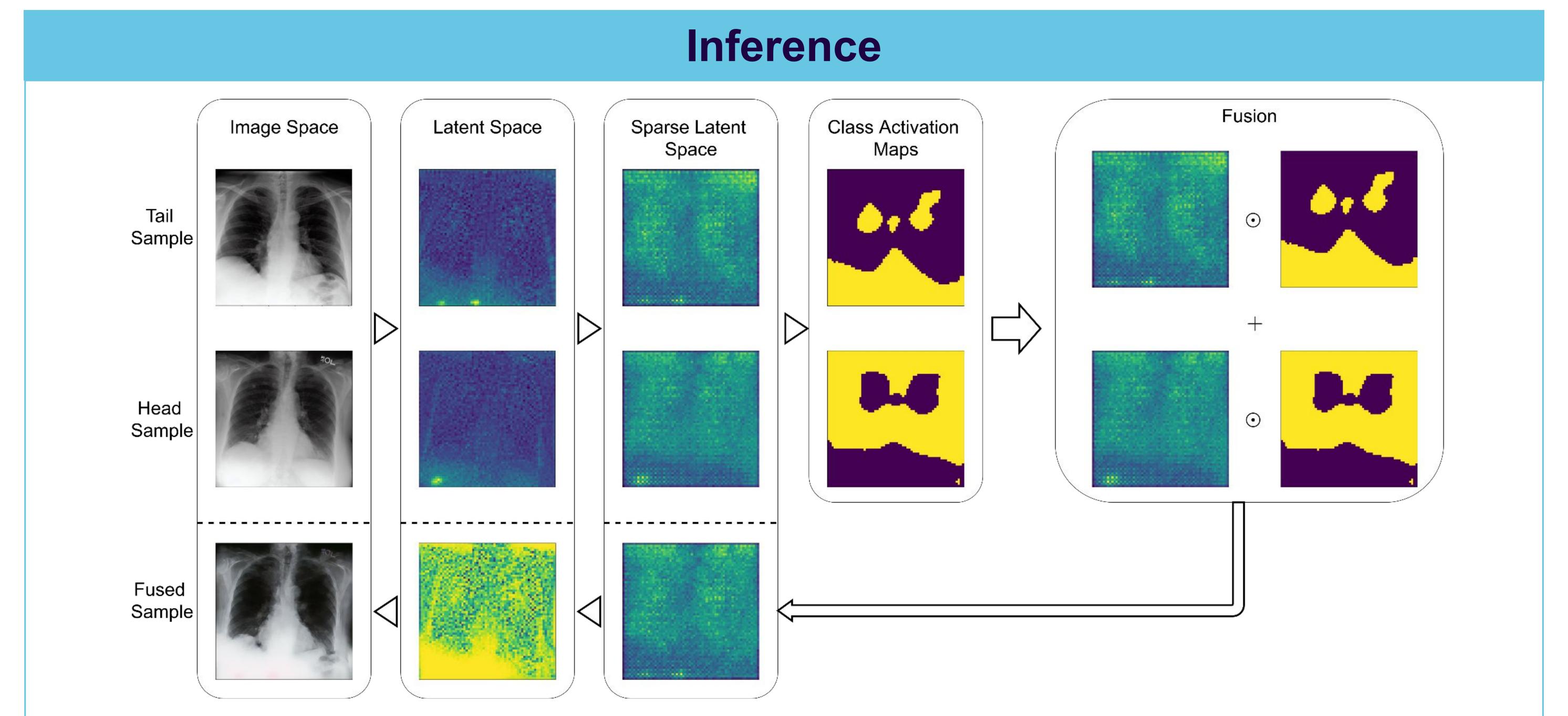
CVPR
JUNE 17-21, 2024
SEATTLE, WA

Introduction

- Specific domains are naturally **Long Tailed** (Medical image domain).
- SOTA image augmentation methods suffer with **long tailed data, reduced sample quality and interference between features.**
- Compositional Learning, **IL** and **SE**.
- We propose creating a **sparse, disentangled** latent space from a pretrained Diffusion Model to **fuse important features from tail examples with unimportant features from head examples**.

Background

- Simplicial Embeddings and Iterated Learning create **Compositional Mappings[1].**
- MIMIC-CXR-LT[2] is a **Long Tailed Dataset** of Chest X Ray images and anomaly labels.
- By separating samples into their components, we can create **new tail class data points** by fusing **class generic features from head samples and class specific features from tail samples[3].**

Examples

Support devices image + No finding image = New support devices image

Tortuous Aorta Image + Atelectasis Near Neighbor Image = New Tortuous Aorta Image

Method

Inference

Experiments

- Classification on a reduced set of **MIMIC-CXR-LT**. We train a DenseNet121 on the original training set, and a set augmented with: **RoentGen[4]** at 75 inference steps, SMOTE[5], and our method.
- We use RoentGEN as our Stable Diffusion Model, testing **0, 1 and 5** inference steps.
- We also evaluate image generation quality (FID).

Results

Model	Generation	Classification	
	Avg Tail FID	Head mAP	Tail mAP
Baseline	-	0.618	0.155
SMOTE	171.864	0.578	0.151
RoentGEN @ 75	138.963	**0.618**	**0.152**
Ours @ 0	191.873	0.607	**0.152**
Ours @ 1	191.646	0.595	0.143
Ours @ 5	**130.110**	0.595	0.144

Conclusions and Future Work

- We obtain **competitive results in image generation** quality using **only 5 inference steps.** Counter intuitively, **more inference steps reduce data augmentation effectiveness.** Our reasoning suggests this is due to the diffusion process not respecting existing labels.
- We plan to experiment with larger datasets and a broader range of domains, and to evaluate on **compositional generalization specific tasks**. We also want to include the mask generation in the learning process, and evaluate different ways to assign labels..

References

[1] Ren, Y., Lavoie, S., Galkin, M., Sutherland, D. J., & Courville, A. (2023, October 28). *Improving compositional generalization using iterated learning and simplicial embeddings.* arXiv.org. https://arxiv.org/abs/2310.18777

[2] Holste, G., Wang, S., Jaiswal, A., Yang, Y., Lin, M., Peng, Y., & Wang, A. (2023). CXR-LT: Multi-Label Long-Tailed Classification on Chest X-Rays (version 1.1.0). PhysioNet. https://doi.org/10.13026/c4tr-kr83.

[3] Chu, P., Bian, X., Liu, S., & Ling, H. (2020, August 9). Feature space augmentation for Long-Tailed data. arXiv.org. https://arxiv.org/abs/2008.03673

[4] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling technique. Journal of Artificial Intelligence Research/the Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

[5] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.