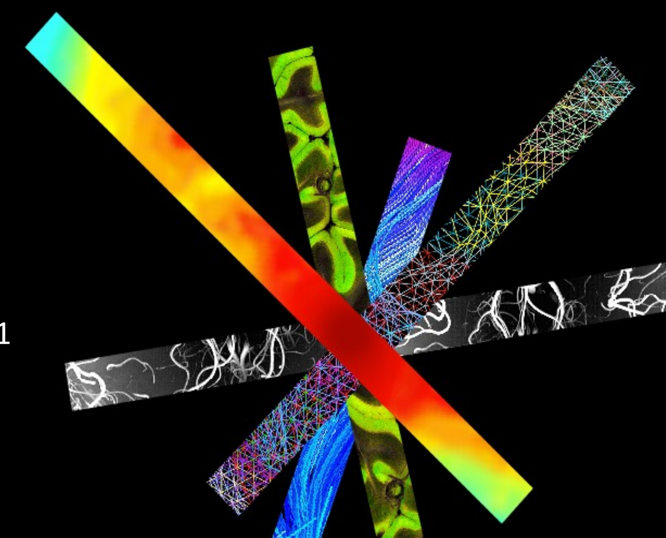
Diffusion-based Zero-shot Medical Image-to-Image Translation for Cross Modality Segmentation

Zihao Wang[1], Yingyu Yang[2], Yuzhou Chen[3], Tingting Yuan[4], Maxime Sermesant[2], Hervé Delingette[2], Ona Wu[1]

1. Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School
2. Inria centre at Université Côte d'Azur
3. Department of Computer and Information Sciences at Temple University
4. Institute of Computer Science at Georg-August-University of Göttingen



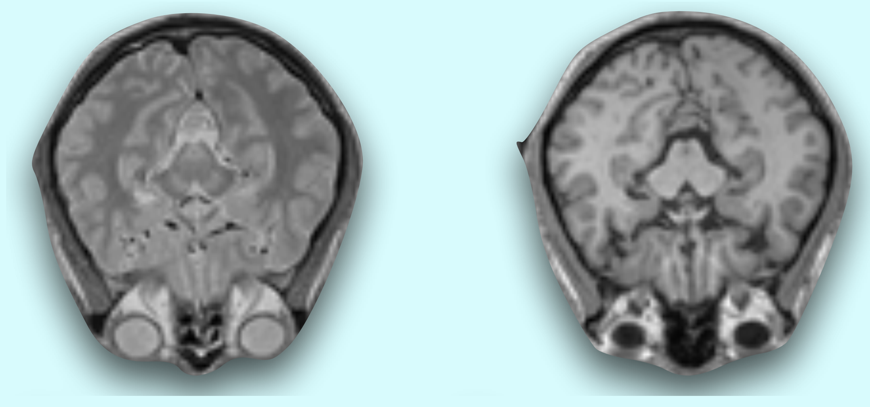
Athinoula A. Martinos Center For Biomedical Imaging

MASSACHUSETTS GENERAL HOSPITAL

HARVARD MEDICAL SCHOOL

Massachusetts Institute of Technology

Introduction: Cross-Modality Image Segmentation



- Diverse set of imaging techniques: MRI, CT, Ultrasound, etc.
- Each modality provides unique information about tissues, structures, or diseases.
- Leveraging existing segmentation tools across modalities presents practical benefits.
- Significance of the Problem: Save money, benefit patients.
- Making the most of available data: **Use existing annotations and knowledge from one modality to assist in another.**

Challenge: Perform **Cross-Modality Image Translation with limited access to the source modality**.

Method: Diffusion Model for Cross-Modality Image Segmentation

Prior Image Translation Frameworks:
Straightforward **pair-to-pair regression**.
GAN-based (Generative Adversarial Network)[1-2] translation.
Use of **Diffusion Model** for generating target data.
Yet, not applicable to zero-/few-shot learning.

Prior Diffusion-based Works:
Condition the step-by-step diffusion process **(supervised)**.
Works like SDEdit[3] has employed **Perturbed Diffusion** approaches (requires source and target modalities are numerically consistent).

Proposed: Statistical Feature Homogeneity

Every imaging modality inherits specific statistical features that remain consistent among different modalities.
These features can be leveraged for image translation, especially when the source modality data is absent.

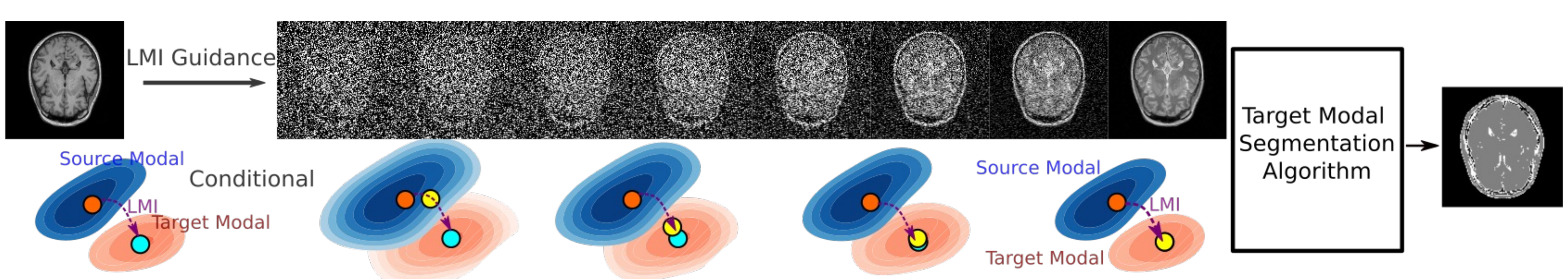


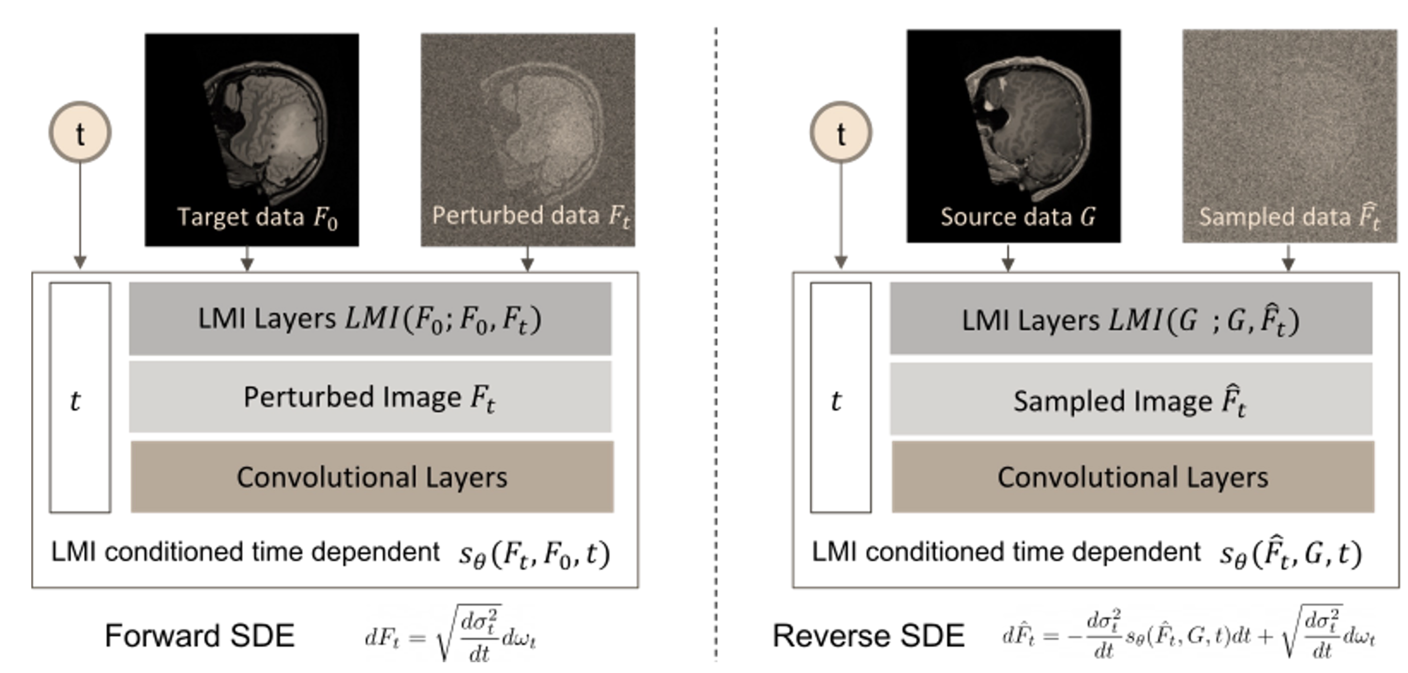
Diagram depicts LMI-guided zero-shot cross-modal segmentation: source (blue) and target (orange) distributions with corresponding data points linked. LMIDiffusion applies LMI for navigation, culminating in an image segmentable by target-modality techniques.

Local Mutual Information (LMI) Layer

The LMI operator **captures the statistical homogeneity** across modalities, making it pivotal for zero-shot translation.

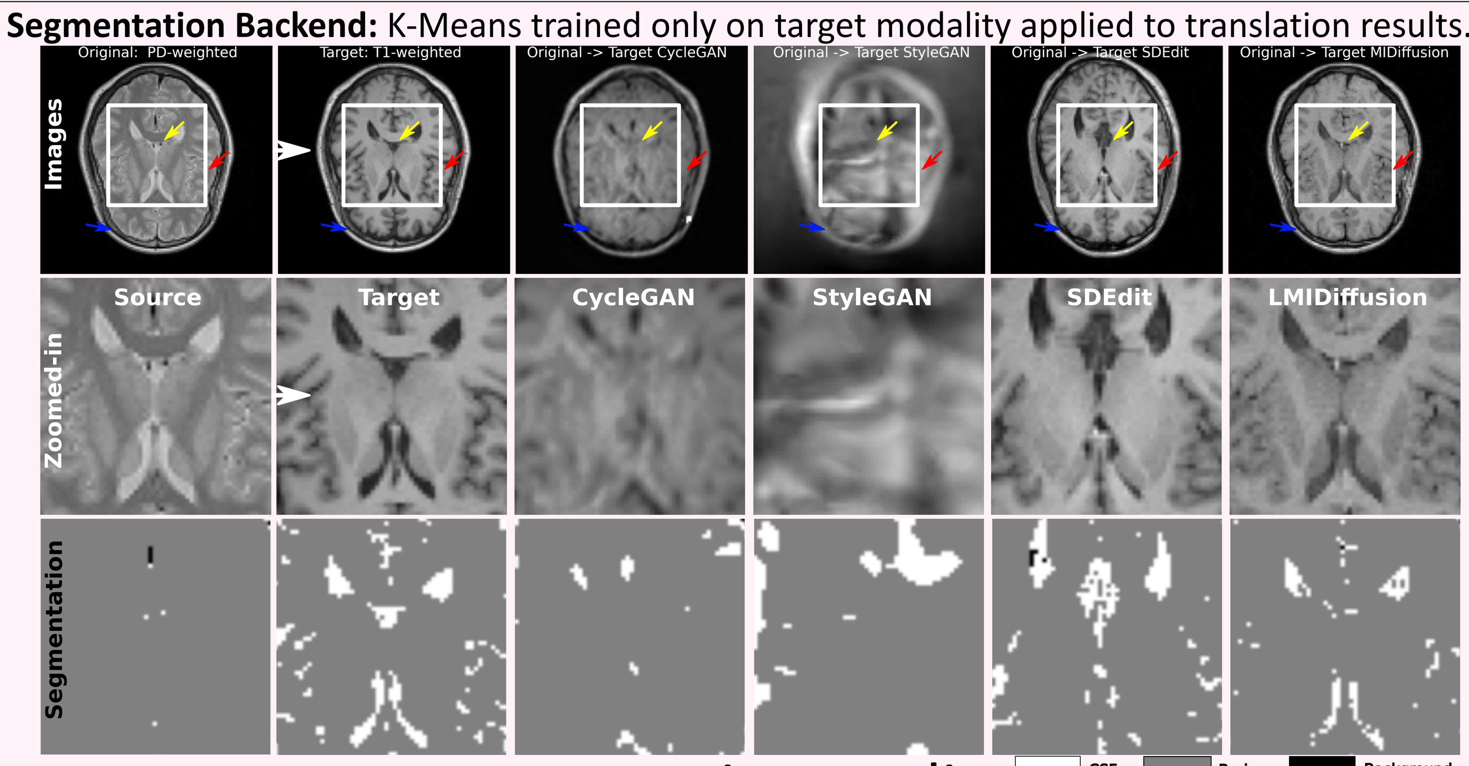
$$LMI_{\delta}(x_i, y_j) = \sup \iint p_{\delta}(x,y) \log \frac{p_{\delta}(x,y)}{p_{\delta_{x_i}}(x) p_{\delta_{y_j}}(y)} dxdy, \forall y_j \in \delta_{x_i}$$

A novel **differentiable** LMI layer (accelerated) designed to extract the inherent statistical features from modalities.



Experiment & Result

Segmentation Backend: K-Means trained only on target modality applied to translation results.



Dataset Details:
- Translation tasks between PD and T1w modalities.
- Training Set: 300 slices. Testing Set: 75

Baselines and Metrics:
- CycleGAN[1]: Used a small part (11%) of the source data.
- StyleGAN2-ADA[2]: Utilized target data, GAN inversion.
- SDEdit[3]: Perturbation guidance, unsupervised.
- Translation performance is measured by SSIM and PSNR, the Dice score reports segmentation quality.

LMIDiffusion Performance:
- Closely mirrors segmentation ground truth.
- Maximally resembles the translation target (T1w).
- Retains superior anatomical features of source (PDw).

PDw→T1	CycleGAN[1]	StyleGAN[2]	SDEdit[3]	LMIDiffusion
Dice Score	0.85 ± 0.07	0.66 ± 0.10	0.82 ± 0.06	**0.88 ± 0.05**
PSNR	18.88 ± 1.26	10.15 ± 1.49	15.96 ± 1.54	**20.22 ± 1.43**
SSIM	0.27 ± 0.04	0.06 ± 0.03	0.50 ± 0.06	**0.69 ± 0.06**

Conclusion

- Novel zero-shot cross-modality image translation method.
- Outperforms existing GAN-based and diffusion-based zero-shot methods in translation quality and segmentation.

References

[1] Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, Zhu et al. 2017
[2] A Style-Based Generator Architecture for Generative Adversarial Networks, Karras et al 2019
[3] SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations, Meng et al. 2021

Acknowledgement

This work was supported in part by the R01NS102574, Massachusetts Life Sciences Center, AHA 19AIML35170037.