PRISM: High-Resolution & Precise Counterfactual Medical Image Generation using Language-guided Stable Diffusion

Amar Kumar [1,2], Anita Kriz [1,2], Mohammad Havaei [3] and Tal Arbel [1,2]

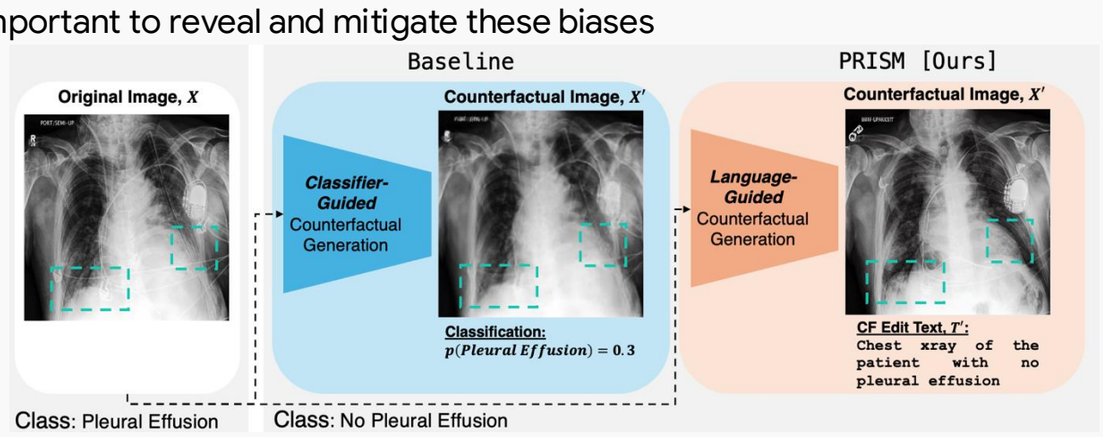
[1] Centre for Intelligent Machines, McGill University, Montreal, Canada
[2] MILA Quebec AI Institute, Montreal, Canada
[3] Google Research, Montreal, Canada

[Oral]

MIDL Salt Lake City 2025

[1] Problem & Motivation

- **Problem:** Deep learning models suffer from spurious correlations and data imbalances[1].
  - ➢ Leading to poor generalization in clinical settings.
- **Explainability via Counterfactual (CF) Image Generation** becomes important to reveal and mitigate these biases



- **Remaining Open Challenges:**
  - ➢ Classifier guided methods depend on biased data they aim to fix.
  - ➢ High Resolution medical image generation at 512x512 and beyond.
  - ➢ Complex image editing tasks such as medical device removal.
- **Our Solution: PRISM**
  - ➢ Language guided CF image generation using fine-tune Stable Diffusion.
  - ➢ High resolution medical image generation and editing via natural language guidance.

[2] Methodology Overview

- Experiments are conducted on the **CheXpert**[5] dataset.
- PRISM utilizes a three-stage pipeline architecture.

1. Tabular -> Text Conversion
2. Fine-tune Stable Diffusion
3. Language-guided CF generation

Pleural Effusion	Healthy
1	0

`Chest X-ray of the patient showing pleural effusion`

Stable Diffusion v1.5 is fine-tuned with image-text pairs of chest x-ray images and tabular information converted to text

Original text: `Chest X-ray of the patient showing pleural effusion`

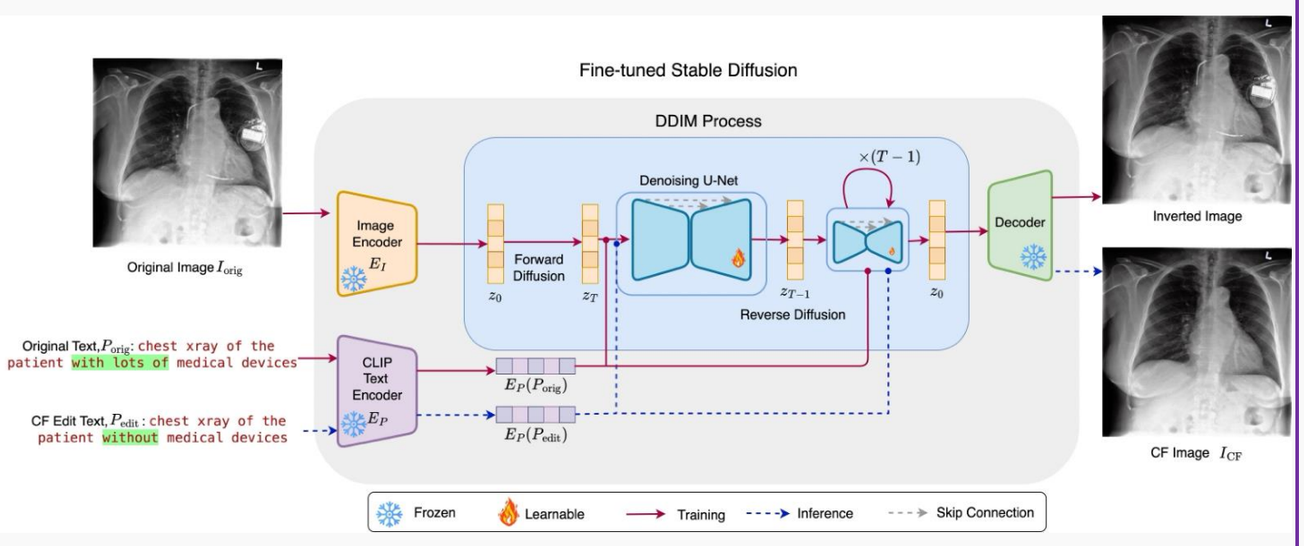
Edit Text: `Chest x-ray of the patient with no significant findings`

- **Key Contributions**
  - ➢ Image Editing with Null –text inversion
  - ➢ Cross attention control for precise modifications using language
  - ➢ Identity-preserving Counterfactual (CF) image generation
  - ➢ High Resolution image synthesis (512x512)

[3] PRISM Framework

PRISM to synthesize high resolution medical images

➔ **Language-guided image editing for counterfactual image generation**

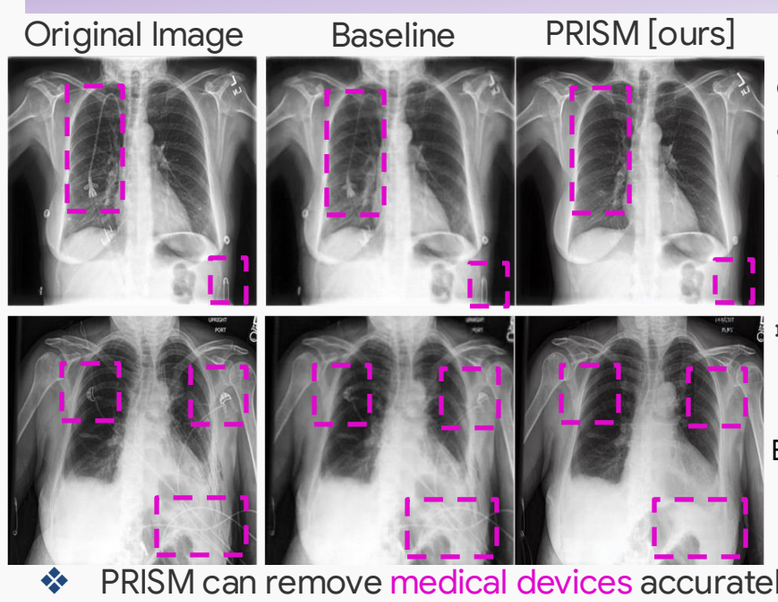


$$S_{\text{CLIP}} = \frac{\Delta I \cdot \Delta P}{\|\Delta I\| \|\Delta P\|} \qquad \Delta I = E_I(I_{\text{CF}}) - E_I(I_{\text{orig}}) \qquad \Delta P = E_P(P_{\text{edit}}) - E_P(P_{\text{orig}})$$

CLIP score

[4] Capabilities of PRISM

I) Image Editing: Device Removal



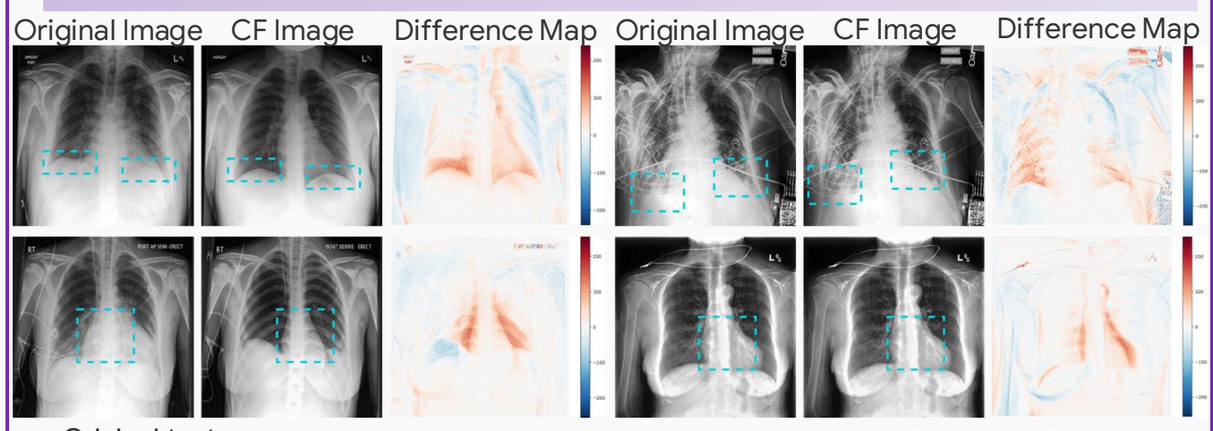
Original text : `Chest X-ray of the patient with lots of medical devices`

Edit text : `Chest x-ray of the patient without medical devices`

Baseline : Classifier Guided GAN

- PRISM can remove medical devices accurately.

II) Explainability of the Disease

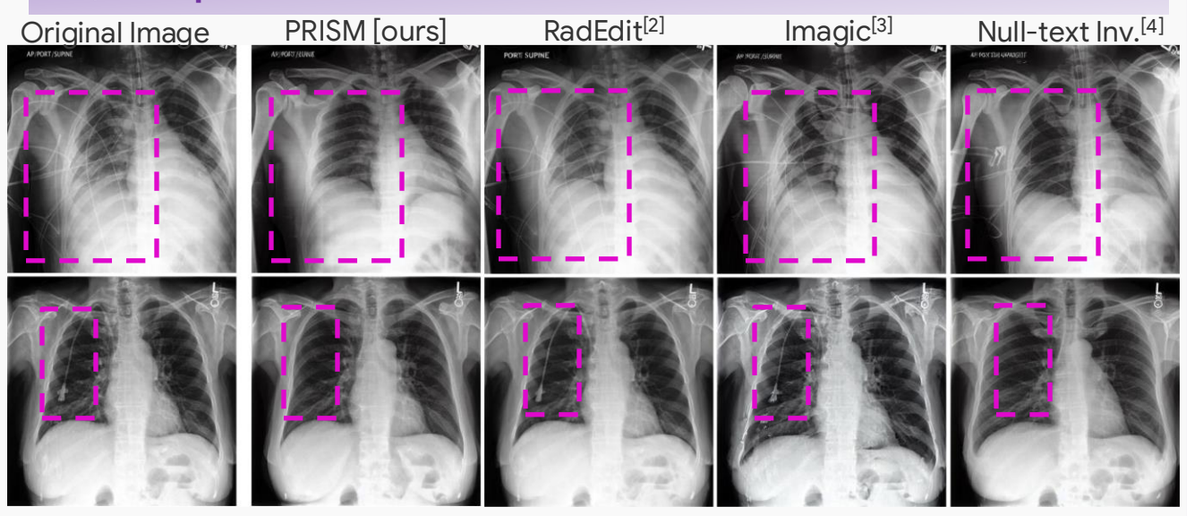


Original text: `Chest X-ray of the patient with severe [disease]`

Edit Text: `Chest x-ray of the patient with no finding`

- PRISM can successfully reveal the disease pathology.

III) Comparison with other SOTA methods for device removal



IV) Accuracies of CF-augmented classifiers

	Pleural Effusion	Cardiomegaly	No Finding	Support Devices
Original Data	0.80	0.87	0.91	0.86
Original Data + SD samples	0.82	0.86	0.91	0.85
Original Data + PRISM CFs	**0.88**	**0.90**	**0.92**	**0.88**

- Augmentation with PRISM CF improves the performance of classifier.

[5] Conclusions

- **PRISM is** the first framework for high-resolution, language guided medical counterfactual generation with unprecedented precision and clinical relevance.
- PRISM's capabilities
  - ➢ High resolution, high fidelity counterfactual generation
  - ➢ Precise medical devices removal using natural language guidance
  - ➢ Disease specific pathology modifications in the image
  - ➢ Enhance classifier robustness via CF augmentation
- Future directions
  - ➢ Out-of-distribution generalization across diverse medical datasets
  - ➢ Multi-modal medical imaging applications - CT, Skin and MRI
  - ➢ Real-time deployment integration within existing clinical workflows

References

[1] A. Kumar *et al*,. "Debiasing Counterfactuals in the Presence of Spurious Correlations", MICCAI 2023 FAIMI Workshop.
[2] F. Perez-Garcia *et al.*, "RadEdit: stress-testing biomedical vision models via diffusion image editing", ECCV 2024.
[3] B. Kawar *et al.*, "Imagic: Text-based real image editing with diffusion models", CVPR 2023.
[4] R. Mokady *et al., "Null-text inversion for editing real images using guided diffusion models", CVPR 2023.*
[5] J. Irvin *et al.*, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison", AAAI 2019.

Acknowledgements

The authors are grateful for funding provided by the Natural Sciences and Engineering Research Council of Canada, the Canadian Institute for Advanced Research (CIFAR) Artificial Intelligence Chairs program, Mila - Quebec AI Institute, Google Research, Calcul Quebec, Fonds de recherche du Quebec (FRQNT) and the Digital Research Alliance of Canada.

kumarama@mila.quebec