CheXplaining in Style: Counterfactual Explanations for Chest X-rays

Klinikum rechts der Isar
Technische Universität München

Matan Atad*, Vitalii Dmytrenko*, Yitong Li*, Xinyue Zhang*, Matthias Keicher,
Jan S. Kirschke, Bene Wiestler, Ashkan Khakzar and Nassir Navab

*Equal contribution

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

In a glance

★ First application of counterfactual explanations for chest X-rays

★ Our EigenFind efficiently finds classifier-specific directions in the decomposed StyleSpace of StyleGAN2

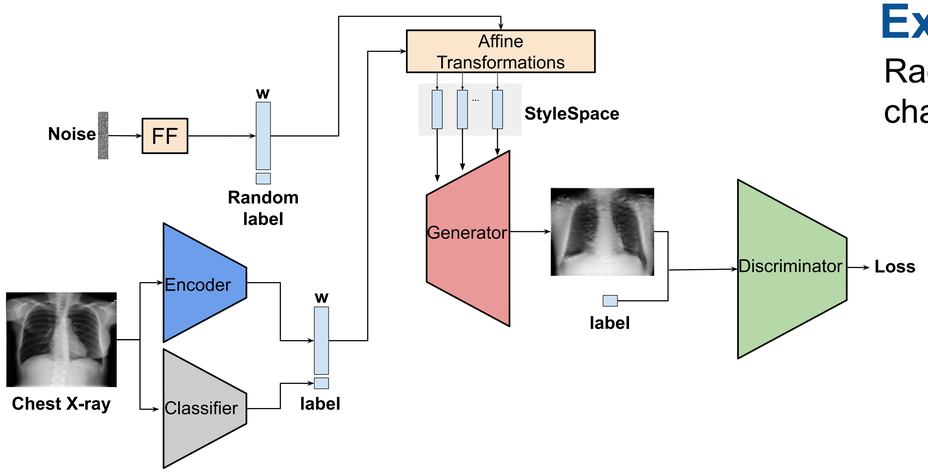
★ Radiologists affirmed the clinical-relevance of the feature changes in our counterfactuals

The problem

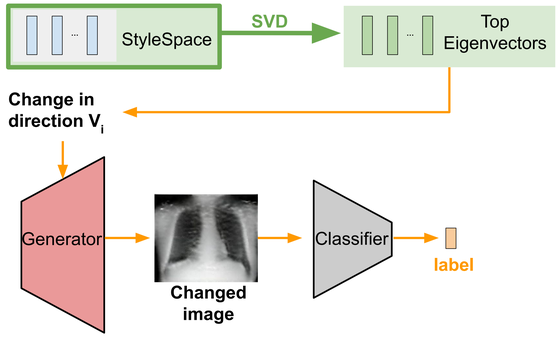
- Essential to know which patterns ML models rely on for clinical routines
- Saliency maps show ***which*** areas are important, but not ***what*** these are

Counterfactual Explanations

Show what changes are required to lead the model to a different prediction



The StyleEx [1] pipeline captures classifier-specific features in the StyleSpace



EigenFind searches for StyleSpace Eigenvectors with significant impact on the classifier decision

Our code:

https://github.com/CAMP-eXplain-AI/Style-CheXplain

Experiments & Evaluation

Radiologists diagnosed which features changed in counterfactuals vs. the originals

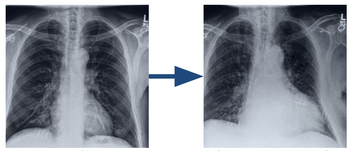
Features of Cardiomegaly	V_{17}	V_2	V_{18}
Increased cardiothoracic ratio	✓	✓	✓
Secondary findings:			
Reduced lung tissue opacity	✓		
Pleural Effusion	✓	✓	
Pacemaker			✓
Older patient		✓	
Features of Pleural Effusion	V_{15}	V_7	V_{12}
Obstruction of the pleural recessus	✓	✓	✓
Opaque lower lungs			✓
Secondary findings:			
Increased cardiac diameter	✓		
Fluid overload	✓		
Pneumonia		✓	

(top 3 significant Eigenvectors)

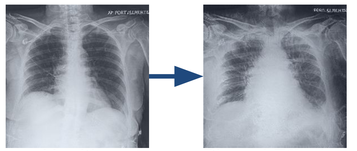
Pathology	**AttFind [1]**	**EigenFind (Ours)**
Atelectasis	94%	94%
Cardiomegaly	96%	95%
Pleural Effusion	94%	91%
Search Time	**12 hours**	**2 minutes**

Percentage of explained images in CheXpert [2] dataset

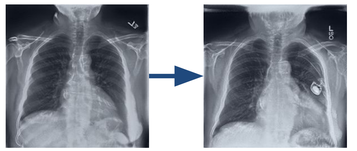
Examples of our Originals vs. Counterfactuals



Increased width of the heart silhouette



Pleural recessus obstruction



Appearance of a pacemaker

References
[1] Lang, Oran, et al. "Explaining in Style: Training a GAN to explain a classifier in StyleSpace." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
[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. 2019.