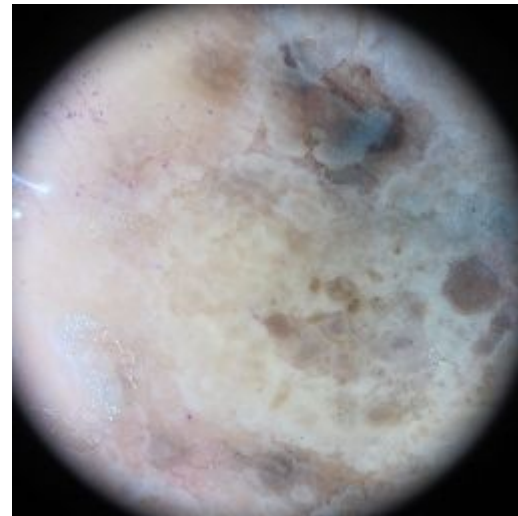
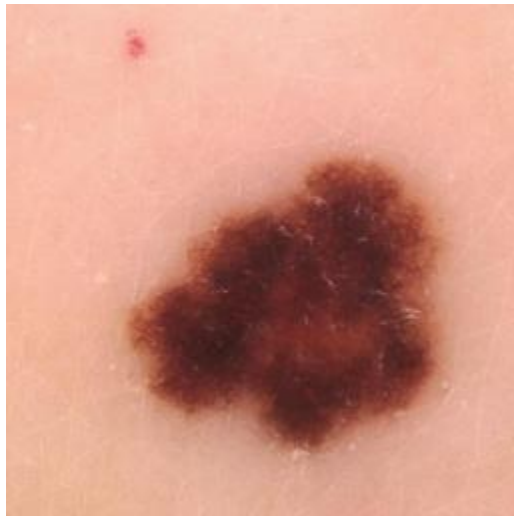


Explanation Evaluation: a melanoma detection case study

Krystian Król, Maciej Pióro, Karol Pustelnik

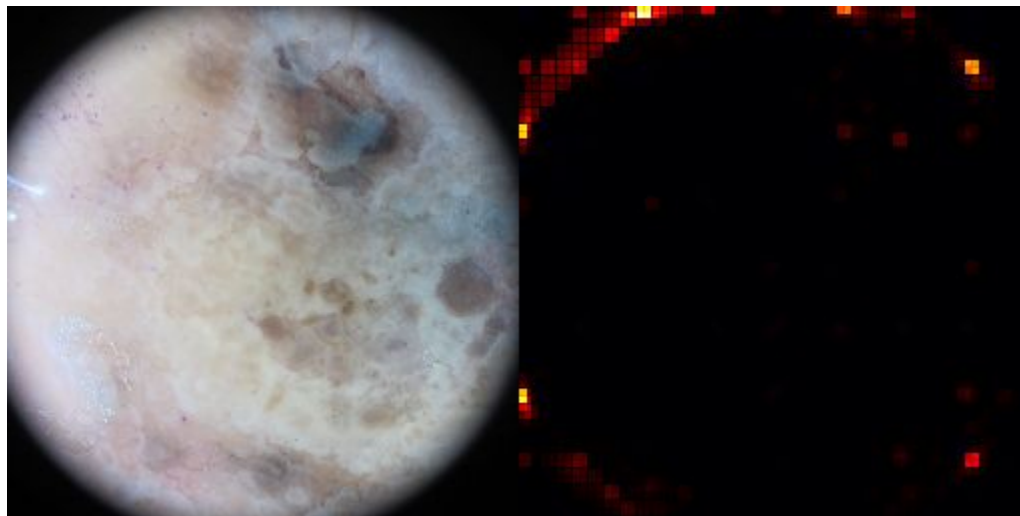
Examples of images from dataset



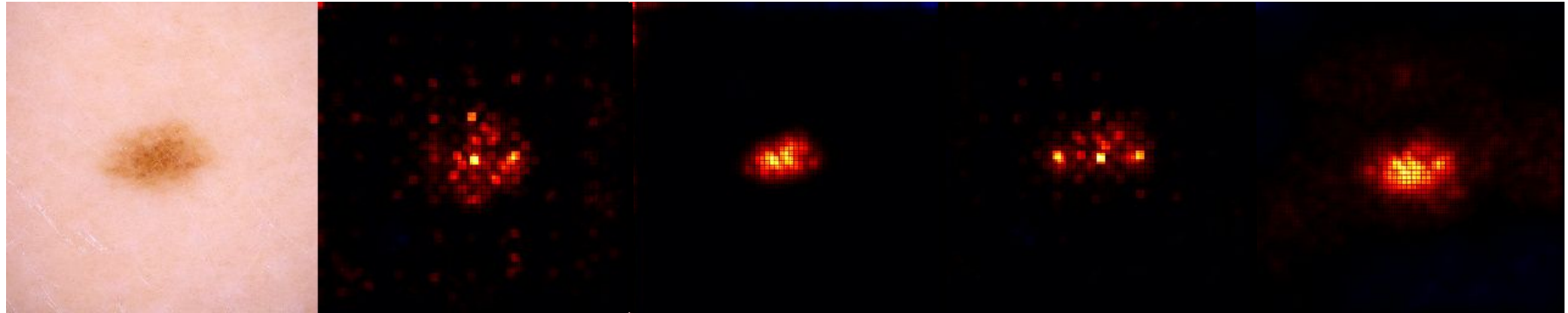
Research topics

- Influence of bias on explanations
- Comparison of **LRP** and **CRP** method for explanation
- Concepts present in model
- ResNet50 vs ResNet50 with attention

Problem discovered using LRP



Examples of explanations – non-melanoma



Examination

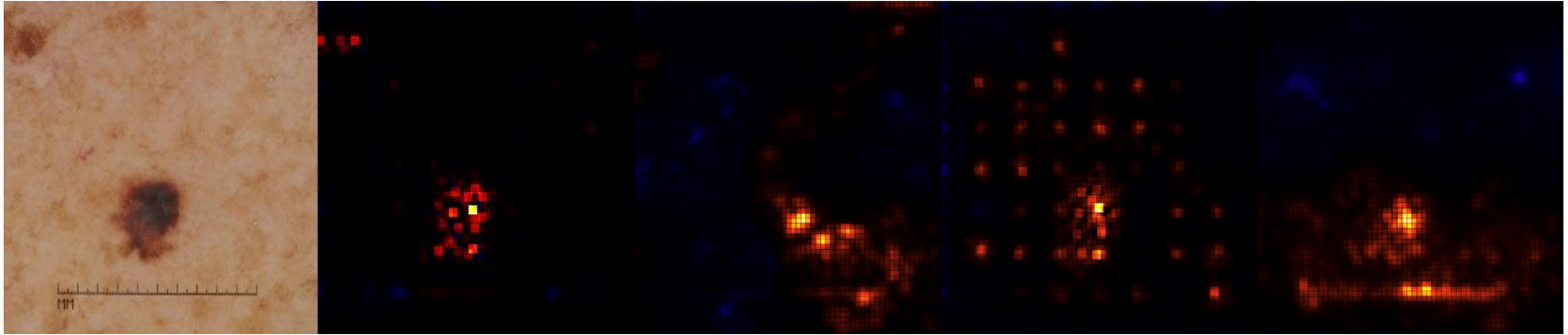
Resnet
Biased

Resnet + Att

Resnet
Unbiased

Resnet + Att
Unbiased

Examples of explanations – melanoma



Examination

Resnet
Biased

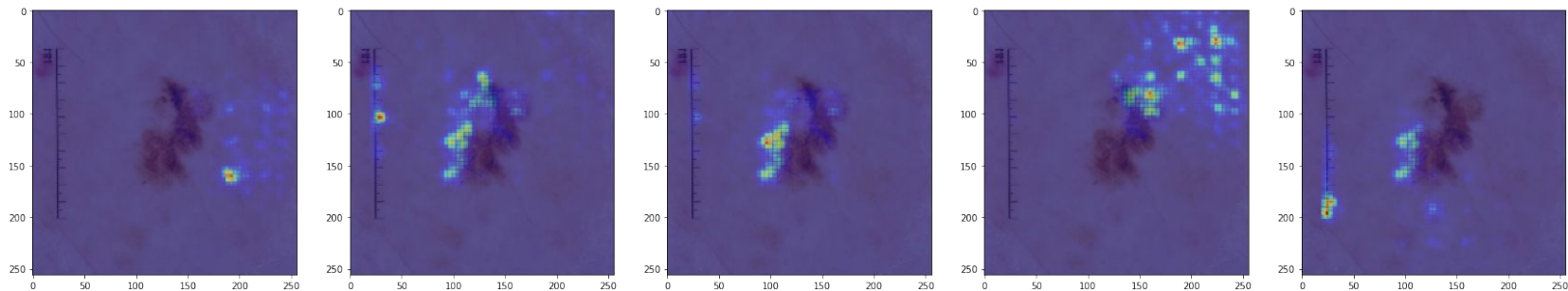
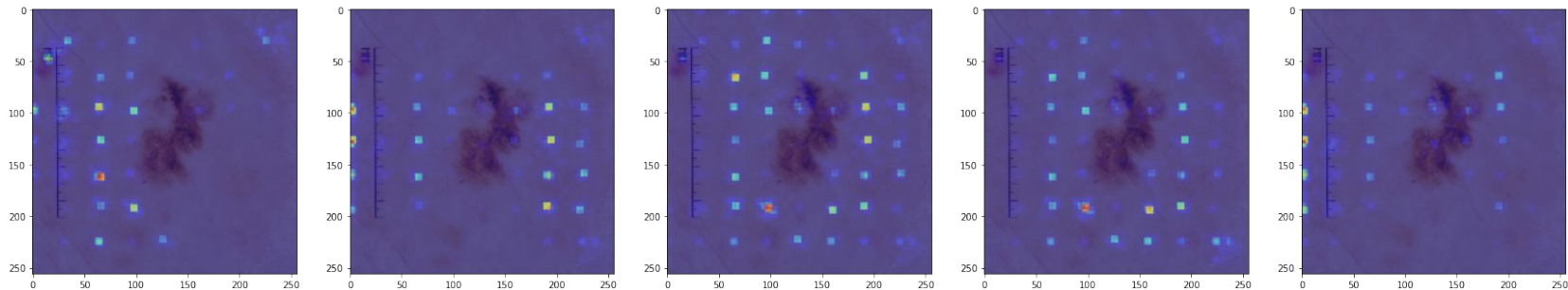
Resnet + Att

Resnet
Unbiased

Resnet + Att
Unbiased

Biased: Resnet w/o att vs Resnet w/ attention

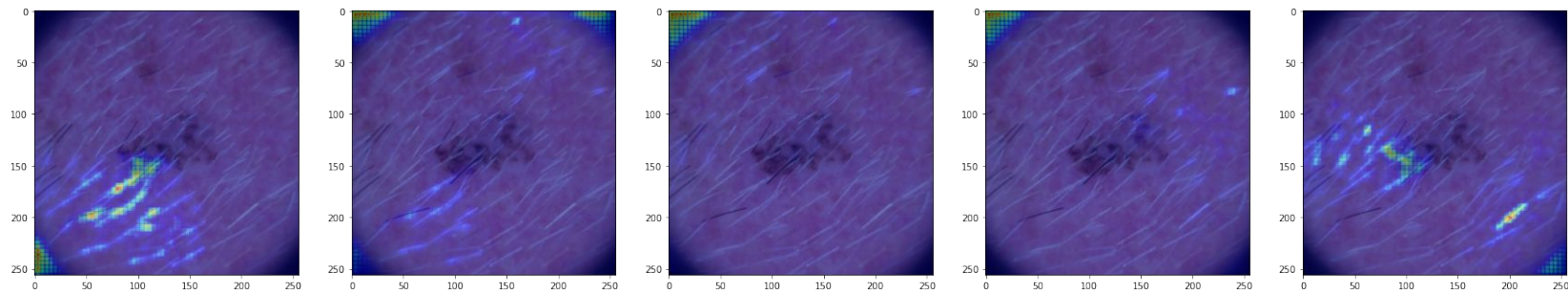
Resnet without attention



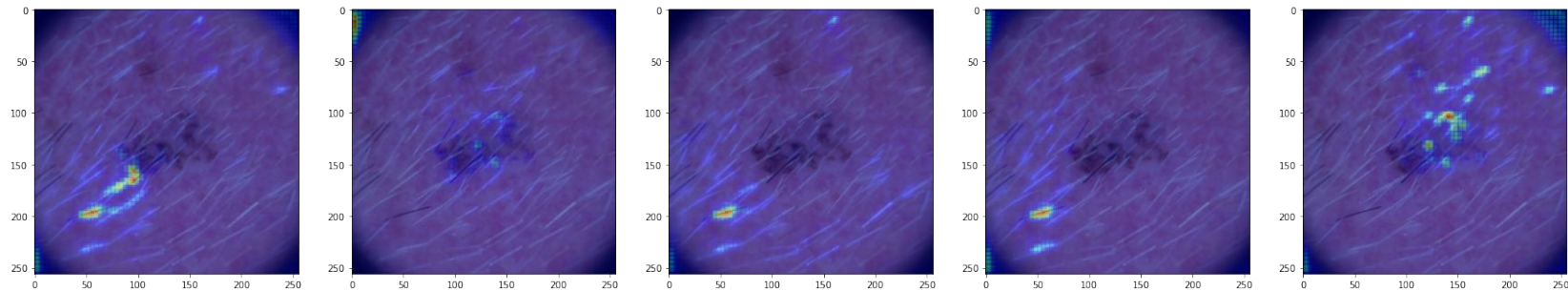
Resnet with attention

Resnet w/o att: Biased vs Unbiased

Biased Resnet

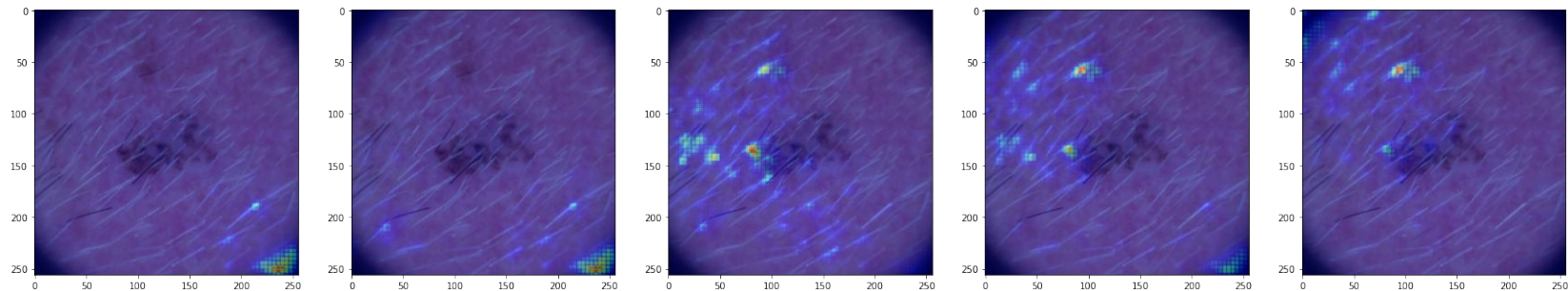
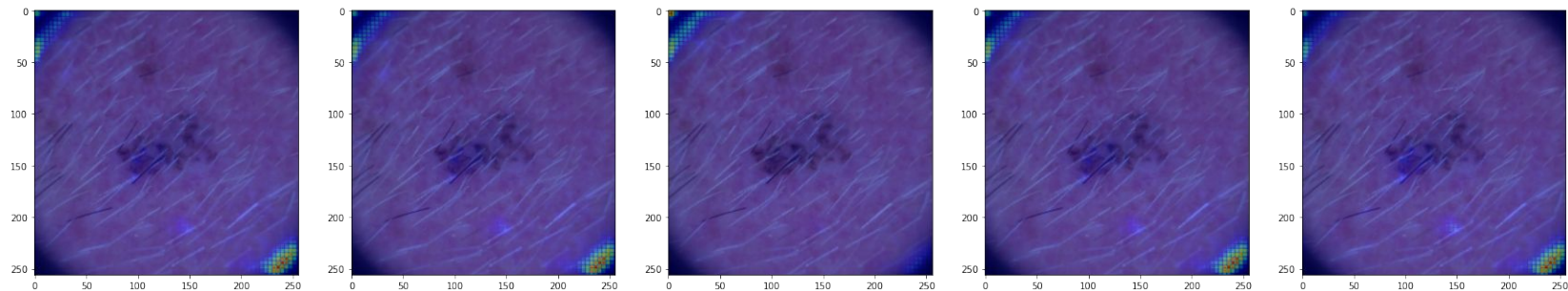


Unbiased Resnet



Resnet w/ att: Biased vs Unbiased

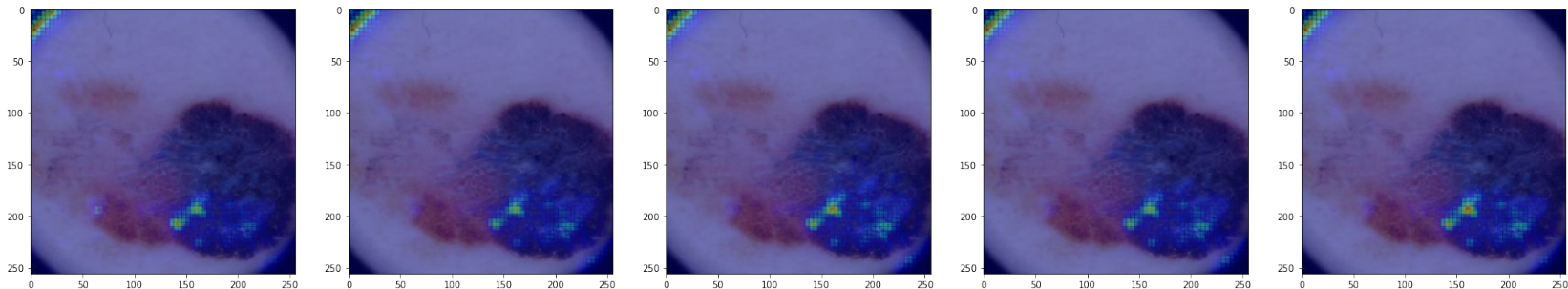
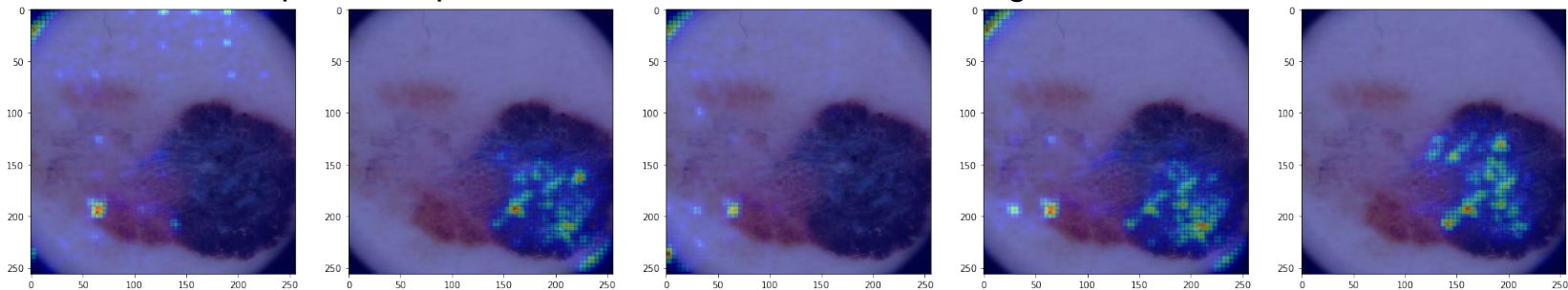
Biased Resnet with Attention



Unbiased Resnet with Attention

Biased Resnet w/ attention: CRP (relevance score) vs CRP (attention score)

Top 5 concepts in Resnet with attention according to CRP Relevance



Top 5 concepts in Resnet with attention according to attention score

Metrics Evaluation

	Resnet	Resnet + Attention	Unbiased Resnet	Unbiased Resnet + Attention
Average Sensitivity	0.014	0.044	0.026	0.056
Infidelity	0.0480	0.0731	0.0808	0.1556

Key takeaways

- Better metrics \neq better explanations
- Our proposed *attention score* fails to generate better CRP concepts than the CRP *relevance score*
- Bias can be detected using XAI methods and then be mitigated with carefully designed augmentations

Thank you for your attention

Questions?