Anatomy-Aware Masked Autoencoders for Hip Osteoarthritis Classification in X-ray Images

Jasper van Beusekom, Jesse Krijthe, Gijs van Tulder

- Technische Universiteit Delft -

1. Introduction

Osteoarthritis (OA) is a growing health issue that causes joint pain and stiffness, often diagnosed using X-ray images—a time-consuming and expensive process. While machine learning offers a faster alternative, it typically needs large amounts of labelled data.

Self-Supervised Learning (SSL) can reduce this need by learning from unlabelled images, but struggles with medical images due to their small regions of interest (ROI).

Masked Autoencoders (MAE) could pose a solution by learning latent features from only part of the image, and steering those features to be based on anatomy. This study explores how guiding MAEs with anatomy-based masks can improve hip OA detection from X-rays.

GOAL: **Explore** the design and effect of medically guided masking strategies for a masked autoencoder, when applied to osteoarthritis detection.

2. Preprocessing & Method

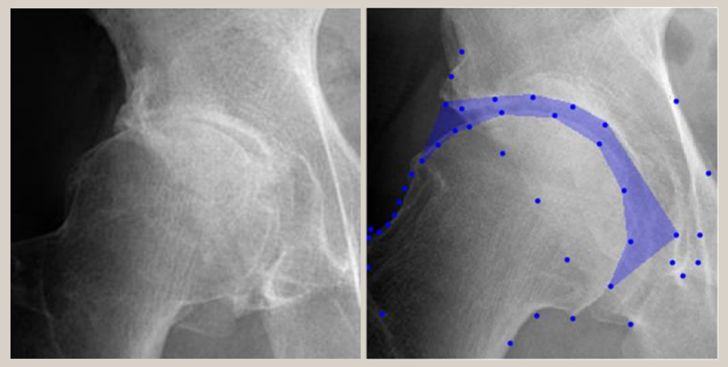
Preprocessing

- The CHECK dataset was used, containing 3.359 images from 1002 patients over 10 years.
- Images are normalized and cropped.
- BoneFinder is used to generate landmarks.
- Each hip is labelled using Kellgren-Lawrence grading.

Full X-ray

Masking

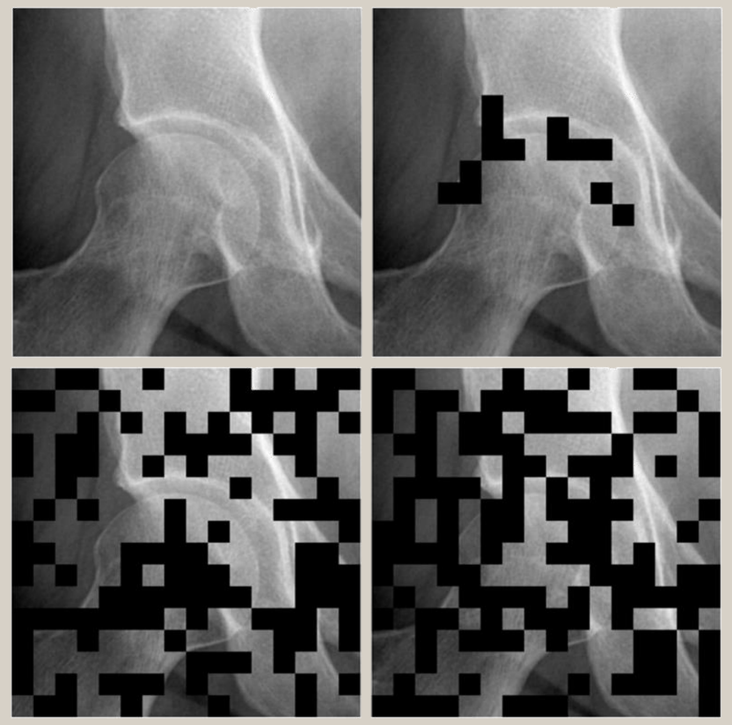
- Landmarks define Region of Interest.
- Four masking strategies:
  1) No masking
  2) ROI masking
  3) Background masking
  4) Random masking (standard)
- 1 and 4 are baselines.
- 2 and 3 are ROI-guided.
- Masking ratio of 50%



Region of Interest

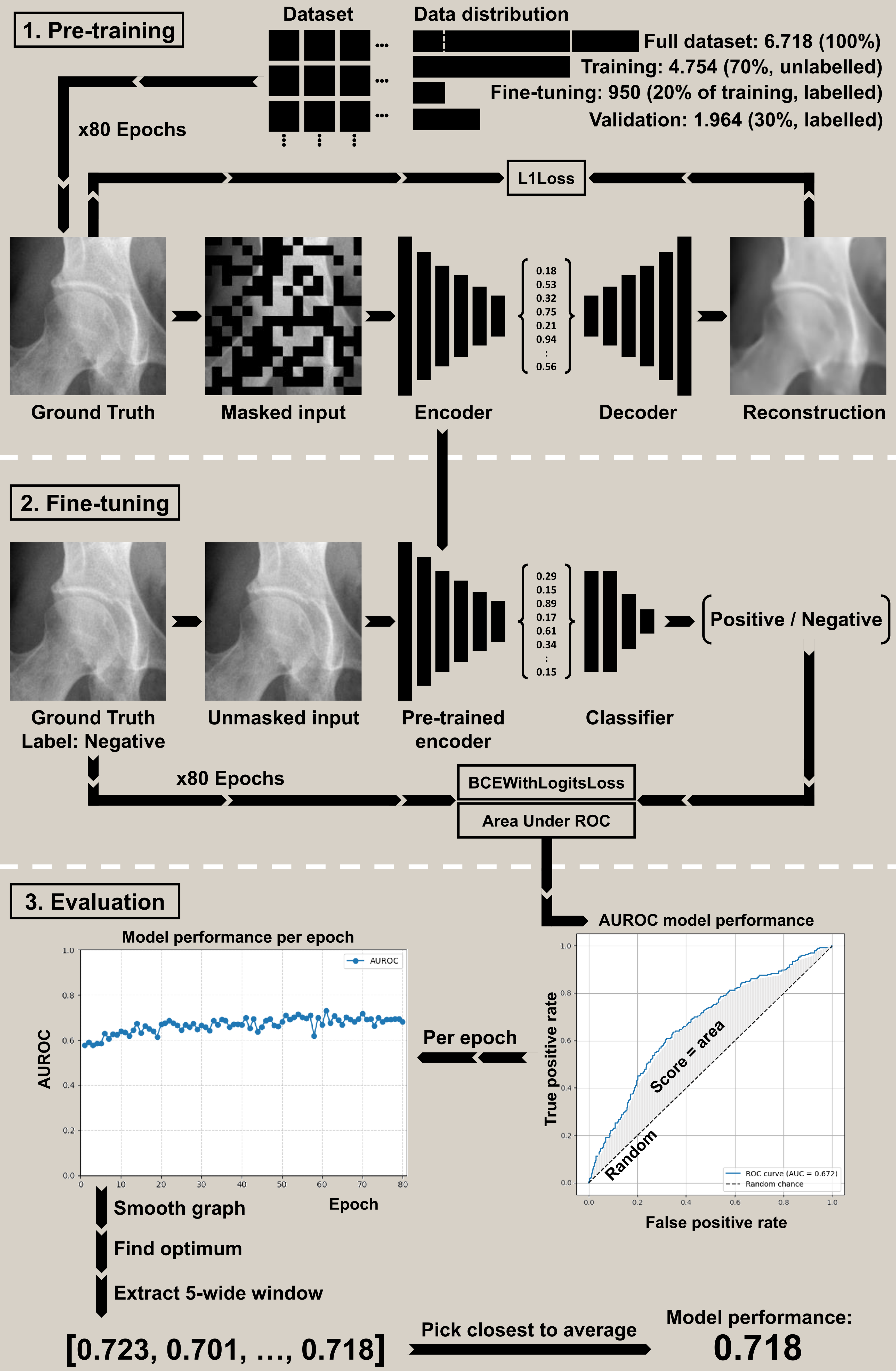
Evaluation

- AUROC is defined by the chance the model rates a positive sample higher than a negative sample.
- Each strategy tested for 3 patch sizes
- Five full runs per combination.



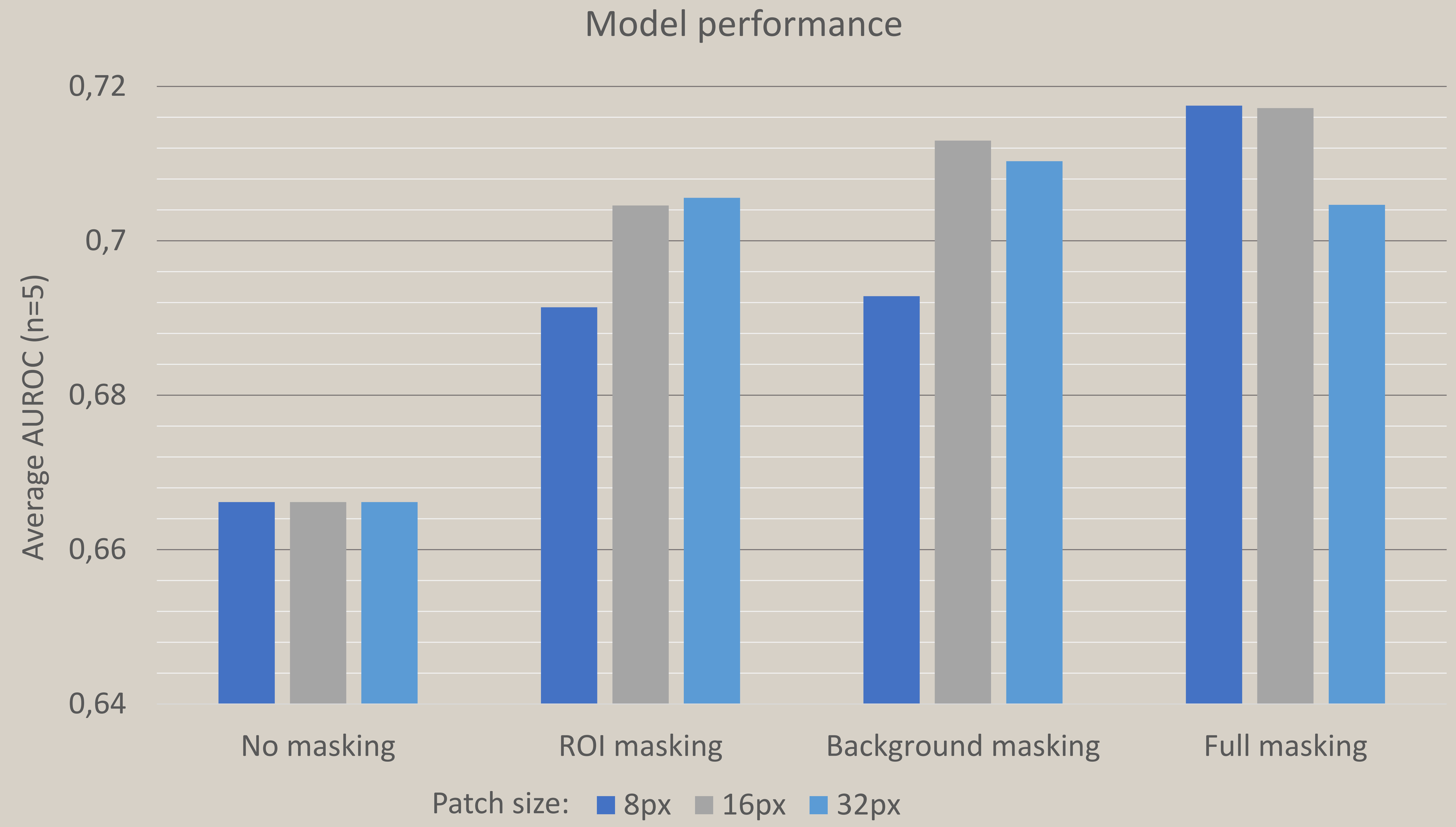
Masking strategies

3. Experiment



Note: Final performance is averaged over five fine-tuning repetitions. The entire pipeline is repeated for 3 patch sizes [8, 16, 32] and each masking strategy, resulting in 10 repeats, as the 'no masking' strategy is not affected by patch size.

4. Results



Average model performance (n=5)

Patch size	No masking	ROI masking	Background masking	Full masking
8px	0,6662	0,6914	0,6928	**0,7175**
16px	0,6662	0,7046	0,7130	**0,7172**
32px	0,6662	0,7056	**0,7103**	0,7047

5. Conclusion

Limitations

- 95% confidence interval per model in range 0.43%-1.15%
- Although reconstruction loss plateaued, more epochs might improve classification performance.
- Testing more masking ratios might reveal more patterns.

Conclusion and Discussion

- Masking outperforms no masking, confirming the promising potential of masked autoencoders in medical classification.
- Random masking outperforms ROI and background masking, suggesting the chosen architecture does not benefit from ROI-guided masking.
- Similar papers with a different architecture observed the opposite, indicating a fundamental difference in what drives performance in different architectures for masked autoencoders.

Broader application

- The observed results can guide the creation of more sophisticated masking strategies, and help build a more fundamental understanding of the factors that affect performance in medical classification tasks.

TUDelft

Name: Jasper Christiaan van Beusekom
Mail: j.c.vanbeusekom-1@student.tudelft.nl