Author

Zhenya Yancheva
z.y.yancheva-1@student.tudelft.nl

Diagnosing hip osteoarthritis in X-ray images

Anatomy-aware data augmentation techniques in Contrastive Self-supervised learning

Supervisors

Gijs van Tulder
Jesse Krijthe

1. Background

- ***Osteoarthritis:*** A degenerative joint disease which causes the protective cartilage that cushions the ends of the bones to wear down over time. Diagnosed from *X-ray images*.
- ***Medical Imaging***: Labeled medical data is not easy to acquire in large amounts. Usually taken under standardized protocols and thus has significantly lower variability.
- ***Contrastive Self-Supervised learning***: SSL is the process of pre-training a model on a hand-crafted task, which can extract meaningful features from the data, without the need for additional labels. Contrastive self-supervised methods are based on the assumption that image transformations do not alter its semantic meaning [1].
- ***Data augmentation methods***: Used to introduce variability in the training dataset. Fundamental part of Contrastive Self-Supervised learning.

2. Research Questions

How does using anatomy-aware data augmentation techniques impact downstream task performance of Contrastive Self-supervised learning models on X-ray images?

Do ROI-excluding data augmentation techniques benefit more than non-ROI-excluding ones when given additional anatomical information?

3. Methodology

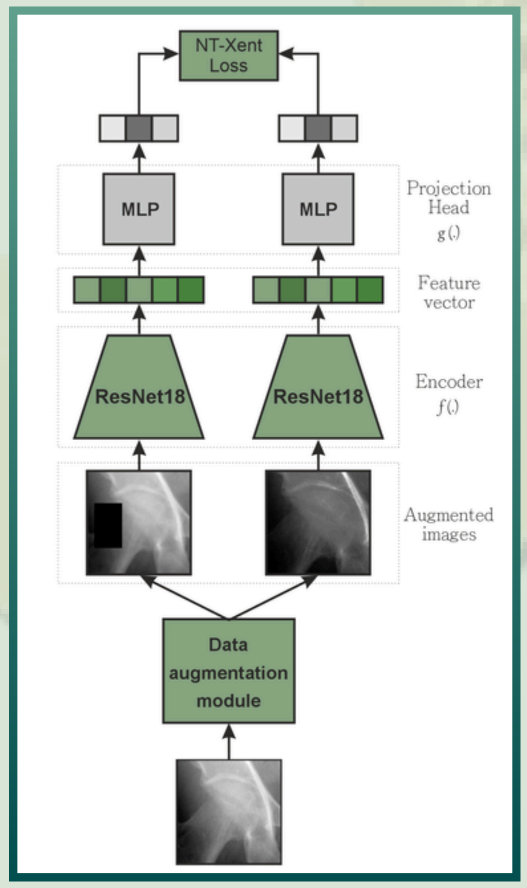

Fig 1: SimCLR training procedure. Two different augmented views are produced for an image. Each view is encoded and the encoding is then projected to the space where the contrastive loss is applied.

Model Architecture: SimCLR framework [2] - minimizes the distance between positive pairs and maximizes it between negative pairs. The two views of an image are considered a positive pair, all others - a negative one. The encoder used is ResNet18, a projection head is added on top of it during pre-training.

Data augmentations: Two types of augmentations are used and compared - Geometric (Crop and Random Erasing) and Appearance-based (Gaussian Blur and Contrast Enhancement). Custom versions are implemented, preserving the joint space.

Joint Space Segmentation: BoneFinder was used to obtain points that trace the curves of the bones. A bounding box is defined around the points outlining the femur head and the acetabular roof, which marks the primarily weight-bearing area of the hip joint.

Experimental approach:

- Four encoders were trained using the same data split and hyperparameters, but different sets of classical and custom anatomy-aware data augmentations. Linear probing is used to evaluate them.
- The dataset used is CHECK [3]. The areas around both hip joints were cropped out, resulting in around 6800 images.
- Evaluation metrics: The AUC ROC score is computed for each classifier and DeLong's test is performed to compare them.

4. Results & Discussion

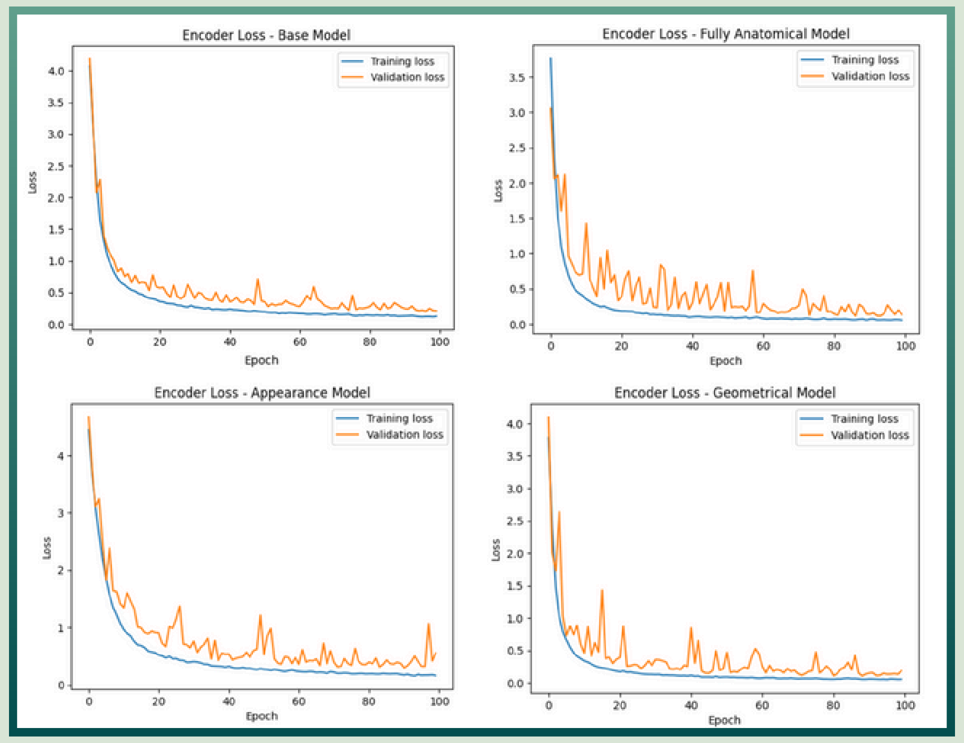

Fig 3: Training and validation loss curves of all 4 encoders.

Compared to the base model, all other models show a less stable validation loss. This indicates problems with generalization, which could be attributed to the small batch size used during training and the reduced strength of the custom data augmentations.

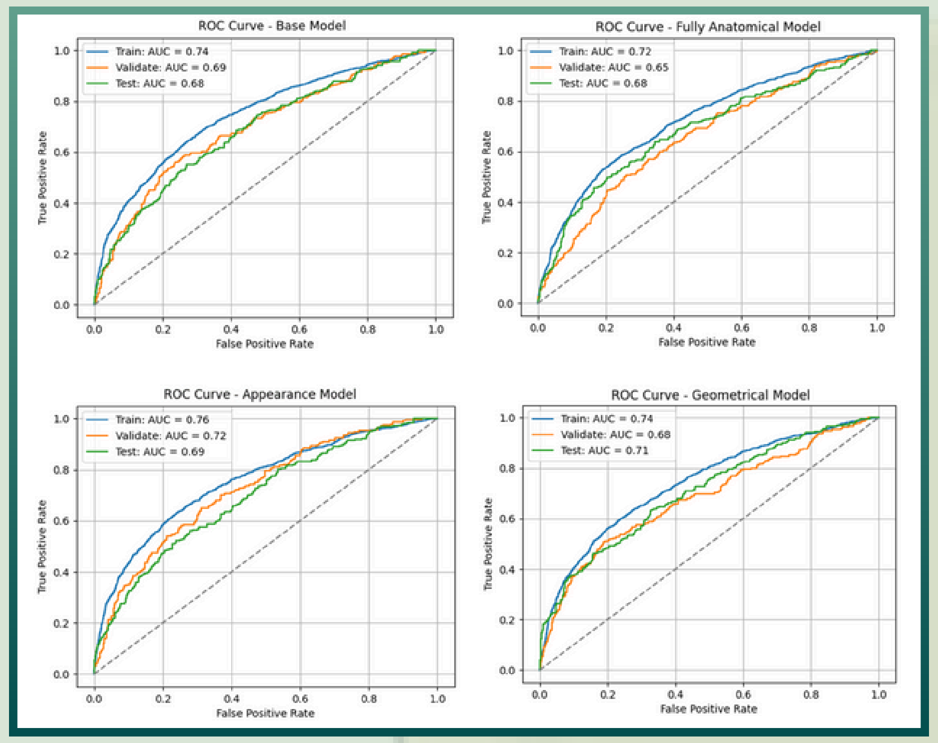

Fig 4: ROC curves and AUC ROC scores of all 4 models.

Since all classifiers, including the Base model one exhibit rather poor performance, it cannot be attributed to the difference in data augmentations, but rather the encoders' feature extracting abilities.

	Accuracy	AUC - ROC
Base Model	76%	0.68
Appearance Model	76%	0.69
Geometrical Model	78%	0.71
Fully Anatomical Model	75%	0.68

Fig 5: Accuracy and AUC - ROC scores for all four classifiers on the testing dataset.

	P-value from performing DeLong's test
Appearance Model	0.3637
Geometrical Model	0.0392
Fully Anatomical Model	0.8340

Fig 6: P-value from performing DeLong's test on all three classifiers, trained on top of the encoders utilizing anatomy-aware data augmentations, compared against Base Model classifier.

Despite the minor difference in accuracy and AUC ROC scores, DeLong's test reveals a statistically significant difference between the Base and Geometrical models.

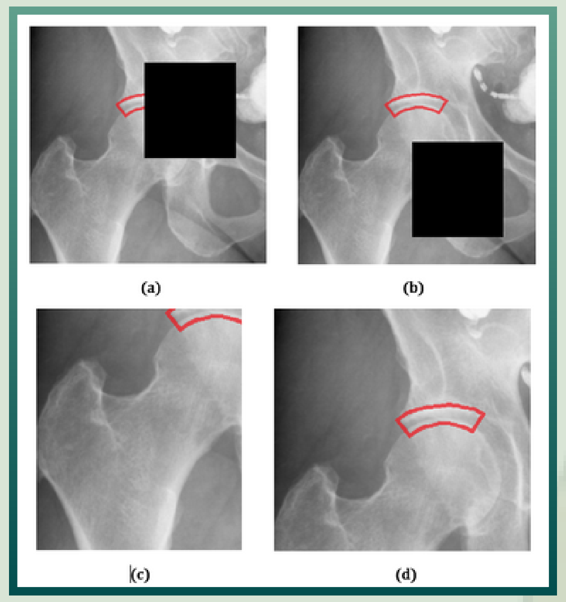

Fig 2: Images (a) and (c) show two views of the same image, obtained by classical augmentation methods (erase and crop respectively) and exclude the ROI. In (b) and (d) our custom anatomy-aware methods lead to views that preserve the ROI.

5. Conclusions

Utilizing our anatomy-aware approach for a larger number of chained image transformations may hinder the learning process and lead to less discriminative representations. However, when used in moderation, this approach could be beneficial, particularly for geometric transformations, such as crop and erase.

6. Limitations & Future Work

- Due to limitations in data availability and computational power, the batch size used is significantly smaller than the one used by the original SimCLR model. This difference notably influences the performance of the model.
- The data augmentations could be made more aggressive.
- The custom anatomy-aware augmentations could be used for augmenting medical datasets and used for training other types of models.

References

[1] Shih-Cheng Huang et al. "Self-supervised learning for medical image classification: a systematic review and implementation guidelines". In: npj Digital Medicine 6.1 (2023). doi: 10.1038/s41746-023-00811-0.

[2] Ting Chen, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. CoRR, abs/2002.05709, 2020

[3] J. Wesseling, Maarten Boers, Max Viergever, Wim Hilberdink, Floris Lafeber, Joost Dekker, and Johannes Bijlsma. Cohort profile: Cohort hip and cohort knee (check) study. International journal of epidemiology, 45, 08 2014