# Towards Reducing the Need for Annotations in Digital Dermatology with Self-Supervised Learning

IJCAI-ECAI-2022 SDAIH Workshop

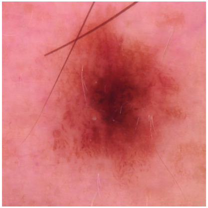
Fabian Gröger[1] , Philippe Gottfrois[2] , Ludovic Amruthalingam[2] , Alvaro Gonzalez-Jimenez[2] , Simone Lionetti[1] , Alexander A. Navarini[2,3] , and Marc Pouly[1]

[1]Lucerne University of Applied Sciences and Arts, Rotkreuz, Switzerland
[2]Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland
[3]Department of Dermatology, University Hospital of Basel, Switzerland

July 23, 2022

# Introduction to Dermatology

- Branch of medicine concerned with the diagnosis and treatment of skin disorders
- Image analysis can have a significant impact since diseases are visible and can easily be photographed
- Two main types of images are dermoscopy and clinical images



(a) dermoscopy (1)

(b) clinical (2)

Figure: Most common image types encountered in dermatology.

# Problems with Deep Learning in Dermatology

- Recordings in patient-files too coarse, more details needed
- Annotations are costly, and images under strict regulations
- Difficult to get experts to agree on annotations (3)
- In the medical domain often the main limitation is the scarce amount of annotated data

# Current Approaches in Medical Imaging

- Model pre-trained on ImageNet and transferred to medical downstream tasks
- Criticised approach since features from natural images may not be ideal representations for medical contexts (4)
- Still requires a sufficiently large dataset to adjust the global features for a specific task

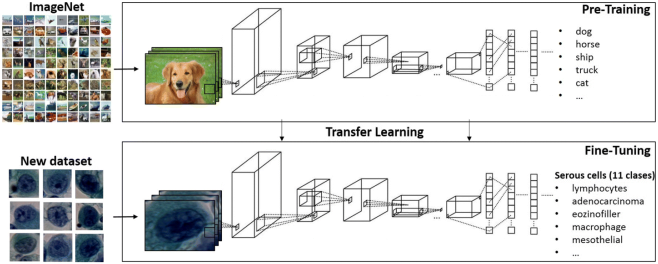



Figure: Illustration of a standard transfer-learning approach for medical image analysis. (5)

# Introduction to Self-Supervised Learning

- Self-supervised learning (SSL) is a hybrid approach that combines supervised and unsupervised learning
- SSL learns semantically useful features by creating a supervised objective from a pool of unlabelled data, called *pretext task*
- Learned features can be used in *downstream tasks* where annotated data is scarce
- SSL became popular in the medical domain as large volumes of unlabelled data are easier to obtain than annotated counterparts

# Project Focus

## Research Question

Can self-supervised domain-specific pre-training help mitigate the need for annotated samples in dermatology?

# Pre-training Data & Downstream Tasks

- Pre-training data includes both dermoscopy and clinical images
  - 242′039 images from both public and private datasets
- Three downstream tasks were used to evaluate the performance of different SSL algorithms
- No downstream data was present in the pre-training

Table: List of the different downstream tasks.

| Name | Task | Type | # Images | Notes |
|---|---|---|---|---|
| PAD-UFES-20 (6) | diagnosis | clinical images | 2′298 | smartphones |
| Fitzpatrick17k (2) | diagnosis | clinical images | 16′577 | multi-skin |
| HAM10000 (1) | diagnosis | dermascopy | 10′015 | multi-skin |

# Self-Supervised Model Architectures

Table: List of backbones used with different SSL methods, along with the respective total number of parameters and the fine-tuning architecture.

| Algorithm | Backbone | # Params | Classification Head |
|---|---|---|---|
| Ⓐ ColorMe (7) | ResNet-50 | 23 Mio. | linear head |
| Ⓑ SimCLR (8) | ResNet-50 | 23 Mio. | linear head |
| Ⓒ BYOL (9) | ResNet-50 | 23 Mio. | linear head |
| Ⓓ DINO (10) | ViT-tiny $16 \times 16$ | 5.4 Mio. | linear head |
| Ⓔ iBOT (11) | ViT-tiny $16 \times 16$ | 5.4 Mio. | linear head |

# Experiment Procedure

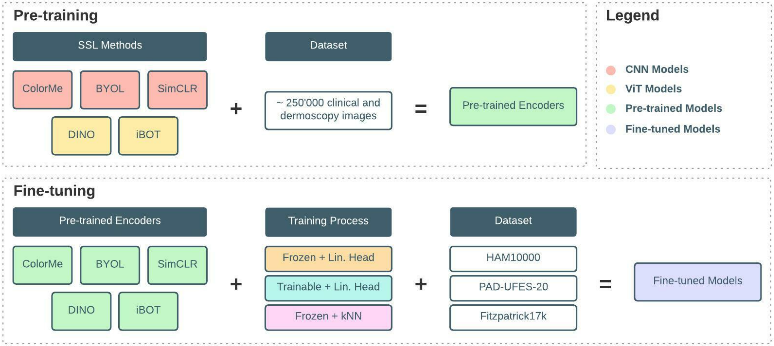



Figure: Illustration of the project experiment procedure.

# Linear Evaluation

Table: Linear performance of various models and a baseline on the hold-out test set of the three open-source dermatology diagnosis downstream tasks.

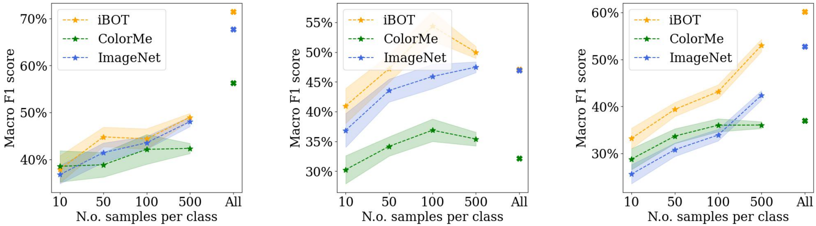
| Type | Method | Test F1 Score Fitzpatrick17k | PAD-UFES-20 | HAM10000 |
|---|---|---|---|---|
| Baselines | Stratified sampling | 33.4 % | 18.0 % | 14.0 % |
| | ImageNet | 51.0 % | 49.7 % | 54.1 % |
| SSL pre-training | A ColorMe | 44.8 % | 42.2 % | 47.0 % |
| | B SimCLR | 37.0 % | 32.7 % | 28.2 % |
| | C BYOL | 48.1 % | 34.4 % | 44.4 % |
| | D DINO | 46.7 % | 44.2 % | 57.2 % |
| | E iBOT | 53.0 % | 58.2 % | 72.0 % |

# Fine-tuning Evaluation

Table: Fine-tuned performance of two self-supervised models and a baseline reported on the hold-out test set of the three open-source dermatology diagnosis downstream tasks.

| | Test F1 Score | | |
|---|---|---|---|
| **Pre-training** | **Fitzpatrick17k** | **PAD-UFES-20** | **HAM10000** |
| ImageNet | 72.1 % | 61.5 % | 79.0 % |
| A ColorMe | 71.0 % | 61.7 % | 73.1 % |
| E iBOT | 73.9 % | 64.1 % | 82.0 % |

# kNN Performance




(a) Fitzpatrick17k (b) PAD-UFES-20 (c) HAM10000

Figure: Results of a kNN classifier when varying the number of samples per class for all three open-source diagnosis tasks.

# Conclusion

- iBOT was able to outperform ImageNet initialisation across multiple downstream tasks
- Best performing self-supervised pre-trained models used up to four times fewer parameters than the supervised baseline
- Uncertainty whether performance difference is due to the pre-training strategy or the architecture
- In the future, we plan ablation experiments to determine what caused the performance gain

# Thank You!
# Q & A

# References I


[1] P. Tschandl, C. Rosendahl, and H. Kittler, ``The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,'' *Scientific Data*, Aug. 2018.

[2] M. Groh, C. Harris, L. Soenksen, *et al.*, ``Evaluating Deep Neural Networks Trained on Clinical Images in Dermatology with the Fitzpatrick 17k Dataset,'', IEEE Computer Society, Jun. 2021.

[3] L. Zhang, R. Tanno, M.-C. Xu, *et al.*, ``Disentangling human error from the ground truth in segmentation of medical images,'' in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS'20, Dec. 2020.

[4] C. Matsoukas, J. F. Haslum, M. Sorkhei, M. Söderberg, and K. Smith, ``What Makes Transfer Learning Work For Medical Images: Feature Reuse & Other Factors,'' Tech. Rep., Mar. 2022.

(5) E. Baykal, H. Dogan, M. E. Ercin, S. Ersoz, and M. Ekinci, ``Transfer learning with pre-trained deep convolutional neural networks for serous cell classification,'' *Multimedia Tools and Applications*, Jun. 2020.

(6) A. G. C. Pacheco, G. R. Lima, A. S. Salomão, *et al.*, ``PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones,'' *Data in Brief*, Oct. 2020.

(7) Y. Li, J. Chen, and Y. Zheng, ``A Multi-Task Self-Supervised Learning Framework for Scopy Images,'' in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2020.

(8) T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton, ``Big Self-Supervised Models are Strong Semi-Supervised Learners,'', Oct. 2020.

(9) J.-B. Grill, F. Strub, F. Altché, *et al.*, ``Bootstrap Your Own Latent - A New Approach to Self-Supervised Learning,'' in *Advances in Neural Information Processing Systems*, 2020.

(10) M. Caron, H. Touvron, I. Misra, *et al.*, ``Emerging Properties in Self-Supervised Vision Transformers,'', 2021.

(11) J. Zhou, C. Wei, H. Wang, *et al.*, ``iBOT: Image BERT Pre-Training with Online Tokenizer,'', Jan. 2022.

(12) T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, ``A Simple Framework for Contrastive Learning of Visual Representations,'' in *Proceedings of the 37th International Conference on Machine Learning*, PMLR, Nov. 2020.

(13) K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, ``Momentum Contrast for Unsupervised Visual Representation Learning,'', Mar. 2020.

# References IV


(14) A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, ``An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,'', Sep. 2020.

(15) O. Ronneberger, P. Fischer, and T. Brox, ``U-net: Convolutional networks for biomedical image segmentation,'' *CoRR*, 2015.

(16) J. Chen, Y. Lu, Q. Yu, *et al.*, ``Transunet: Transformers make strong encoders for medical image segmentation,'' *CoRR*, 2021.