Self-supervised learning for medical imaging

https://github.com/PolLM/Self-supervised-learning-for-medical-imaging

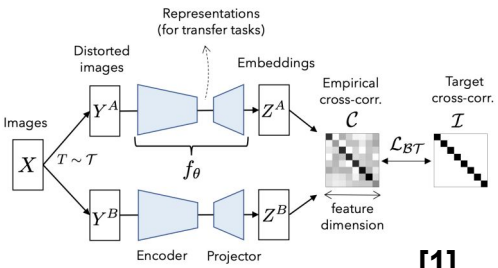
Advisor: Kevin McGuinness

Students: Francesc Garcia, Pol Llopart, Sergio Rodriguez

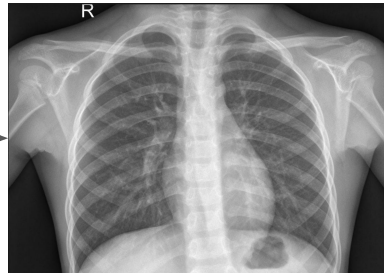
Introduction - Motivation

- Labelling task in medical field is expensive.
  - Requires expert professionals + Plenty of samples

- Barlow twins in medical field



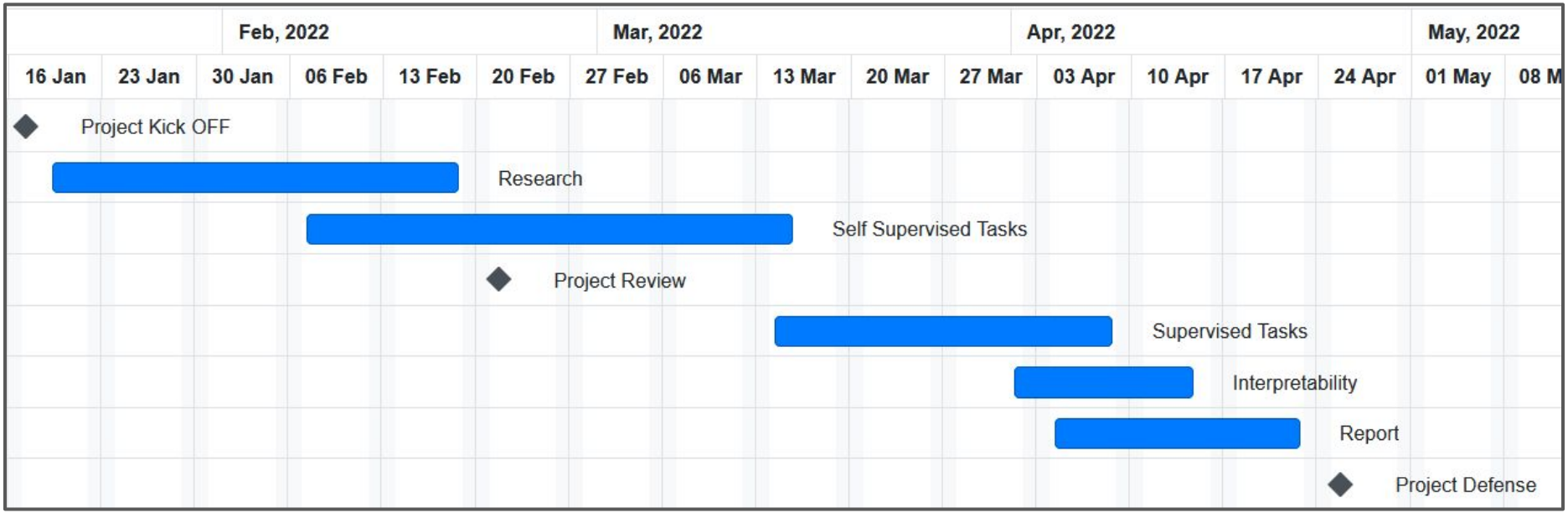
[1]



Introduction - Goals

- Pre-train a model using Barlow Twins architecture on Chest X-ray images.
- Use the self-supervised pre-trained model in several downstream tasks and compare its performance to untrained models.
- Interpret how the model works: understand what parts of the image (visual patterns) are key to determine the predictions.
- Observe whether Barlow Twins could be applied to medical images for a classification and evaluate its label efficiency.
- Compare Barlow Twins' performance on medical images with SOA and similar self-supervised models.

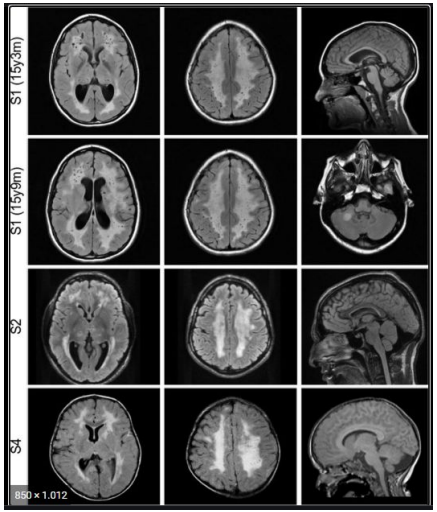
Project Plan



Dataset

1- Data types

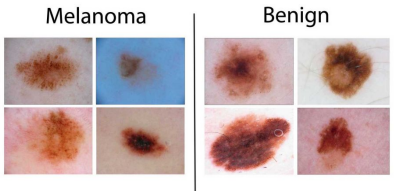
X-RAY



MRI

CT scan

Echography



Pictures

2- Sources of information

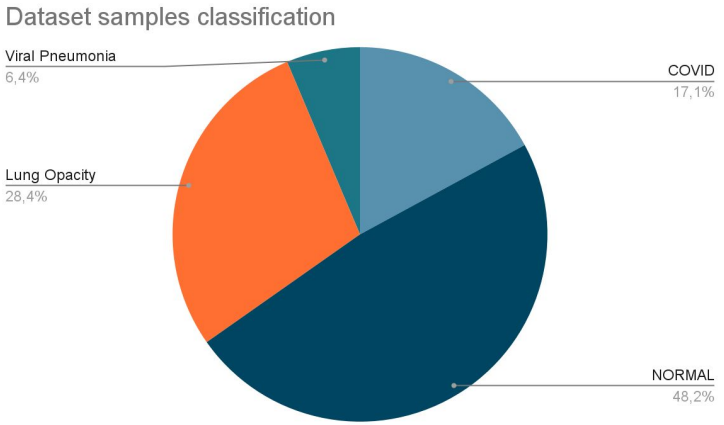
kaggle [2]

StanfordML [9]

Introduction - State of the art

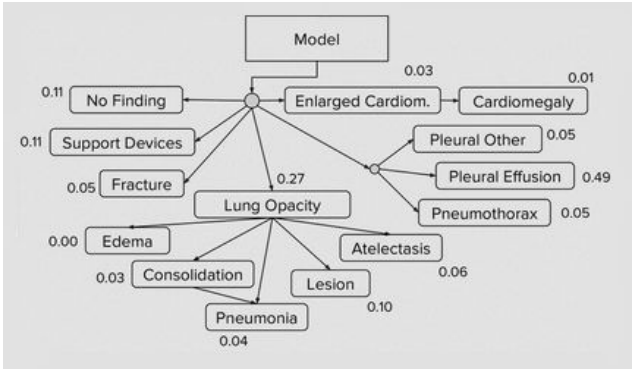
State of the art (Covid Dataset)

Accuracy: 90% - 99%



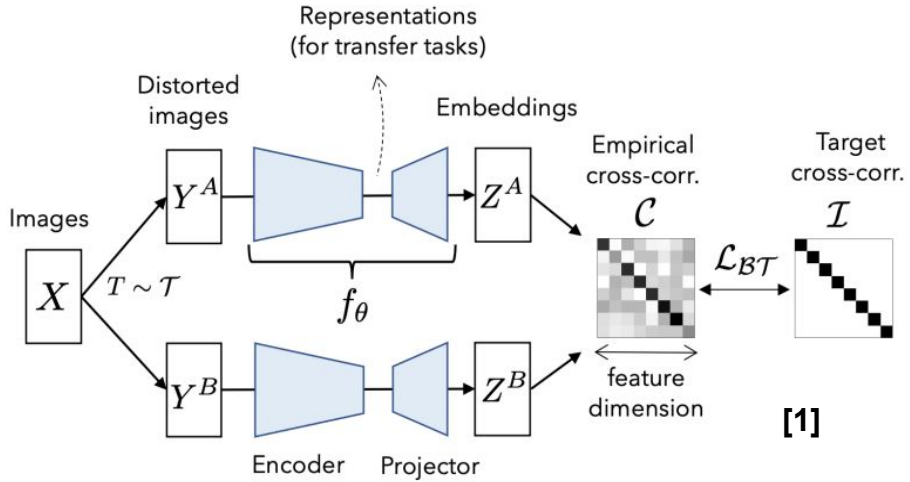
State of the art (CheXpert)

Accuracy: 92.8% - 93%



Self Supervised - Model architecture

- Barlow Twins is a Self-supervised Deep Learning method. Its objective function measures the cross-correlation matrix between the embeddings of two identical networks fed with distorted versions of a batch of samples.



[1]

$$\mathcal{L}_{\mathcal{BT}} \triangleq \underbrace{\sum_i (1 - \mathcal{C}_{ii})^2}_{\text{invariance term}} + \lambda \underbrace{\sum_i \sum_{j \neq i} \mathcal{C}_{ij}{}^2}_{\text{redundancy reduction term}}$$

Barlow Twins Lambda from the original paper:

$$\lambda_{BT} = 5e - 3$$

Self Supervised - Resources study

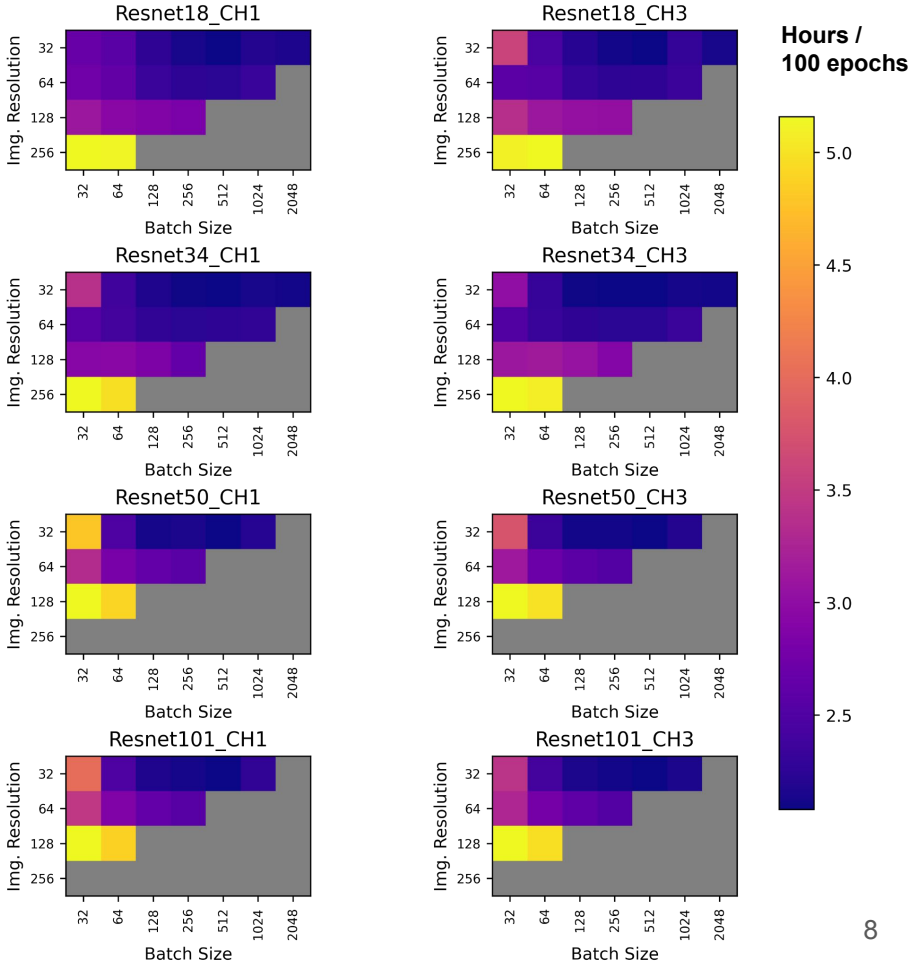
Barlow Twins Architectures:
- Resnet
- Efficientnet

COVID-19_Radiography Dataset (21.165 imgs.) **[2]**
Training for 1 epoch
Projector layers: 512 x 512
GPU: Nvidia RTX 2070 super

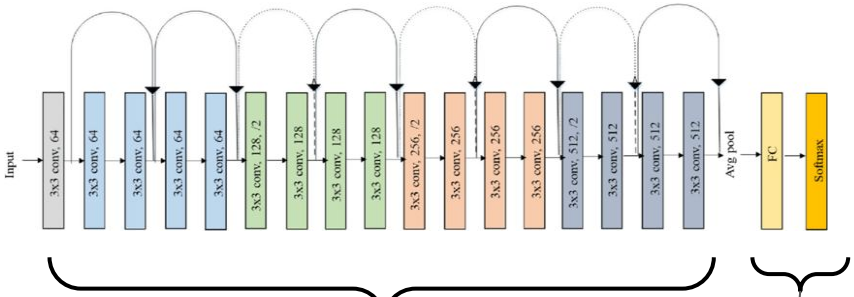
Selected parameters

- Resolution of the chest X-ray images: **224**
- Batch size: **128**
- Encoder/backbone model used in the Barlow Twins architecture: **Resnet- 18**
- Number of input channels (1 or 3): **1 single channel**

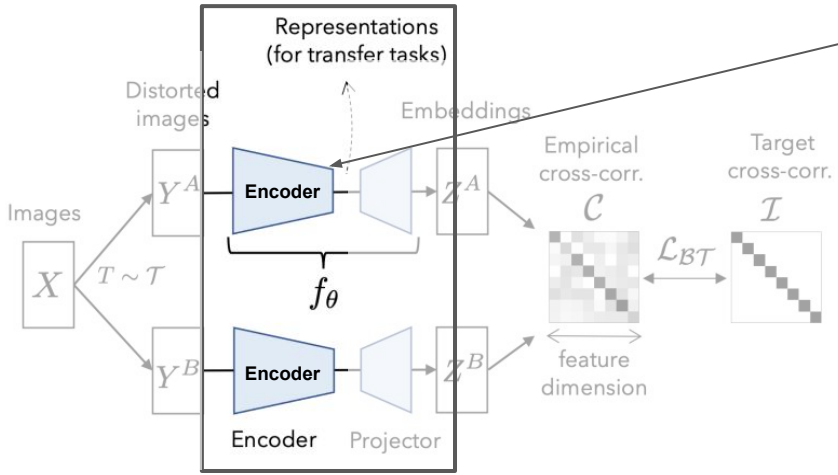
Img. Resolution vs. Batch Size colormaps

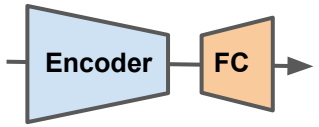


Self Supervised - Backbone model: Resnet 18



Self-supervised training (BT Loss)



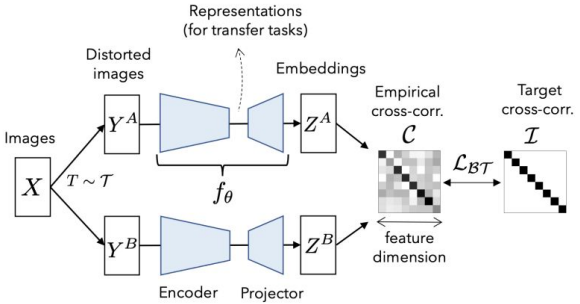


Supervised training (CCE Loss)

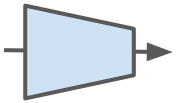
Self Supervised - Hyperparameter tuning pipeline

SELF-SUPERVISED TRAINING

Self-supervised training (BT Loss)



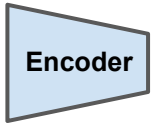
Saving encoder/backbone state dict



SUPERVISED TRAINING

Loading pre-trained encoder/backbone

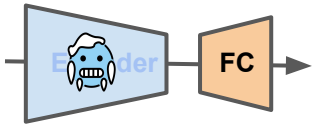
Pre-trained Resnet-18



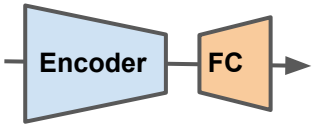
Linear projector 512 x 4
(num. of classes from Covid dataset)

Supervised training (CCE Loss)

Linear projector

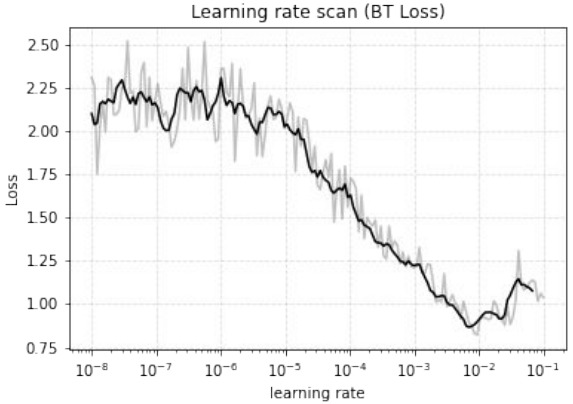


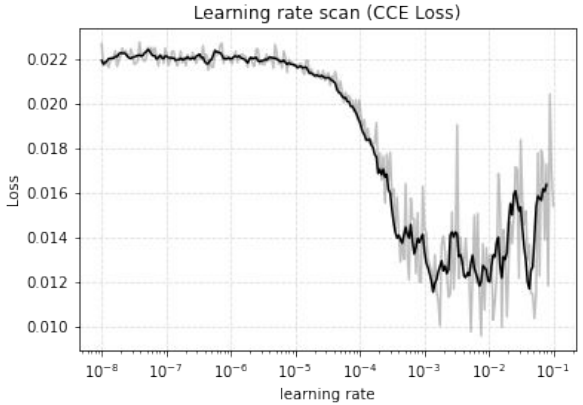
Entire network



Performance validation (Accuracy)

Self Supervised - LR scan & Sceduler

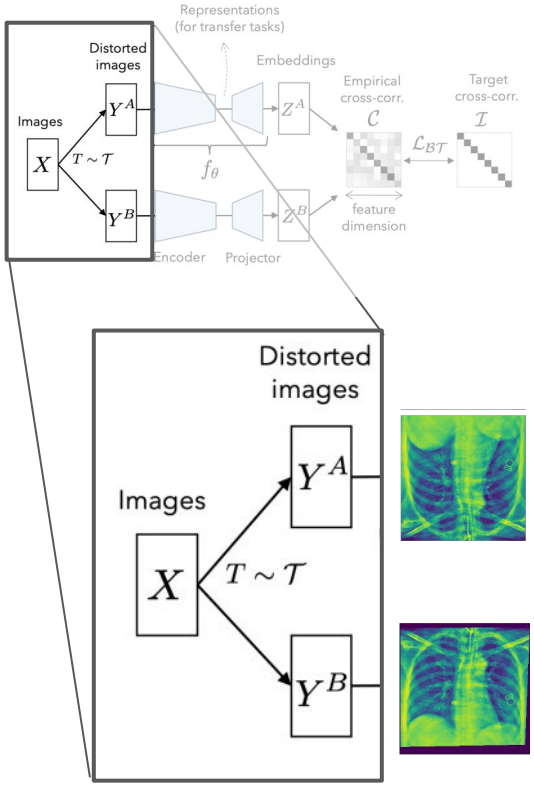




- Learning rate scan for self-supervised and supervised tasks:
  - Initial learning rate for the self-supervised experiments **lr = 2e−3**
  - Initial learning rate for the supervised experiments **lr = 1e−4**
- Learning rate scheduler: **Cosine annealing**

	Constant lr	*Linear lr*	*Cosine annealing*
Final lr (after 10 epochs)	32.8685	37.4496	32.0759

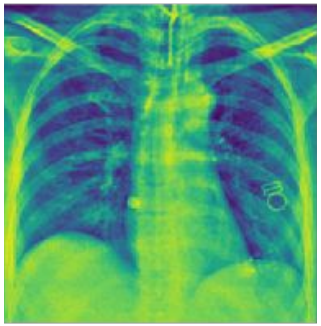
Self Supervised - Image transformations



Name	Transformations
COVID-19 prognosis via self-supervised representation learning and multi-image prediction [6]	**Each random augmentation was applied with probability p = 0.5** **Random resizing/cropping** (224x224), **Random horizontal flipping,** Random vertical flipping, Random Gaussian blur, Random Gaussian noise addition, Histogram normalization.
Momentum Contrastive Learning for Few-Shot COVID-19 Diagnosis from Chest CT Images [7]	**Each random augmentation was applied with probability p = 0.5** **Random resizing/cropping** (512x512), **Random horizontal flipping.**
Big Self-Supervised Models Advance Medical Image Classifications [8]	**Each random augmentation was applied with probability p = 0.5** Random rotation, **Random resizing/cropping** (224x224), **Random left-right flipping**, Random additive brightness modulation Random multiplicative contrast modulation

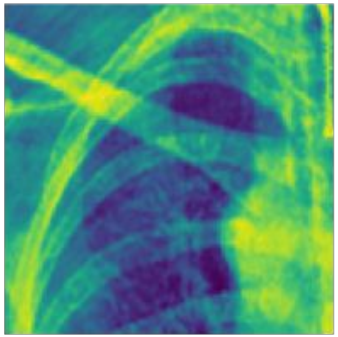
Self Supervised - Image transformations

Original



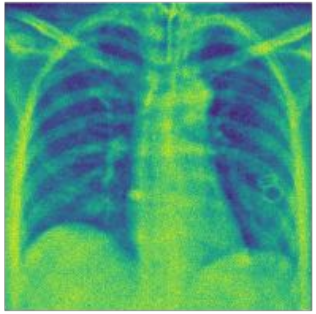
Base transformations

Random resizing/cropping(224x224),
Random horizontal flipping,

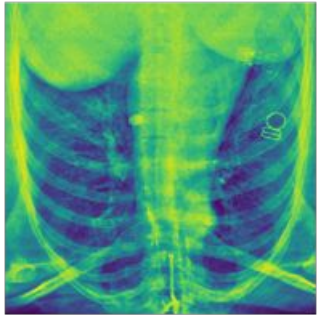


+

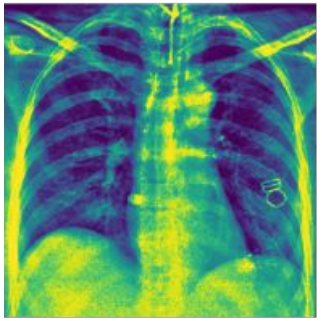
Gaussian Blur & Noise



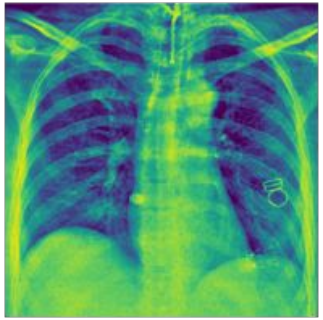
Random Vertical Flip



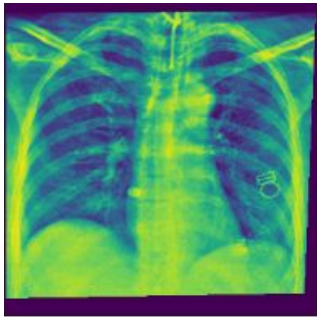
Random Equalize



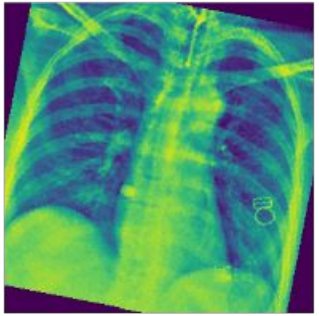
Brightness & Contrast Modulation



Random Perspective

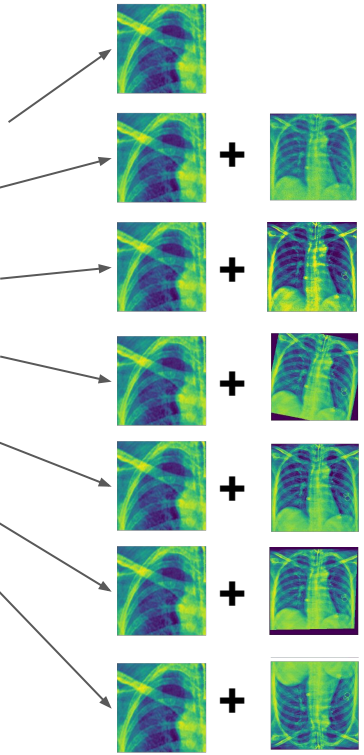


Random Rotation



Self Supervised - Transformations 1st scan

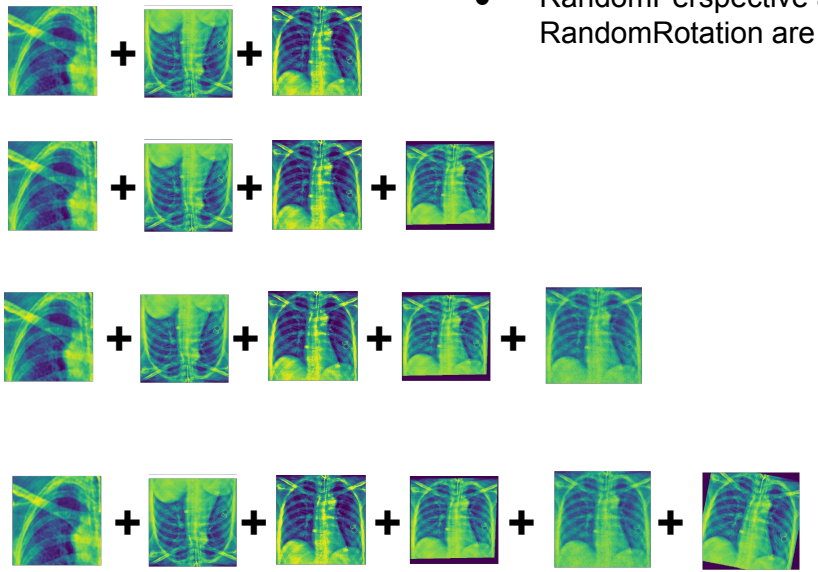
1st Transformations scan Validation Accuracy	*Linear projector*	*Full network*
Base transformations	0.6044	**0.8879**
Base transformations + GaussianBlur + GaussianNoise	0.6053	0.8833
Base transformations + RandomEqualize	**0.6136**	0.8686
Base transformations + RandomRotation	0.6030	0.8860
Base transformations + Brightness & Contrast modulation	0.5782	0.7786
Base transformations + RandomPerspective	0.6049	0.8686
Base transformations + RandomVerticalFlip	0.6131	0.8824



- 20 epochs of self-supervised training + 5 epochs of supervised training, using Covid dataset.
- Decided to discard Brightness & Contrast modulation due to their poor performance.

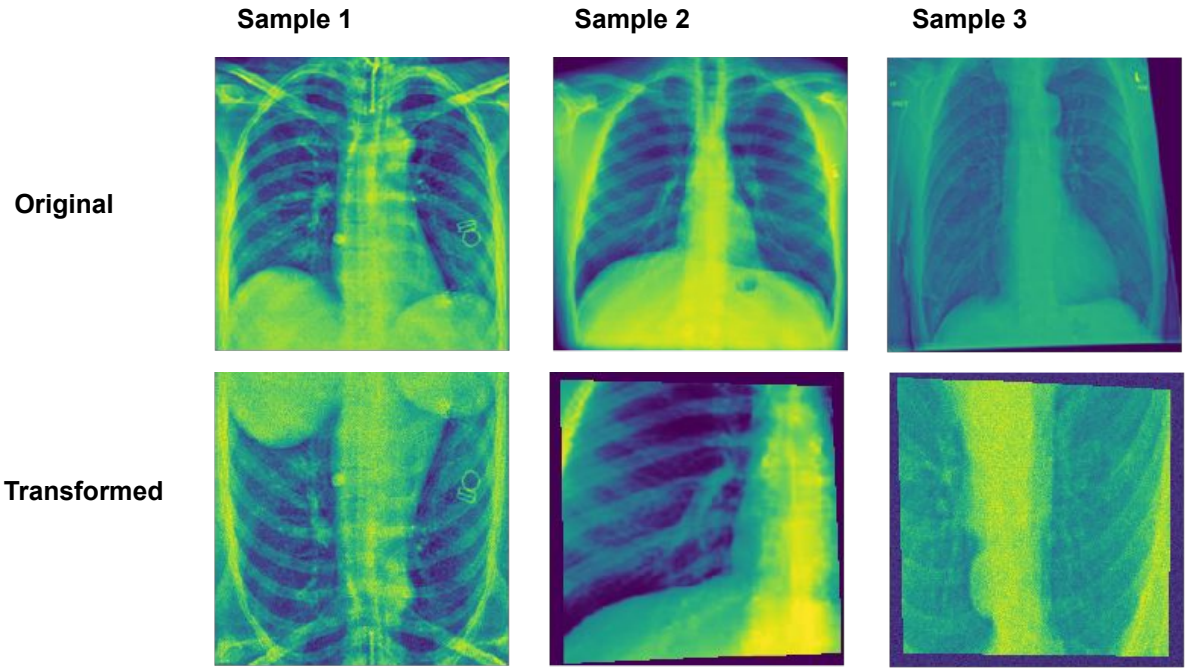
Self Supervised - Transformations 2nd scan

2nd Transformations scan Validation Accuracy	*Linear projector*	*Full network*
Base transformations + RandomVerticalFlip + RandomEqualize	0.6646	0.8952
Base transformations + RandomVerticalFlip + RandomEqualize + RandomPerspective	0.6619	0.8934
Base transformations + RandomVerticalFlip + RandomEqualize + RandomPerspective + GaussianBlur + GaussianNoise	0.6596	0.8920
Base transformations + RandomVerticalFlip + RandomEqualize + RandomPerspective + GaussianBlur + GaussianNoise + RandomRotation	0.6586	0.8879



- Big increase in the linear projector's Accuracy when adding more than two extra transformations.
- RandomPerspective and RandomRotation are redundant.

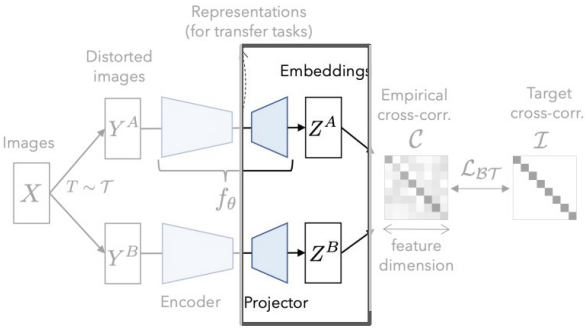
Self Supervised - Final set of transformations



Final set of transformations used in the self-supervised training:

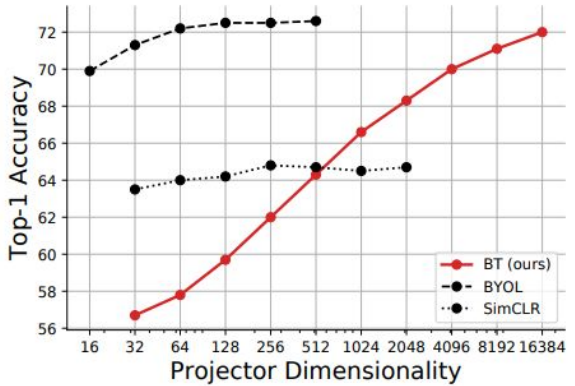
Base transformations +
RandomVerticalFlip +
RandomEqualize +
RandomPerspective +
GaussianBlur +
GaussianNoise

Self Supervised - Projection head



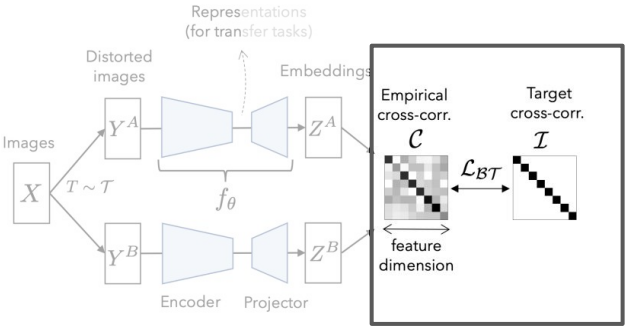
Projector FC dimensionality	*Linear projector*	*Full network*
512	**0.6908**	**0.8975**
1024	0.6848	0.8718
2048	0.6320	0.8263
512-512	0.5888	0.8548
512-1024	0.6550	0.8814
512-2048	0.6481	0.8676

- 20 epochs of self-supervised training + 5 epochs of supervised training, using Covid dataset.
- Optimal projection head: single layer with 512 units. Small dataset, too few epochs.



[1]

Self Supervised - Barlow Twins Lambda



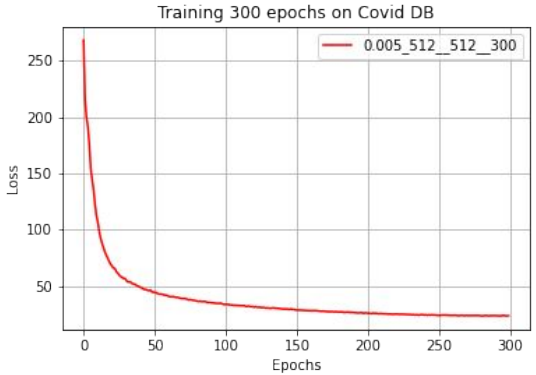
Projector FC dimensionality	*Linear projector*	*Full network*
512	**0.6908**	**0.8975**
1024	0.6848	0.8718
2048	0.6320	0.8263
512-512	0.5888	0.8548
512-1024	0.6550	0.8814
512-2048	0.6481	0.8676

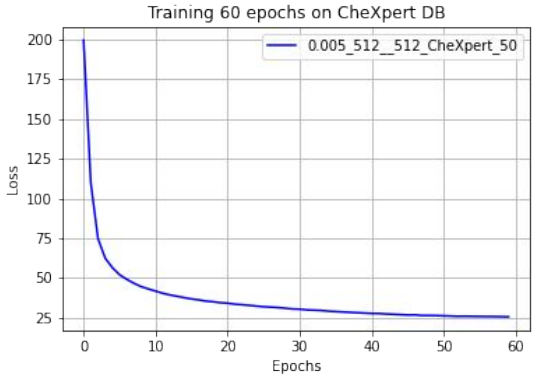
- 20 epochs of self-supervised training + 5 epochs of supervised training, using Covid dataset.
- Optimal projection head: single layer with 512 units. Small dataset, too few epochs.
- Decided to stick to the lambda from the original Barlow Twins paper.

Barlow Twins Lambda λ_{BT}	*Linear projector*	*Full network*
1e-4	0.6311	0.8755
5e-4	0.6398	0.8814
1e-3	0.6476	0.8796
5e-3	0.6550	0.8548
1e-2	**0.6632**	**0.8851**
5e-2	0.5667	0.7835
1e-1	0.4938	0.7546

Self Supervised - Final training

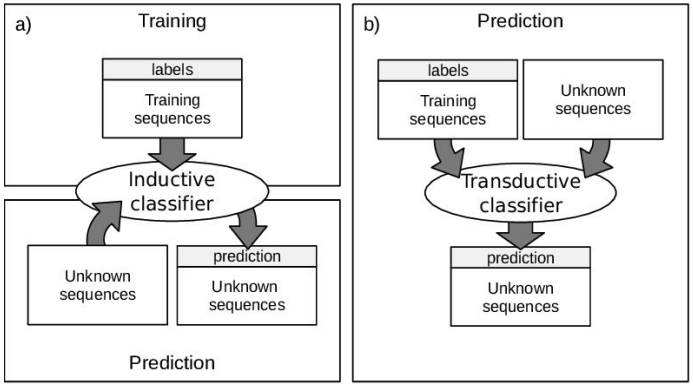
Covid 19 public dataset





With the transformations and hyperparameters found in previous slides we have trained:

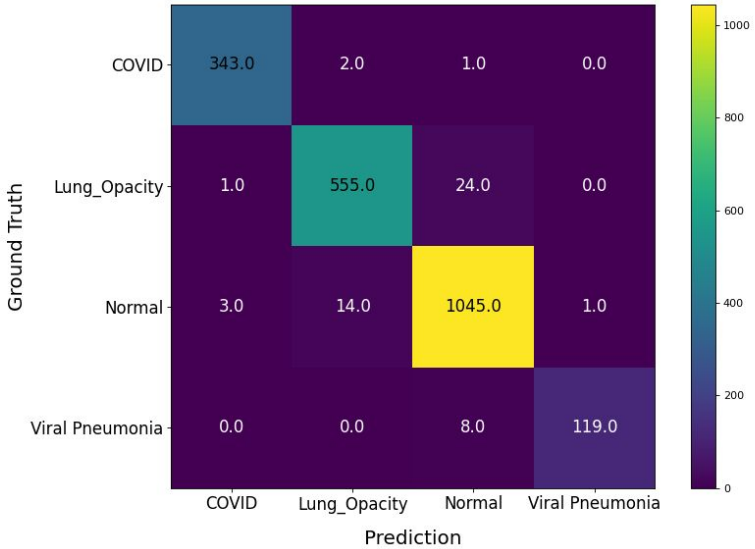
- Barlow Twins model for 300 epochs on the Covid dataset (approx. 20k images, training for 15 hours).
- Barlow Twins model for 60 epochs on the CheXpert dataset (approx. 190k images, training for 32 hours).

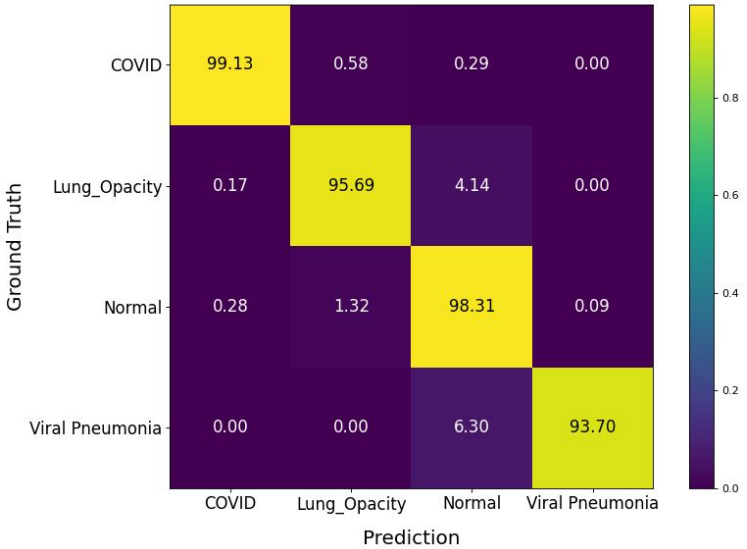


[4]

Downstream task - Confusion matrix

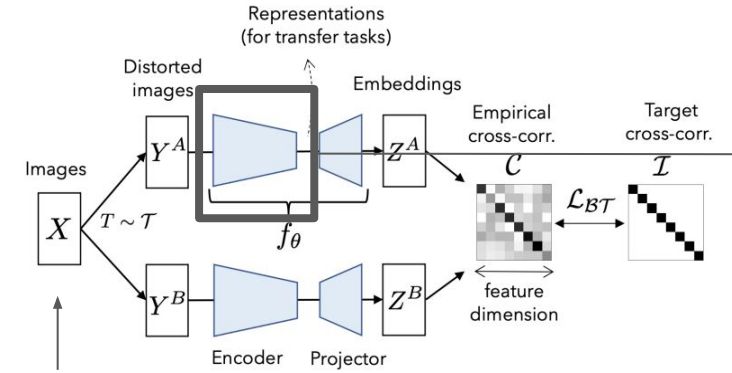
COVID dataset self supervised and supervised training





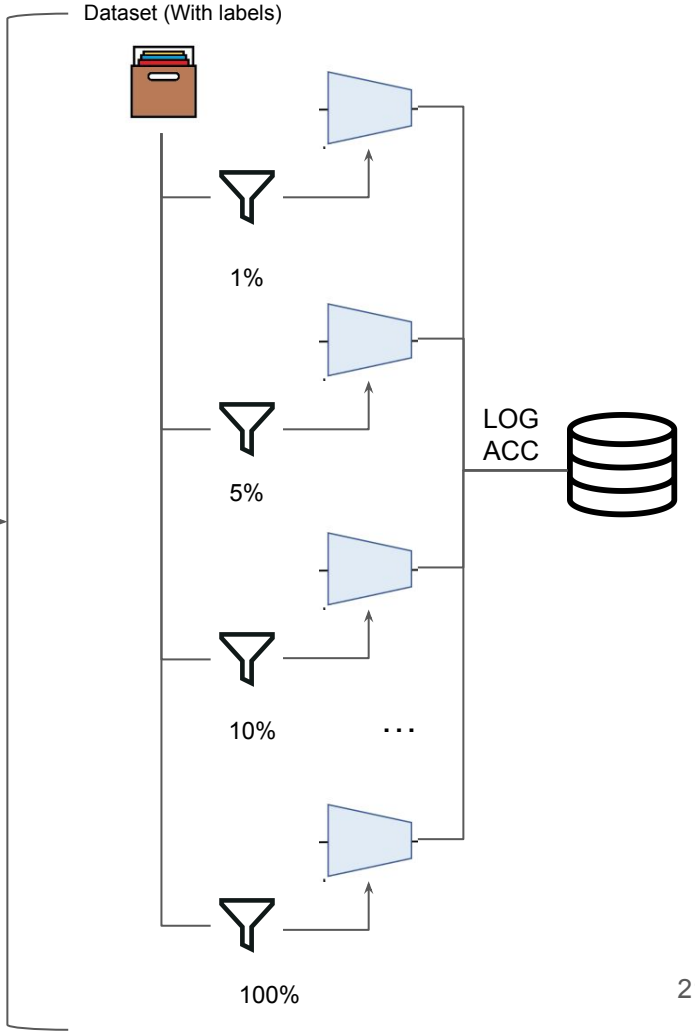
Note: Normalized percentage by the number of samples from each category

Downstream task - Label scan

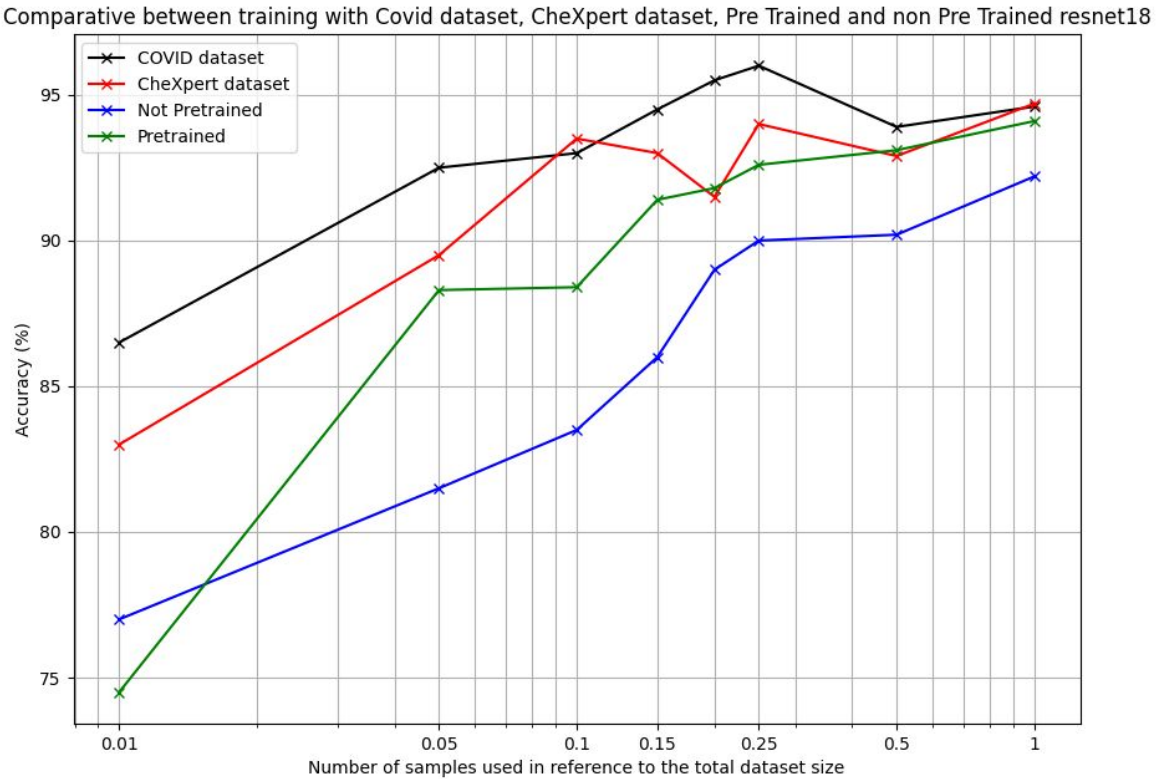


Dataset (Without labels)

Export

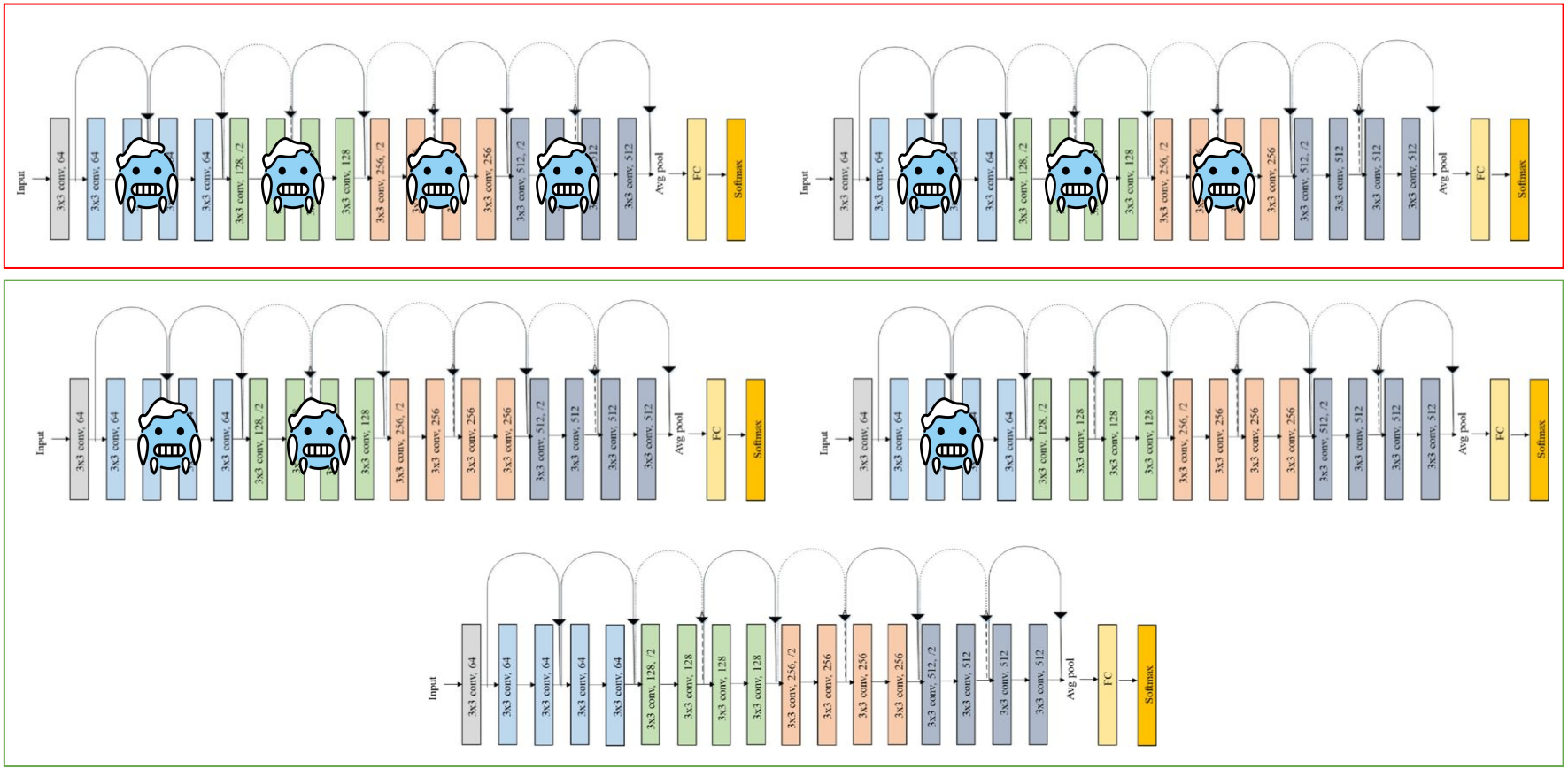


Self supervised results vs Transfer Learning

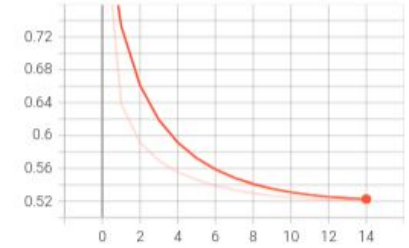


Note: Due to lack of resources, the results for 50% and 100% of the samples were only trained for 15 epochs.

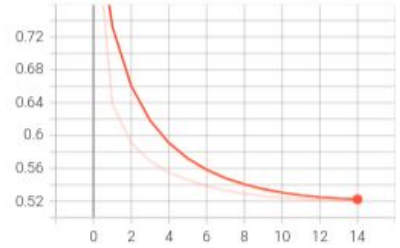
Downstream task - Freezing parts of the model



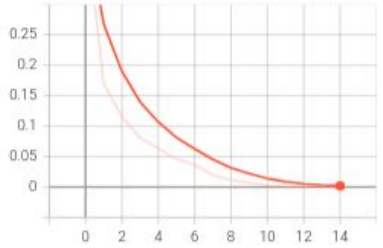
Loss/train:model fr except fc layers



Loss/train:model fr except layer4 + fc layers

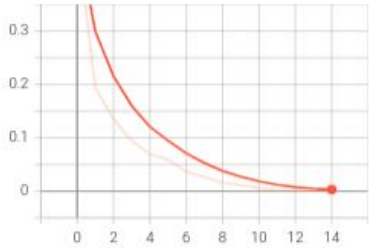


Loss/train:model fr except layers 1,2,3,4 + fc layers

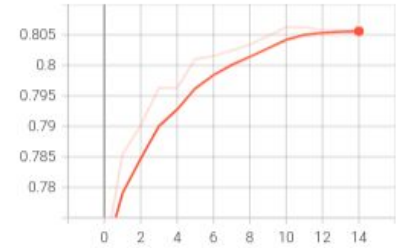


Loss/train:model fr except layers 2,3,4 + fc layers

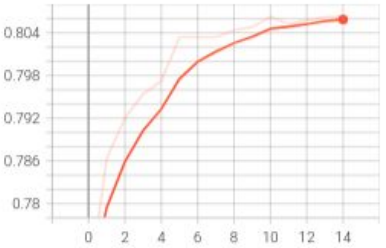
Loss/train:model fr except layers 3,4 + fc layers



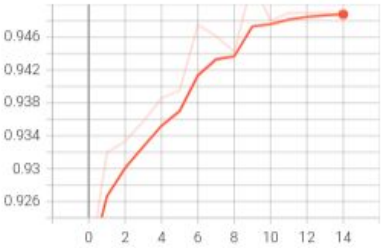
Acc/valid:model fr except fc layers



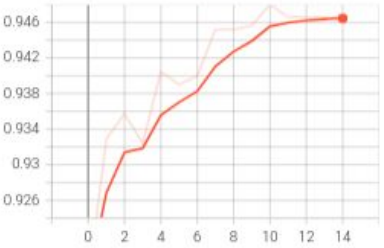
Acc/valid:model fr except layer4 + fc layers



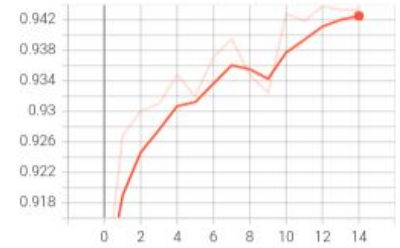
Acc/valid:model fr except layers 1,2,3,4 + fc layers



Acc/valid:model fr except layers 2,3,4 + fc layers



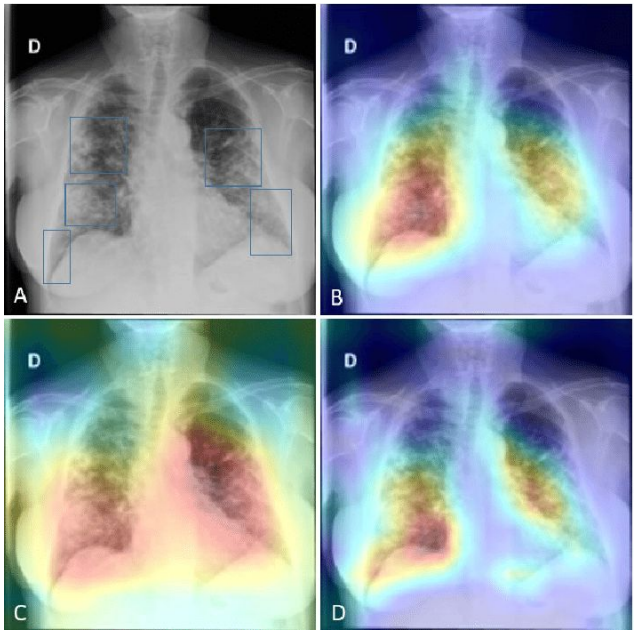
Acc/valid:model fr except layers 3,4 + fc layers



Models with frozen architectures that do not freeze the first three convolutional layer blocks were performing better than the ones that did so.

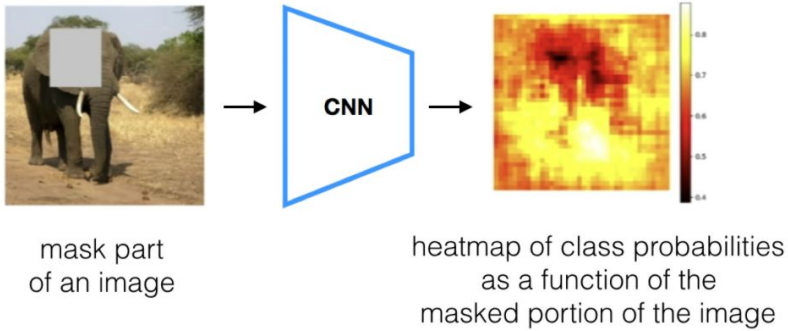
Interpretability

Grad-CAM



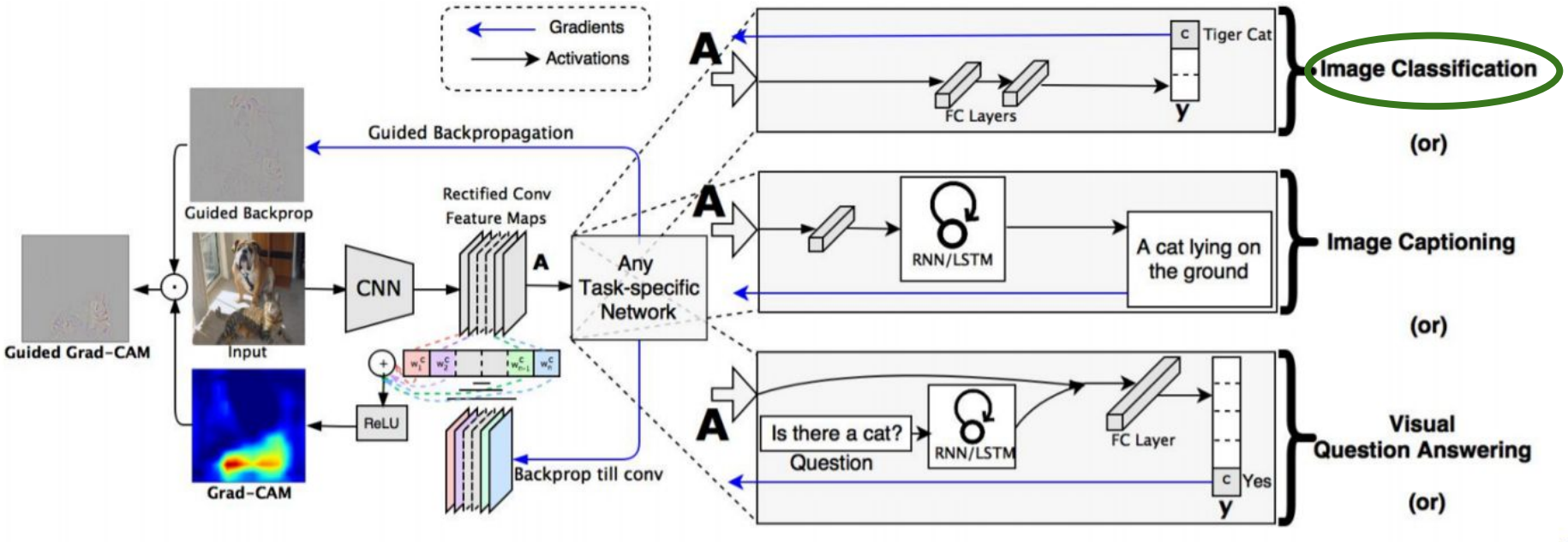
[3]

Occluding image regions



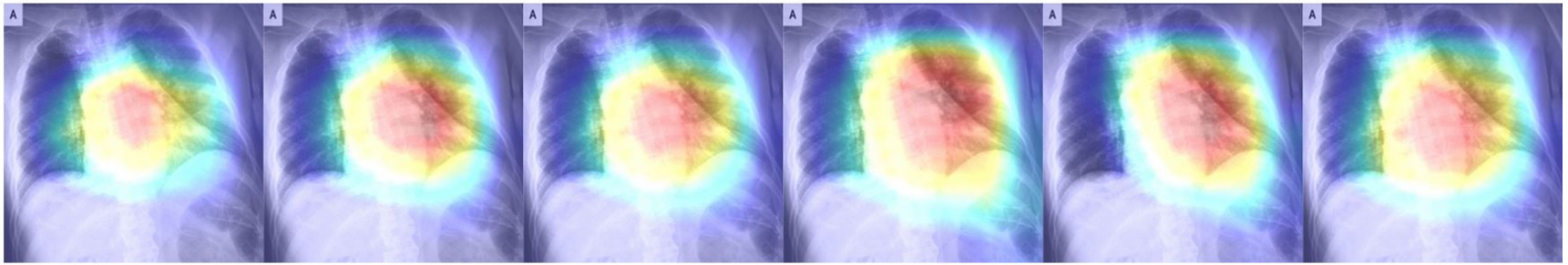
Occlusion experiment with an image of an elephant.

Interpretability - GradCAM

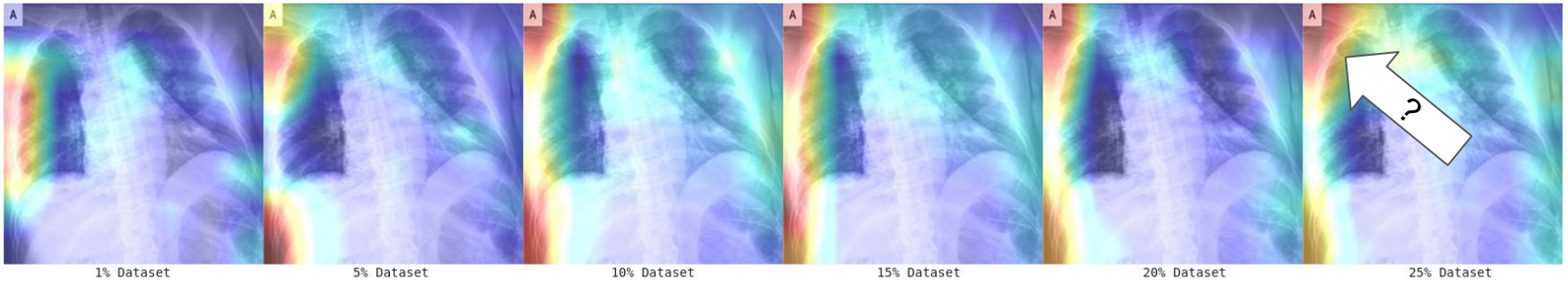


Grad-CAM results

Pretrained Resnet18 model using Barlow Twins

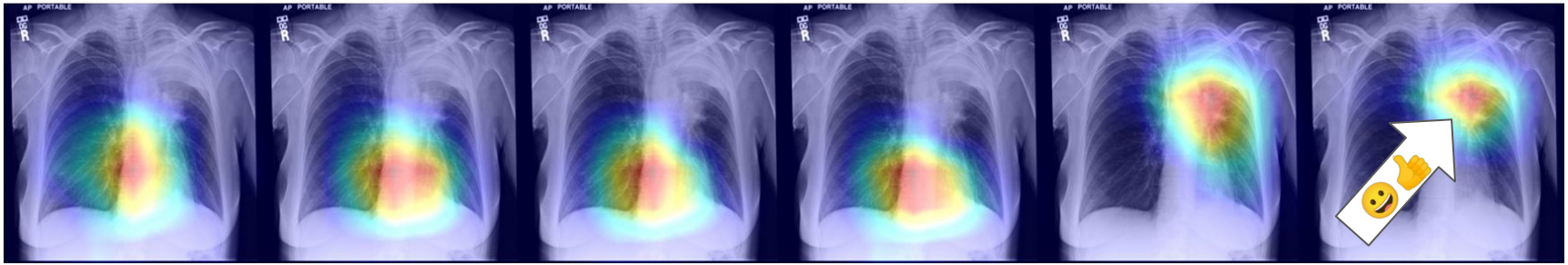


Not pretrained Resnet18 model

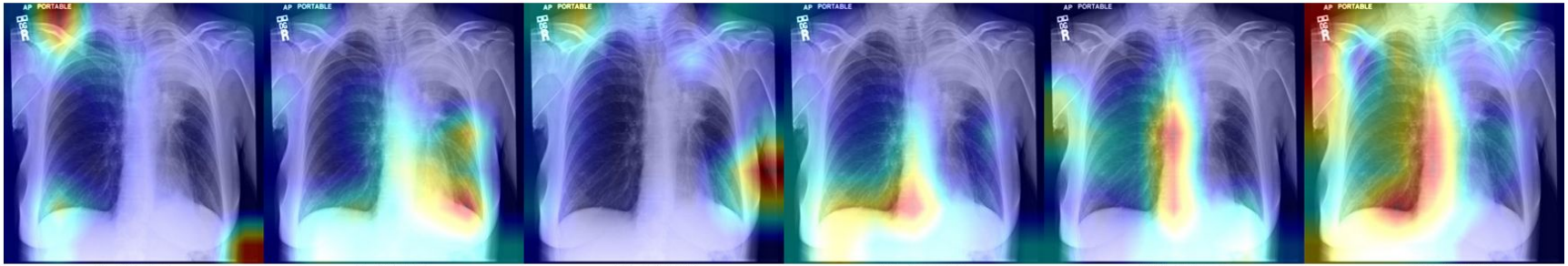


COVID sample

Pretrained Resnet18 model using Barlow Twins



Not pretrained Resnet18 model



Lung Opacity sample

Pretrained model focuses on visual patterns at the lungs while the non-pretrained model does not seem to clearly focus on them

Conclusions

- This architecture is well suited for achieving state of the art results without major resources.
- When adding more than one extra image transformations, the accuracy of the linear projector task improves substantially. Contrary to the conclusions of the barlow twins paper, we have found that the best architecture for the projector head is a single 512 neuron linear layer.
- The self supervised pre-trained models do not improve drastically the accuracy of the classification problem when we use all the labels available on the dataset (~16k). However, when the number of samples used to train is scarce (~1.6k-2k) the pre-trained models outperform the not pre-trained model and the ImageNet pre-trained models.
- The accuracy for the pre-trained models (one with Covid dataset and the other one with the CheXpert dataset) does not change significantly.
- Freezing different layers of the model architecture we observed that the performance of the model is affected remarkably.
- Observing the results of Grad-CAM, in the pre-trained models we can observe the significant patterns in the area of the lungs even at lower percentage of samples.

Thank you

References

[1] Zbontar, Jure, et al. "Barlow twins: Self-supervised learning via redundancy reduction." *International Conference on Machine Learning*. PMLR, 2021

[2] COVID-19 Radiography Database https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

[3] Rajaraman, Sivaramakrishnan & Siegelman, Jen & Alderson, Philip & Folio, Lucas & Folio, Les & Antani, Sameer. (2020). Iteratively Pruned Deep Learning Ensembles for COVID-19 Detection in Chest X-rays. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3003810.

[4] https://www.researchgate.net/figure/Inductive-and-transductive-learning-a-Traditional-inductive-scheme-with-two-separate_fig4_32041423

[5] https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html

[6] Sriram, Anuroop, et al. "Covid-19 prognosis via self-supervised representation learning and multi-image prediction." arXiv preprint arXiv:2101.04909 (2021).

[7] Chen, Xiaocong, et al. "Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images." Pattern recognition 113 (2021): 107826.

[8] Azizi, Shekoofeh, et al. "Big self-supervised models advance medical image classification." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[9] Irvin, Jeremy and Rajpurkar, Pranav and Ko et al. (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. Retrieved from https://arxiv.org/abs/1901.07031

Appendix - Supervised hyperparameter tuning

Once trained the self-supervised models, for the downstream tasks we performed a hyperparameter scan by running 55 trials with different configurations of parameters. The Covid dataset was used for this task.

- Learning rate: **0.00153**
- Batch Size: **96**
- Number of epochs: **15**
- Weight decay: **9.870e-6**
- Augmentations: **Horizontal flip with p=0.5**

Appendix - Supervised hyperparameter tuning

Once trained the self-supervised models, for the downstream tasks we performed a hyperparameter scan by running 55 trials with different configurations of parameters. The Covid dataset was used for this task.

- Learning rate: **0.00153**
- Batch Size: **96**
- Number of epochs: **15**
- Weight decay: **9.870e-6**
- Augmentations: **Horizontal flip with p=0.5**

Appendix - State of the art results for COVID and CheXpert datasets

Study	Precision
Automatic detection of coronavirus disease [...]	96%-99%
CovXNet	89.6% - 97.4%
Lightweight Neural Network for COVID-19 Detection	95.8%
Covid-19: automatic detection[...] transfer learning	95%
Multi-Channel Transfer Learning [...]	94%
COVID-Net	93.3%
Using X-ray images and deep learning ...	92.18%
CoroNet	89.6%
Automated detection of COVID-19 [...]	87.0%

Table 6: State of the art using COVID-19 dataset

Study	Precision
DeepAUC-v1 ensemble	93%
Hierarchical-Learning-V1 (ensemble)	93%
Conditional-Training-LSR ensemble	92.9%
Hierarchical-Learning-V4 (ensemble)	92.9%
YWW(ensemble)	92.9%
Conditional-Training-LSR-V1 ensemble	92.9%
Hierarchical-Learning-V0 (ensemble)	92.9%
Multi-Stage-Learning-CNN-V3 (ensemble)	92.8%

Table 7: State of the art using CheXpert dataset