SELF-SUPERVISED DEEP LEARNING APPROACHES FOR MEDICAL IMAGE ANALYSIS

Cristian Simionescu

Supervisor: Dr. Adrian Iftene

September 22, 2022

Faculty of Computer Science, Iasi

SIMIONESCU

SELF-SUPERVISED LEARNING

MOTIVATION

- From 2010 to 2020, the amount of data created, captured, copied, and consumed in the world increased from 1.2 trillion gigabytes to 59 trillion gigabytes
- By extrapolation from 2021 to 2024, more data will be created than the past 30 years
- Human labeling can't keep up with the exponential growth curve of data collection

MOTIVATION

- Percentage of labeled data will rapidly decline compared to unlabeled data
- Hardware is continuing to provide increased parallel and single threaded performance
- Deep Learning model which can scale with more data and compute are becoming state of the art in all fields (e.g Transformers)

As a result, Self-Supervised Learning will probably be the most important driver of future AI SOTA improvements

SELF-SUPERVISED LEARNING - CATEGORIES

- Predictive Methods
- Generative Methods
- Contrastive Methods
- Clustering Methods
- Distillation Methods
- Information Maximization Methods

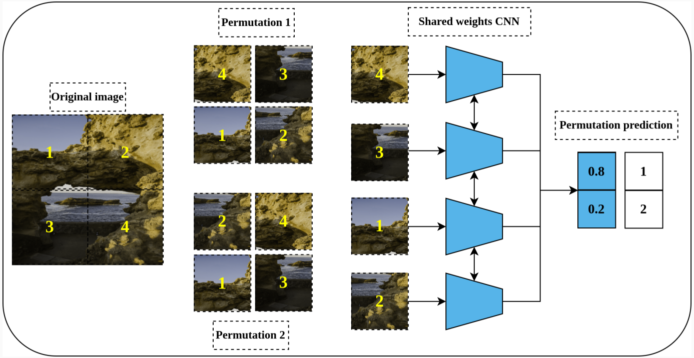


Figure: Example of Jigsaw pretext task

SELF-SUPERVISED LEARNING - GENERATIVE METHODS

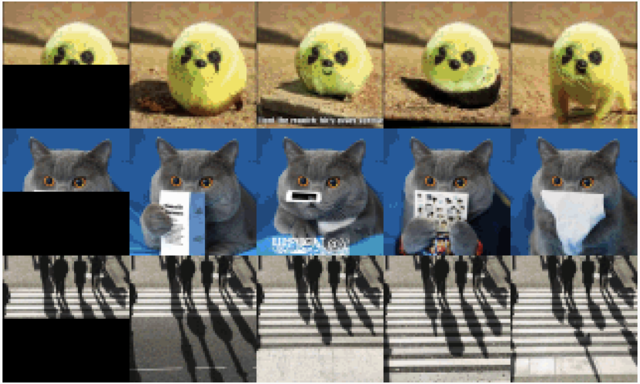


Figure: Example of Autoregressive Generation, Image GPT

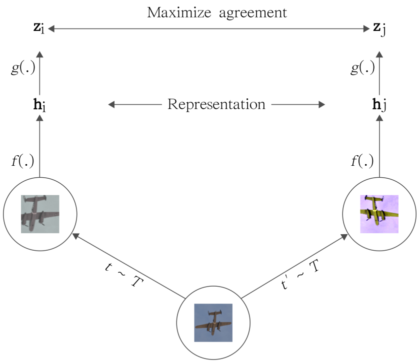


Figure: SimCLR Architecture

Same idea as contrastive models, but forms positive/negative contrastive pairs at a cluster level, not individual sample level

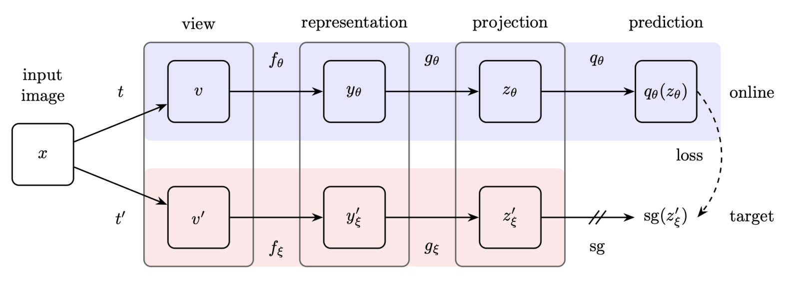


Figure: BYOL Architecture

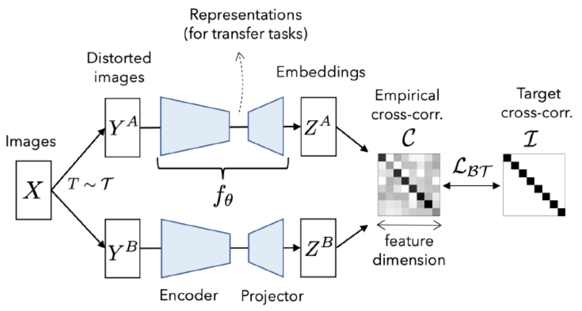


Figure: Barlow's Twins Architecture - Note: its so simple, it is surprising it took so long to invent

SELF-SUPERVISED LEARNING FOR MEDICAL DATA

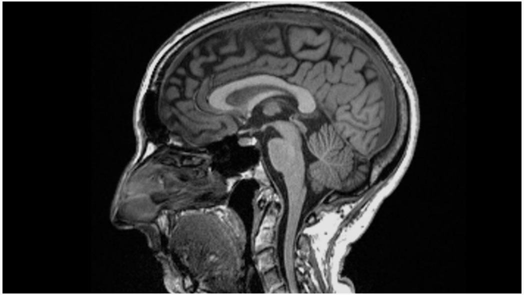
SIMIONESCU

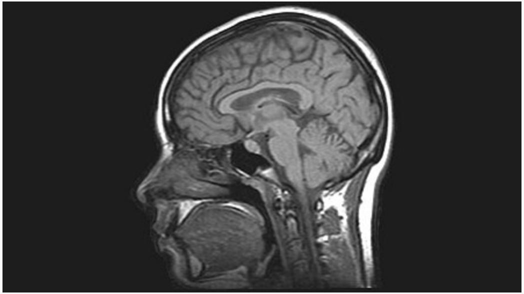
MEDICAL IMAGE ANALYSIS

- Quality labeling can only be done by highly specialized medical professionals, slower, smaller capacity
- Non-standardization of medical practices, terminology, technology
- Individual medical datasets are small
- Research delay from natural images methods to medical imaging methods
- Smaller researcher base
- Improvements on medical tasks can have a bigger societal impact compared to natural image tasks

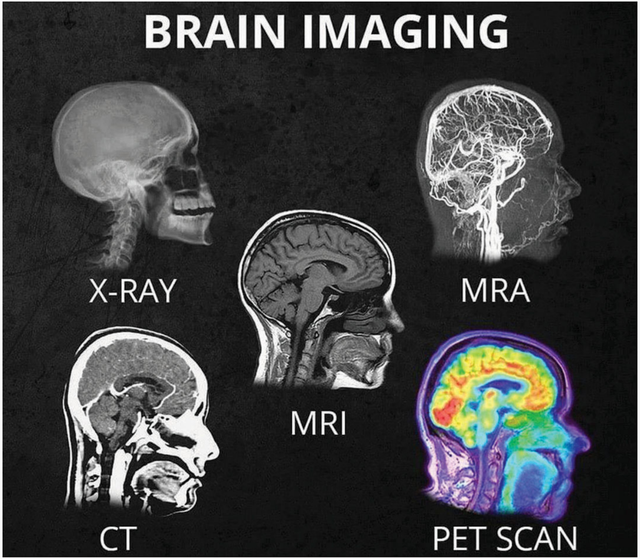
Medical images share a lot of information between them.

Scans with the same technology (CT, MRI) on the same organ, will share a lot of features, even if the pathology of the two patients is completely unrelated.

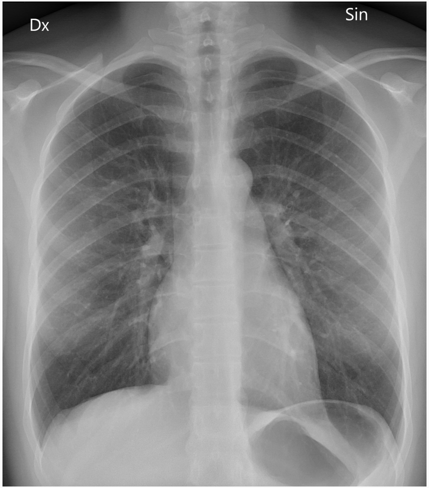




Scans with a different technology, of the same organ, will share many structural features as well as textural information.



Scans of different organs with the same technology will also share features since the scans work by capturing a small set of frequencies which translate to types of tissues.



OPPORTUNITY

Medical tasks tend to be more difficult in isolation due to the them requiring both global and very small detail, local feature modeling to work (Why U-Net is so good)

Medical image models will hit overfitting problem sooner than natural image equivalents, due to fundamental reduced variance in underlying data

Developing self-supervised pretraining methods which can leverage multi-modality, multi-organ, multi-task data, has the potential of using much larger models and unlocking big SOTA improvements

COMPLETED WORK

DYNAMIC BATCH ADAPTATION – SELECTION

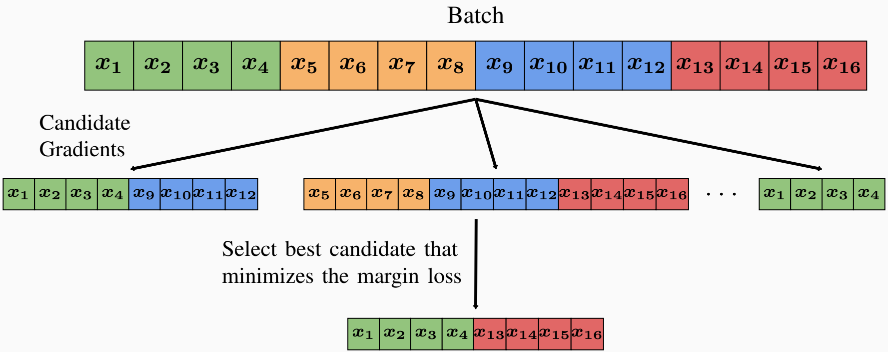


Figure: Gradient Subset Selection

Procedure `DBA::step(`$loss$`):`

$modelMetric \leftarrow ModelMetric(loss)$

for $layer \in model.layers$ **do**

$selectionStrides \leftarrow SplitStrides(layer.gradSamples, strideSize)$

$chosen \leftarrow SelectionStrategy(selectionStrides)$

$layer.gradient \leftarrow Mean(chosen)$

end

optimizer.step($loss$)

Algorithm 1: Optimizer step

Figure: Gradient Subset Selection

DYNAMIC BATCH ADAPTATION - BATCH CHANGING

$$\text{Next} = \text{Current} + \begin{cases} \text{delta}, q > 0.8 \cdot \text{Current} \\ -\text{delta}, q < 0.2 \cdot \text{Current} \\ 0, \text{otherwise} \end{cases} \tag{1}$$

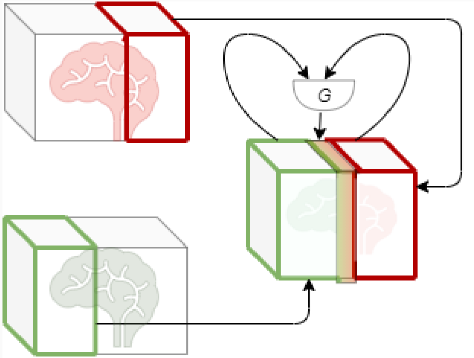
Where q is the 50th percentile batch size selected in the current epoch.

$dba \leftarrow DBA(model, optimizer, strideSize, minBatchSize, maxBatchSize)$
while $epochs < maxEpochs$ **do**
  $batches \leftarrow SplitInBatches(data, batchSize)$
  for $batch \in batches$ **do**
    $loss = model.fit(batch)$
    $dba.step(loss)$
  end
  $batchSize \leftarrow dba.nextBatchSize()$
end

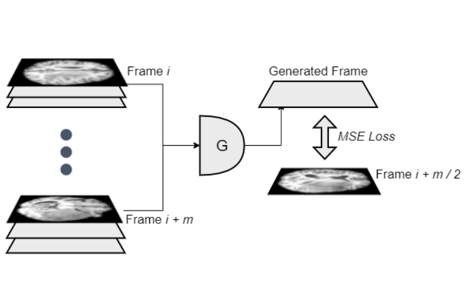
Algorithm 4: Training process

Table: Results on MNIST using x% of the data

		Test Accuracy %		
Optimizer	Variant	100% of Data	10% of Data	1% of Data
SGD	batch size=64	**98.164 $\pm$ 0.077**	95.148 $\pm$ 0.128	87.628 $\pm$ 0.339
SGD	batch size=128	98.148 $\pm$ 0.087	94.990 $\pm$ 0.076	87.574 $\pm$ 0.401
SGD	batch size=256	98.068 $\pm$ 0.036	94.834 $\pm$ 0.137	87.380 $\pm$ 0.386
Adam	batch size=64	97.786 $\pm$ 0.037	95.262 $\pm$ 0.108	88.350 $\pm$ 0.330
Adam	batch size=128	97.770 $\pm$ 0.068	95.480 $\pm$ 0.093	88.250 $\pm$ 0.452
Adam	batch size=256	97.872 $\pm$ 0.085	95.466 $\pm$ 0.180	88.314 $\pm$ 0.545
DBA+SGD **(ours)**	gradient norm	98.122 $\pm$ 0.074	95.546 $\pm$ 0.073	96.772 $\pm$ 0.420
DBA+SGD **(ours)**	variance norm	98.140 $\pm$ 0.032	**97.422 $\pm$ 0.198**	**97.830 $\pm$ 0.079**
DBA+Adam **(ours)**	gradient norm	95.038 $\pm$ 0.118	93.730 $\pm$ 0.233	90.846 $\pm$ 1.749
DBA+Adam **(ours)**	variance norm	94.952 $\pm$ 0.181	93.864 $\pm$ 0.214	92.198 $\pm$ 0.550

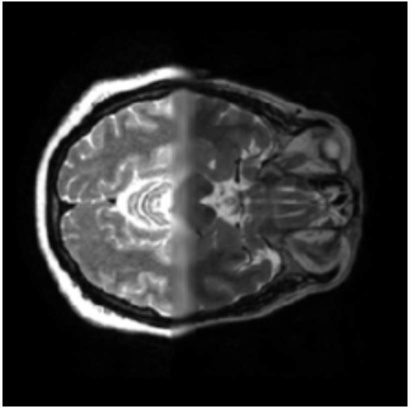


(a) Brain Mixing Idea

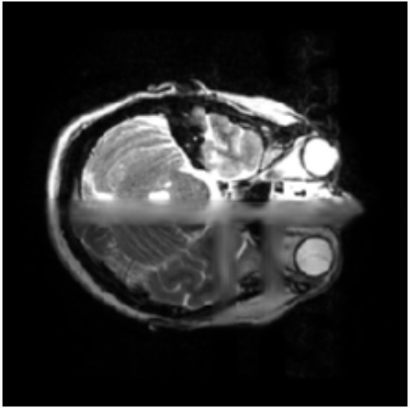


(b) Generative Model Training Idea

Figure: Overview.



(a) Generated brain by fusing two patients.



(b) Generated brain by fusing eight patients.

Method	Top1 Acc	Top3 Acc	Top5 Acc
NT (Normal Training)	$32.1 \pm 0.27\%$	$53.8 \pm 0.18\%$	$64.21 \pm 0.32\%$
NT + BT (Barlow/Twins)	$40.1 \pm 0.77\%$	$62.8 \pm 0.11\%$	$73.01 \pm 0.1\%$
NT + BT + BrainFuse	$50.11 \pm 0.3\%$	$71.16 \pm 0.12\%$	$82.9 \pm 0.22\%$
NT + BT + RawBrainFuse	$45.1 \pm 0.94\%$	$67.5 \pm 0.41\%$	$80.01 \pm 0.1\%$
NT + BT + BrainFuse + Consistency Regularisation	**51.4** $\pm 0.84\%$	**72.1** $\pm 0.61\%$	**83.8** $\pm 0.52\%$
NT + BT + RawBrainFuse + Consistency Regularisation	$46.6 \pm 0.9\%$	$68.8 \pm 0.78\%$	$81.1 \pm 0.31\%$

Table: Results for the different training procedures. Each procedure has been run 3 times

MISCELLANEOUS

Separately from medical image analysis research and connected issues, together with Software Engineering student teams we have worked on different AI problems, resulting in 4 publications.

- Spotlight Charging - National semi-final at InnovationLabs
- ElectroWay - INISTA 2022 Paper
- Social Media Post Impact Prediction - INISTA 2022 Paper
- Music Generation - INISTA 2022 Paper
- Renewable Energy Investment Calculator - INISTA 2022 Paper

ONGOING WORK

SIMIONESCU

EFFICIENT DYNAMIC BATCH ADAPTATION - CHANGES

- Only does per-sample gradient selection at the last layer
- Simpler random candidate sampling selection algorithm
- Batch changing decision based on underlying metric variation, not the secondary effect of candidate batch size
- New margin loss minimizing function
- Can scale to larger models and datasets

EFFICIENT DYNAMIC BATCH ADAPTATION - RESULTS

Table: Results using x% of the data

		Test Accuracy %			
Dataset	Method	100% of Data	10% of Data	1% of Data	100 Samples
CIFAR10	Baseline	92.014 ± 0.115	71.654 ± 0.087	40.516 ± 1.358	23.796 ± 1.092
CIFAR10	EDBA	**92.196 ± 0.219**	**91.648 ± 0.171**	**91.514 ± 0.224**	**90.688 ± 0.168**
CIFAR100	Baseline	67.308 ± 0.197	29.628 ± 0.862	8.404 ± 0.403	—
CIFAR100	EDBA	**70.550 ± 0.097**	**69.138 ± 0.264**	**68.852 ± 0.341**	—

Our attempt at a multi-modal, multi-organ, multi-task self-supervised learning pretraining framework for medical image analysis

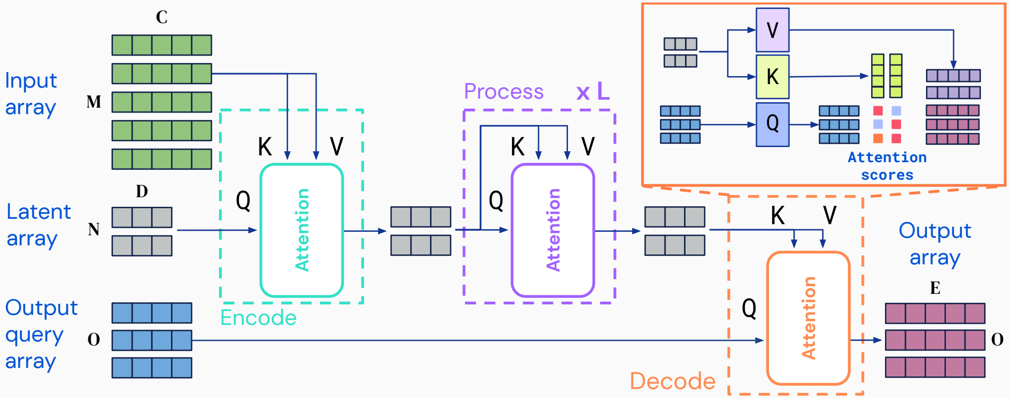


Figure: Perceiver IO Architecture

Collections of latent embeddings to store information for the following categories:

- Organ (Brain, Chest, Knee, Abdomen)
- Modality (MRI, CT, PET, Xray)
- Dimensionality (2D, 3D)
- Task (Nodule segmentation, Pathology classification, SSL tasks)

Table: MedMNIST dataset distribution

Data Modality	Tasks	# Samples
Colon Pathology (2D)	MC	107, 180
Chest X-Ray (2D)	ML, BC	117, 976
Dermatoscope (2D)	MC	10, 015
Retinal OCT (2D)	MC	109, 309
Fundus Camera (2D)	OR	1, 600
Breast Ultrasound (2D)	BC	780
Blood Cell Microscope (2D)	MC	17, 092
Kidney Cortex Microscope (2D)	MC	236, 386
Abdominal CT (2D)	MC	107, 731
Abdominal CT (3D)	MC	1, 743
...	...	...

There is a general understanding of the existence of ICS, but there is disagreement in the literature about the exact effects if ICS, and how some methods (BatchNorm) interact with it

We want to analyze the effect of Internal Covariate Shift on deep neural network training

Backforward is a method of completly eliminating ICS from training, at a very high computational cost, making it usable only for analysis, not an method to be actively used turing training

BACKFORWARD PROPAGATION

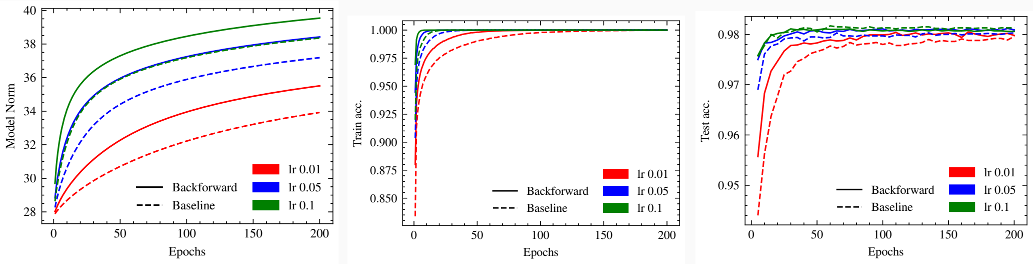


Figure: Backforawrd results using different learning rates

Current ideas and questions to find out:

- It seems that ICS has a regularization effect on weight norms, see if it scales with deeper models
- no-ICS runs tend to overfit a bit quicker, check if the problem persists
- BatchNorm has long been assumed to work **because** it reduces ICS, run a model with BatchNorm and Backforward and see if BatchNorm's effect still exists or not
- Since ICS seems to have a regularization effect, test if Backforward can be used as an effective for a short warmup with high lr, or as a fine-tune algorithm with moderate lr and a big batch size

FUTURE WORK

SIMIONESCU

FUTURE WORK

- Foundational Medformer - After validating Medformer, train a foundational model on a very large collection of medical datasets (more hardware might be needed)
- BrainFuse 2.0 - Take a larger dataset (ADNI) and train the self-supervised algorithm to convergence, try a 2.5D or 3D model
- Meical Image Translation - An idea which has seen great interest from InnovationLabs judges and mentors, using Medformer or similar, train a model to approximate a scan result from another. For example, approximate brain PET's from brain MRI's, can be very important, since PET's are expensive, very few and have long wait times

FUTURE WORK

- Medical Image Super Resolution & Artefact Correction - Use Medformer or similar to experiment with self-supervised tasks with real-world applicability, such as noise/artefact removal, or resolution enchacement for medical images

TENSOR RELOADED

We formed an informal, unofficial research group focused on machine learning, **Tensor Reloaded**

Expand member pool with more grad students or interested faculty members

Prepare an undergrad training program, in order to bring them up to a sufficient level to be able to be incorporated in teams that participate in competitions, workshops or research projects

Setup a Machine Learning paper presentation weekly event for interested parties

/cristian.simionescu.92

@CrisSimi2525

simi2525

cristian.simionescu.me

/cristian-simionescu/

Cristian Simionescu