Brain-Aware Replacements for Supervised Contrastive Learning in Detection of Alzheimer's Disease

Mehmet Saygin Seyfioglu

Advisors: Prof. Linda Shapiro, Prof. Sheng Wang, Prof. Thomas Grabowski

05/18/2022

Outline

- Motivation
- Self-supervised Learning
  - Self-supervised contrastive learning
    - Self-supervised contrastive learning
    - Issues with Self-supervised contrastive learning in Alzheimer’s Disease prediction
- Supervised Contrastive Learning
  - Issues with Supervised Contrastive Learning
- Supervised Contrastive Learning with Synthetic Samples
  - CutMix
  - Brain Aware Region Replacements (BAR)
- Results
- Discussions
- Future Directions

Motivation

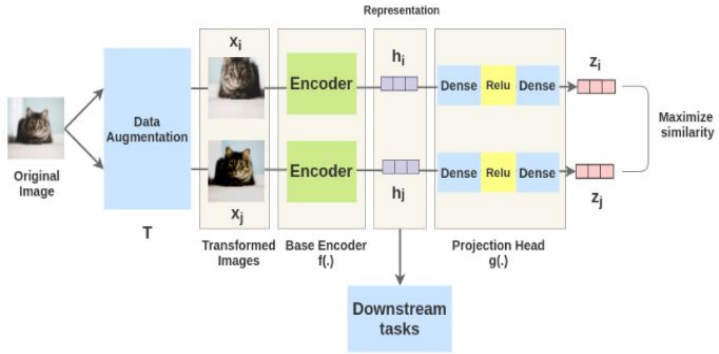
- We want to detect Alzheimer’s Disease from structural MRIs.
  - However, having low training sample support limits the complexity of the models that we can train.
    - Pre-training is the key!

Self-supervised Learning

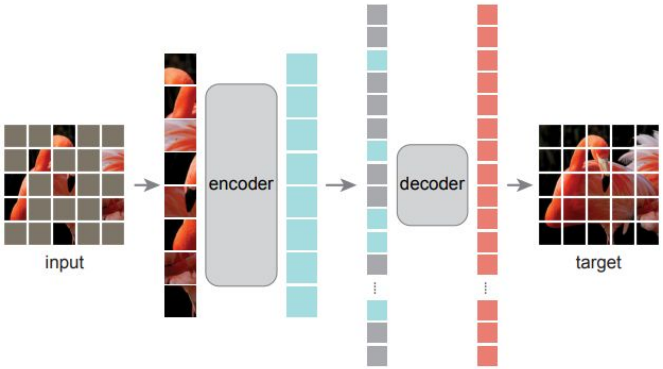
- The idea is to create tasks without using the human annotations (labels) and pre-train the model on those tasks.
  - A simple example task could include predicting the original version of the image by providing its grayscale one to the model.
- By pre-training, the hope is that the model will learn useful representations. Then, we can fine-tune the model to downstream tasks such as classification, segmentation etc.

Self-supervised Learning

1) Self-Supervised Contrastive Training [2]:

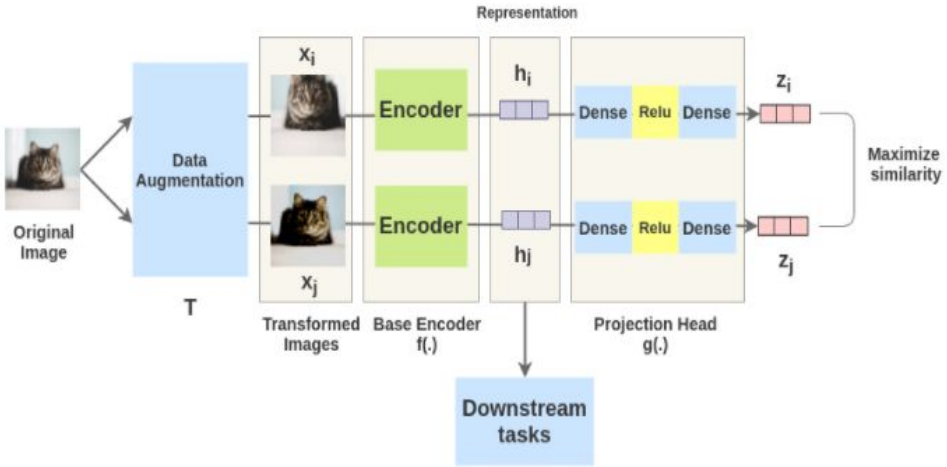


2) Patch Reconstruction [3]:



Self-supervised Contrastive Training

The idea is to create two differently augmented copies (positives) of the anchor image, while considering the rest of the samples within the batch as negatives. Augmentations are a set of parametric transformations, such as random crops, rotations, etc. that aim to preserve semantics of the data while altering them.

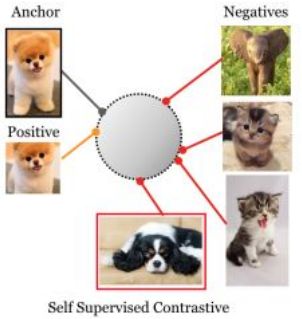


These positives are then mapped closer in the latent space, while the negatives become further away. This approach is shown to be very effective in natural images [5].

Self-supervised Contrastive Loss

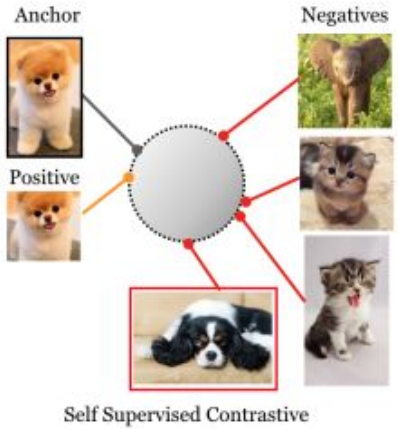
$$L_{NCE} = -\sum_{i=1}^{n} log \frac{e^{\theta(t_1^i, t_2^i)}}{\frac{1}{b}\sum_{j=1}^{b} e^{\theta(t_1^i, t_2^j)}}$$

- Here, t_1^i and t_2^i are the two augmented copies of the anchor image X_i, $t_2^j|_{j \neq i}$ is a negative sample, n is the number of samples, θ denotes an encoder, and b is the number of samples within the batch.



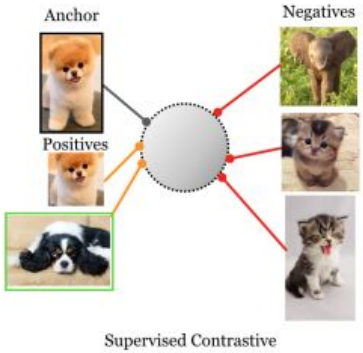
Issues with Self-supervised Contrastive Loss with AD Prediction

- Every sample for $t_2^j|_{j \neq i}$ is considered equally different from the anchor X_i. This generally holds pretty well for tasks with a large number of classes. (ImageNet has 1000 classes)
- However, for classification problems with a small number of classes such as ours, this approach has its flaws since it is highly probable that $t_2^j|_{j \neq i}$ contains false negatives, i.e., the samples that are from the same class as the anchor. Or even worse, it could contain samples coming from the same subject.



Supervised Contrastive Loss

- One way to fix negative sampling issue is to use supervised-contrastive learning [4] during pre-training, which leverages hard labels to embed features.



- However, this approach has its limitations as using hard labels during pre-training exhausts the entropic capacity of labels, thus leading to sub-optimal fine-tuning performance.

Summing up

- Self-supervised Contrastive Learning is problematic because of faulty negative-sampling.
- Supervised Contrastive Learning exhausts all label information
- What to do?

Mixture Prediction with Synthetic Samples

- We can reformulate the contrastive objective as a mixture detection problem where we create synthetic samples and soft-labels by mixing two MRIs and semantically group similarly mixed samples.
- To that end, we need two main components
  - A way to generate mixtures (synthetic samples)
  - A soft-label capable contrastive loss

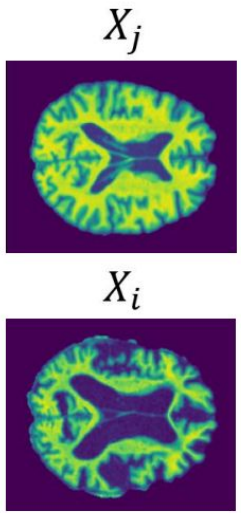
CutMix Strategy

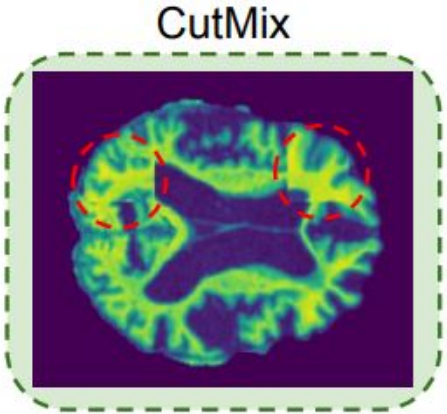
- CutMix [5] is a technique known to be very effective in creating soft labels by non-linearly combining images to create synthetic images and labels.
- Given a set of 3D brain MRIs $X_i|_{i=1}^n$ and their binary annotations $y_i|_{i=1}^n$ it is possible to generate synthetic images $X_i^p|_{i=1}^n$ and soft labels $y_i^p|_{i=1}^n$ by transferring a 3D region from X_j into X_i and modifying the label y_i to be a linear combination of y_i and y_j.

$$X_i^p = (1 - M) \odot X_i + M \odot X_j$$

$$y_i^p = \lambda y_i + (1 - \lambda) * y_j$$

- However, CutMix creates “non-realistic looking” samples. We can do better.
  - How?



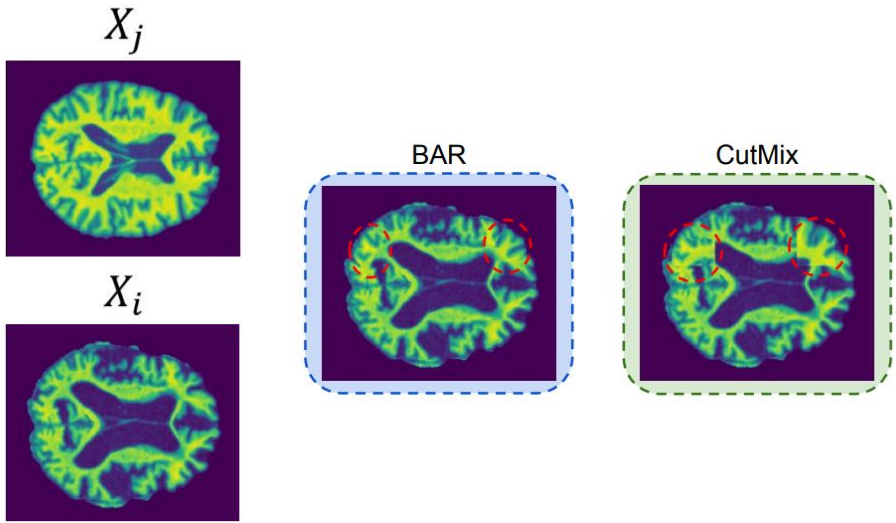


Brain-Aware Replacements (BAR)

- We propose an augmentation technique for brain MRIs that we call BrainAware Replacements (BAR), which utilizes anatomically relevant regions from the Automated Anatomical Labeling Atlas (AAL) for non-linear replacements from a randomly picked MRI into an anchor MRI
- AAL has 62 distinct brain regions when the left and right lobes are merged.

CutMix vs BAR

- For BAR, Superior frontal gyrus, medial orbital and Superior frontal gyrus, dorsolateral are selected.
- Notice how BAR produces more realistic looking synthetic MRIs as random patches often are too bulky and cutting/replacing regions from lateral ventricle.
    - Compared to CutMix, BAR leads to less local distribution shift and harder to solve examples



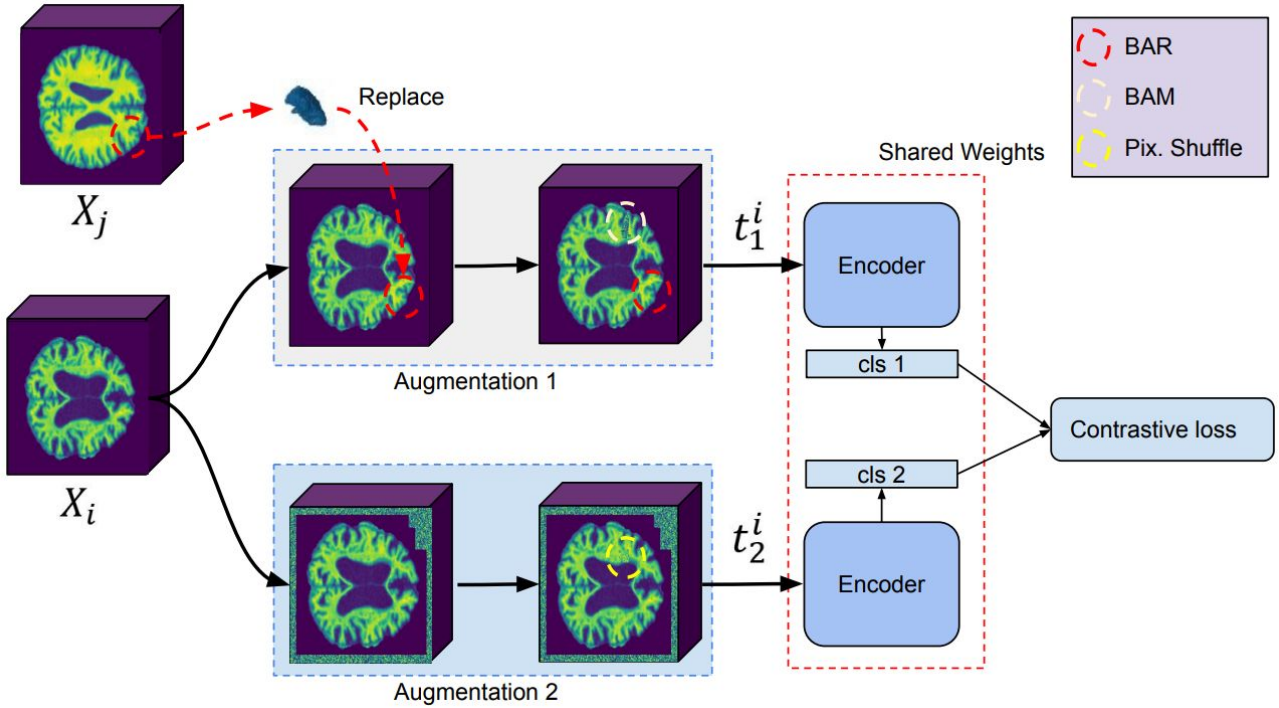
Soft-Label Capable Contrastive Loss

- With a slight modification on the supervised contrastive loss, soft-labels can be exploited to learn the relative similarity between pairs.

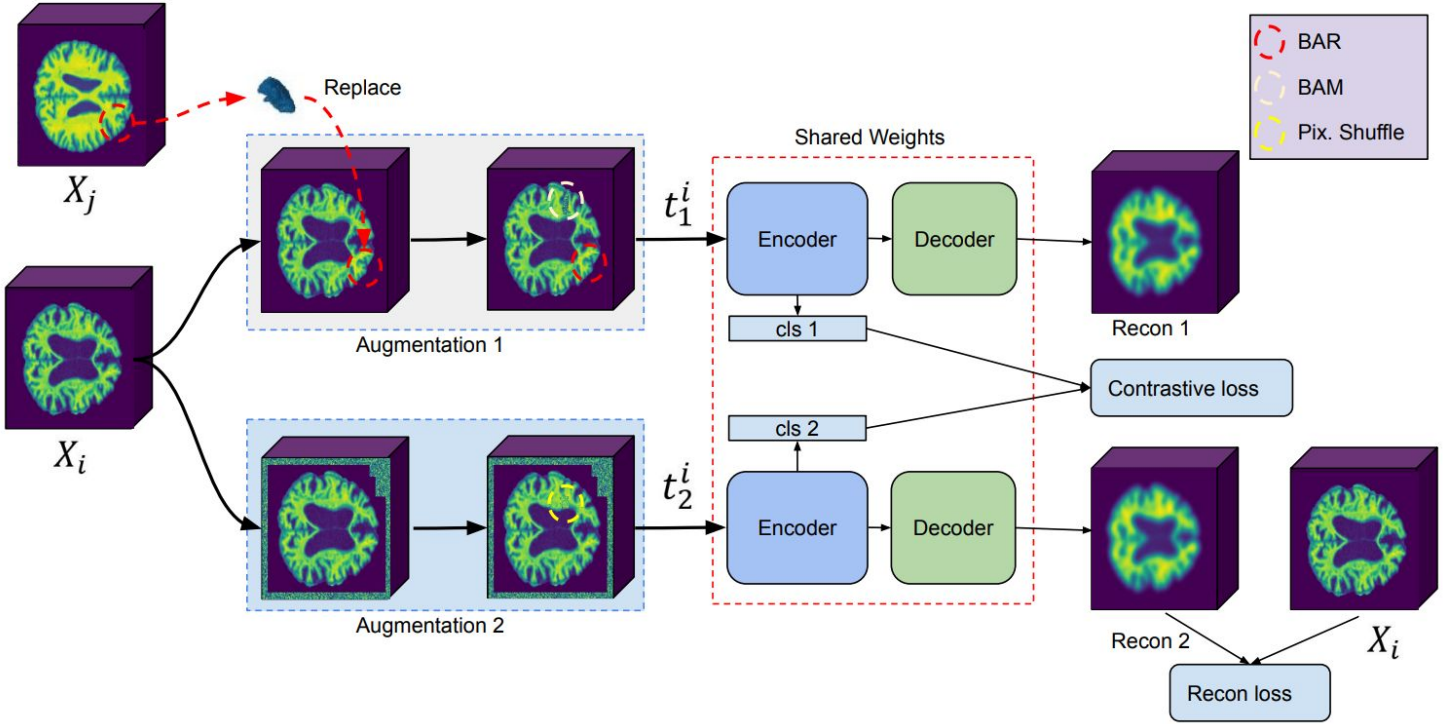
$$L_{NCE}^{c} = -\sum_{k=1}^{n} \frac{\varphi(y_k^p, y_i^p)}{\sum_{j=1}^{b} \varphi(y_j^p, y_i^p)} log \frac{e^{\theta(t_1^i, t_2^k)}}{\frac{1}{b}\sum_{j=1}^{b} e^{\theta(t_1^i, t_2^j)}}$$

- Here φ denotes a distance kernel between two labels, which in our case are the soft labels of mixtures. Hence, this objective explicitly force the model to learn the relative similarity of the augmented versions, and bring similarly mixed MRIs together.

BAR Pre-training

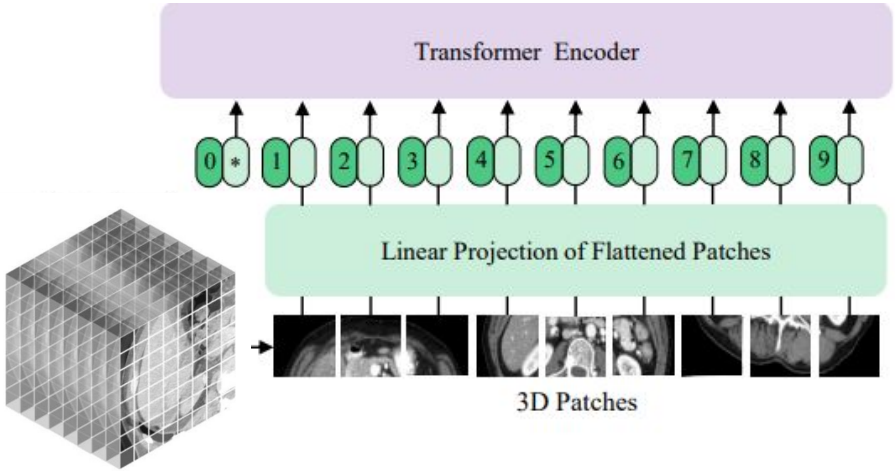


Proposed Pre-training Framework



Encoder

- We used a 3D enabled Vision Transformer with 10 layers and 12 attention heads. Also, input MRI size is selected as 96x96x96 and patch size of 16x16x16 is employed.

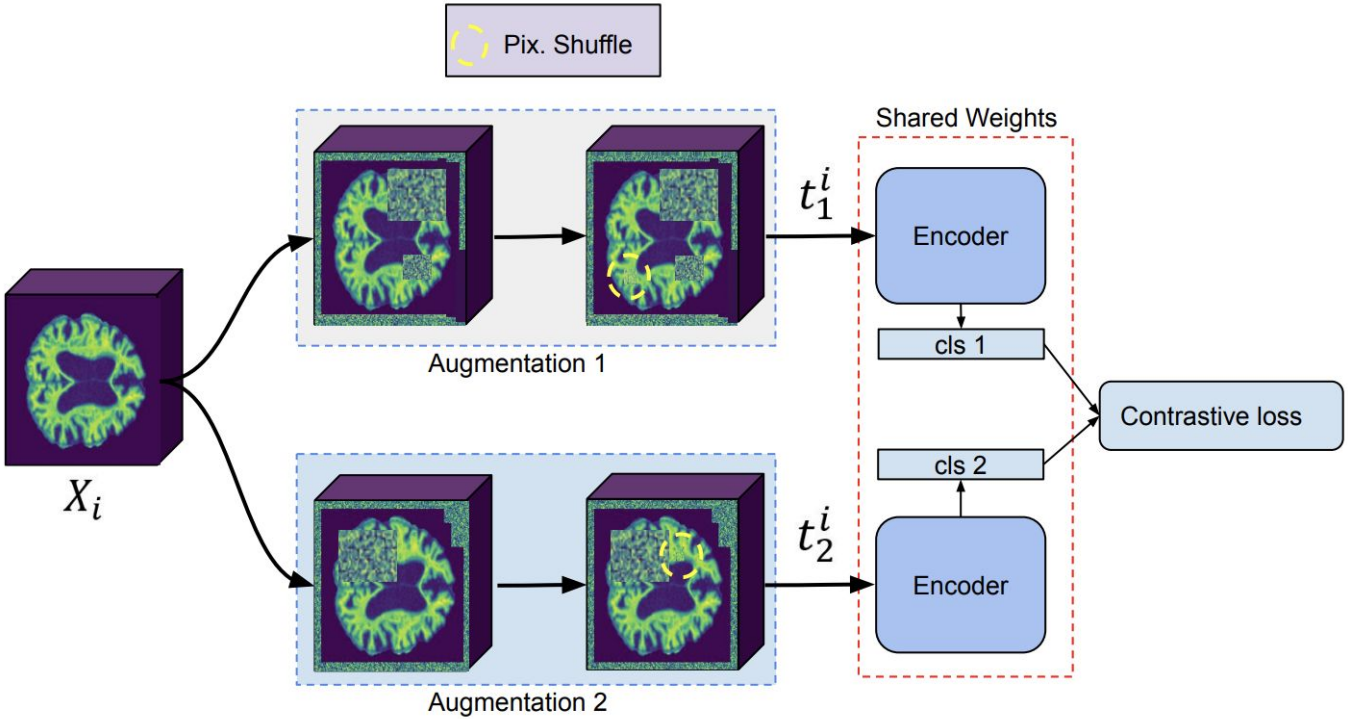


Experiments

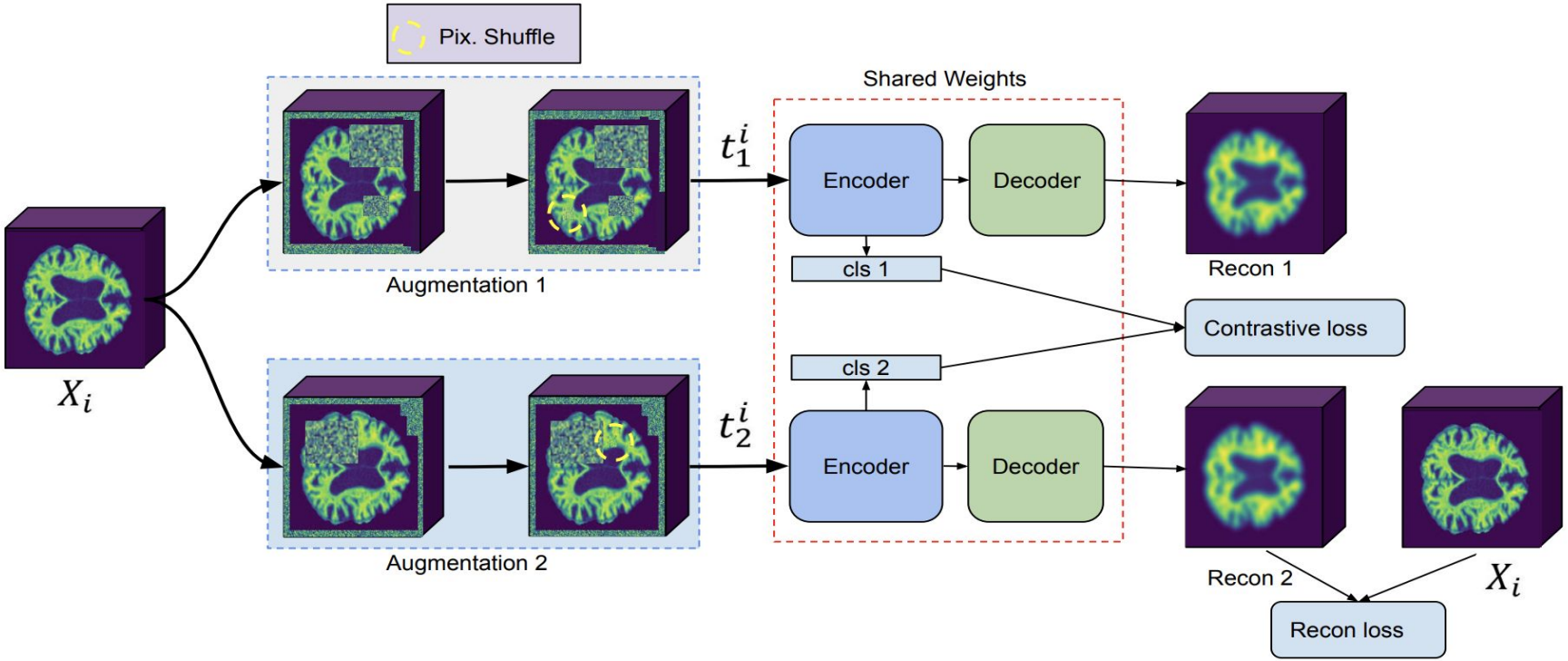
We compared the performance of our proposed framework against:

1. Training a ViT (as our Encoder) from scratch
2. Self-Supervised pre-training + fine-tuning
    a. Contrastive only
    b. Recon only
    c. Contrastive + Recon
3. CutMix based supervised pre-training + fine tuning
    a. Contrastive
    b. Contrastive + Recon

Self-supervised Contrastive Learning for AD



Self-supervised Contrastive Learning + Reconstruction for AD



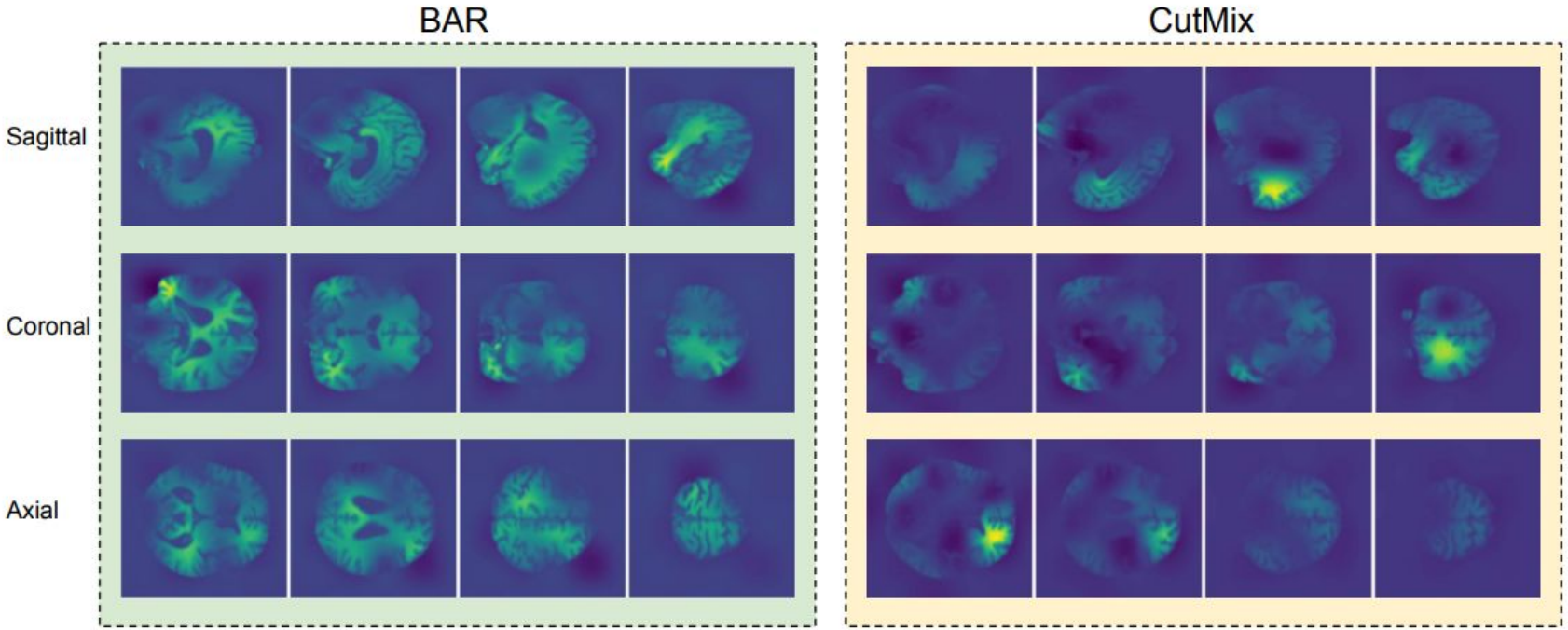
Results

Framework	Method	Precision	Recall	Accuracy
No Pre Training	ViT from scratch	74.38±7	85.6±3.1	80.83±3
Self Supervised Pre-Training + Fine Tuning	Contrastive	78.42±4.5	81.18±1.6	80.1±1.9
	Recon	78.6±5	85.57±1.1	82.69±2.5
	Contrastive + Recon	80.2±4.1	85.77±2	83.4±1.7
Supervised Pre-Training +Fine Tuning	CutMIX	83.06±4.8	87.08±3.5	85.29±2.8
	CutMIX + Recon	84.6±3.8	87.9±2.2	86.4±1
	BAS	84.7±3.3	87.6±2.1	86.3±1.1
	BAS + Recon	**86.24±3**	**88.08±2.3**	**87.22±0.8**

Advantages of Our Framework

- Mixture Learning during pre-training provides a great insight for the model especially when combined with the recon loss.
  - This way of pre training do not exhaust entropic capacity of our hard labels, since we do not utilize them “directly” but only use them to create synthetic mixtures, thus the same labels could be exploited during fine-tuning.
- BAR produces more realistic-looking synthetic MRIs, which leads to higher local variability, thus harder-to-solve synthetic samples.

AD case



Selection of Beta Distribution for BAR

- We tried two different beta distributions for sampling brain regions in BAR, a left skewed one with parameters of beta(0.2, 0.8) and a uniform distribution with beta(1,1). We obtained an overall accuracy of 86.9±1.5 with the left skewed one as opposed to 87.2 ± 1.3 with the uniform one.
- We argue that the replacement ratio sampled from the left skewed beta distribution makes somewhat of an easier objective with less replacements and thus is easier to solve. However, more research is needed to find the optimal replacement ratio.

Further Transferability

- We froze the ViT encoder in the BAR framework and trained an MLP.
  - Obtained an accuracy of 85.2 which shows that there is further room for the same features to be used for fine tuning, as the fine-tuned model yields about 87.22.
  - We argue that this is the case because we do not directly use our hard labels during pre-training but use them for creating soft labels and realistic looking synthetic images instead, thus their entropic capacity is not fully exhausted during the pre-training phase.

Directly Using the Hard-labels During Pretraining

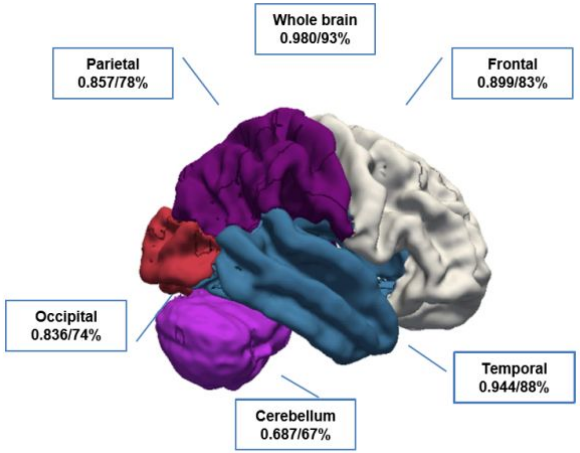
- We compared our soft-label supervised contrastive learning + fine-tuning approach against hard-label supervised contrastive learning + fine-tuning approach.
  - We utilized hard-labels and no replacements during pre-training of supervised contrastive loss + recon loss, using inner outer cuts and pixel shuffling for both $t_1^i \text{ and } t_2^i$. This approach produces lower quality embeddings compared to the soft-label approach as it yields an accuracy around 83.7% by training an MLP on top of the frozen encoder, and its fine tuning results are 84.7%.

Anybody wants to collaborate for the Summer?

- We want to expand our approach in some other Brain-Related tasks like Autism or Schizophrenia.

Future Projects: Guided Soft-Attention

- Use a region prominence mapping tensor to guide VALUE vector in Transformer training.
- Then slowly relax the importance of prominence mapping as the training continues based on the derivative of the loss



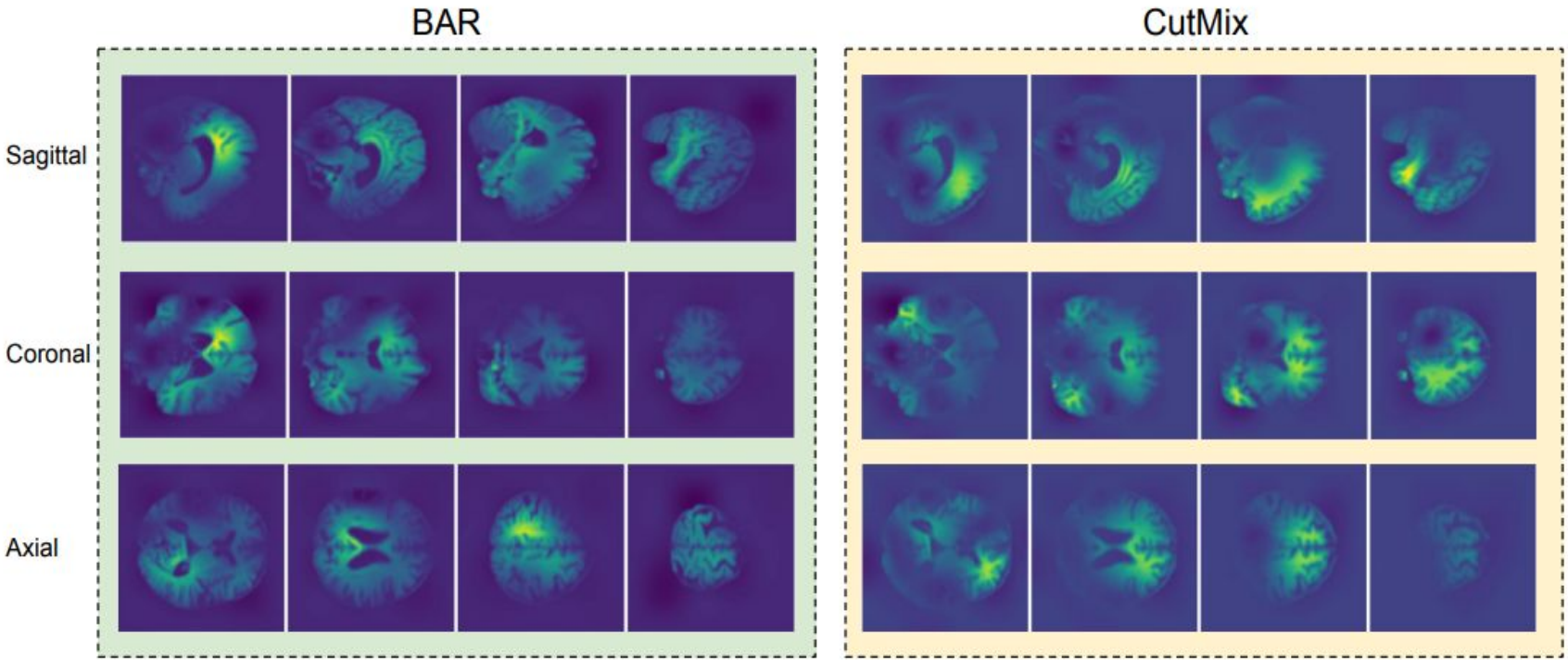
[6], They trained models using different parts of the brain and mapped their results, stating that representation strength of Temporal lobe is the greatest and only slightly worse than using the whole brain

Thank you!

References

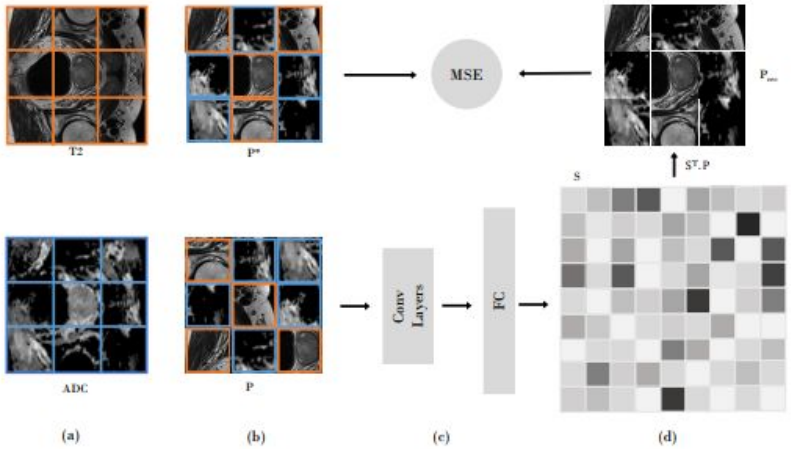
1. Alzheimer’s Disease Brain MRI Classification: Challenges and Insights
2. A Simple Framework for Contrastive Learning of Visual Representations
3. Masked Autoencoders Are Scalable Vision Learners
4. Supervised Contrastive Learning
5. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features
6. Deep Learning on MRI Affirms the Prominence of the Hippocampal Formation in Alzheimer’s Disease Classification

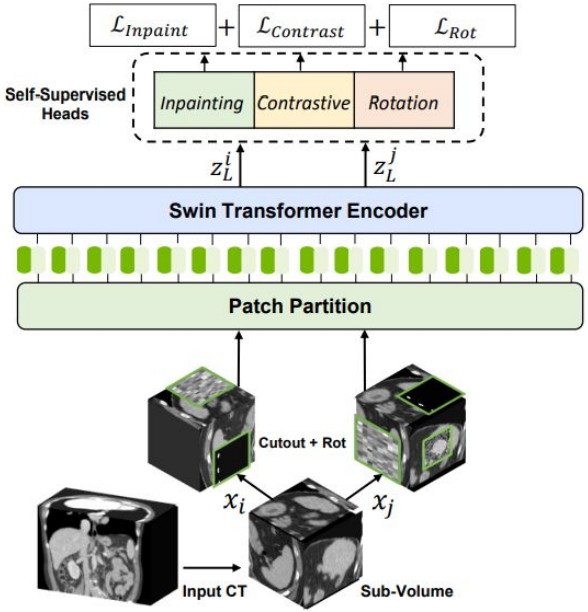
CN Case



Self-supervised learning in medical domain

- Since self supervised objectives help to alleviate problems coming from having low training data, many papers published in medical field too.
  - Mostly based on 3d jigsaw puzzles, 3d rotation predictions, patch location, reconstruction etc.





Conclusions

- We proposed a new framework for AD detection that combines a novel augmentation strategy, BAR, which leverages 3D anatomical brain regions to create synthetic MRIs and labels.
- We showed that, when pre-trained with the synthetic samples, a continuous valued supervised contrastive loss is very effective for the AD detection task.
- We experimented on the public dataset ADNI and showed that our approach outperforms training from scratch as well as self-supervised approaches.
- Furthermore, we compared BAR with (CutMix), and showed that BAR creates realistic looking samples, which leads to better embedding learning during pre-training.

Brain Aware Masking in Self-Supervised Case

- We test the performance of Brain Aware Mapping, (i.e., we randomly selected and filled 3D anatomical brain regions with noise) against the use of inner and outer cuts in a self-supervised manner. When fine-tuned, the performance is comparable to inner outer cuts with an overall accuracy of 83.54 ± 1.8 when trained with Contrastive + Recon with a similar drop ratio used in inner-outer cuts.