Researching Forgetting on "CLEAR" using Elastic Weight Consolidation

Ioan Daniel Savu & Rusu Andrei-Cristian | Supervisor : Florin Gogianu

University of Bucharest

Key findings

- The accuracy for the pretrained model nearly match the results in the original CLEAR experiment, even though our model was pretrained on Imagenet
- In the *CLEAR* dataset, we don't encounter catastrophic forgetting
- We can see the phenomenon of *forward* and *backward knowledge transfer*
- EWC seems to improve the overall performance of the models, not only prevent catastrophic forgetting
- For the **pretrained** model, the mean accuracy on sequential training **without** using **EWC** was **81%**, while using **EWC** we obtained **86%**.

Introduction and Related Work

The idea of **Elastic Weight Consolidation** was first discussed in [1], and it's a regularization method, inspired from neuroscience - "*synaptic consolidation might enable continual learning by reducing the plasticity of synapses that are vital to previously learned tasks*". **EWC** is an algorithm which performs a similar operation, by constraining important parameters of the neural network to stay close to their old values, when being trained on new tasks. The new loss formula is:

$$\mathcal{L}(\theta) = \mathcal{L}_B(\theta) + \sum_i \frac{\lambda}{2} F_i(\theta_i - \theta^*_{A,i})^2$$

where:

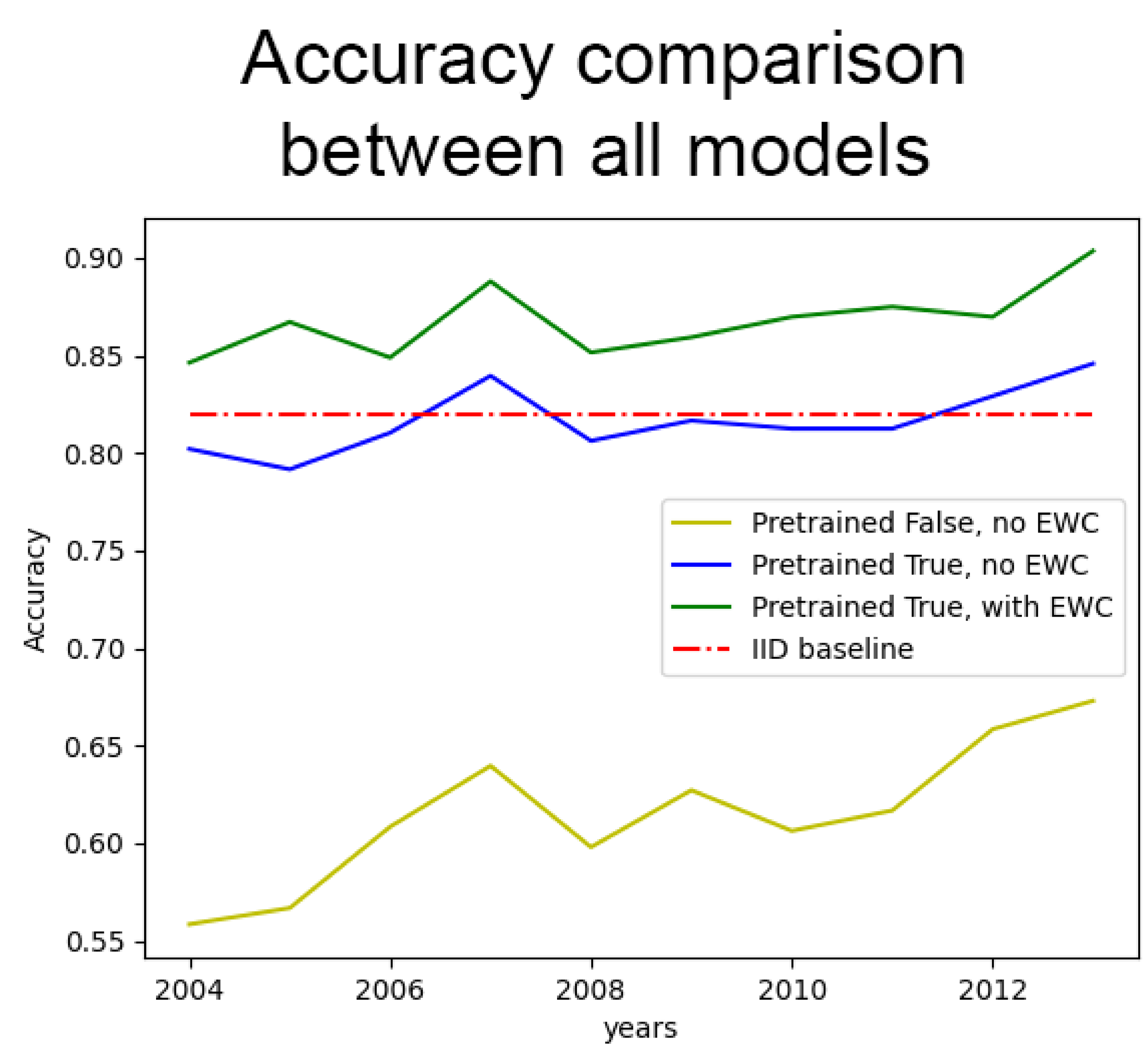
1. $\mathcal{L}(\theta) =$ The updated loss function
2. $\mathcal{L}_B(\theta) =$ The loss function for the new task B
3. $\lambda =$ Importance for the old tasks
4. $F_i =$ Fisher Information Matrix for each parameter i

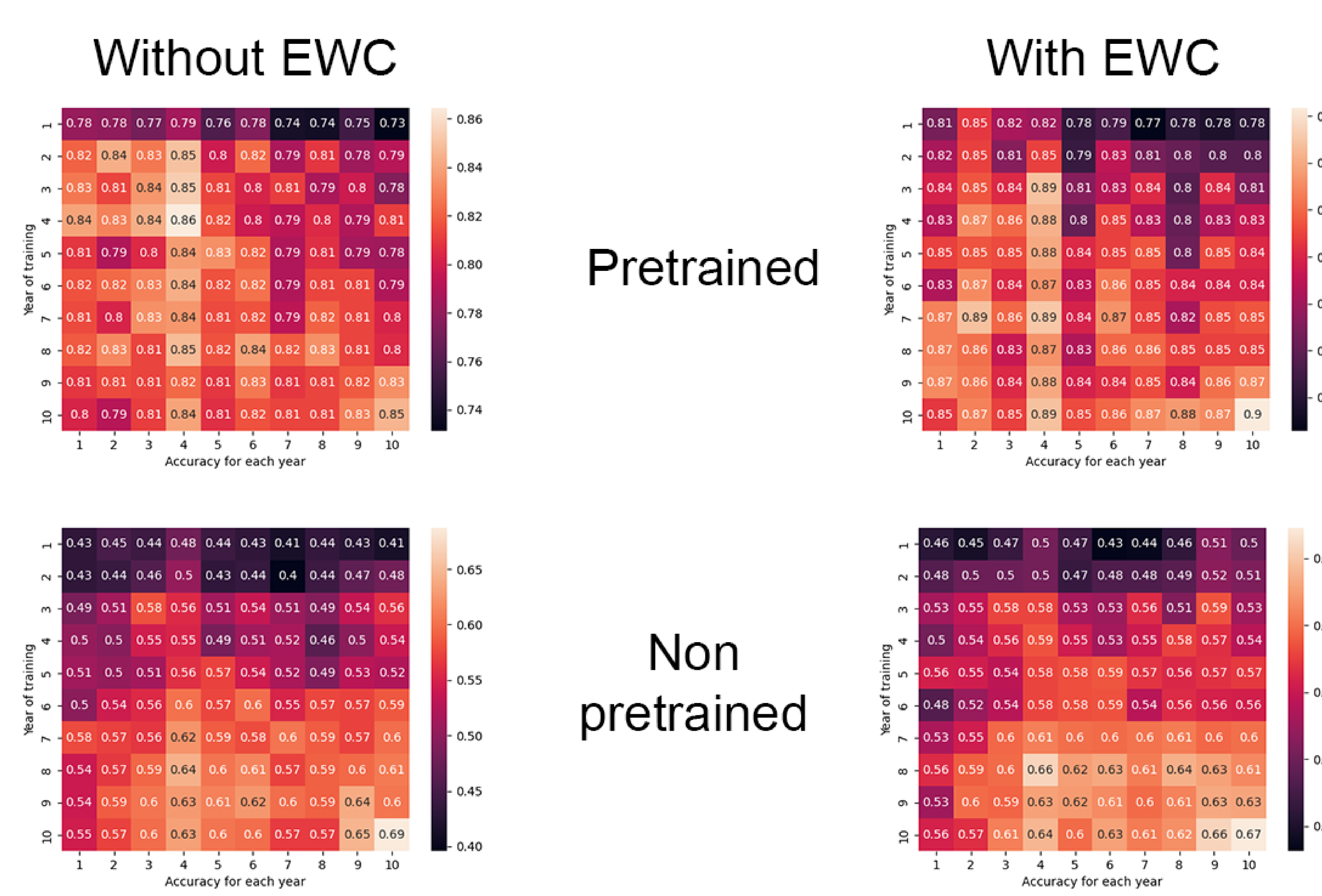
The *CLEAR* [2] dataset is specifically made for Continuous Learning on Real World Imagery. It contains 300 images for each of the 11 classes, and the data is distributed across 10 years, from 2004 to 2014. As can be seen, the continuous learning aspect of this dataset is the classes distribution shift over the years.

Setup & hyperparameters

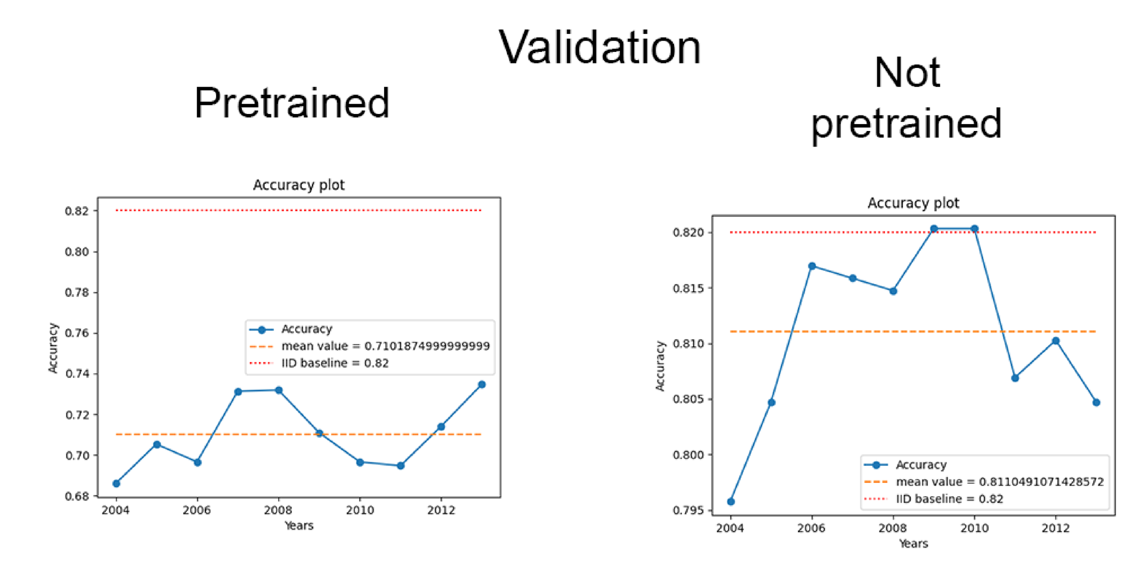
- We used data augmentation such as: Random Crop, Gaussian Blur, Random Rotation and Random Flip
- We capped the number of epochs to 50 and used *early stopping*
- We used a *learning rate scheduler*
- Patience : 3
- Learning rate: $1\mathrm{e}{-3}$, but we also tried values $1\mathrm{e}{-2}, 2\mathrm{e}{-3}, 2\mathrm{e}{-4}$
- Momentum: 0.9
- Batch size: 128, but we also tried $256, 64, 32$
- Image size: $112x112$

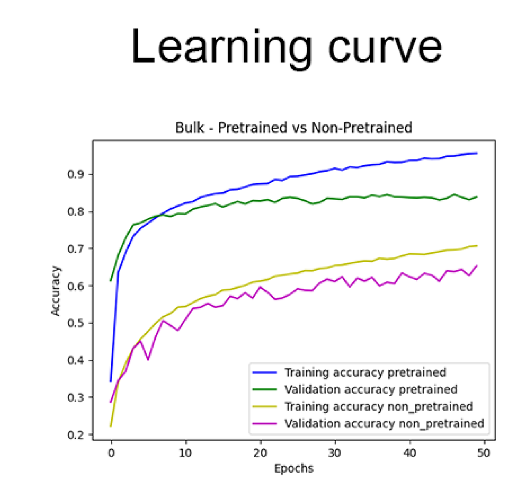
Results





We trained various models sequentially on the datasets, both with and without *Elastic Weight Consolidation*. A relevant metric for testing the performance is the accuracy matrix in which we compute the accuracy on the past and future datasets after training on each year. This gave us insight to how the model transfers or forgets the knowledge. The first thing we notice after having a look at the results is opposed to what we've been expecting - rather than forgetting, the models learn, for each year, new knowledge that propagates both forward and backwards, which improves the performance. This phenomenon suggests that the datasets are sharing features and that the distribution shift between them is smooth.





Key findings suggest that **EWC** may be able to do more than just prevent catastrophic forgetting. Since this method of regularization helps preserving important weights in the model, it might help in the training process on datasets that have a relative small number of datapoints and a gentle distribution shift between them to generalize better and thus give better results overall.

Conclusions & future work

We have observed an unexpected behaviour - even though *EWC* is thought to be a regularization method for catastrophic forgetting, it seems to improve the overall performance of the models. In the future, we intend to analyze if this property of *EWC* stands for other datasets with small distribution shifts, as this would mean that *EWC* is not just a solution for catastrophic forgetting, but also a good regularization method for improving performance.

References

References

[1] James Kirkpatrick et al. "Overcoming catastrophic forgetting in neural networks". In: *Proceedings of the national academy of sciences* 114.13 (2017), pp. 3521–3526.

[2] Zhiqiu Lin et al. "The CLEAR Benchmark: Continual LEArning on Real-World Imagery". In: *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. 2021.