A Memory Transformer Network for Incremental Learning
Ahmet Iscen, Thomas Bird, Mathilde Caron, Alireza Fathi, Cordelia Schmid

Motivation
- Catastrophic forgetting - Incremental Learning (IL) model sees tasks sequentially, cannot store all the data.
- Rehearsal - keep subset of the data (exemplars) to prevent forgetting
- Can we also explicitly incorporate the exemplars into the prediction process to further prevent forgetting?

Experiments
- We use a variety of off-the-shelf, publicly available feature extractors for MTN - SimCLR and CLIP.
- We compare to both end-to-end learned methods and methods with a feature extractor.
- We perform experiments on the large scale datasets ImageNet-1k and Landmarks-1k.
- MTN shows significantly less forgetting than SOTA competing methods.

Method
1. Embed query image with a fixed feature extractor, f.
2. Find k nearest neighbors from the memory, i.e. the set of embedded exemplars.
3. Input query embedding and k nearest memory embeddings into the Memory Transformer Network (MTN), h.
4. Obtain class probabilities using a linear head, g.

The MTN can attend directly to saved examples from memory - so, unlike other rehearsal methods, our mechanism to prevent forgetting is not solely via changes to the optimization procedure.