A Memory Transformer Network for Incremental Learning

Ahmet Iscen, Thomas Bird, Mathilde Caron, Alireza Fathi, Cordelia Schmid

BMVC 2022

Google Research

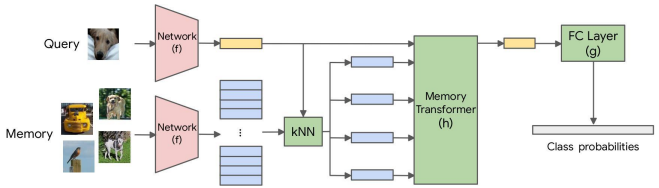
UCL

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?

Method

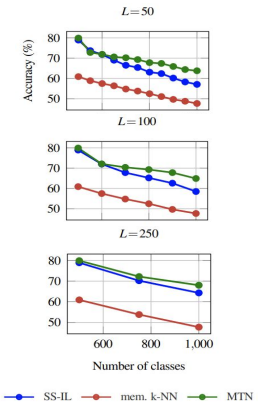
1. Embed query image with a fixed feature extractor, *f*.
2. Find k nearest neighbors from the memory, i.e. the set of embedded exemplars.
3. Input query embedding and k nearest memory embeddings into the *Memory Transformer Network* (MTN), *h*.
4. Obtain class probabilities using a linear head, *g*.



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.

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	Backbone	Memory usage: Train.	Pred.	ImageNet-1k 50	100	250	Landmarks-1K 50	100	250
PODNet [8]	R18$_{(Scratch)}$	✓		64.1	67.0	-	-	-	-
CCIL [23]†	R18$_{(Scratch)}$	✓		**65.8**	**67.6**	**69.1**	40.1	49.7	55.4
SS-IL [8]†	R18$_{(Scratch)}$	✓		57.0	62.5	67.1	54.9	60.0	64.7
MTN	🔒R18$_{(SS-IL)}$	✓	✓	60.8	64.8	68.4	**59.4**	**63.1**	**66.5**
mem. k-NN†	🔒-R50$_{(SimCLR)}$		✓	53.8	53.9	54.2	30.0	30.1	30.2
BIC [33]†	🔒-R50$_{(SimCLR)}$	✓		34.3	43.9	56.9	17.4	26.5	38.0
iCARL [21]†	🔒-R50$_{(SimCLR)}$	✓	✓	56.5	57.0	58.4	27.3	27.7	28.5
LUCIR [18]†	🔒-R50$_{(SimCLR)}$	✓		60.1	63.0	68.7	48.4	50.9	55.3
SS-IL [8]†	🔒-R50$_{(SimCLR)}$	✓		66.1	67.5	71.2	46.2	46.6	47.9
MTN	🔒-R50$_{(SimCLR)}$	✓	✓	**69.5**	**70.7**	**73.3**	**51.0**	**52.3**	**55.4**
mem. k-NN†	🔒-ViT-B$_{(CLIP)}$		✓	47.7	47.7	48.1	35.8	35.9	36.1
BIC [33]†	🔒-ViT-B$_{(CLIP)}$	✓		32.9	43.2	57.5	25.3	35.6	47.2
iCARL [21]†	🔒-ViT-B$_{(CLIP)}$	✓	✓	52.3	53.2	54.7	36.5	36.9	37.8
LUCIR [18]†	🔒-ViT-B$_{(CLIP)}$	✓		51.6	55.7	63.3	46.2	48.8	53.8
SS-IL [8]†	🔒-ViT-B$_{(CLIP)}$	✓		63.9	65.7	69.7	52.4	52.8	54.4
MTN	🔒-ViT-B$_{(CLIP)}$	✓	✓	**67.4**	**68.9**	**71.2**	**54.1**	**55.5**	**57.6**