# Continual HyperTransformer: A Meta-Learner for Continual Few-Shot Learning


**Max Vladymyrov, Andrey Zhmoginov, Mark Sandler**

{mxv, azhmogin, sandler}@google.com

Google Research


## Goal

Propose a few-shot continual hypernetwork model:

- Few-shot : learning from few samples
- Continual : learning without forgetting.
- Hypernetwork : learning on the fly (no training!).

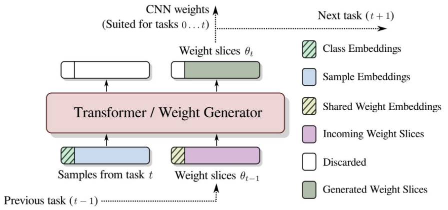
## Motivation

While useful for many machine learning applications (e.g. robotics, privacy-preserving training), we argue that the combination above suggests a an appealing framework for modeling the biological learning systems, such as the brain [1]. As humans, we are able to learn directly (hypernetwork) from few examples (few-shot) without forgetting what we have learned before (continual).

## Model

HyperTransformer [2] is a few-shot hypernetwork that is able to generate weights for the custom CNN model on the fly from a few labeled examples. It works by decoupling the complexities of the model generator (via a Transformer) and the generated model (via a CNN).

We want to extend it to incremental setting, by *using the weights generated for the previous tasks as input when trained for the new task*.




## Learning with Prototypes

In order to separate learning classes from different tasks, use prototypical loss:

- Accumulate prototypes for the support set

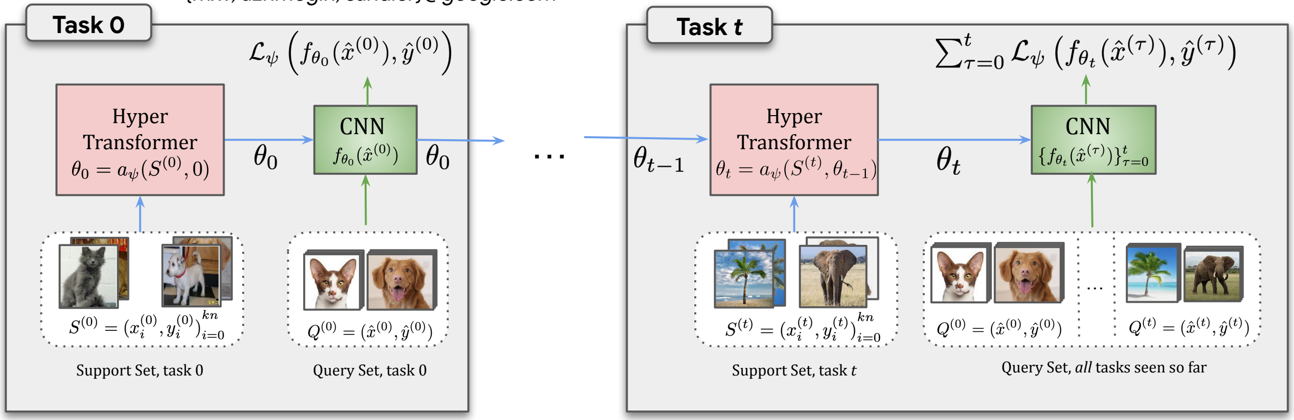
$$c_{\tau k} = \frac{1}{N}\sum_{(x,y)\in S^{(\tau)}} f_{\theta_\tau}(x)\mathbf{1}_{y=k}$$

- Compute the softmax over the query set
  - Task-incremental learning:

$$p(\hat{y} = k|\hat{x},\tau) = \frac{\exp(-\|f_{\theta_t}(\hat{x}) - c_{\tau k}\|^2)}{\sum_{k'}\exp(-\|f_{\theta_t}(\hat{x}) - c_{\tau k'}\|^2)}$$

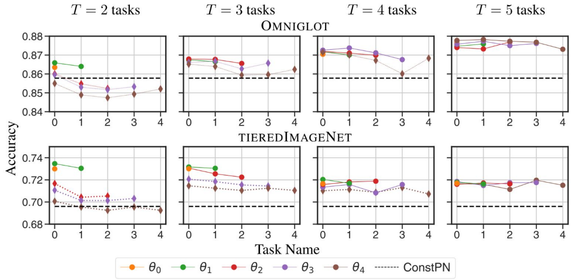
  - Class-incremental learning:

$$p(\hat{y} = k,\tau|\hat{x}) = \frac{\exp(-\|f_{\theta_t}(\hat{x}) - c_{\tau k}\|^2)}{\sum_{\tau' k'}\exp(-\|f_{\theta_t}(\hat{x}) - c_{\tau' k'}\|^2)}$$



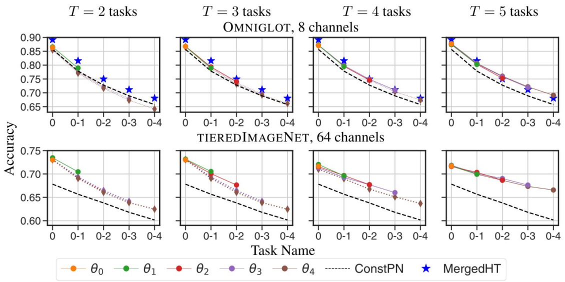

## Task-incremental learning

Given *task_id*, predict *class_id*.

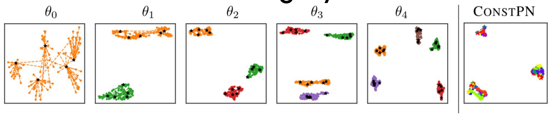



## Class-incremental learning

Predict both *task_id* and *class_id*.




## UMAP of the embedding layer




## References


[1] Miller, E.K. and Cohen, J.D. *An integrative theory of prefrontal cortex function*. Annual review of neuroscience, 24(1), 2001, pp.167-202.

[2] Zhmoginov, A., Sandler, M. and Vladymyrov, M., *Hypertransformer: model generation for supervised and semi-supervised few-shot learning*, ICML 2022.