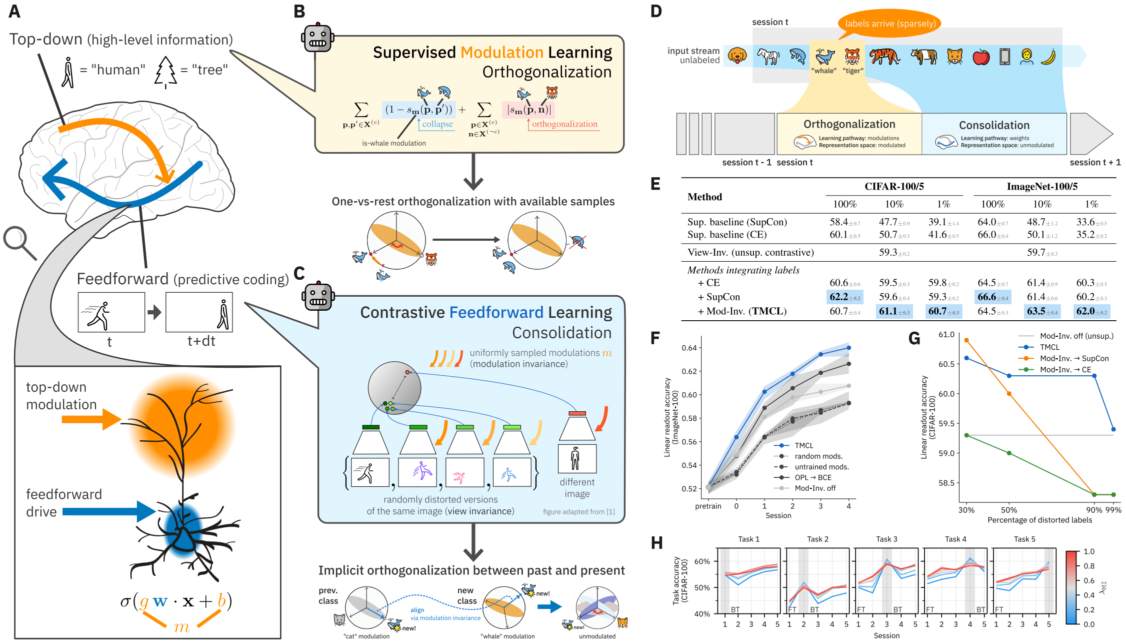
Continual integration of top-down information via modulation-invariant predictive learning

Viet Anh Khoa Tran(1), Emre O. Neftci(1,2) and Willem A. M. Wybo(1)

(1) PGI-15, Research Center Jülich, (2) RWTH Aachen University



Method	CIFAR-100/5			ImageNet-100/5		
	100%	10%	1%	100%	10%	1%
Sup. baseline (SupCon)	58.4 ±0.7	47.7 ±0.9	39.1 ±1.4	64.0 ±0.7	48.7 ±1.2	33.6 ±0.5
Sup. baseline (CE)	60.1 ±0.5	50.7 ±0.3	41.6 ±0.5	66.0 ±0.4	50.1 ±1.2	35.2 ±0.2
View-Inv. (unsup. contrastive)		59.3 ±0.2			59.7 ±0.3	
Methods integrating labels						
+ CE	60.6 ±0.6	59.5 ±0.3	59.8 ±0.2	64.5 ±0.7	61.4 ±0.9	60.3 ±0.5
+ SupCon	**62.2** ±0.2	59.6 ±0.4	59.3 ±0.2	**66.6** ±0.4	61.4 ±0.6	60.2 ±0.3
+ Mod-Inv. (**TMCL**)	60.7 ±0.4	**61.1** ±0.3	**60.7** ±0.3	64.5 ±0.3	**63.5** ±0.4	**62.0** ±0.2

Input data streams encountered by animals or humans consist largely of unsupervised data, but are interspersed with infrequent supervisory episodes (e.g. a parent telling their child that the object is called an apple, Fig. **D**). It is unclear how such sparse supervisory episodes can improve the learned representation; in machine learning, naive integration of supervised labels via fine-tuning leads to catastrophic forgetting as weight updates interfere with previous tasks and contexts. Therefore, the learning dilemma that arises is how learning mechanisms can benefit from sparse high-level information without negatively affecting generalizability of representations.

In this work, we follow the idea that predictive learning principles govern unsupervised plasticity in the neocortex (Fig. **A**), which we implement via contrastive learning, building a view-invariant representation space. Our *task-modulated contrastive learning (TMCL)* paradigm co-opts the predictive learning machinery to integrate high-level information from top-down modulations continually and without supervision (Fig. **C**) In our class-incremental learning setup, whenever labeled samples of a new class occur, new affine modulations are learned that improve separation of the new class from all others, without affecting feedforward weights (Fig. **B**). By co-opting the view invariance learning mechanism, we then train feedforward weights to match the unmodulated representation of a data sample to its modulated counterparts. This introduces modulation invariance into the representation space, and, by also using past modulations, implicitly separates past from present classes (Fig. **C, bottom**).

Our simulations show improvements in standard class-incremental representation learning benchmarks over state-of-the-art unsupervised approaches, as well as over comparable supervised approaches in classification accuracy (Fig. **E**) and also despite noisy labels (Fig. **G**). Ablations show that the specific design of the modulations is crucial (Fig. **F**). Our work suggests that top-down modulations play a crucial role in balancing stability and plasticity, as adapting the stength of the modulation-invariant loss balances forward and backward transfer against task adaptation (Fig. **H**), while also offering a new perspective on the role of modulatory inputs in shaping the canonical predictive learning machinery.