ECCV'20 ONLINE
23-28 AUGUST 2020

Adversarial Continual Learning

Sayna Ebrahimi
UC Berkeley

Franziska Meier
Facebook AI Research

Roberto Calandra
Facebook AI Research

Trevor Darrell
UC Berkeley

Marcus Rohrbach
Facebook AI Research

facebook Artificial Intelligence Research

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

What is Continual Learning?

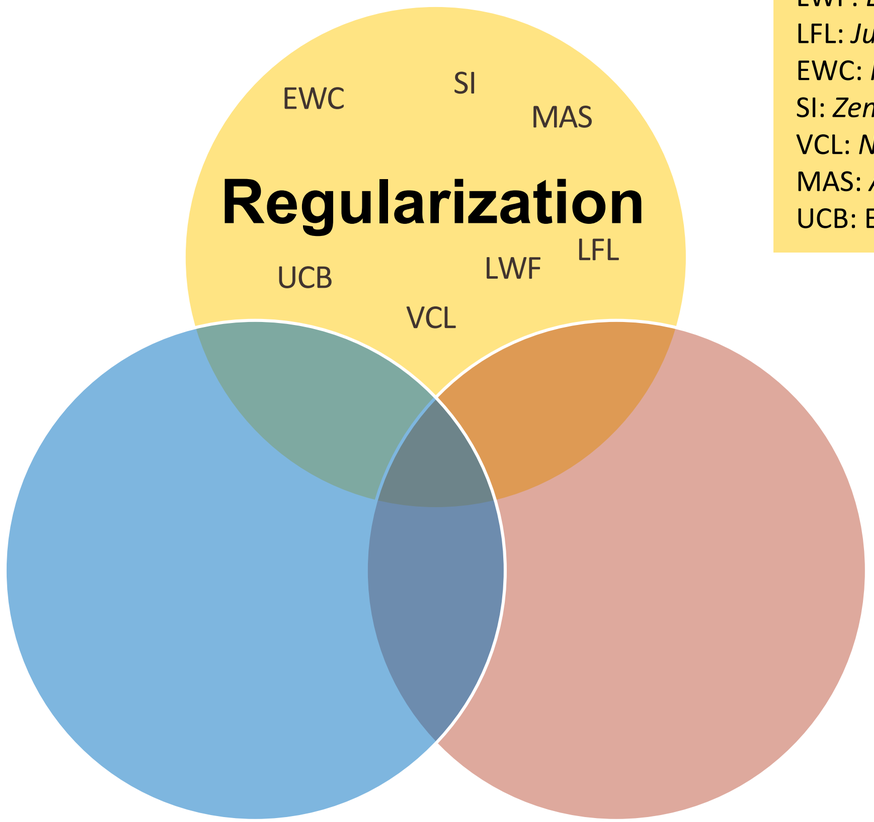
Definition:

learning a sequence of tasks and performing well on all of them

Objectives:

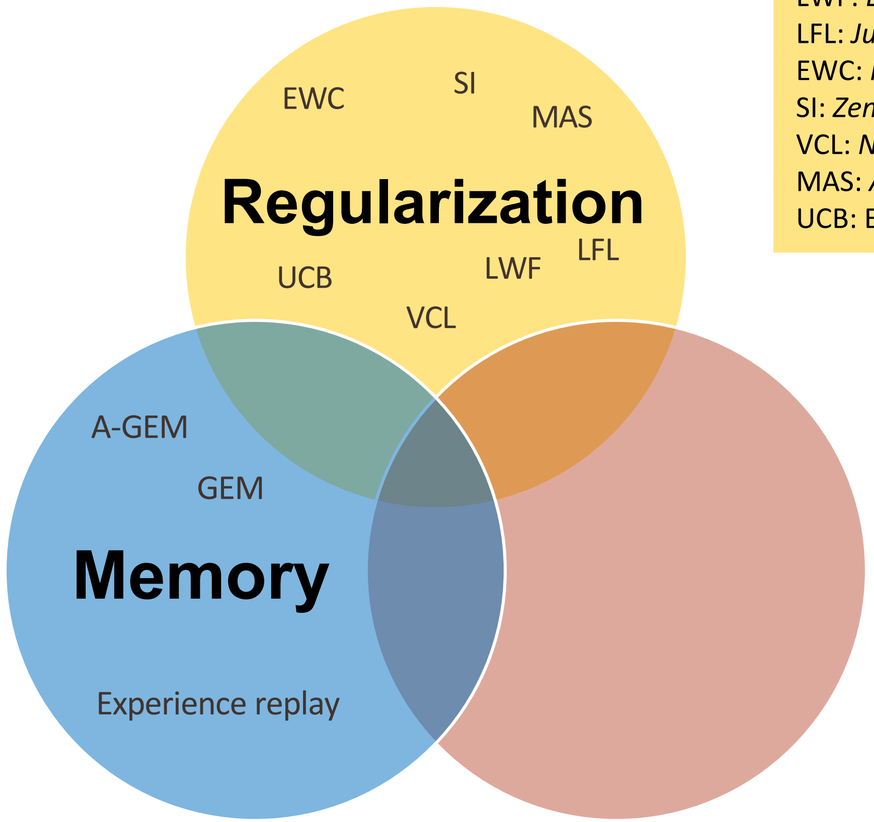
- No forgetting
- Data streams and revisiting is not allowed/limited
- High knowledge transferability
- Efficiency and scalability
- No/limited task information at test time
- etc.

Approaches in Continual Learning



LWF: *Li & Hoiem, 2016*
LFL: *Jung et al., 2016*
EWC: *Kirkpatrick et al., 2016*
SI: *Zenke et al., 2017*
VCL: *Nguyen et al., 2017*
MAS: *Aljundi, 2018*
UCB: Ebrahimi *et al., 2020*

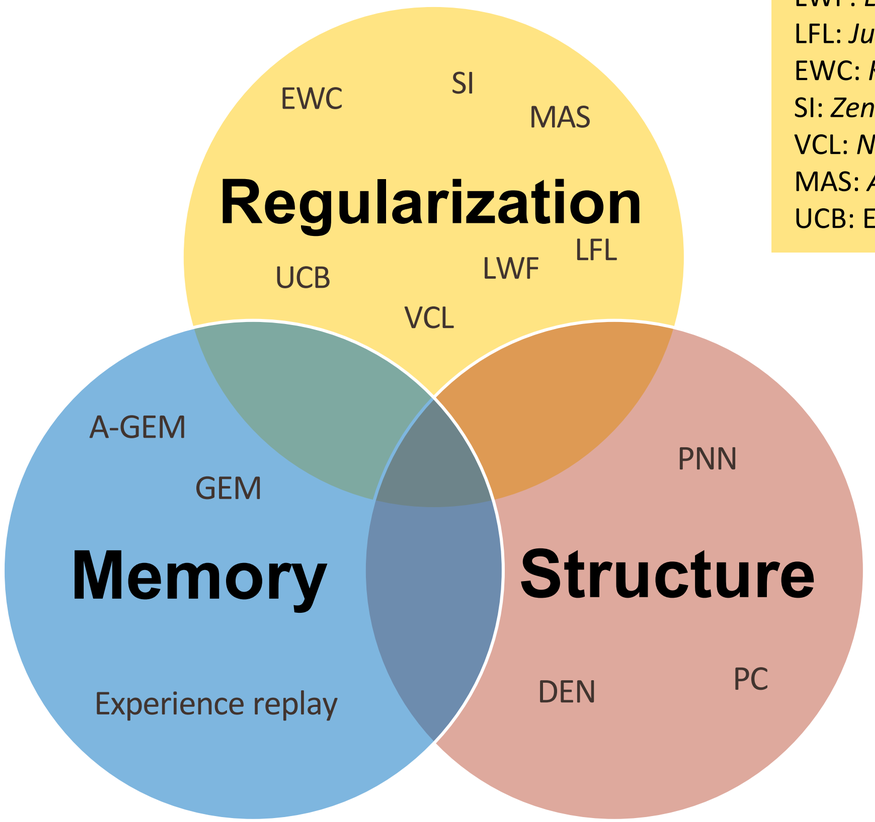
Approaches in Continual Learning



LWF: *Li & Hoiem, 2016*
LFL: *Jung et al., 2016*
EWC: *Kirkpatrick et al., 2016*
SI: *Zenke et al., 2017*
VCL: *Nguyen et al., 2017*
MAS: *Aljundi, 2018*
UCB: Ebrahimi *et al., 2020*

Experience replay: *Robins, 1995*
GEM: *Lopez-Paz & Ranzato, 2017*
A-GEM: *Chaudhry et al., 2019*

Approaches in Continual Learning

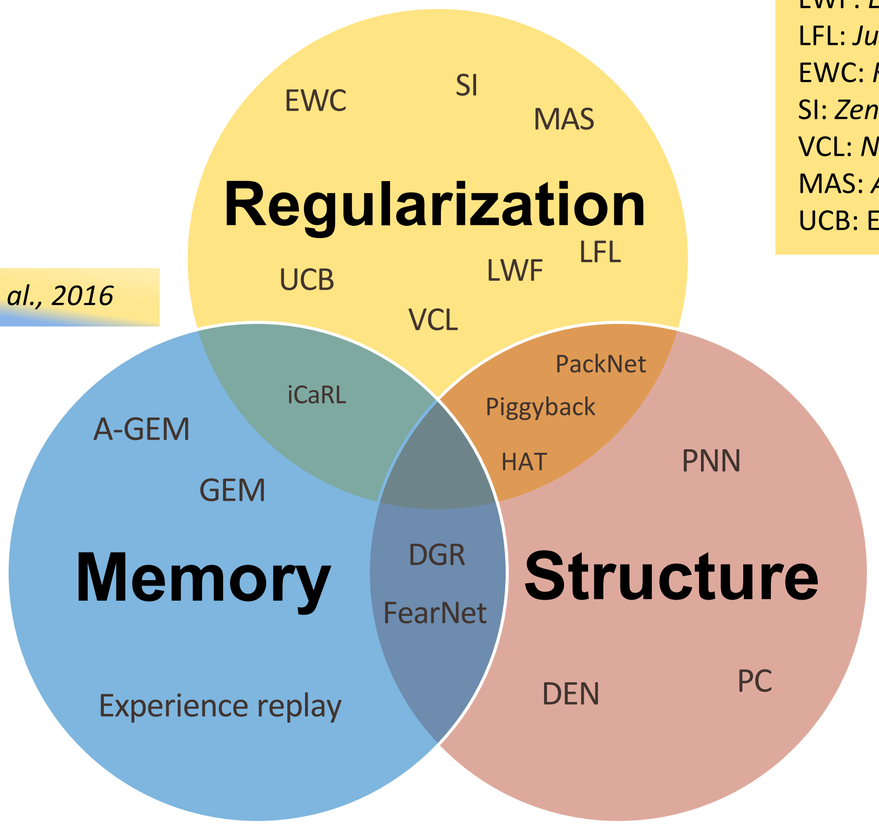


LWF: *Li & Hoiem, 2016*
LFL: *Jung et al., 2016*
EWC: *Kirkpatrick et al., 2016*
SI: *Zenke et al., 2017*
VCL: *Nguyen et al., 2017*
MAS: *Aljundi, 2018*
UCB: Ebrahimi *et al., 2020*

Experience replay: *Robins, 1995*
GEM: *Lopez-Paz & Ranzato, 2017*
A-GEM: *Chaudhry et al., 2019*

PNN*: Rusu et al., 2016*
DEN*: Yoon et al., 2018*
PC: *Schwarz et al., 2018*

Approaches in Continual Learning



LWF: *Li & Hoiem, 2016*
LFL: *Jung et al., 2016*
EWC: *Kirkpatrick et al., 2016*
SI: *Zenke et al., 2017*
VCL: *Nguyen et al., 2017*
MAS: *Aljundi, 2018*
UCB: Ebrahimi *et al., 2020*

iCaRL: *Rebuffi et al., 2016*

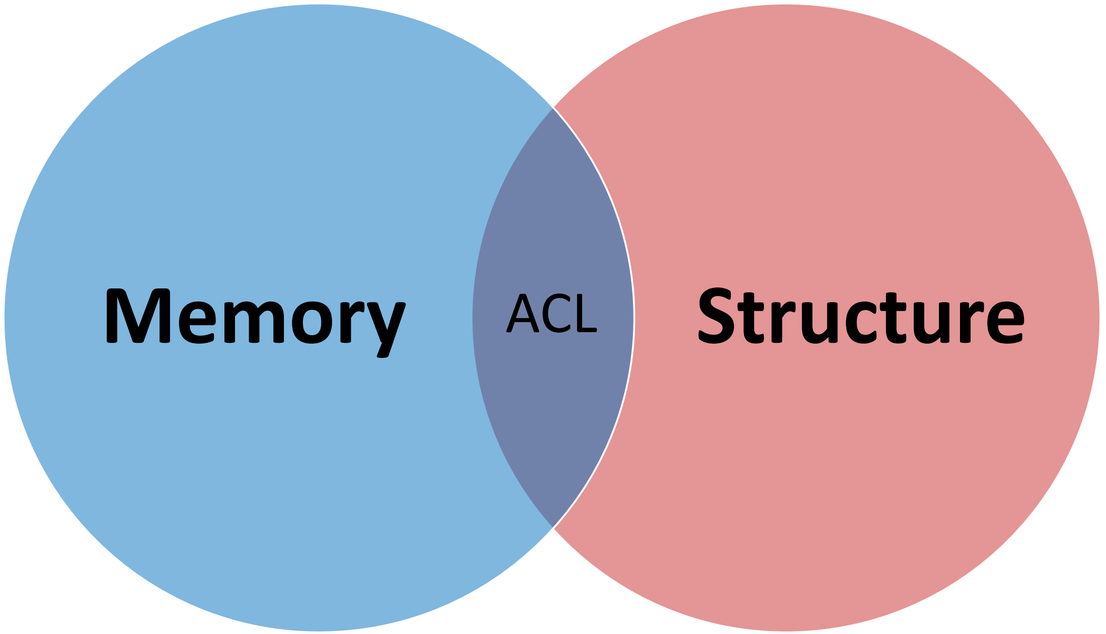
PackNet: *Mallya, 2018*
Piggyback: Mallya 2018
HAT: *Serrà et al., 2018*

Experience replay: *Robins, 1995*
GEM: *Lopez-Paz & Ranzato, 2017*
A-GEM: *Chaudhry et al., 2019*

PNN*: Rusu et al., 2016*
DEN*: Yoon et al., 2018*
PC: *Schwarz et al., 2018*

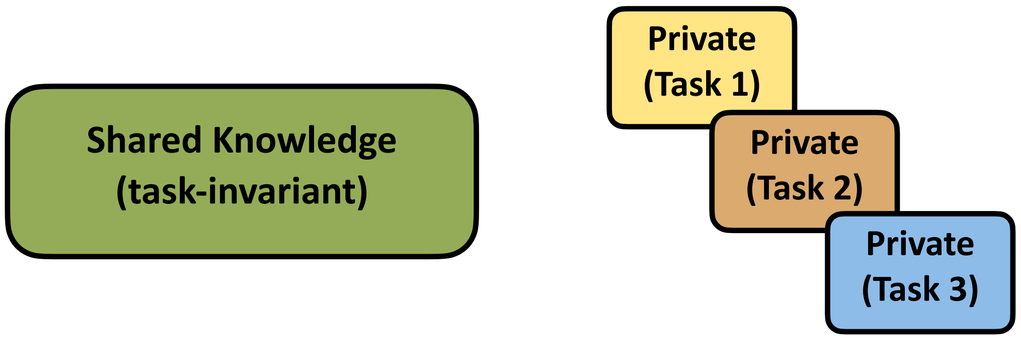
DGR: Shin *et al., 2017*
FearNet: Kemker *et al., 2017*

Where does our approach (ACL) stand?



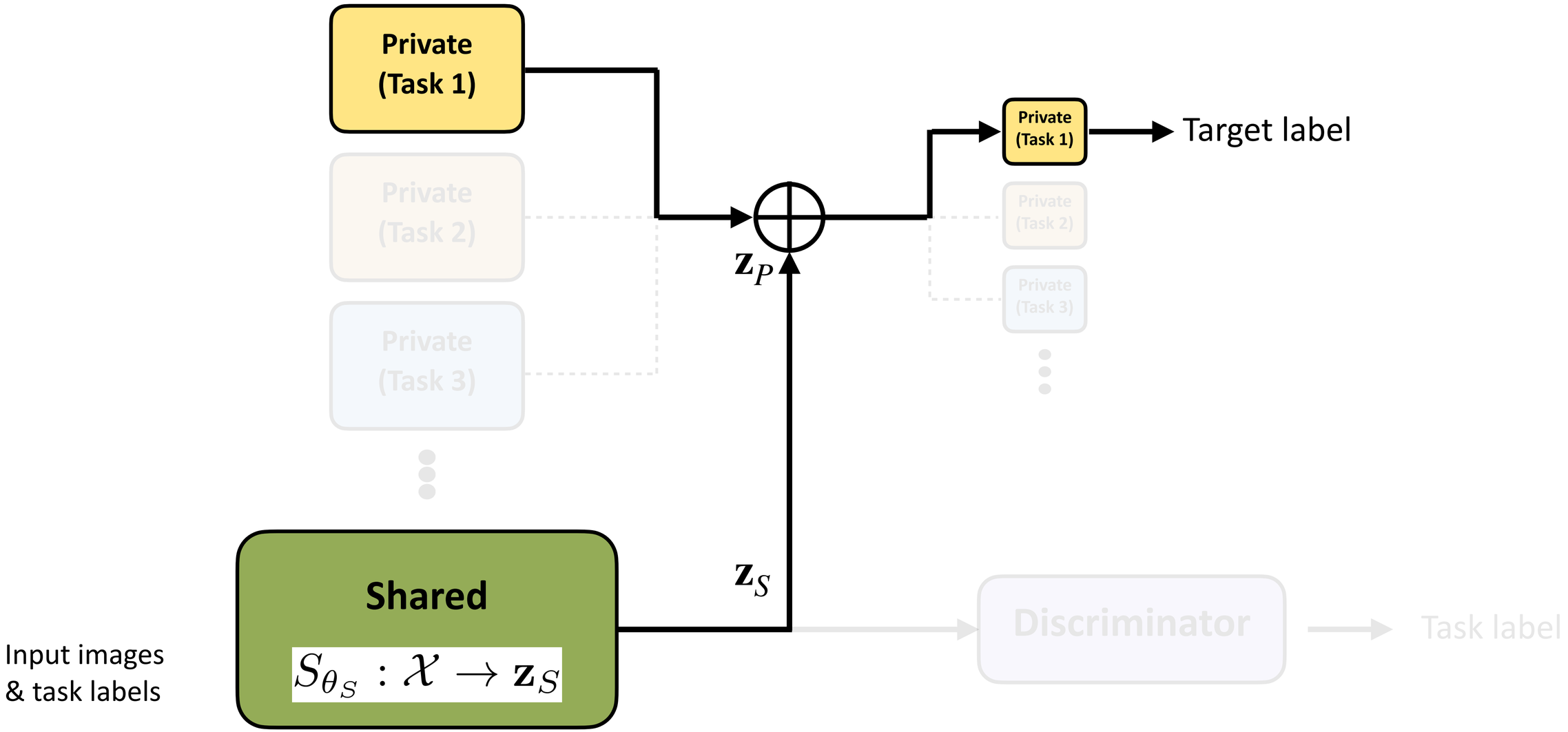
Intuition

- Tasks in a sequence
  - Have *task-invariant (shared)* knowledge in common
  - Require task-specific *(private)* features to master them

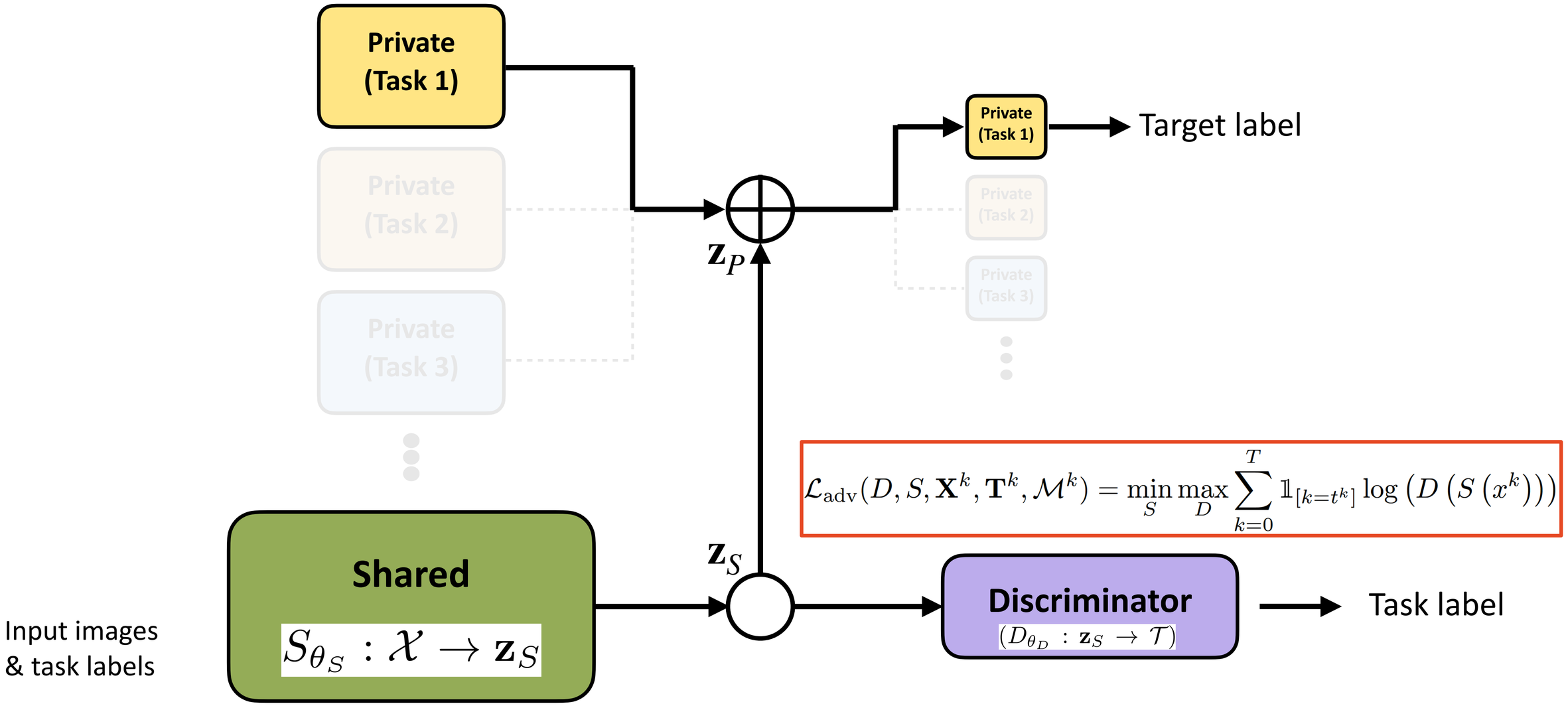


How can we factorize *task-invariant* from *task-specific* features?

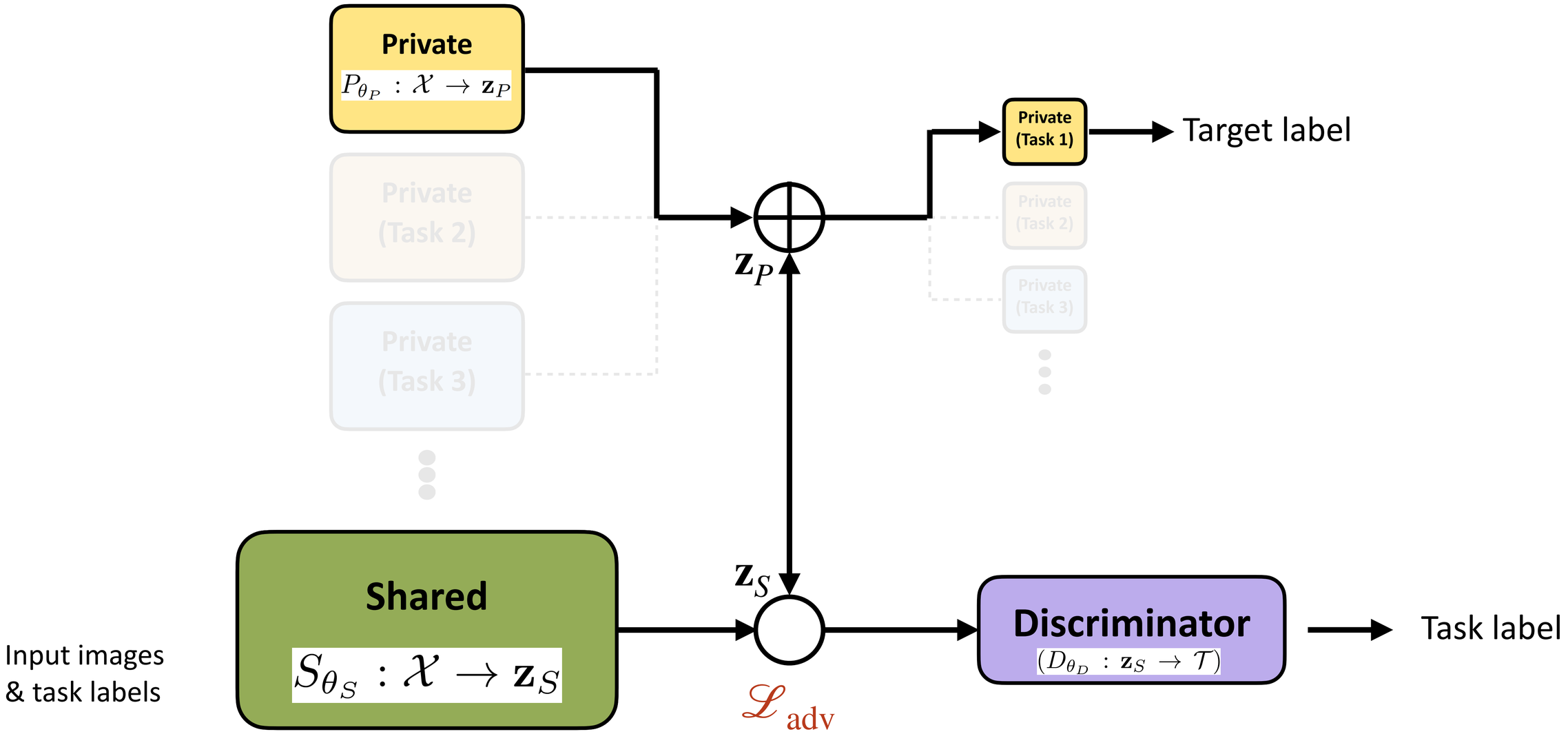
Our Approach: Adversarial Continual Learning



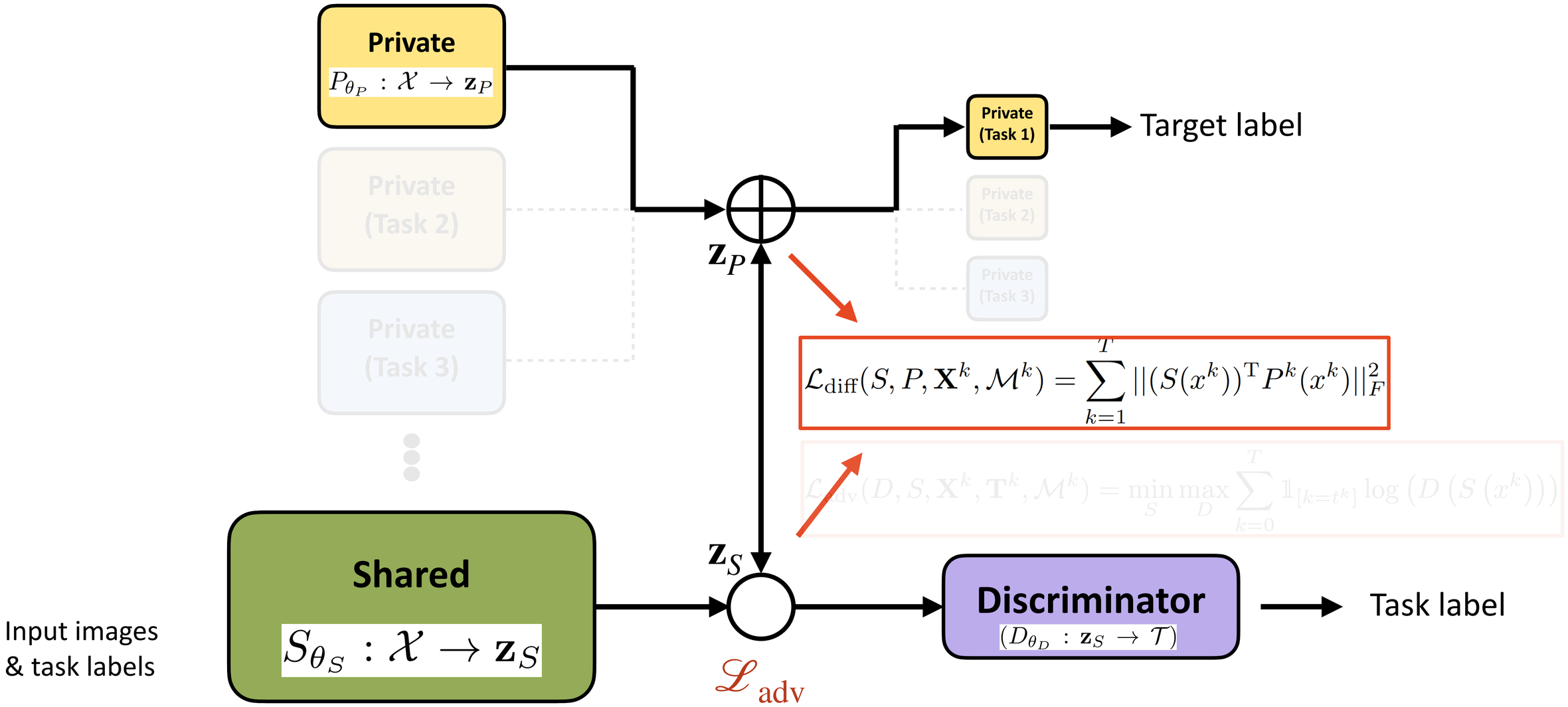
Our Approach: Adversarial Continual Learning



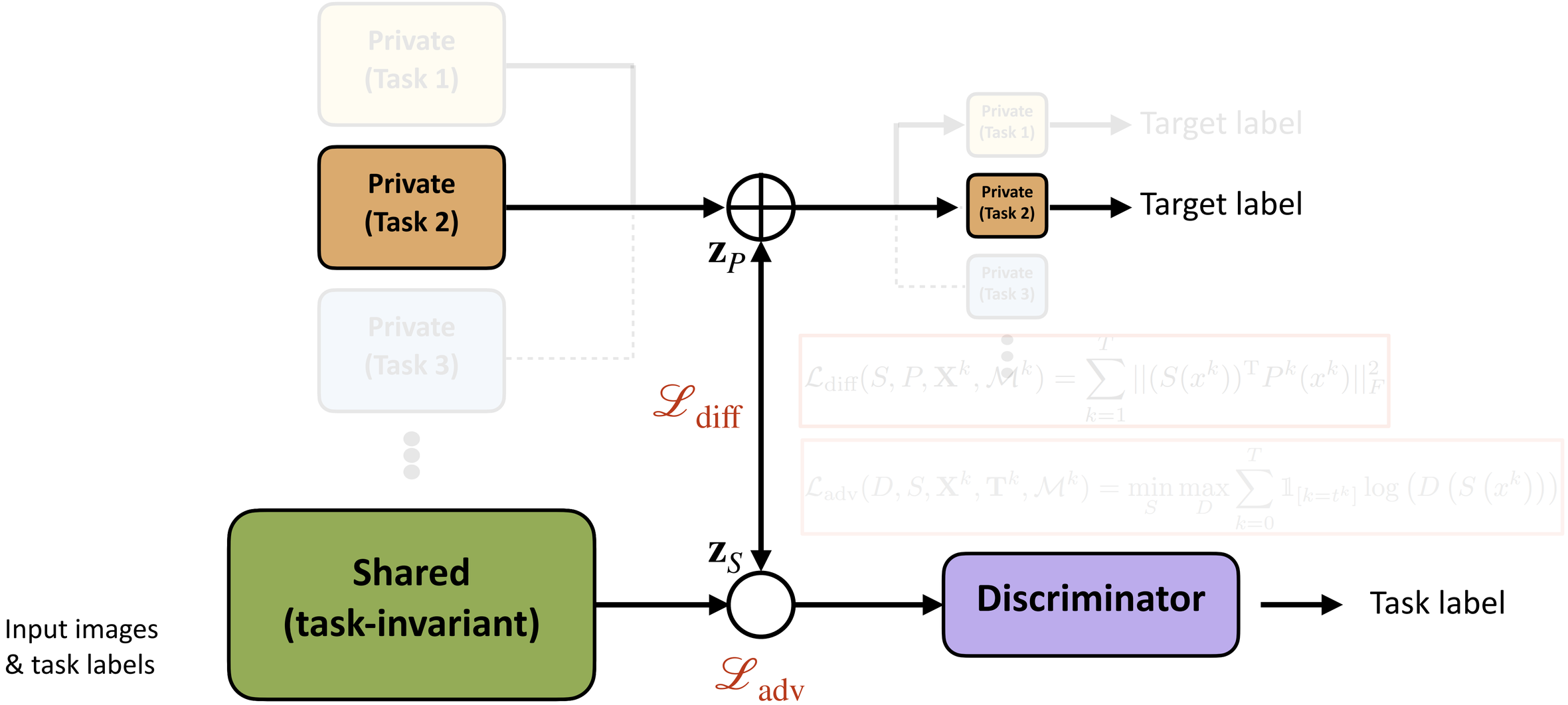
Our Approach: Adversarial Continual Learning



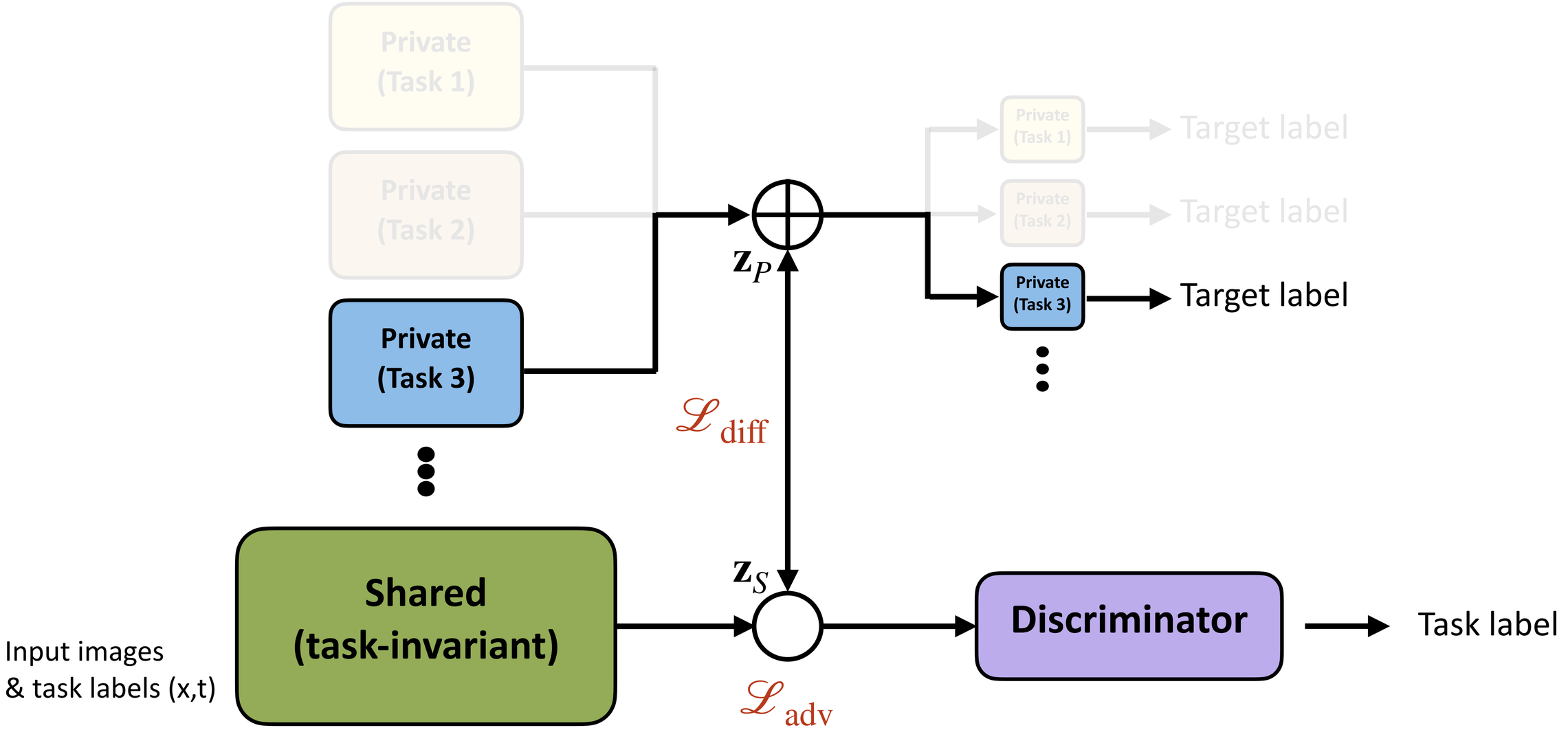
Our Approach: Adversarial Continual Learning



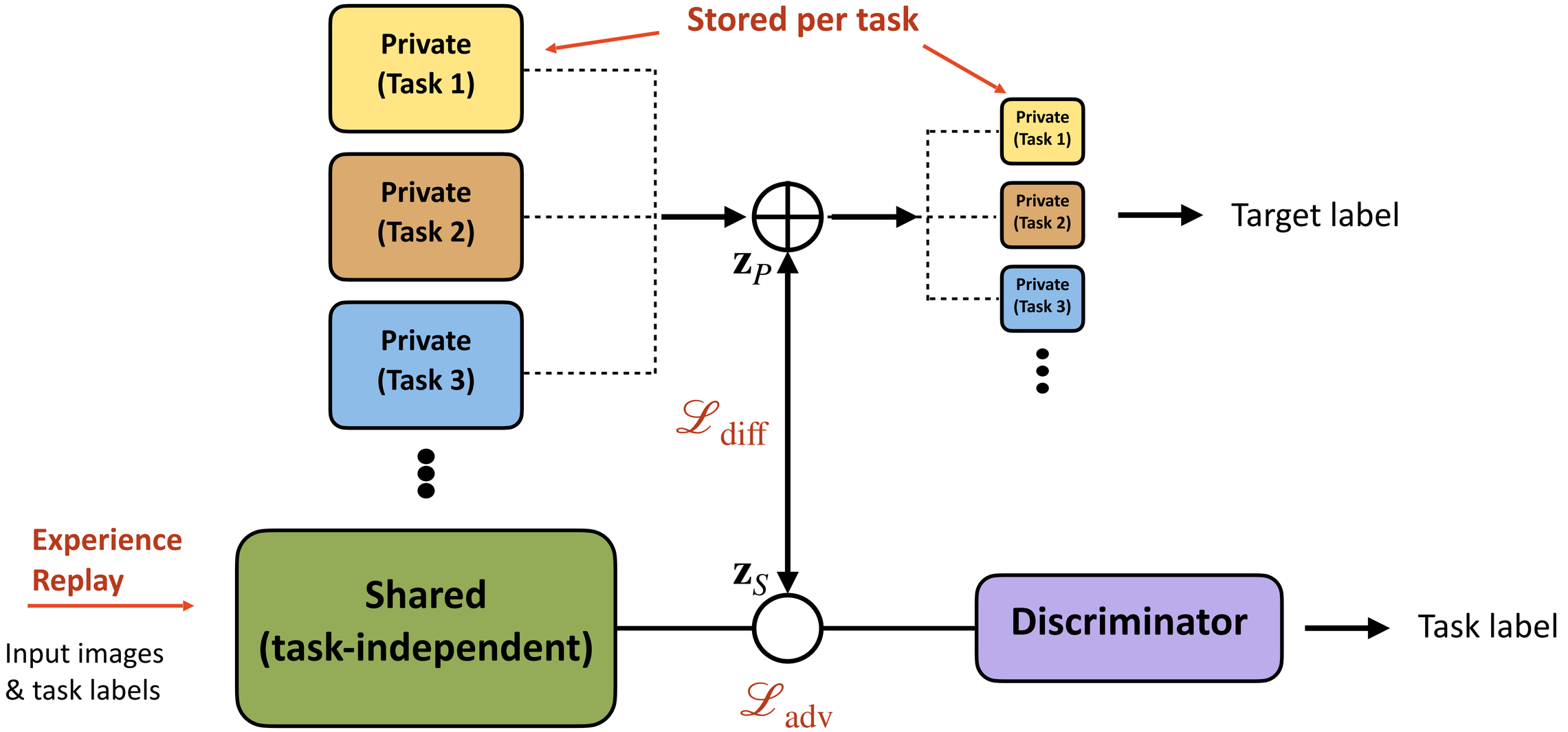
Our Approach: Adversarial Continual Learning



Our Approach: Adversarial Continual Learning



Avoiding Forgetting in ACL



Experiments

miniImageNet (20 Tasks)

CIFAR100 (20 Tasks)

Split MNIST (5 Tasks)

Permuted MNIST (10/20/30/40 Tasks)

Sequence of 5 datasets:

(SVHN, CIFAR10, MNIST, FashionMNIST, NotMNIST)

Datasets

Average Accuracy

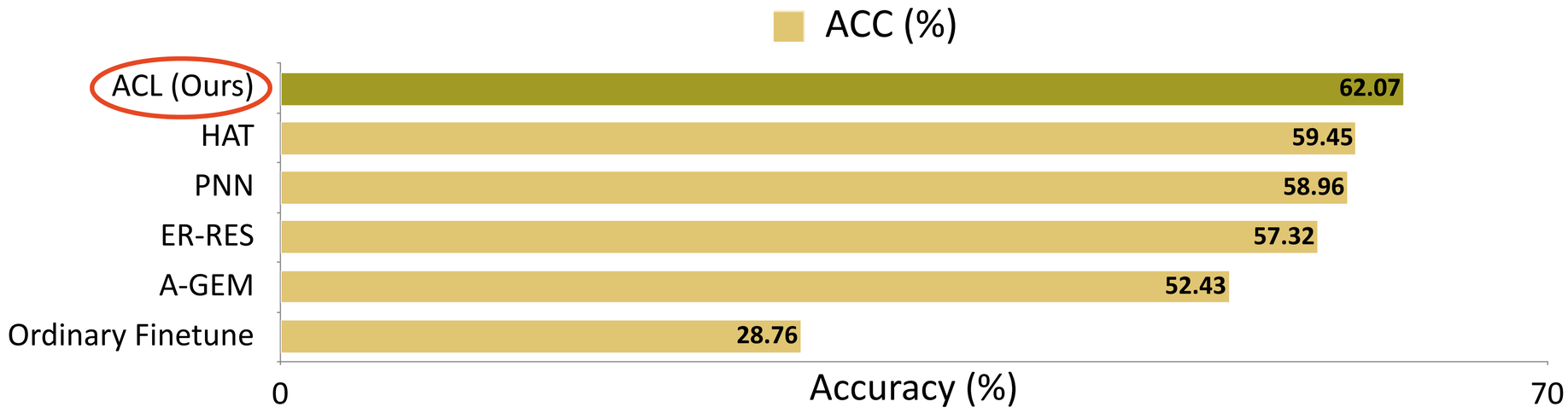
$$\mathrm{ACC} = \frac{1}{n}\sum_{i=1}^{n} R_{i,n}$$

Backward Transfer:

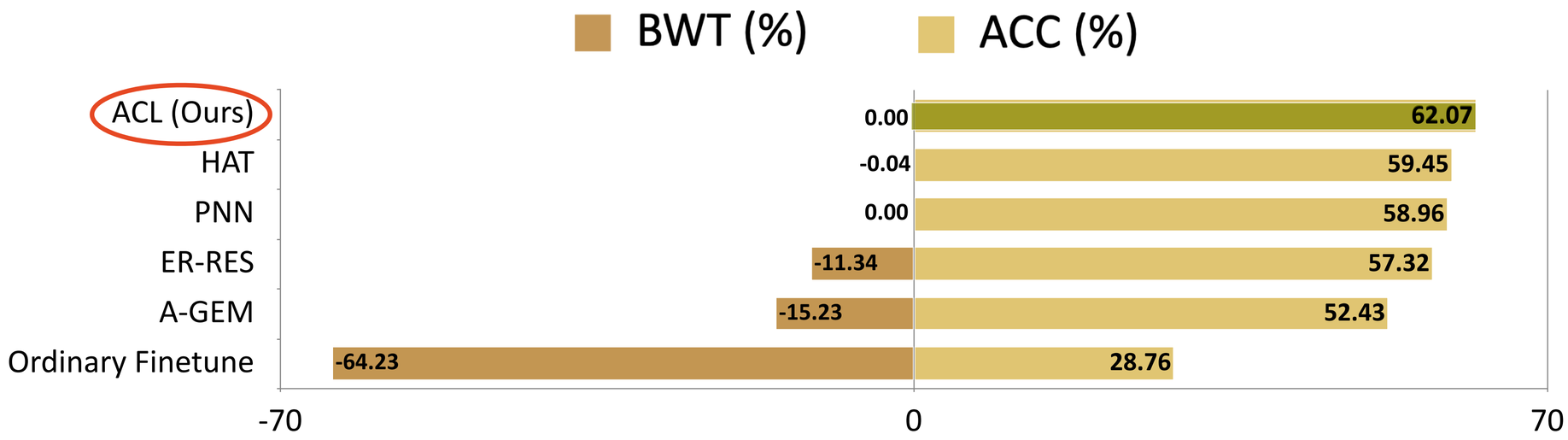
$$\mathrm{BWT} = \frac{1}{n-1}\sum_{i=1}^{n} R_{i,n} - R_{i,i}$$

Evaluation metrics

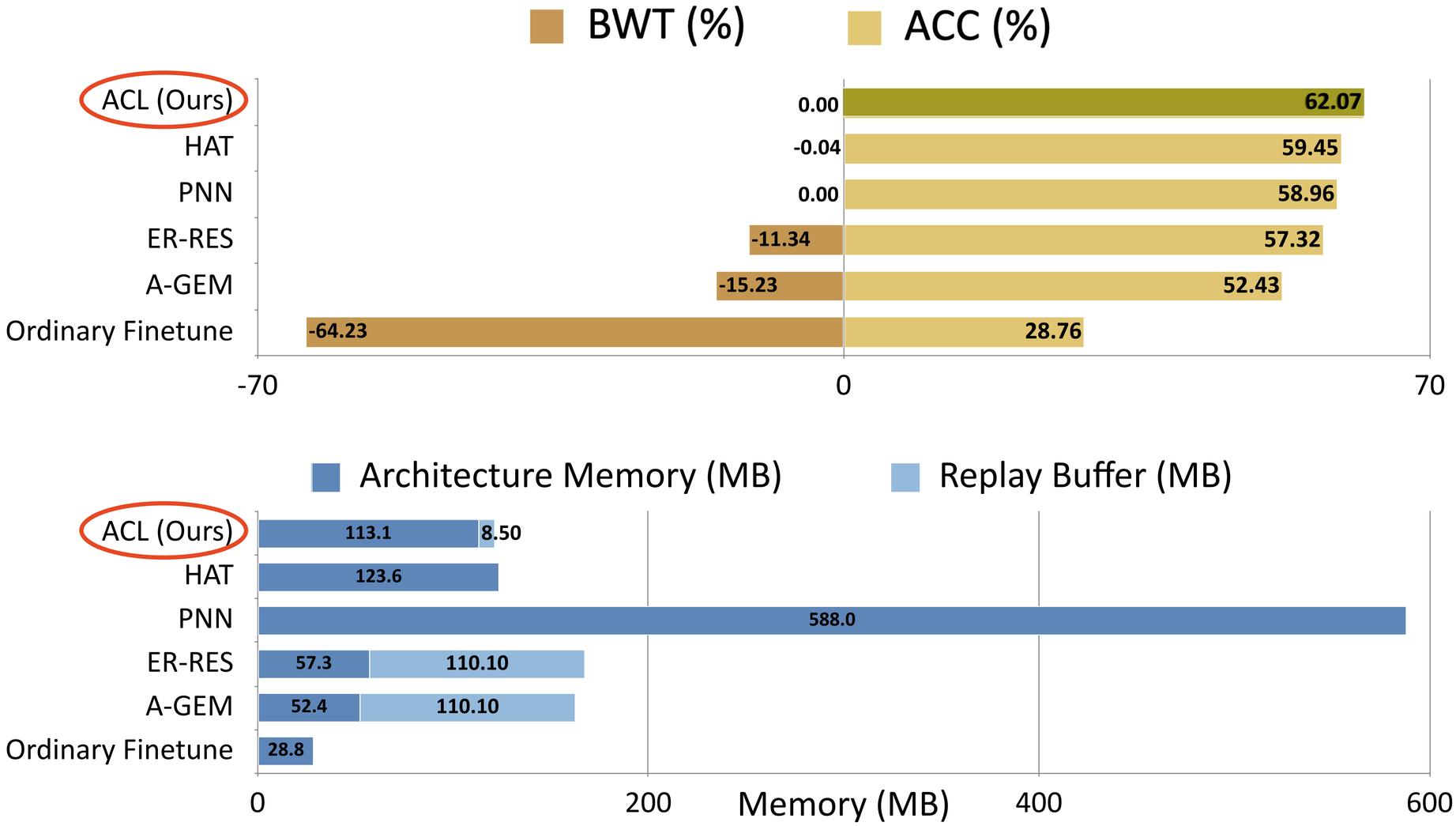
Results on 20-Split MiniImageNet



Results on 20-Split MiniImageNet



Results on 20-Split MiniImageNet

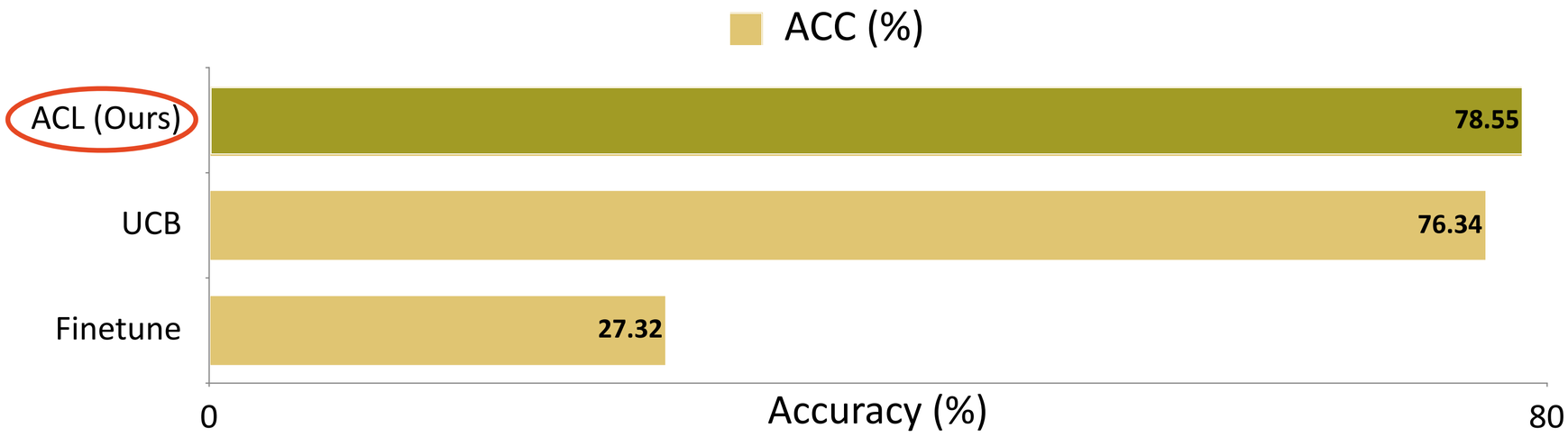


Results on Sequence of 5-Datasets

(SVHN, CIFAR10, MNIST, FashionMNIST, NotMNIST)

# Classes	Training	Test
50	212,785	48,365

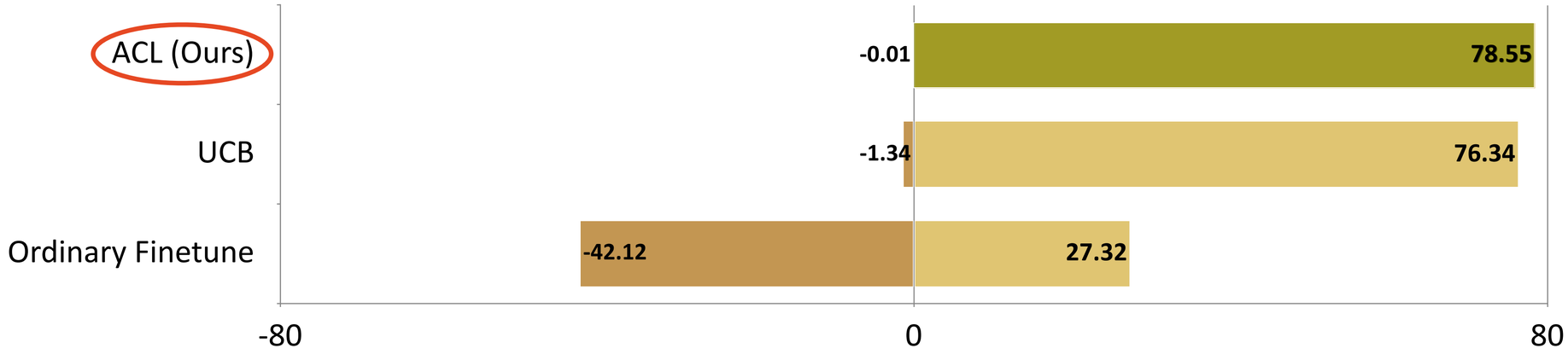
Results on Sequence of 5-Datasets

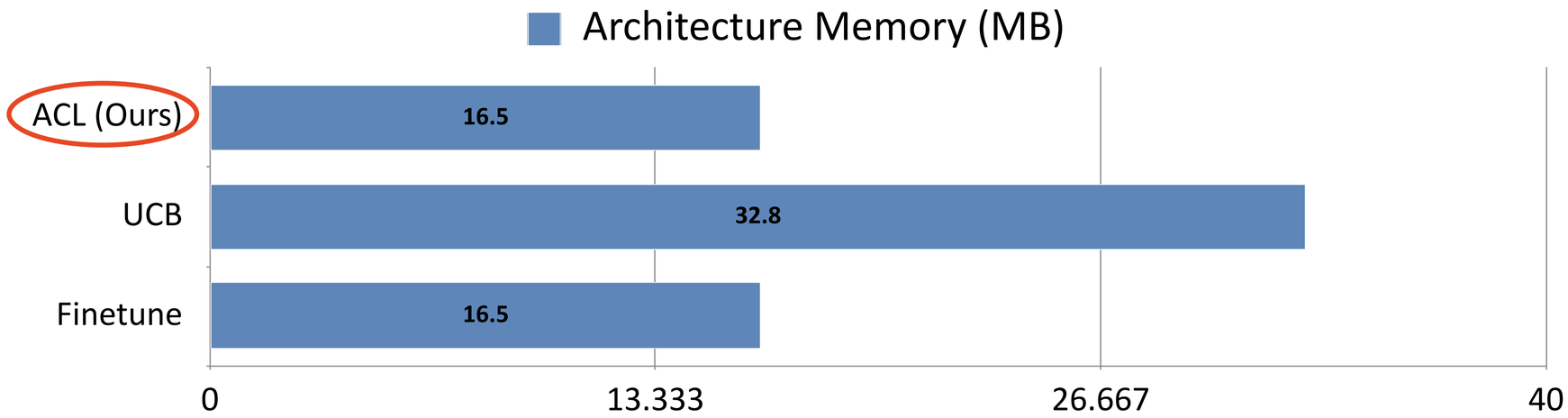


(SVHN, CIFAR10, MNIST, FashionMNIST, NotMNIST)

# Classes	Training	Test
50	212,785	48,365

Results on Sequence of 5-Datasets





Ablation Study on 20-Split miniImageNet

Discriminator	Replay buffer	$\mathscr{L}_{\text{diff}}$	ACC (%)	BWT (%)
X	X	X	62.07	0.00

Ablation Study on 20-Split miniImageNet

Discriminator	Replay buffer	$\mathscr{L}_{\text{diff}}$	ACC (%)	BWT (%)
-	X	X	52.07	-0.01
X	**X**	**X**	**62.07**	**0.00**

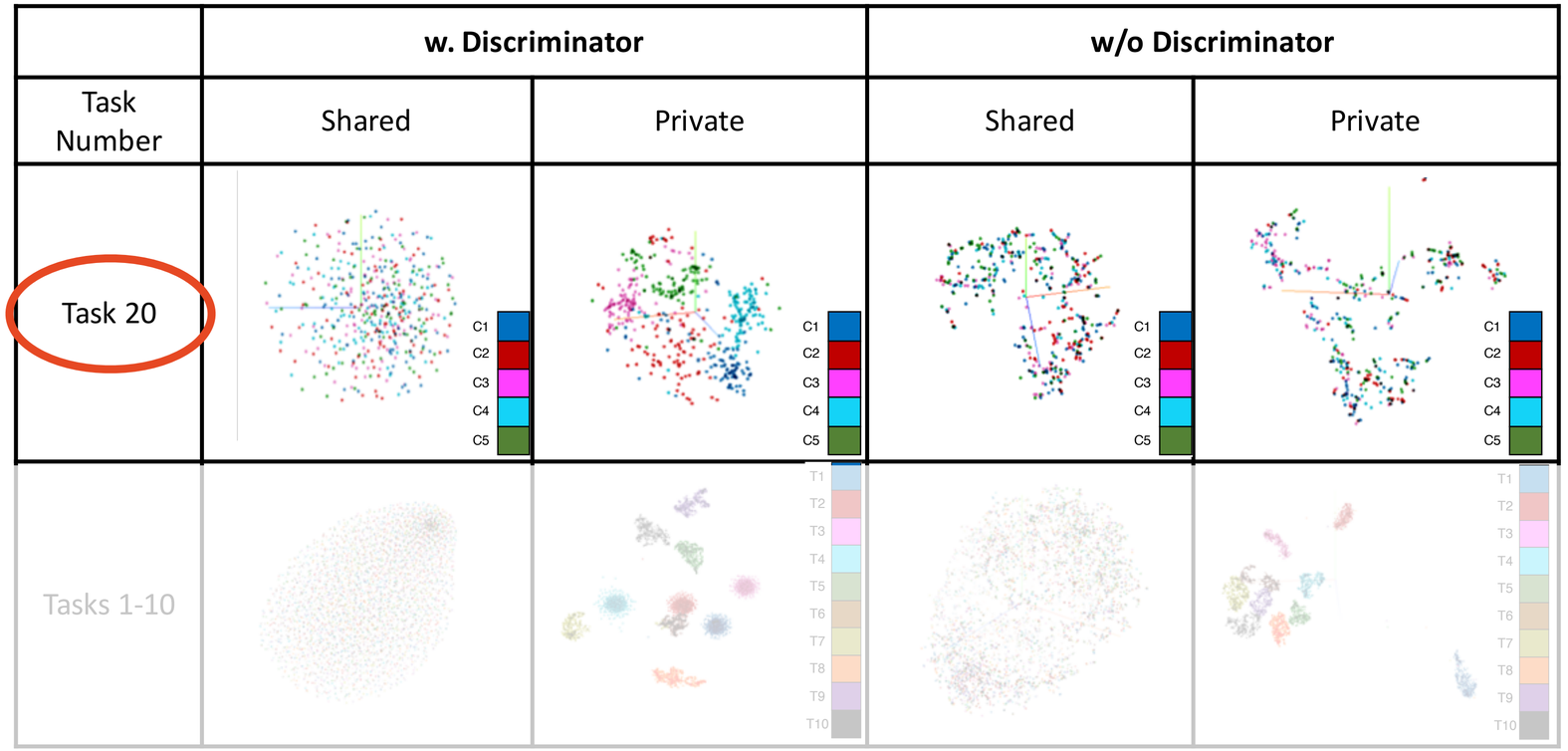
Ablation Study on 20-Split miniImageNet

Discriminator	Replay buffer	$\mathscr{L}_{\text{diff}}$	ACC (%)	BWT (%)
-	X	X	52.07	-0.01
X	-	X	57.66	-3.71
X	**X**	**X**	**62.07**	**0.00**

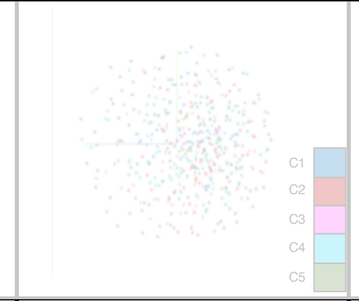
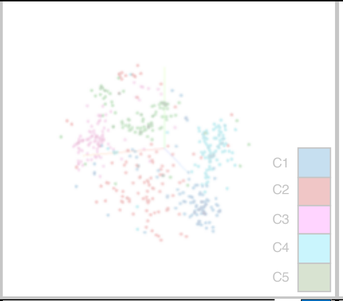
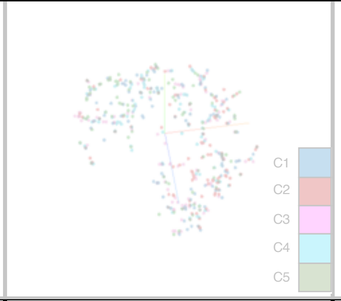
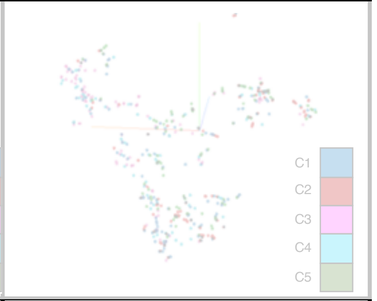
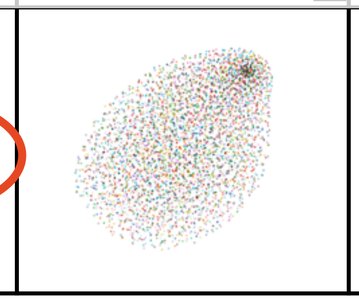
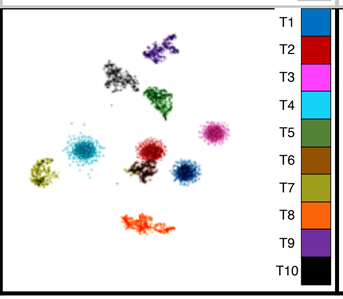
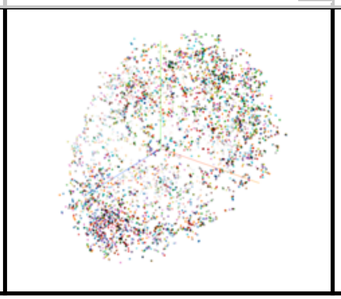
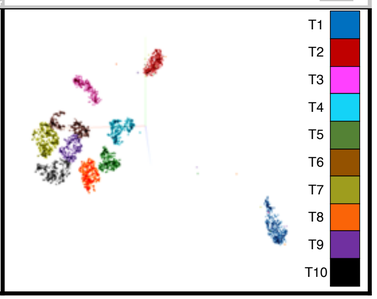
Ablation Study on 20-Split miniImageNet

Discriminator	Replay buffer	$\mathscr{L}_{\text{diff}}$	**ACC (%)**	**BWT (%)**
	X	X	52.07	-0.01
X	X		57.66	-3.71
X	X	-	60.28	0.00
X	**X**	**X**	**62.07**	**0.00**

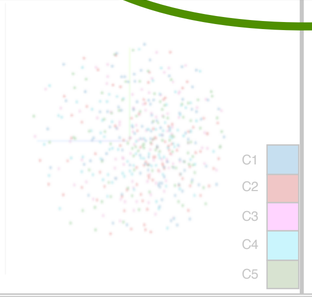
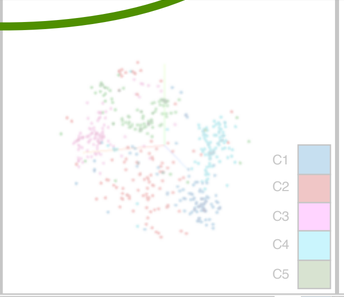
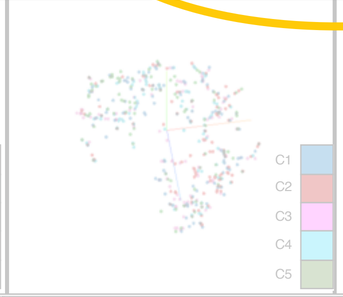
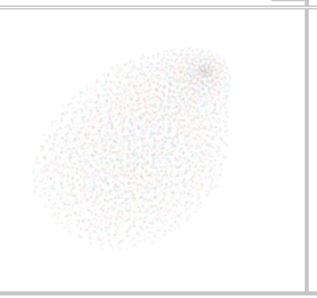
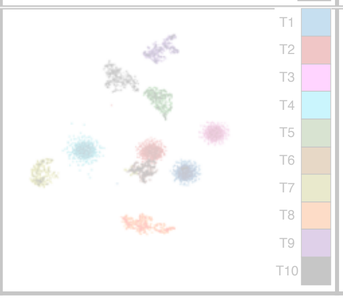
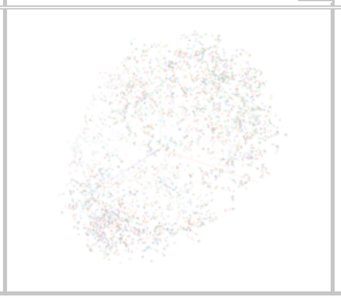
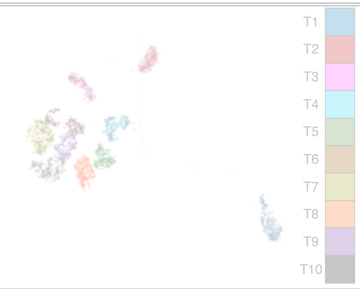
Visualizing Adversarial Learning Effect
(20-Split miniImageNet)



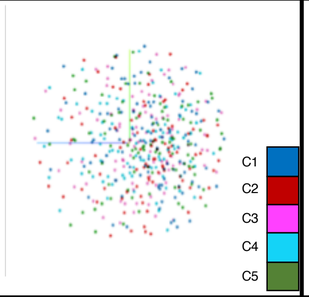
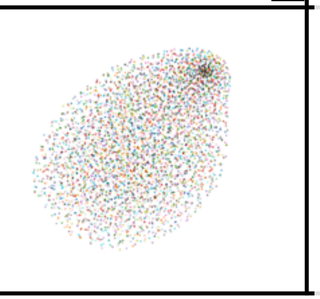
Visualizing Adversarial Learning Effect
(20-Split miniImageNet)

	w. Discriminator		w/o Discriminator	
Task Number	Shared	Private	Shared	Private
Task 20				
Tasks 1-10				

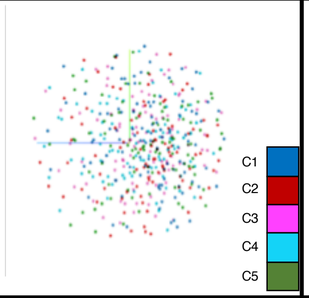
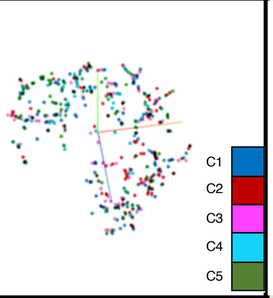
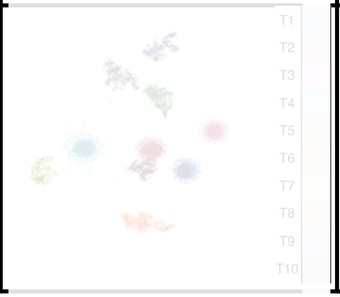
Visualizing Adversarial Learning Effect
(20-Split miniImageNet)

	w. Discriminator		w/o Discriminator	
Task Number	Shared	Private	Shared	Private
Task 20				
Tasks 1-10				

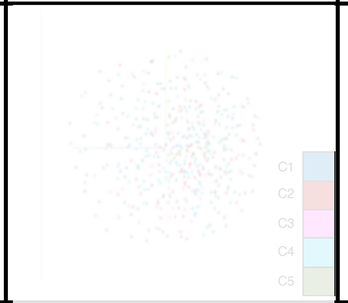
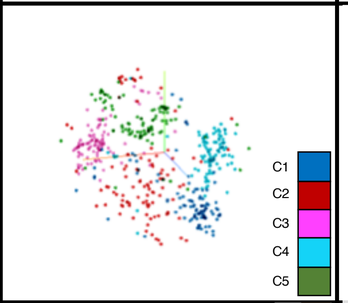
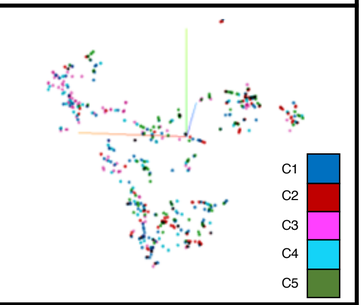
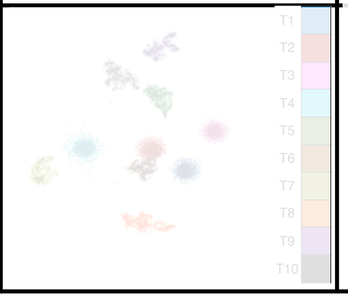
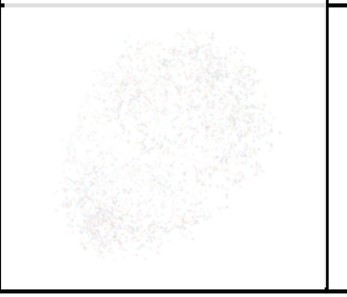
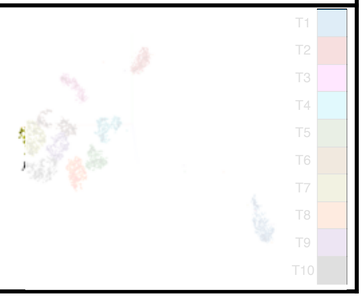
Visualizing Adversarial Learning Effect
(20-Split miniImageNet)

	w. Discriminator		w/o Discriminator	
Task Number	Shared	Private	Shared	Private
Task 20				
Tasks 1-10				

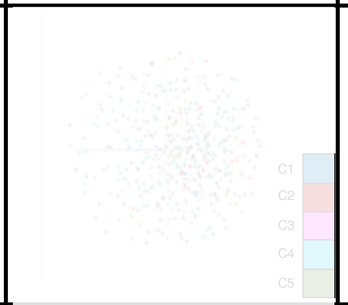
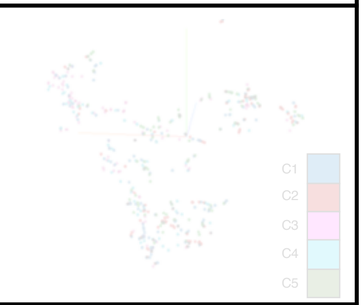
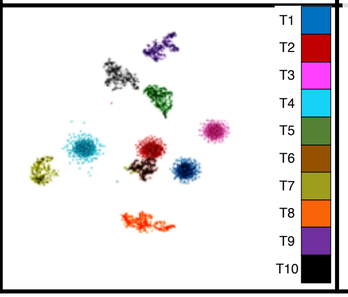
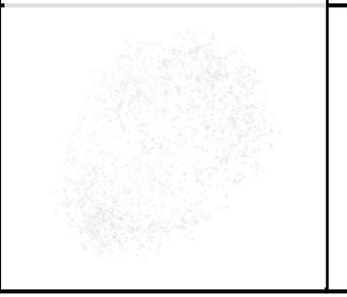
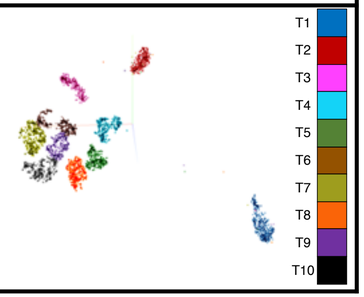
Visualizing Adversarial Learning Effect
(20-Split miniImageNet)

	w. Discriminator		w/o Discriminator	
Task Number	Shared	Private	Shared	Private
Task 20				
Tasks 1-10				

Visualizing Adversarial Learning Effect
(20-Split miniImageNet)

	w. Discriminator		w/o Discriminator	
Task Number	Shared	Private	Shared	Private
Task 20				
Tasks 1-10				

Visualizing Adversarial Learning Effect
(20-Split miniImageNet)

	w. Discriminator		w/o Discriminator	
Task Number	Shared	Private	Shared	Private
Task 20				
Tasks 1-10				

10 well-separated clusters

9-entangled clusters

Conclusion

- ACL is primarily an architecture-based method but can benefit from experience replay if need be
- Uses adversarial learning and an orthogonality constraint to disentangle task-specific and task-invariant features
- Achieves near zero forgetting and state-of-the-art accuracy on image classification benchmarks

ECCV'20 ONLINE
23-28 AUGUST 2020

Paper: https://arxiv.org/pdf/2003.09553.pdf
Code: https://github.com/facebookresearch/Adversarial-Continual-Learning

Questions: sayna@berkeley.edu

Sayna Ebrahimi
UC Berkeley

Franziska Meier
Facebook AI Research

Roberto Calandra
Facebook AI Research

Trevor Darrell
UC Berkeley

Marcus Rohrbach
Facebook AI Research

facebook Artificial Intelligence Research

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH