Methods for comparison: EWC, GEM, AGEM, MER, iCaRL, HAL, ER

The original experiment setting:

1. HAL (from paper Using Hindsight to Anchor Past Knowledge in Continual Learning)
1.1 Compared methods:
EWC, iCaRL, VCL, AGEM, MER, MIR, ER-Ring
1.2 Datasets:
Permuted MNIST (10 classes, 23 tasks, 1000 samples per task)
Rotated MNIST (10 classes, 23 tasks, 1000 samples per task)
Split Cifar100 (random 5 classes per task, 20 tasks, 250 samples per class)
Split miniImageNet (random 5 classes per task, 20 tasks, 250 samples per class)
1.3 Neural Network:
For MNISTs, 2 hidden layers with 256 ReLU neurons; for Cifar100 & ImageNet, ResNet18
1.4 Hyperparameter:
Batch size: 10, Memory size: 1-5 samples per class

2. iCaRL (from paper iCaRL: Incremental Classifier and Representation Learning)
2.1 Compared methods:
None
2.2 Datasets & Network:
Cifar100 (100 classes, 32-layer ResNet, batch size:2, 5, 10, 20, 50)
ImageNet ILSVRC (100/1000 classes, 10/100 batch size, 18-layer ResNet)

3. MER (from paper Learning to Learn without Forgetting by Maximizing Transfer and Minimizing Interference)
3.1 Compared methods:
Online, Independent, Task Input, EWC, GEM
3.2 Datasets:
MNIST Permutations (1000 samples per task, 20 tasks, 5120 memory size)
MNIST Rotations (1000 samples per task, 20 tasks, 5120 memory size)
Many Permutations (100 tasks, 200 samples per task, variant of MNIST P)
Omniglot (90% for training, 10% for testing)
3.3 Neural Network:
For MNISTs, 2 hidden layers of 100 neurons each
For Omniglot, the image scale to 28x28 and 4 modules before a fully connected softmax layer, each module contains a 3x3 convolution with 64 filters, a ReLU and 2x2 max-pooling
3.4 Hyperparameter searching:
Learning rate, regularization (EWC), memory strength (GEM), batch size (MER)

4. ER (from paper Continual Learning with Tiny Episodic Memories)
4.1 Compared methods:
Finetune, EWC, AGEM, MER
4.2 Datasets:

Permuted MNIST (23 tasks)
Split Cifar100 (20 tasks, random 5 classes per task without replacement)
Split miniImageNet (100 classes, 600 images per class, 20 tasks)
Split CUB (20 tasks)
4.3 Neural Network:
For MNISTs, 2 hidden layers with 256 ReLU neurons
For Cifar100 & ImageNet, reduced ResNet18
For CUB, standard ResNet18

Our experiment:
(1) Compared methods:
EWC, GEM, AGEM, MER, iCaRL, HAL, ER
(2) Datasets:
MNIST P, MNIST R, Split Cifar100, Split miniImageNet
(3) Network:
For MNISTs, 2 hidden layers with 256 ReLU neurons
For Cifar100 & ImageNet, reduced ResNet18