Appendix

A. Dataset Details

Miniimagenet [56] A subset dataset from ImageNet with 100 different classes, each class with 600 images. The meta train/validation/test splits are 64/16/20 classes respectively, following the same splits of [44].

Omniglot [34] An image dataset handwritten characters from 50 different alphabets, with each class of 20 examples, following the same setup and data split in [56].

CUB [61] A dataset consisting of 200 bird species. Following the same split of [16], the meta train/validation/test splits are of 100/50/50 classes respectively.

AIRCRAFT [39] An image dataset for aircraft models consisting of 102 categories, with 100 images per class. Following the split in [58], the dataset is split into 70/15/15 classes for meta- training/validation/test.

Quickdraw [28] An image dataset consisting of 50 million black-and-white drawings with 345 categories. Following [35], the dataset is split into 241/52/52 classes for meta-training/validation/test.

Necessities ***Necessities*** Logo images from the large-scale publicly available dataset Logo-2K+ [59]. The dataset is randomly split into 100/41/41 classes for meta- training/validation/test.

B. Implementation Detail

We use 750 evaluation tasks from each domain for meta testing. $m = 5$ for constructing the projected space. $\delta = 1.64$ (corresponds to confidence level of 95%) and window size $B = 100$ for domain change detection. The meta batch size (number of training tasks at each iteration) is 2.

We approximate each $||\nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T})||_2 \sim ||\nabla_{x_i} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T})||_2 = G_i$; where x_i is the pre-activation of last layer output of the network as in [29].

C. Theorem proof

Proof

Let $\mu = \mathbb{E}_{\mathbf{p}(\mathcal{T})} \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T})$

$$Tr(\mathbb{V}_{\mathbf{q}}[\Omega]) = \mathbb{E}_{\mathbf{q}(\mathcal{T})}[(\frac{\mathbf{p}(\mathcal{T})}{\mathbf{q}(\mathcal{T})} \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T}) - \mu)$$

$$(\frac{\mathbf{p}(\mathcal{T})}{\mathbf{q}(\mathcal{T})} \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T}) - \mu)^T] = \mathbb{E}_{\mathbf{q}(\mathcal{T})}[||\frac{\mathbf{p}(\mathcal{T})}{\mathbf{q}(\mathcal{T})} \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T})||_2^2] - ||\mu||_2^2 \quad (15)$$

Table 5: Effect of memory size with order 1

Algorithm	5-Way 1-Shot ACC	5-Way 5-Shots ACC
PNet-RS ($n = 100$)	36.82 ± 1.32	54.83 ± 1.12
PNet-Ours ($n = 100$)	42.35 ± 0.91	59.85 ± 0.79
PNet-RS ($n = 200$)	37.30 ± 1.21	55.23 ± 0.87
PNet-Ours ($n = 200$)	43.61 ± 0.86	61.32 ± 0.68
PNet-RS ($n = 300$)	37.76 ± 1.19	55.79 ± 0.91
PNet-Ours ($n = 300$)	44.32 ± 0.83	61.67 ± 0.57
PNet-RS ($n = 500$)	38.82 ± 1.27	55.95 ± 0.98
PNet-Ours ($n = 500$)	44.81 ± 0.63	62.08 ± 0.61

By Jensen's inequality:

$$\mathbb{E}_{\mathbf{q}(\mathcal{T})}[||\frac{\mathbf{p}(\mathcal{T})}{\mathbf{q}(\mathcal{T})} \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T})||_2^2] \geq \mathbb{E}_{\mathbf{q}(\mathcal{T})}[||\frac{\mathbf{p}(\mathcal{T})}{\mathbf{q}(\mathcal{T})} \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T})||_2]^2$$
$$= (\mathbb{E}_{\mathbf{p}(\mathcal{T})}[||\nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T})||_2])^2 \quad (16)$$

The equality holds at $\mathbf{q}^*(\mathcal{T}) = \frac{\mathbf{p}(\mathcal{T})||\nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T})||_2}{\int \mathbf{p}(\mathcal{T})||\nabla_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\theta}}(\mathcal{T})||_2}$. by plugging the above $\mathbf{q}^*(\mathcal{T})$ into the covariance expression. ■

D. Additional Results

D.1. New ordering

Order 1: Omniglot, Aircraft, Necessities, CUB, Quickdraw, MiniImagenet

To simulate imbalanced domains in streaming setting, each domain on this sequence is trained on 3000, 2000, 4000, 2000, 4000, 40000 steps respectively.

Order 2: Necessities, CUB, Omniglot, Aircraft, MiniImagenet, Quickdraw

To simulate imbalanced domains in streaming setting, each domain on this sequence is trained on 6000, 2000, 6000, 3000, 3000, 24000 steps respectively.

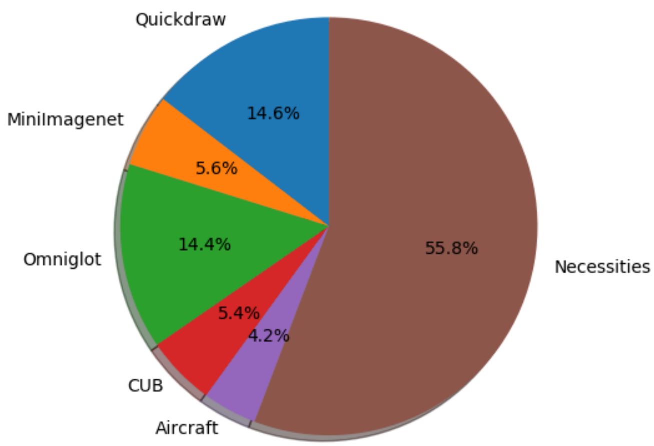
Order 1. Table 5 shows the results. **Order 2**. Table 6 shows the results.

D.2. Effect of domain revisiting

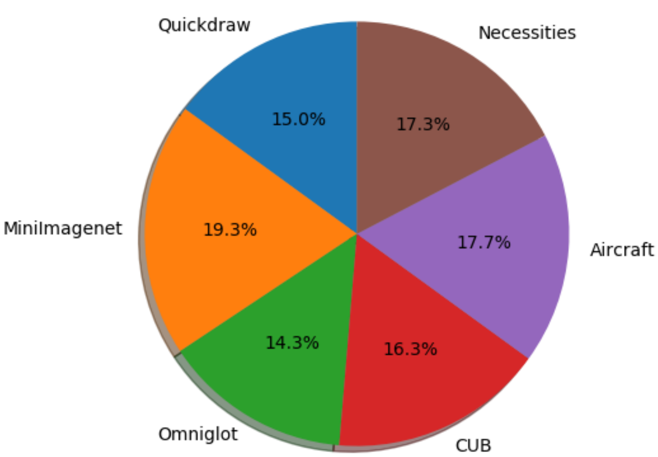
This section shows the results of effect of domain revisiting with domain ordering, **Quickdraw, MiniImagenet, Omniglot, CUB, Quickdraw, Aircraft, Necessities**. The domain Quickdraw is revisited. To simulate imbalanced domains in streaming setting, each domain on this sequence is trained on 5000, 2000, 6000, 2000, 3000, 2000, 24000 steps respectively.

Table 6: Effect of memory size with order 2

Algorithm	5-Way 1-Shot ACC	5-Way 5-Shots ACC
PNet-RS ($n = 100$)	43.08 ± 0.79	57.97 ± 0.87
PNet-Ours ($n = 100$)	46.67 ± 0.61	62.14 ± 0.50
PNet-RS ($n = 200$)	43.36 ± 0.72	58.23 ± 0.72
PNet-Ours ($n = 200$)	46.95 ± 0.52	62.83 ± 0.58
PNet-RS ($n = 300$)	44.16 ± 0.80	58.65 ± 0.79
PNet-Ours ($n = 300$)	47.64 ± 0.45	63.21 ± 0.46
PNet-RS ($n = 500$)	45.29 ± 0.82	59.36 ± 0.85
PNet-Ours ($n = 500$)	47.16 ± 0.49	63.02 ± 0.46



(a) reservoir sampling



(b) our memory management mechanism

Figure 5: Results of different domain proportion in the memory of our memory management methods and reservoir sampling when meta learning on an imbalanced task stream from three latent domains.

Table 7: Comparisons with PNet-based baselines with domain revisiting

Algorithm	5-Way 1-Shot ACC	5-Way 5-Shots ACC
PNet-Sequential	32.02 ± 0.50	49.60 ± 0.45
PNet-RS	37.31 ± 1.56	56.29 ± 1.35
PNet-Ours	40.25 ± 0.98	60.36 ± 0.83
Joint-training	52.96 ± 0.45	68.56 ± 0.57
Independent-training	58.25 ± 0.36	72.23 ± 0.29

Table 8: Comparisons with PNet-based baselines with different imbalanced ratio of each domain

Algorithm	5-Way 1-Shot ACC	5-Way 5-Shots ACC
PNet-Sequential	29.91 ± 0.71	46.97 ± 0.65
PNet-RS	34.97 ± 1.52	54.79 ± 0.69
PNet-GSS	35.65 ± 1.28	56.65 ± 0.81
PNet-AGEM	34.53 ± 1.36	54.91 ± 0.73
PNet-MIR	35.09 ± 1.29	54.56 ± 0.90
PNet-MER	35.16 ± 1.32	55.71 ± 0.78
PNet-Ours	40.57 ± 0.68	61.53 ± 0.58
Joint-training	52.96 ± 0.45	68.56 ± 0.37
Independent-training	58.25 ± 0.36	72.23 ± 0.29

D.3. Effect of different ratios of domains

D.4. Ablation Study

Effect of PETS Figure 6 shows the effect of sampling tasks with PETS.

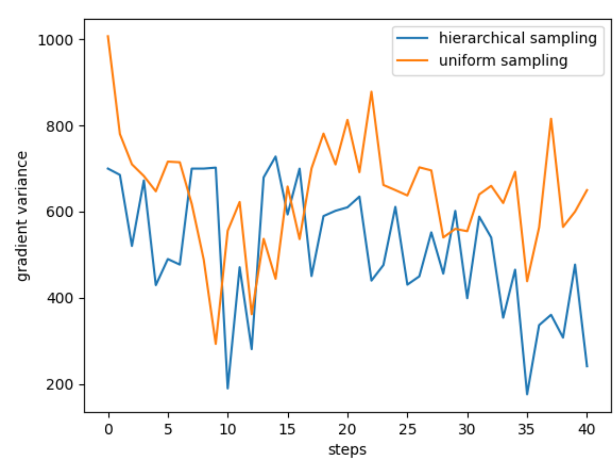


Figure 6: Gradient variance comparison between uniform sampling and PETS, each step (1000 iterations).

Table 9: Effect of Adaptive memory task sampling

Algorithm	5-Way 1-Shot ACC	5-Way 5-Shots ACC
PNet-RS	34.68 ± 1.96	53.69 ± 0.76
PNet-Ours (without PETS)	38.85 ± 0.79	57.95 ± 0.67
PNet-Ours (with PETS)	41.10 ± 0.42	60.37 ± 0.32

Effect of memory management mechanism Table 9 shows the effect of the proposed memory management mechanism.