7 Appendix

7.1 The Details of Benchmark Datasets

To achieve a comprehensive study, we conduct extensive experiments in the main paper, including datasets CIFAR-100 (Krizhevsky et al., 2009), ImageNet-Subset (Deng et al., 2009), and ImageNet-Full (Deng et al., 2009).

- CIFAR-100 is a widely-used image classification dataset, consisting of 60,000 color images with dimensions of 32×32 pixels, across 100 different classes. Each class in the dataset is designed to represent a distinct object category (*e.g.*, animals, vehicles, everyday objects). The dataset is split into a training set of 50,000 images, with 500 images per class, and a validation (or test) set of 10,000 images, with 100 images per class.
- ImageNet-subset (ImageNet-100) is a smaller, 100-class subset derived from the larger ImageNet dataset. It is frequently used for tasks like transfer learning and incremental learning, offering a balance between dataset size and complexity. Each class in ImageNet-Subset contains approximately 1,300 training images and 50 validation images, making it a more computationally manageable version of the full ImageNet dataset while still providing substantial class diversity and variability in visual content.
- ImageNet-Full (ImageNet-1000) refers to the subset of ImageNet containing 1,000 classes. It is the most commonly used version of ImageNet for tasks such as image classification, pretraining, and benchmarking deep learning models. This dataset includes around 1.2 million training images and 50,000 validation images, with approximately 50 images per class in the validation set. Each class in ImageNet-full represents a distinct object category, ranging from animals to everyday objects.

7.2 Proof Proposition 1

Proposition 2. *Consider the incremental model θ_t and oracle θ_t^*, both initialized from the old model θ_{t-1}^*, with optimization objectives defined in Eqs. 7 and 8. If θ_i and θ_i^* are searched within the neighborhood set $\bigcup_{i=1}^{t-1} \mathcal{N}_i$, where $\mathcal{N}_i = \{\theta : d(\theta, \hat{\theta}_i) < \delta_i\}$, then θ_t^* can be approximately expressed as the sum of θ_{t-1}^* and an increment vector $(\theta_t - \theta_{t-1})$ transformed by the term $(\bar{H}_{t-1} + \bar{H}_t)^{-1}\bar{H}_t$, which is shown below:*

$$\theta_t^* \approx \theta_{t-1} + (\bar{H}_{t-1} + \bar{H}_t)^{-1}\bar{H}_t(\theta_t - \theta_{t-1})$$

Proof. We begin by stating the stationarity conditions for both the incremental model θ_t and the oracle θ_t^*, which are derived from setting the derivatives of the objectives in Eqs. 7 and 8 to zero:

$$\bar{H}_{t-1}(\theta_t - \theta_{t-1}) = -\nabla\mathcal{L}_t(\theta_t), \tag{12}$$

$$\bar{H}_{t-1}(\theta_t^* - \theta_{t-1}) = -\sum_{i=1}^{t} \nabla\mathcal{L}_i(\theta_t^*), \tag{13}$$

Next, we subtract Eq. 13 from Eq. 12, yielding:

$$\bar{H}_{t-1}(\theta_t^* - \theta_t) = -\sum_{i=1}^{t-1} \nabla\mathcal{L}_i(\theta_t^*) - \left[\nabla\mathcal{L}_t(\theta_t^*) - \nabla\mathcal{L}_t(\theta_t)\right]. \tag{14}$$

To proceed, we apply a first-order Taylor approximation to approximate the difference between the gradients:

$$\nabla\mathcal{L}_t(\theta_t^*) - \nabla\mathcal{L}_t(\theta_t) = H_t(\theta_t^* - \theta_t). \tag{15}$$

Substituting Eq. 15 into Eq. 14, we obtain:

$$\bar{H}_{t-1}(\theta_t^* - \theta_t) = -\sum_{i=1}^{t-1} \nabla\mathcal{L}_i(\theta_t^*) - H_t(\theta_t^* - \theta_t). \tag{16}$$

We then move the term $H_t(\theta^t - \theta_t)$ to the left-hand side and divide the entire expression by $\bar{H}t^{-1}$:

$$\theta_t^* - \theta_t = -\bar{H}_t^{-1} \sum_{i=1}^{t-1} \nabla \mathcal{L}_i(\theta_t^*) \tag{17}$$

Now, by approximating θ_i^* for each i as in Eq. 15, we express:

$$\theta_t^* - \theta_t = -\bar{H}_t^{-1} \sum_{i=1}^{t-1} [\nabla \mathcal{L}_i(\theta_i) + H_i(\theta_t^* - \theta_i)] \tag{18}$$

Since the gradient $\nabla \mathcal{L}i$ is close to zero for the converged old model, it can be neglected in practice, leading to:

$$\theta_t^* - \theta_t \approx -\bar{H}_t^{-1} \sum_{i=1}^{t-1} H_i(\theta_t^* - \theta_i) \tag{19}$$

Assuming the parameters are searched within the neighborhood set $\bigcup_{i=1}^{t-1} \mathcal{N}_i$, where $\mathcal{N}_i = \{\theta : d(\theta, \hat{\theta}_i) < \delta_i\}$, we follow the approximation from (Huszár, 2018):

$$\sum_{i=1}^{j-1} H_i(\theta - \theta_i) \approx (\sum_{i=1}^{j-1} H_i)(\theta - \theta_{j-1}) \tag{20}$$

Substituting Eq. 20 into Eq. 19 and rearranging with respect to θ_t, we recover Eq. 9:

$$\theta_t^* \approx \theta_{t-1} + (\bar{H}_{t-1} + \bar{H}_t)^{-1} \bar{H}_t(\theta_t - \theta_{t-1}) \tag{21}$$

□

7.3 DETAILED COMPARATIVE RESULTS

For a fair comparison with subsequent work, we provide the detailed comparative results in Tab. 3 and Tab. 4.

Table 5: Classification accuracy (%) on CIFAR-100 for 5 increments.

Method	Step					
	1	2	3	4	5	6
PODNet	79.56	69.726	65.25	60.22	54.74	54.47
PODNet *w/* IVT	79.56	70.80	66.15	61.82	57.22	56.62
AFC	79.71	71.57	67.09	62.00	56.44	56.24
AFC *w/* IVT	79.71	71.74	67.13	62.54	57.90	56.62

Table 6: Classification accuracy (%) on CIFAR-100 for 10 increments.

Method	Step										
	1	2	3	4	5	6	7	8	9	10	11
PODNet	79.56	73.89	68.45	64.94	63.30	60.97	58.72	56.96	53.46	53.97	52.89
PODNet *w/* IVT	78.76	72.99	68.83	65.43	64.12	62.15	59.86	58.92	55.48	55.99	55.41
AFC	79.71	74.49	70.05	67.01	65.48	63.52	60.86	58.57	54.59	55.30	54.37
AFC *w/* IVT	79.71	74.49	69.84	66.94	65.83	63.48	61.12	59.16	56.38	56.83	55.99

Table 7: Classification accuracy (%) on CIFAR-100 for 25 increments.

Method	*Step*													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
PODNet	79.56	74.62	72.46	70.65	66.82	65.49	63.76	62.17	61.25	60.51	60.57	59.74	59.42	57.72
PODNet *w/* IVT	79.56	74.73	73.37	71.47	67.72	66.61	65.45	64.11	63.16	62.45	62.20	61.08	60.97	59.82
AFC	79.71	75.81	74.59	72.79	69.03	67.86	67.33	66.04	64.60	64.30	63.53	62.61	62.34	60.95
AFC *w/* IVT	79.71	76.06	74.64	72.90	69.45	68.63	67.93	66.66	65.30	65.08	64.12	63.13	63.21	62.16

Method	*Step*											
	15	16	17	18	19	20	21	22	23	24	25	26
PODNet	57.07	55.95	55.77	54.98	54.37	54.12	51.16	52.38	51.75	51.34	50.92	50.71
PODNet *w/* IVT	59.47	58.77	58.29	57.16	57.30	56.69	53.83	55.02	54.68	53.95	53.81	53.43
AFC	59.39	58.95	58.28	57.02	56.84	56.62	54.33	55.15	54.92	54.64	54.27	53.86
AFC *w/* IVT	60.37	60.21	59.40	58.02	58.26	57.83	55.68	56.51	56.14	55.74	55.36	54.77

Table 8: Classification accuracy (%) on ImageNet-Subset for 5 increments.

Method	*Step*					
	1	2	3	4	5	6
PODNet	84.60	78.00	72.49	70.47	65.82	63.06
PODNet *w/* IVT	84.60	78.67	73.66	71.88	67.53	65.10
AFC	83.60	80.43	77.14	74.70	70.84	70.20
AFC *w/* IVT	83.60	80.53	77.69	75.22	71.76	70.68

Table 9: Classification accuracy (%) on ImageNet-Subset for 10 increments.

Method	*Step*										
	1	2	3	4	5	6	7	8	9	10	11
PODNet	84.64	80.11	74.63	72.28	70.31	69.39	67.95	65.20	62.18	60.63	59.28
PODNet *w/* IVT	84.60	80.58	76.73	74.06	71.09	70.43	68.97	66.42	64.56	63.98	62.76
AFC	83.84	82.00	78.47	77.11	75.17	74.03	73.00	70.64	69.22	68.99	66.88
AFC *w/* IVT	83.60	83.02	78.77	76.83	75.46	75.15	73.52	71.53	69.87	69.07	67.68

Table 10: Classification accuracy (%) on ImageNet-Subset for 25 increments.

Method	*Step*													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
PODNet	84.60	72.69	69.89	68.68	65.17	64.43	63.03	61.78	60.97	59.06	57.80	58.17	58.41	58.21
PODNet *w/* IVT	84.60	80.04	78.41	77.25	74.17	73.80	72.03	71.41	69.36	67.18	64.91	66.94	65.95	64.97
AFC	83.60	80.65	80.78	80.04	78.48	75.70	75.61	73.94	72.88	72.47	72.94	71.50	72.14	70.97
AFC *w/* IVT	83.60	83.00	81.81	81.32	79.90	76.77	76.55	74.94	74.36	73.00	73.69	71.56	72.22	71.92

Method	*Step*											
	15	16	17	18	19	20	21	22	23	24	25	26
PODNet	57.85	56.63	56.10	54.07	54.63	53.93	51.80	51.39	50.68	49.83	48.88	48.04
PODNet *w/* IVT	65.05	63.90	63.95	62.24	61.12	61.50	60.24	59.26	57.91	57.04	56.49	55.64
AFC	69.82	69.37	67.76	67.74	66.47	66.41	64.67	65.26	63.66	63.02	62.65	62.36
AFC *w/* IVT	70.23	70.00	69.17	68.24	67.47	67.36	66.18	66.43	64.83	64.44	63.41	63.46

Table 11: Classification accuracy (%) on ImageNet-Full for 10 increments.

Method	*Step*										
	1	2	3	4	5	6	7	8	9	10	11
PODNet	76.83	72.85	69.68	67.20	64.72	62.87	61.10	59.52	57.96	56.80	55.57
PODNet *w/* IVT	76.91	73.16	70.43	68.04	65.60	63.79	62.23	60.97	59.48	58.25	56.95
AFC	76.82	72.02	69.21	67.06	64.91	63.16	61.32	60.18	58.74	57.71	56.86
AFC *w/* IVT	76.81	72.28	69.73	67.53	65.19	63.46	62.10	60.97	59.55	58.55	57.36