In Appendix A, we provide training details for each of our fine-tuning run. In Appendix B, we provide details on how to compute the concatenated feature vector $\Phi_\theta(\mathbf{x})$ from a ViT-B/32 CLIP model for LDIFS. In Appendix C, we provide additional results on **i)** observing how fine-tuning can lead to concept forgetting, **ii)** our investigations on why LP-init-L2SP is a better baseline compared to others in avoiding concept forgetting and **ii)** performance of our proposed LDIFS regularizer.

A TRAINING DETAILS

Training datasets and hyper-parameters: We fine-tune the CLIP ViT-B/32 model on each of the 9 image classification datasets: **a)** Stanford Cars, **b)** CIFAR-10, **c)** CIFAR-100, **d)** DTD, **e)** EuroSAT, **f)** GTSRB, **g)** MNIST, **h)** RESISC45, **i)** SVHN. For each dataset, we have a separate fine-tune run for each of the baselines discussed in §2. For each run, we train using the AdamW Loshchilov & Hutter (2017) optimizer, with an initial learning rate of $1e-5$, a weight decay of 0.1 and a cosine learning rate scheduler with a warmup length of 500. For all the runs, we use a batch-size of 128. Following the code of Ilharco et al. (2022a), we use the following number of epochs to fine-tune each dataset: **a)** Stanford Cars: 35 epochs, **b)** CIFAR-10/100: 10 epochs, **c)** DTD: 76 epochs, **d)** EuroSAT: 12 epochs, **e)** GTSRB: 11 epochs, **f)** MNIST: 10 epochs, **g)** RESISC45: 15 epochs and **h)** SVHN: 10 epochs. For each dataset, we keep a minimum of 10 epochs for fine-tuning.

Compute and number of experiments: Each of our fine-tuning run is done on a single NVIDIA A100 GPU. As there are a total of 9 classification datasets and a 8 fine-tuning baselines (including our proposed LDIFS), we perform a total of 72 fine-tune training runs from a pre-trained CLIP ViT-B/32. Furthermore, in order to produce the plots capturing $\mathcal{A}_{\text{ZS}}$, $\mathcal{A}_{\text{LP}}$, as well as ℓ_2 distance in parameter and feature space, we store intermediate checkpoints over the course of fine-tuning. Particularly, for each run, we evaluate checkpoints after every 20% of training completion. Finally, in order to obtain the full set of results in Table 6, for the 72 fine-tuned models, we evaluate each of them on 9 datasets (i.e., including the test set of the fine-tuned task), thereby having a total of 648 evaluations. Similar to training, each evaluation is also performed on a single NVIDIA A100 GPU.

Choosing λ_{LDIFS}: One hyper-parameter which the LDIFS regularizer introduces is λ_{LDIFS}. A higher value of λ_{LDIFS} encourages the model to preserve features of the original foundation model and vice versa. For each classification task, we performed a grid search over $\lambda_{\text{LDIFS}} \in \{0.01, 0.05, 0.1, 0.5, 1, 10, 100\}$ and cross-validated this hyper-parameter on a held-out validation set, choosing the value which produces the best performance on the validation set. We found $\lambda_{\text{LDIFS}} = 10$ to produce the best performance over datasets in general, so all the results we present in this paper are with λ_{LDIFS} set to 10.

B COMPUTING $\Phi_\theta(\mathbf{x})$

In this section, we discuss how we compute the concatenated feature vector $\Phi_\theta(\mathbf{x})$ given input $\mathbf{x}$ and model parameters θ, specifically for a ViT-B/32 model. This is used for computing LDIFS, our proposed regularizer for fine-tuning.

Let the feature output from layer l in the network for input $\mathbf{x}$ be $\Phi_{\theta(l)}(\mathbf{x}) \in \mathbb{R}^l$. Then the normalized feature vector for layer l can be represented as $\frac{\Phi_{\theta(l)}(\mathbf{x})}{||\Phi_{\theta(l)}(\mathbf{x})||}$. In order to form the concatenated feature vector, one can take technically take features from every intermediate layer of the network. However, storing all the features is memory intensive. Thus, we follow a similar approach to the LPIPS Zhang et al. (2018) implementation for VGG and AlexNet, and choose 5 intermediate points in the ViT-B/32 architecture to collect features from. For a single input image $\mathbf{x} \in \mathbb{R}^{3\times 224\times 224}$, each of the intermediate feature representations has a dimension of $\mathbb{R}^{50\times 768}$ with 50 tokens and 768 dimensional representation for each token. When normalizing the feature vector, we flatten this vector out to a single 38400 dimensional vector. Thus the full concatenated feature vector $\Phi_\theta(\mathbf{x})$ has dimensionality $5 \times 38400 = 192000$. However, note that there can be other ablations to this design and we leave that for future exploration.

LDIFS vs LPIPS Zhang et al. (2018): One can find similarities between our proposed LDIFS regulariser and the LPIPS metric Zhang et al. (2018) used for measuring perceptual similarity between images. However, while LPIPS uses feature space distance on a *pre-trained, frozen* model to find perceptual similarity between pairs or sets of images, LDIFS instead feature space distance between a pair of pre-trained and fine-tuned model for the same image, to preserve the input-output behaviour of the pre-trained model.

C ADDITIONAL RESULTS

In this section, we present additional empirical results to supplement our observations and conclusions in the main paper.

C.1 CRIPPLING EFFECT OF FINE-TUNING

In §3, we observed how most existing fine-tuning methods, while gaining state-of-the-art performance on fine-tuned tasks, can lead to concept forgetting in models. In this section, we provide additional empirical results to further strengthen those observations.

In Figure 7 and Figure 15, we present the $\mathcal{A}_{\mathrm{ZS}}$ and $\mathcal{A}_{\mathrm{LP}}$ accuracy for models fine-tuned on each of the 9 classification tasks, on their respective test sets. This shows expected behaviour as models broadly achieve very high test set accuracy, which increases over the course of fine-tuning, on their respective fine-tuned tasks. Note that LP-init baselines seem to have relatively lower performance on $\mathcal{A}_{\mathrm{ZS}}$ accuracy compared to their ZS-init counter-parts. However, this reduced performance is only limited to $\mathcal{A}_{\mathrm{ZS}}$ as this does not translate to a reduced performance in $\mathcal{A}_{\mathrm{LP}}$. Moreover, L2SP baselines (both ZS and LP-init) obtain relatively lower fine-tuned accuracy both in case of $\mathcal{A}_{\mathrm{ZS}}$ and $\mathcal{A}_{\mathrm{LP}}$.

In Figure 8 and Figure 16, we present the $\mathcal{A}_{\mathrm{ZS}}$ and $\mathcal{A}_{\mathrm{LP}}$ for models fine-tuned on EuroSAT (first row), GTSRB (second row) and SVHN (third row) as captured on 8 classification datasets different from their respective fine-tuned set. Broadly, the ZS and LP performance for all models drops on other datasets as they are fine-tuned, thereby capturing concept forgetting in the fine-tuned models. Among the baselines, we consistently observe LP-init-L2SP to perform better than others in avoiding concept forgetting. This is evident through the distinctly higher $\mathcal{A}_{\mathrm{ZS}}$ and $\mathcal{A}_{\mathrm{LP}}$ accuracies over fine-tuning for the LP-init-L2SP baselines.

C.2 ANALYSING L2SP

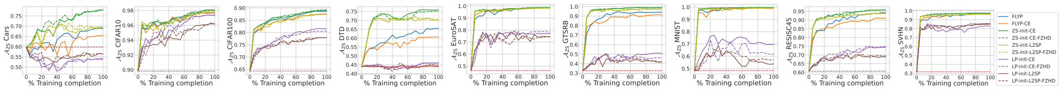
In this section, we provide additional results to supplement our observations related to investigating the L2SP baseline in §4 of the main paper. In Figure 19, we present the ℓ_2 distance in parameter space $||\theta - \theta_0||_2^2$ between the pre-trained and current models captured over fine-tuning. For all the datasets, we observe that L2SP baselines while initially diverging slightly in the parameter space, converge back to a model having low ℓ_2 parameter space distance from the original foundation model. Other baselines on the other hand, completely diverge away from the original model in the parameter space.

To further investigate the change in input-output behaviour of the model over fine-tuning, we measure the distance in feature space (see Equation (3)) over fine-tuning. In Figure 20, we present the ℓ_2 distance in feature space captured for models fine-tuned on EuroSAT. Again, consistent with our previous observation, we find that unlike other baselines, L2SP first diverges away from the original foundation model and then converges back to the original input-output behaviour, as is indicative through a decreasing L2 feature space distance in the later stages of fine-tuning. This observation is consistent on the EuroSAT train set, the EuroSAT test set as well as on other datasets, thereby providing our motivation for the LDIFS regularizer.

C.3 PERFORMANCE OF LDIFS

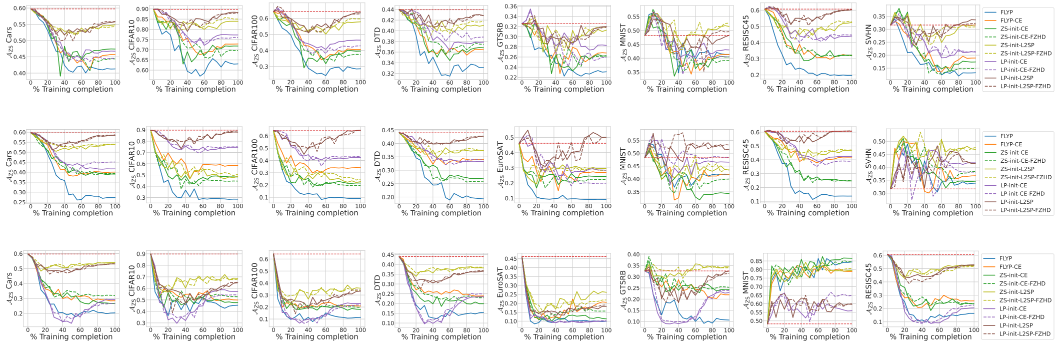
In this section, we provide additional results to supplement observations related to the performance of the proposed LDIFS regularizer.

Analysing LDIFS on parameter space and feature space distance: In Figure 9 and Figure 10, we plot the ℓ_2 distance in parameter/weight and feature space respectively (same as Figure 3 and Figure 4 in the main paper), but with the LDIFS baselines added in. In the weight space, LDIFS, while lower than other fine-tuning methods, has a relatively higher ℓ_2 distance from the pre-trained model compared to L2SP. On the contrary, in the feature space, LDIFS consistently gets the lowest distance from the pre-trained model. This is expected and was indeed the purpose behind LDIFS's



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 7: **Test set ZS Accuracy $\mathcal{A}_{\text{ZS}}$ for different fine-tuning methods** on 9 image classification tasks: (a) Stanford Cars, (b) CIFAR-10, (c) CIFAR-100, (d) DTD, (e) EuroSAT, (f) GTSRB, (g) MNIST (h) RESISC45 and (i) SVHN. $\mathcal{A}_{\text{ZS}}$ generally rises over the course of fine-tuning. However, for the ZS-init-L2SP and LP-init-L2SP baselines, the gain in $\mathcal{A}_{\text{ZS}}$ is relatively lower. Furthermore, for LP-init baselines, the performance is consistently lower compared to other baselines



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 8: **ZS Accuracy $\mathcal{A}_{\text{ZS}}$ for models fine-tuned on EuroSAT (first row), GTSRB (second row), and SVHN (third row) and evaluated on 8 different datasets different from their fine-tuning dataset.** Most fine-tuning methods show a drop in $\mathcal{A}_{\text{ZS}}$ performance over the course of fine-tuning indicating a reduction in the model's transferability.

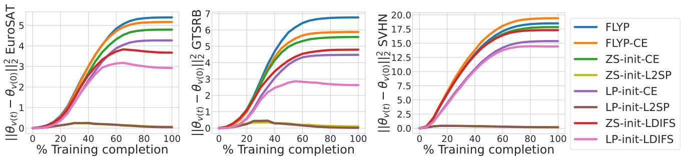


Figure 9: ℓ_2 **distance in weight space** $||\theta_{v(t)} - \theta_{v(0)}||_2^2$ between the image encoder fine-tuned to the current time-step $f_{\theta_{v(t)}}$ and the pre-trained image encoder $f_{\theta_{v(0)}}$ over the course of fine-tuning.

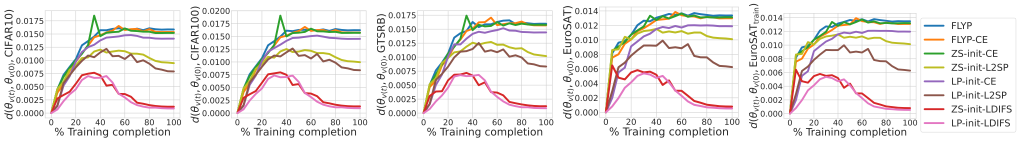


Figure 10: ℓ_2 **distance in feature space** $d(\theta_{v(t)}, \theta_{v(0)}, \mathcal{D})$, between image encoders, $f_{\theta_{v(t)}}$ and $f_{\theta_{v(0)}}$ computed over different fine-tuning methods for models fine-tuned on EuroSAT.

design which is to minimize the difference in input-output behaviour, captured through feature space distance between the fine-tuned and pre-trained models and not necessarily constrain the parameter space of the fine-tuned model to lie close to the pre-trained model. Furthermore, note from Figure 10, that although we are applying the LDIFS regularizer only on the EuroSAT samples during fine-tuning, the preservation of features extends even to other datasets like CIFAR10, CIFAR100 and GTSRB, as is indicated by the low feature space distance even on these datasets.

Results and observations: Firstly, in Figure 13 and Figure 17, we present the test set accuracy of all fine-tuning methods including LDIFS on all 9 classification tasks. We see that LDIFS performs competitively with all other baselines and improves on L2SP consistently. In Figure 14 and Figure 18, we present the $\mathcal{A}_{\text{ZS}}$ and $\mathcal{A}_{\text{LP}}$ accuracy respectively, for models fine-tuned on EuroSAT (first row),

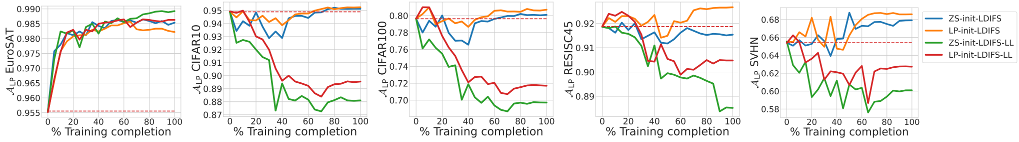


Figure 11: $\mathcal{A}_{\text{LP}}$ **for models trained with LDIFS on EuroSAT using the full concatenated feature vector vs just the last layer (LL) feature vectors.** Full feature vectors with earlier features is crucial for LDIFS's performance.

GTSRB (second row) and SVHN (third row) over the course of fine-tuning. Furthermore, we also provide the Δ_{ZS}, Δ_{LP} and the $\mathcal{A}_{\text{LP}}$ values for fully fine-tuned models on all 9 classification tasks and all baselines in Table 6.

As is crystal clear from our results, LP-init-LDIFS consistently minimizes concept forgetting without compromising on performance on the fine-tuned task. This is evident from its high $\mathcal{A}_{\text{ZS}}$, $\mathcal{A}_{\text{LP}}$ (see Figure 14 and Figure 18) and high Δ_{ZS} and Δ_{LP} (see Table 6) when evaluated on tasks other than the fine-tuning task at hand. Note from Table 6 that even when LP-init-LDIFS is not achieving the best performance on a certain dataset pair, it is often a close second with very little difference compared to the best performing baseline. In addition to this advantage, its $\mathcal{A}_{\text{LP}}$ accuracy on the fine-tuned task itself is very competitive with the top scores obtained by other fine-tuning baselines and provides a consistent improvement on L2SP.

Finally, in Table 7, we present the full results of the 2 task sequence setup for continual fine-tuning. Even in this setting, we find LP-initLDIFS to not only preserve performance of the first fine-tuning task during the second fine-tuning stage but also to preserve performance of all other tasks during the entire sequence of end-to-end fine-tuning.

Ablation with last layer features: In Figure 11, we present an ablation where we fine-tune on EuroSAT using LDIFS but using just the last layer features as opposed to the full concatenated feature vector. Clearly, including earlier features when computing the LDIFS regularizer is crucial to its performance as preserving just the last layer features leads to a significant performance drop compared to using the full feature vector.

C.4 COMPARISON WITH ADAPTERS AND PROMPT TUNING

In Table 5, we present results of fine-tuning a CLIP ViT-B/32 model using linear probing, CoOp, CLIP-Adapter, Tip-Adapter and full end-to-end fine-tuning using ZS-init-CE loss on 9 downstream image classification tasks. We consistently observe end-to-end fine-tuning to produce the best test set accuracy across all 9 tasks. Furthermore, we also note that in case of tasks where the gap between linear probe performance and end-to-end fine-tuning performance is significant (e.g. SVHN), adapter and prompt tuning approaches don't cover this performance gap. Hence, if the pre-trained encoder is not performant on a downstream task, the only way to achieve state-of-the-art results on the task requires end-to-end fine-tuning the model on the task itself. This observation thus necessitates study into better end-to-end fine-tuning methods for foundation models.

C.5 COMPARISON WITH WISE-FT

In this section, we perform an ablation to see the effect of Wise-FT Wortsman et al. (2022b) on preventing concept forgetting. Wise-FT is a weight interpolation method which forms a linear combination between the pre-trained θ_0 and fine-tuned θ_f model parameters:

$$\theta_{\text{wse}} = \alpha\theta_0 + (1 - \alpha)\theta_f \tag{6}$$

Since this can work with any fine-tuned model, we can apply it on all the end-to-end fine-tuning baselines in this work. Specifically, we tune the hyperparameter α on a held-out validation set of the downstream task at hand to maximum validation accuracy. We evaluate concept forgetting on 3 downstream tasks: CIFAR-10, EuroSAT and SVHN using the Δ_{LP} metric and report the results in Table 4. For each dataset, we evaluate the mean Δ_{LP} on 5 other downstream tasks: Cars, DTD, GTSRB, RESISC45 and MNIST. Our observations show a consistent reduction in concept forgetting across all fine-tuning baselines. However, the order of performance between the fine-tuning baselines does not change and LDIFS still maintains superior performance over other baselines both pre and

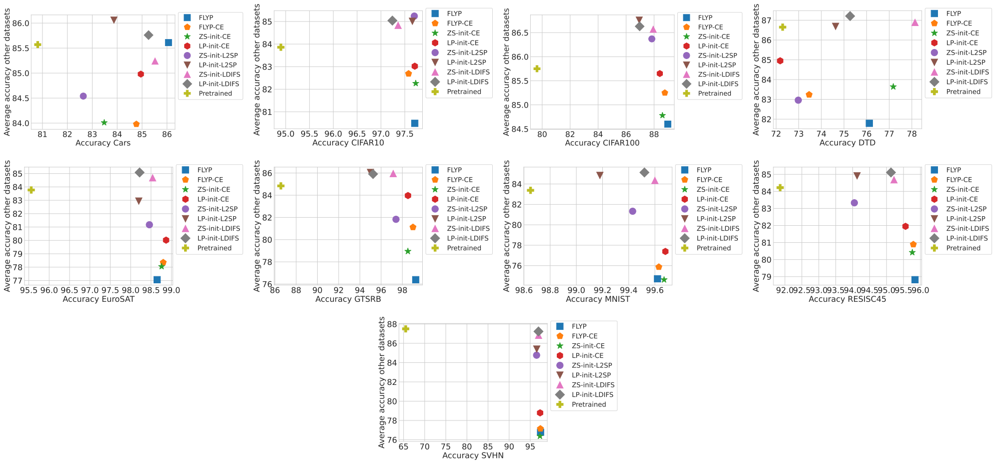


Figure 12: **Visualizing test LP accuracy of fine-tuning baselines on different datasets.** For each dataset, we plot its test set LP accuracy on the x-axis and the average test set LP accuracy of the remaining datasets on the y-axis.

Dataset	Fine-tuning baselines															
	FLYP		FLYP-CE		ZS-init-CE		LP-init-CE (LP-FT)		ZS-init-L2SP		LP-init-L2SP		ZS-init-LDIFS (Ours)		LP-init-LDIFS (Ours)	
	Δ_{LP}	$\Delta_{\text{LP}}(+\text{wse})$	Δ_{LP}	$\Delta_{\text{LP}}(+\text{wse})$	Δ_{LP}	$\Delta_{\text{LP}}(+\text{wse})$	Δ_{LP}	$\Delta_{\text{LP}}(+\text{wse})$	Δ_{LP}	$\Delta_{\text{LP}}(+\text{wse})$	Δ_{LP}	$\Delta_{\text{LP}}(+\text{wse})$	Δ_{LP}	$\Delta_{\text{LP}}(+\text{wse})$	Δ_{LP}	$\Delta_{\text{LP}}(+\text{wse})$
CIFAR10	-5.14	-0.11	-1.33	-0.16	-1.67	-0.1	-1.49	-0.01	1.58	1.64	1.49	1.96	1.53	1.99	**1.72**	**2.01**
EuroSAT	-6.5	-1.32	-5.53	-1.04	-5.62	-1.23	-4.08	-0.76	-1.77	-0.12	-0.87	0.08	0.83	1.2	**1.24**	**1.96**
SVHN	-10.82	-6.24	-10.18	-6.37	-10.98	-7.2	-9.2	-4.11	-2.65	-0.66	-2.13	-0.23	-0.57	0.12	$\mathbf{-0.18}$	**0.33**

Table 4: Δ_{LP} without and with Wise-FT Wortsman et al. (2022b) for all fine-tuning baselines on CIFAR-10, EuroSAT and SVHN.

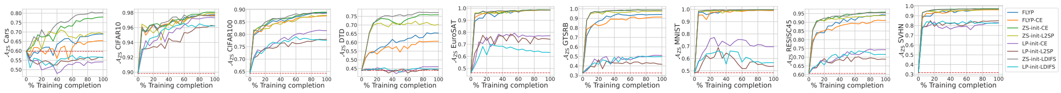


Figure 13: **Comparing test set ZS Accuracy $\mathcal{A}_{\text{ZS}}$ for different fine-tuning methods with LDIFS** on 9 image classification tasks. While LP-init baselines generally underperform on $\mathcal{A}_{\text{ZS}}$, we find ZS-init-LDIFS to be competitive with the best baselines.

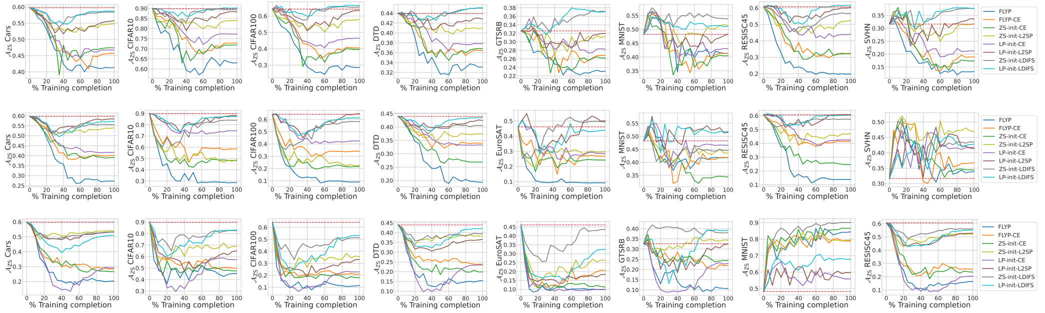
post application of Wise-FT. Thus, the results show that Wise-FT can be combined with any E2E fine-tuning method to further minimize concept forgetting.

C.6 VISUALIZING TABLE 1

In this section, we present the results of Table 1 using an alternate visualization. For each of the 9 downstream datasets, we plot the LP accuracy on the dataset itself on the x-axis and the average LP accuracy on all other datasets on the y-axis. The best baselines should lie on the top right corner of this plot as that indicates that as the model is fine-tuned on the dataset, it still retains performance on other datasets. Results are in Figure 12. Clearly, we consistently see the LDIFS baselines to lie at the top right, again indicative of their performance in not just minimizing concept forgetting but also obtaining excellent downstream task performance.

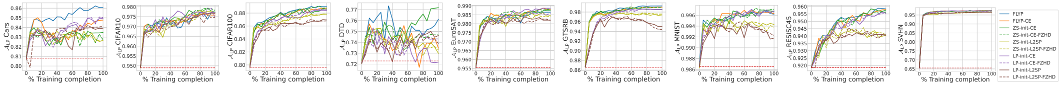
C.7 GENERALITY OF CONCEPT FORGETTING BEYOND CLIP VIT-B/32

To study the relevance of concept forgetting beyond a CLIP ViT model, in Table 8 and Table 9, we apply the end-to-end fine-tuning methods to fine-tune a CLIP RN-50 model and a FLAVA Singh et al. (2022) ViT-B/16 model respectively on EuroSAT. From both these tables, we observe both concept forgetting to exist and LP-init-LDIFS to minimize it, thereby validating our assumptions on models beyond a CLIP ViT model.



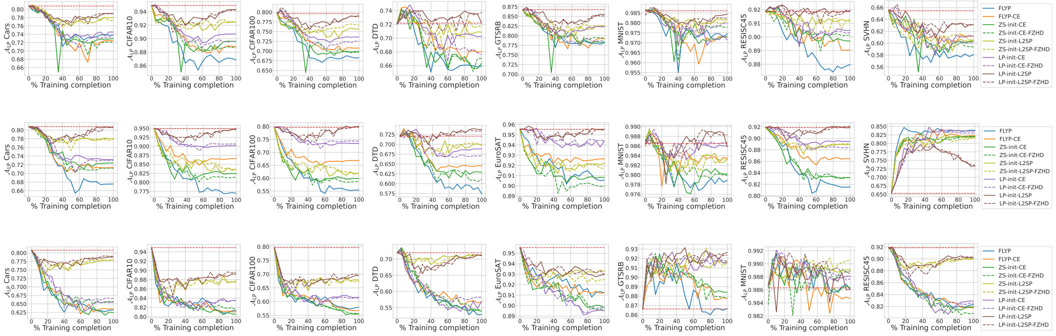
(a) (b) (c) (d) (e) (f) (g) (h)

Figure 14: **Comparing ZS Accuracy $\mathcal{A}_{\text{ZS}}$ of different fine-tuning methods with LDIFS for models fine-tuned on EuroSAT (first row), GTSRB (second row) and SVHN (third row) and evaluated on 8 datasets different from their fine-tuning dataset.** LP-init-LDIFS almost consistently beats all other baselines including LP-init-L2SP in preserving the model's transferability.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 15: **Test set LP Accuracy $\mathcal{A}_{\text{LP}}$ for different fine-tuning methods** on 9 image classification tasks: (a) Stanford Cars, (b) CIFAR-10, (c) CIFAR-100, (d) DTD, (e) EuroSAT, (f) GTSRB, (g) MNIST (h) RESISC45 and (i) SVHN. $\mathcal{A}_{\text{LP}}$ generally rises over the course of fine-tuning. However, for the ZS-init-L2SP and LP-init-L2SP baselines, the gain in $\mathcal{A}_{\text{LP}}$ is relatively lower. Furthermore, for LP-init baselines, the performance is consistently lower compared to other baselines



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 16: **LP Accuracy $\mathcal{A}_{\text{LP}}$ for models fine-tuned on EuroSAT (first row), GTSRB (second row), and SVHN (third row) and evaluated on 8 different datasets different from their fine-tuning dataset.** Most fine-tuning methods show a drop in $\mathcal{A}_{\text{LP}}$ performance over the course of fine-tuning indicating a reduction in the model's transferability.

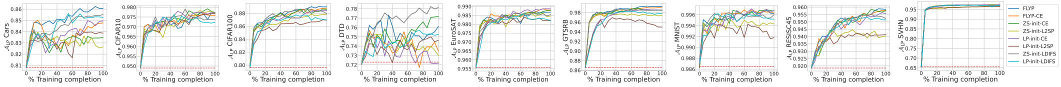
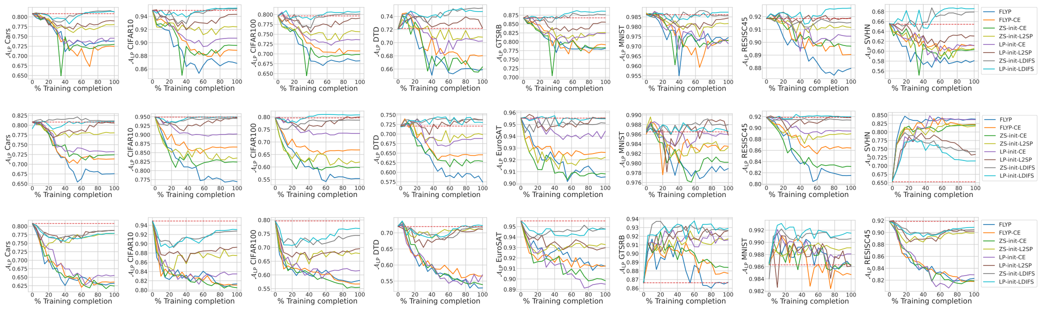


Figure 17: **Comparing test set LP Accuracy $\mathcal{A}_{\text{LP}}$ for different fine-tuning methods with LDIFS** on 9 image classification tasks.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 18: **Comparing LP Accuracy $\mathcal{A}_{\mathrm{LP}}$ of different fine-tuning methods with LDIFS for models fine-tuned on EuroSAT (first row), GTSRB (second row) and SVHN (third row) and evaluated on 8 datasets different from their fine-tuning dataset.** LP-init-LDIFS almost consistently beats all other baselines including LP-init-L2SP in preserving the model's transferability.

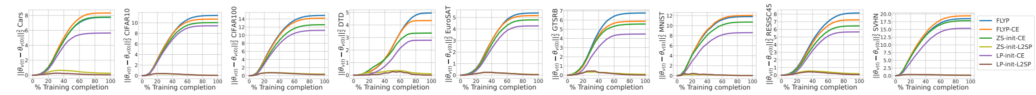


Figure 19: ℓ_2 **distance in weight space** $||\theta - \theta_0||_2^2$ between pre-trained image encoder f_{θ_0} and fine-tuned image encoder f_θ over the course of fine-tuning. Apart from ZS-init-L2SP and LP-init-L2SP, all fine-tuning baselines diverge in weight space over the course of fine-tuning.

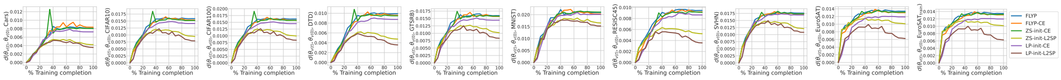


Figure 20: ℓ_2 **distance in feature space** $d(\theta, \theta_0, \mathcal{D})$, between fine-tuned and pre-trained image encoders, f_θ and f_{θ_0} computed over different fine-tuning methods for models fine-tuned on EuroSAT.

Dataset	**LP**	**CoOp**	**CLIP-Adapter**	**Tip-Adapter**	**E2E Fine-tuning**
Cars	80.80	80.88	81.24	81.32	**83.48**
CIFAR10	94.92	95.02	94.87	95.13	**97.73**
CIFAR100	79.62	80.12	80.06	80.82	**88.6**
DTD	72.08	72.01	71.87	72.62	**77.18**
EuroSAT	95.56	95.14	95.38	96.23	**98.76**
GTSRB	86.70	86.81	87.45	88.04	**98.52**
MNIST	98.65	98.84	99.03	99.11	**99.67**
RESISC45	91.86	91.79	91.82	91.80	**95.76**
SVHN	65.47	67.28	69.76	69.23	**97.3**

Table 5: Test set accuracy obtained from linear probing (LP), CoOp, CLIP-Adapter, Tip-Adapter and end-to-end (E2E) fine-tuning using ZS-init-CE loss.

Dataset		Fine-tuning baselines																							
Fine-tune	Eval	FLYP			FLYP-CE			ZS-init-CE			LP-init-CE (LP-FT)			ZS-init-L2SP			LP-init-L2SP			ZS-init-LDIFS (Ours)			LP-init-LDIFS (Ours)		
		$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$
Cars	CIFAR10	−3.55	−0.68	94.24	−9.0	−1.74	93.18	−9.81	−1.52	93.4	−10.42	−0.93	94.0	−2.78	−0.89	94.02	**−0.26**	**0.06**	**94.97**	−2.98	−0.47	94.45	−2.85	−0.22	94.69
	CIFAR100	−6.91	−0.76	78.89	−7.41	−2.74	76.91	−8.62	−3.58	76.04	−6.84	−2.12	77.51	−3.4	−2.28	77.34	**0.22**	**0.17**	**79.81**	−1.6	−0.37	79.26	−1.76	0.07	79.71
	DTD	−2.61	1.28	73.46	−4.2	−1.44	70.9	−3.99	0.53	72.5	−3.99	2.13	74.26	−3.09	−0.64	71.44	**−1.17**	**2.39**	**74.63**	−1.54	−0.32	71.76	−2.34	0.43	72.55
	EuroSAT	−16.94	−0.2	95.33	−5.89	−0.54	95.0	−6.69	−0.8	94.74	2.74	−0.02	95.52	−4.8	−0.56	94.98	−2.46	**0.02**	**95.56**	1.0	−0.13	95.41	**4.41**	**0.02**	**95.56**
	GTSRB	−3.62	0.55	87.09	−4.28	−1.39	85.24	−1.1	−0.66	85.96	−3.58	−1.01	85.62	2.19	−0.4	86.21	1.85	**0.64**	**87.21**	**2.36**	−0.02	86.56	0.74	0.44	87.18
	MNIST	4.28	−0.19	98.56	2.26	−0.52	98.13	**4.29**	−0.43	98.24	2.21	−0.14	98.48	1.46	−0.34	98.33	2.45	−0.07	98.59	2.33	0.01	98.71	2.15	**0.06**	**98.62**
	RESISC45	−6.19	−1.21	90.67	−7.16	−1.57	90.3	−10.7	−2.13	89.75	−4.81	−0.87	91.0	−4.48	−1.4	90.48	**−0.21**	−0.25	91.62	−4.13	−0.48	91.4	−2.7	**−0.13**	**91.75**
	SVHN	0.33	**1.14**	**66.63**	−5.4	−3.38	62.15	−1.63	−3.91	61.47	−0.7	−2.09	63.49	**2.93**	−2.01	63.51	1.54	0.78	66.08	−0.05	−1.13	64.35	1.03	0.53	66.01
	Cars	9.2	**5.25**	**86.06**	5.73	3.95	84.77	18.24	2.67	83.48	−5.62	4.14	84.95	9.44	1.83	82.64	−3.31	3.06	83.87	**20.52**	4.71	85.52	−3.15	4.45	85.26
CIFAR10	Cars	−58.0	−14.58	66.24	−45.32	−7.04	73.75	−46.08	−5.68	75.12	−36.86	−8.27	72.53	−15.81	0.52	81.32	−4.19	−0.55	80.26	−13.78	**0.53**	**81.33**	**−3.95**	−1.31	79.51
	CIFAR100	−19.22	0.97	80.62	−0.33	1.86	81.52	−2.39	1.86	81.48	0.23	2.85	82.46	8.79	4.81	84.45	**8.97**	**5.22**	**84.85**	3.08	0.28	79.92	3.31	2.12	81.74
	DTD	−14.31	−7.93	64.15	−9.52	−4.52	67.71	−9.73	−6.06	66.01	−9.52	−4.36	67.77	−1.86	**1.38**	**73.46**	−2.77	−0.74	71.33	−0.8	0.27	72.34	**−0.69**	0.05	72.18
	EuroSAT	−16.93	−0.07	95.46	−14.76	−0.3	95.24	−18.06	−0.19	95.35	−17.04	−0.33	95.2	**4.2**	−0.07	95.46	−1.09	0.44	95.98	0.17	0.07	95.61	2.83	**0.46**	**96.0**
	GTSRB	−2.72	−1.24	85.3	−0.8	0.35	87.05	−2.16	−0.48	86.25	−0.33	1.03	87.63	1.63	1.15	87.82	−0.87	1.58	88.23	**5.81**	2.94	89.45	5.38	**3.21**	**89.99**
	MNIST	−9.31	−0.45	98.23	−10.42	−0.07	98.59	−12.21	−0.1	98.54	−4.02	**0.22**	**98.73**	**2.74**	−0.1	98.55	2.25	−0.13	98.52	−0.4	0.04	98.62	0.21	−0.17	98.49
	RESISC45	−26.32	−4.83	87.05	−14.44	−2.54	89.33	−15.49	−2.51	89.37	−16.32	−1.97	89.9	−0.29	−0.25	91.62	−1.49	−0.1	91.78	0.25	**0.17**	**92.05**	**1.03**	0.13	92.0
	SVHN	−3.49	1.36	66.84	−7.19	2.9	68.33	−5.45	0.4	65.98	0.62	4.34	69.95	7.15	3.71	69.22	7.08	3.59	69.13	6.17	3.97	69.34	**8.13**	**4.92**	**70.4**
	CIFAR10	7.87	2.79	97.71	7.8	2.65	97.58	**8.25**	**2.83**	**97.73**	7.5	2.79	97.71	8.05	2.8	97.7	6.38	2.74	97.66	7.72	2.46	97.36	6.39	2.33	97.24
CIFAR100	Cars	−27.76	−3.15	77.66	−12.57	−3.36	77.44	−21.22	−3.58	77.23	−12.42	−4.38	76.42	−6.74	−0.12	80.69	**−1.32**	**0.15**	**80.96**	−12.29	−0.72	80.08	−6.32	−1.07	79.73
	CIFAR10	−8.17	1.05	95.96	0.22	1.33	96.25	−0.64	1.5	96.41	1.06	1.52	96.42	2.65	**1.96**	**96.87**	**3.24**	1.62	96.53	0.33	1.34	96.25	0.43	0.88	95.79
	DTD	−10.37	−6.44	65.64	−8.83	−5.53	66.54	−9.36	−5.11	66.97	−6.54	−3.3	68.83	−4.52	−0.74	71.33	−2.39	**0.16**	**72.23**	−4.41	−1.44	70.64	**−1.97**	−1.97	70.11
	EuroSAT	−23.19	−0.06	95.48	−12.96	−0.26	95.28	−16.59	−0.48	95.06	−7.11	−0.17	95.37	**−0.46**	0.19	95.72	−7.13	0.46	96.0	−3.33	0.26	95.8	−2.19	**0.56**	**96.09**
	GTSRB	−6.46	0.17	86.79	−0.43	2.52	89.11	−4.64	0.99	87.62	2.98	2.69	89.3	1.27	1.8	88.42	3.2	2.65	89.18	4.88	2.98	89.65	**6.08**	**4.04**	**90.7**
	MNIST	−8.99	−0.35	98.36	−7.26	−0.15	98.51	−1.37	−0.3	98.36	−6.36	0.11	98.79	3.09	0.08	98.52	2.85	0.04	98.75	**8.1**	−0.07	98.68	0.25	**0.2**	**98.95**
	RESISC45	−20.65	−3.54	88.33	−16.51	−3.56	88.32	−17.65	−3.3	88.57	−14.14	−1.95	89.92	−5.78	−1.3	90.59	**−2.81**	−0.57	91.3	−7.68	−1.1	90.78	−6.89	**−0.44**	**91.43**
	SVHN	2.46	3.06	68.59	6.05	5.04	70.58	2.43	2.57	68.03	5.7	4.64	70.18	7.89	3.3	68.84	6.98	3.7	69.12	6.02	**5.32**	**70.67**	**8.24**	4.68	70.21
	CIFAR100	24.4	9.34	88.98	23.16	9.12	88.77	**24.63**	8.97	88.6	17.26	8.78	88.41	23.33	8.2	87.84	13.77	7.31	86.94	24.24	8.31	87.94	13.3	**9.35**	**88.99**
DTD	Cars	−18.09	−3.36	77.44	−10.19	−2.87	77.94	−9.85	−2.54	78.25	−5.97	−2.43	78.36	−11.63	−4.15	76.66	**−0.06**	0.21	81.01	−1.94	**0.77**	**81.58**	−0.2	0.36	81.17
	CIFAR10	−36.94	−6.55	88.37	−13.28	−4.32	90.59	−13.4	−4.74	90.17	−9.33	−1.81	93.11	−12.64	−4.99	89.92	**0.57**	−0.44	94.48	−0.76	−0.22	94.7	0.44	**0.21**	**95.12**
	CIFAR100	−38.48	−10.03	69.61	−18.82	−7.08	72.57	−20.99	−8.23	71.41	−8.12	−3.3	76.34	−17.73	−8.45	71.18	**2.39**	0.8	80.45	−3.41	−0.72	78.93	1.61	**0.84**	**80.49**
	EuroSAT	−32.67	−0.57	94.96	−20.46	−1.54	94.0	−12.56	−0.48	95.06	−11.02	−0.52	95.02	−22.26	−1.33	94.2	**4.87**	0.15	95.69	−7.19	**0.43**	**95.96**	2.76	0.0	95.54
	GTSRB	−18.13	−7.02	79.69	−7.89	−3.52	83.15	−13.89	−2.21	84.51	−4.41	−1.61	85.01	−5.76	−2.62	83.97	0.34	−0.39	86.18	**2.11**	0.69	87.32	1.43	**1.3**	**87.93**
	MNIST	−24.64	−0.95	97.75	3.59	−0.51	98.16	−7.74	0.02	98.35	−5.74	−0.32	98.34	−7.92	−0.14	98.49	**5.03**	−0.02	98.69	−2.84	**0.17**	**98.86**	2.6	0.15	98.84
	RESISC45	−34.51	−5.0	86.87	−11.0	−2.52	89.37	−13.37	−2.65	89.22	−10.52	−1.48	90.4	−19.27	−4.48	87.38	0.33	−0.13	91.75	−1.86	−0.38	91.49	**1.21**	**0.14**	**92.02**
	SVHN	−17.89	−5.9	59.64	−8.6	−5.14	60.18	0.31	−3.28	62.14	−3.49	−2.61	63.03	0.55	−3.54	61.9	1.84	−0.11	65.3	**7.88**	0.76	66.31	3.81	**1.23**	**66.69**
	DTD	21.44	4.04	76.12	16.7	1.33	73.46	31.91	4.89	77.18	2.13	0.05	72.18	26.38	0.9	72.98	0.16	2.45	74.63	**33.14**	**6.12**	**78.14**	0.74	3.19	75.27
EuroSAT	Cars	−18.44	−7.0	73.8	−14.04	−8.34	72.45	−12.34	−7.85	72.95	−13.0	−6.27	74.53	−4.89	−2.79	78.01	−3.95	−1.84	78.97	−1.37	0.42	81.22	**−1.08**	**0.53**	**81.33**
	CIFAR10	−26.75	−8.04	86.86	−17.0	−6.07	88.84	−17.99	−5.32	89.59	−12.57	−4.24	90.68	−5.87	−2.49	92.41	−1.59	−0.6	94.31	−0.53	0.23	95.13	**0.39**	**0.34**	**95.25**
	CIFAR100	−35.66	−11.37	68.26	−23.68	−8.87	70.76	−24.64	−9.78	69.85	−17.79	−6.03	73.59	−6.43	−4.0	75.62	−0.83	−0.56	79.08	1.23	0.46	80.09	**2.27**	**1.06**	**80.7**
	DTD	−10.85	−6.17	65.9	−7.5	−4.04	68.03	−7.13	−5.85	66.33	−6.12	−1.81	70.27	−2.93	−1.22	70.85	−1.17	0.05	72.23	1.01	**3.19**	**75.27**	**1.12**	2.66	74.73
	GTSRB	−9.49	−8.53	78.14	−5.78	−7.54	79.16	−6.39	−8.39	78.35	−4.29	−4.35	82.31	−1.27	−4.05	82.6	−0.58	−1.2	85.53	**4.56**	0.94	87.67	4.51	**1.98**	**88.6**
	MNIST	−6.85	−1.2	97.45	−6.92	−1.35	97.32	−7.76	−1.32	97.29	−5.24	−1.48	97.14	3.21	−0.54	98.16	0.04	−0.2	98.44	**5.54**	−0.07	98.57	2.82	**0.13**	**98.74**
	RESISC45	−40.95	−3.89	87.97	−28.6	−2.83	89.05	−28.44	−2.16	89.7	−16.65	−1.4	90.48	−8.29	−0.7	91.16	−0.46	0.05	91.9	−4.08	−0.33	91.54	**0.9**	**0.79**	**92.67**
	SVHN	−18.42	−7.38	58.09	−12.7	−4.44	61.17	−14.24	−5.12	60.32	−10.42	−4.45	61.13	0.07	−4.86	60.56	1.99	−2.51	63.06	**6.08**	2.5	67.92	6.07	**3.03**	**68.59**
	EuroSAT	52.11	3.09	98.65	51.96	3.26	98.8	**52.37**	3.2	98.76	30.93	**3.31**	**98.87**	52.19	2.91	98.46	28.04	2.67	98.2	52.15	2.98	98.54	17.02	2.67	98.22
GTSRB	Cars	−32.45	−13.23	67.57	−19.66	−9.45	71.36	−20.69	−8.39	72.42	−18.07	−7.61	73.2	−5.77	−2.75	78.05	−1.24	−0.25	80.55	−2.66	0.31	80.94	**−0.88**	**1.82**	**81.11**
	CIFAR10	−61.3	−17.99	76.92	−31.36	−8.35	86.57	−41.57	−12.51	82.4	−15.19	−4.71	90.21	−40.89	−11.41	83.51	−1.77	−0.2	94.72	−1.71	−0.38	94.53	**−0.14**	**0.0**	**94.92**
	CIFAR100	−55.13	−24.28	55.36	−30.12	−12.83	66.83	−42.57	−19.81	59.82	−21.7	−6.2	73.43	−41.4	−17.66	61.97	0.46	0.39	80.01	−2.16	−0.91	78.73	**1.81**	**1.21**	**80.85**
	DTD	−24.68	−14.63	57.45	−10.0	−7.5	64.57	−17.18	−9.2	62.87	−10.74	−3.51	68.83	−6.6	−2.13	69.84	−1.54	**1.76**	**73.72**	−2.55	1.65	73.62	**−0.32**	1.17	73.03
	EuroSAT	−36.94	−5.0	90.56	−19.31	−2.91	92.63	−21.91	−4.72	90.81	−17.7	−1.11	94.43	−16.5	−3.31	92.22	**3.94**	**0.0**	**95.54**	1.17	−0.43	95.11	2.57	−0.06	95.48
	MNIST	−6.53	−0.81	97.89	−6.43	−0.25	98.35	−13.95	−0.63	98.02	−1.72	−0.09	98.6	−4.72	−0.39	98.33	3.33	−0.06	98.58	−2.23	**0.13**	**98.8**	**6.78**	0.02	98.66
	RESISC45	−46.86	−10.35	81.52	−19.24	−5.44	86.43	−35.92	−8.73	83.14	−18.17	−2.43	89.44	−13.68	−2.92	88.95	0.1	−0.02	91.86	−2.6	−0.48	91.4	**0.32**	0.11	91.98
	SVHN	2.22	**18.35**	**83.82**	5.07	16.66	82.25	2.74	16.81	82.08	10.11	18.12	83.65	**15.22**	16.15	81.75	9.68	7.79	73.39	10.08	8.67	74.17	7.94	5.9	71.45
	GTSRB	61.47	**12.64**	**99.26**	58.8	12.4	99.0	66.1	11.96	98.52	18.42	11.9	98.53	64.74	10.8	97.4	9.19	8.27	95.0	**66.27**	12.03	98.45	19.33	11.2	97.81
MNIST	Cars	−29.96	−17.16	63.64	−29.3	−16.44	64.38	−29.55	−18.58	62.24	−44.11	−15.91	64.92	−4.56	−2.18	78.65	**−2.47**	**−0.41**	**80.38**	−3.89	−0.9	79.93	−3.25	−0.58	80.23
	CIFAR10	−59.44	−15.91	79.0	−39.13	−13.25	81.66	−58.04	−15.62	79.29	−50.68	−10.78	84.14	−15.31	−5.61	89.3	**−1.19**	**0.32**	**95.23**	−1.49	−0.93	93.98	−3.11	−0.26	94.66
	CIFAR100	−56.3	−22.92	56.72	−42.41	−18.2	61.44	−50.72	−21.05	58.57	−55.09	−16.71	62.94	−23.77	−9.44	70.19	−5.18	0.11	79.76	−5.64	−1.26	78.37	**−4.43**	**0.46**	**80.11**
	DTD	−28.03	−7.5	64.41	−15.21	−10.59	61.49	−21.91	−9.95	62.13	−35.59	−11.81	60.37	−3.14	−0.32	71.7	−3.3	1.55	73.57	−4.95	0.11	72.18	**−0.37**	**1.6**	**73.62**
	EuroSAT	−32.02	−3.09	92.44	−34.48	−3.44	92.09	−27.09	−2.98	92.56	−34.72	−1.94	93.59	**−2.46**	−0.72	94.81	−4.15	−0.17	95.37	**−2.46**	**0.3**	**95.83**	−12.63	0.02	95.56
	GTSRB	−25.71	−6.02	80.59	−15.46	−4.22	82.45	−20.58	−6.25	80.39	−23.1	−0.02	86.63	−5.57	−2.45	84.19	1.26	2.77	89.33	−0.91	1.85	88.47	**3.71**	**2.89**	**89.49**
	RESISC45	−42.9	−6.68	85.19	−39.11	−7.29	84.59	−37.94	−7.05	84.84	−51.0	−5.94	85.94	−5.46	−1.56	90.32	**−1.68**	0.24	92.11	−4.68	−0.06	91.81	−3.19	**0.26**	**92.13**
	SVHN	17.26	10.15	75.71	16.11	13.21	78.82	15.42	11.42	76.94	13.47	14.99	80.57	12.9	6.06	71.45	12.41	7.51	72.85	**22.94**	15.31	80.84	19.95	**16.76**	**82.31**
	MNIST	51.19	0.88	99.62	50.56	0.98	99.63	**51.28**	**1.02**	**99.67**	21.34	0.98	99.68	51.02	0.8	99.43	5.51	0.5	99.18	51.18	0.98	99.6	8.35	0.83	99.52
RESISC45	Cars	−17.17	−5.3	75.51	−10.88	−6.02	74.78	−12.42	−4.5	76.3	−11.42	−6.12	74.7	−4.32	−0.21	80.6	−0.68	0.49	81.28	−1.17	0.56	81.37	**0.26**	**0.66**	**81.47**
	CIFAR10	−33.88	−8.43	86.49	−15.74	−5.1	89.81	−18.47	−5.84	89.08	−10.76	−3.44	91.47	−7.58	−2.04	92.88	−0.89	0.0	94.91	−0.6	−0.3	94.61	**−0.2**	**0.13**	**95.04**
	CIFAR100	−35.9	−11.43	68.19	−17.78	−6.75	72.89	−23.07	−8.85	70.78	−12.54	−5.25	74.38	−7.25	−3.25	76.39	1.55	0.57	80.19	0.37	−0.43	79.2	**1.57**	**0.63**	**80.26**
	DTD	−12.66	−4.47	67.77	−6.44	−2.39	69.52	−8.56	−3.24	68.83	−3.51	−0.48	71.6	−3.4	−0.74	71.33	−0.59	2.07	74.15	−1.81	1.01	73.4	**−0.43**	**2.98**	**74.79**
	EuroSAT	2.5	1.06	96.61	8.85	1.07	96.61	5.22	0.74	96.28	9.15	1.07	96.61	12.3	1.13	96.67	13.11	**1.26**	**96.8**	**13.74**	0.91	96.44	12.76	0.76	96.3
	GTSRB	−9.58	−6.32	80.42	−6.98	−3.82	82.76	−11.79	−4.01	82.62	−4.41	−1.08	85.51	−1.76	−1.06	85.66	0.1	0.34	86.95	−2.09	0.81	87.51	**1.96**	**0.85**	**87.55**
	MNIST	−15.45	−0.85	97.75	−7.76	−0.37	98.33	−8.0	−0.07	98.57	−0.94	−0.23	98.44	2.03	−0.19	98.49	2.19	−0.12	98.58	4.04	−0.35	98.37	**5.47**	**−0.1**	**98.6**
	SVHN	−21.2	−7.58	57.82	−11.01	−3.14	62.38	−8.6	−4.57	60.93	−4.06	−2.6	62.88	3.55	−0.88	64.61	**4.9**	0.68	66.23	4.16	1.04	66.36	2.9	**1.27**	**66.89**
	RESISC45	33.49	**3.98**	**95.84**	30.4	3.94	95.79	**35.1**	3.9	95.76	13.65	3.68	95.56	33.14	2.19	94.05	8.14	2.25	94.13	34.76	3.37	95.22	10.92	3.25	95.13
SVHN	Cars	−39.37	−17.56	63.23	−31.22	−17.22	63.59	−33.06	−18.38	62.43	−30.44	−15.28	65.51	**−5.66**	−2.92	77.9	−6.63	**−2.0**	**78.8**	−6.5	−2.04	78.76	−8.73	−2.89	77.91
	CIFAR10	−44.46	−13.57	81.34	−39.98	−13.84	81.06	−42.24	−14.41	80.5	−32.49	−11.44	83.49	−21.11	−7.47	87.45	−24.95	−5.66	89.25	−7.25	−2.16	92.75	**−6.83**	**−1.86**	**93.06**
	CIFAR100	−52.73	−21.64	58.0	−43.38	−22.93	56.72	−46.16	−24.14	55.51	−41.28	−18.07	61.59	−28.52	−11.84	67.8	−30.95	−10.09	69.56	−12.88	−5.26	74.38	**−11.03**	**−2.73**	**76.9**
	DTD	−28.51	−19.2	52.87	−20.11	−15.48	56.6	−24.1	−18.09	53.88	−20.48	−15.11	56.97	−5.8	−1.06	71.17	−7.45	−0.9	71.12	−4.47	−0.48	71.76	**−1.86**	**0.21**	**72.29**
	EuroSAT	−36.17	−4.31	91.22	−25.59	−4.3	91.24	−34.87	−5.7	89.83	−36.09	−6.04	89.5	−19.98	−2.26	93.28	−27.78	−2.56	92.98	**−2.61**	**−0.78**	**94.76**	−13.74	**−0.78**	**94.76**
	GTSRB	−21.91	0.0	86.68	−10.66	0.95	87.72	−8.27	1.78	88.41	−9.62	4.95	91.57	1.8	5.04	91.7	−0.06	6.03	92.64	5.49	6.03	92.66	**6.45**	**6.23**	**92.9**
	MNIST	36.16	−0.04	98.61	30.61	−0.19	98.47	38.35	0.0	98.63	7.66	0.15	98.81	31.19	0.23	98.91	11.39	−0.06	98.61	**41.91**	**0.66**	**99.15**	19.55	0.49	99.06
	RESISC45	−44.29	−9.59	82.29	−34.94	−10.1	81.78	−37.16	−10.0	81.87	−40.17	−8.97	82.9	−8.44	−1.95	89.92	−7.84	−1.65	90.22	**−4.56**	−1.38	90.49	−5.56	**−1.0**	**90.87**
	SVHN	65.0	31.92	97.44	64.38	31.86	97.4	**65.58**	31.79	97.3	51.28	**31.98**	**97.5**	64.84	31.02	96.5	53.39	31.04	96.54	65.25	31.36	96.97	50.92	31.38	96.95

Table 6: **Δ_{ZS}, Δ_{LP} and $\mathcal{A}_{\text{LP}}$ for models fully fine-tuned on 9 different classification datasets.** LP-init-LDIFS outperforms other fine-tuning baselines and even LP-init-L2SP in minimizing concept forgetting, while also outperforming LP-init-L2SP on the fine-tuned task performance.

Dataset		Fine-tuning baselines																							
Fine-tune	Eval	FLYP			FLYP-CE			ZS-init-CE			LP-init-CE (LP-FT)			ZS-init-L2SP			LP-init-L2SP			ZS-init-LDIFS (Ours)			LP-init-LDIFS (Ours)		
		$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$
SVHN → C10	SVHN	52.84	27.98	93.5	53.48	27.68	93.23	57.21	27.48	92.99	43.77	28.52	93.88	51.54	26.11	91.59	40.15	26.44	91.94	56.91	29.36	94.98	58.75	29.68	95.04
	CIFAR10	7.81	2.77	97.68	7.7	2.52	97.42	7.81	2.57	97.48	7.31	2.8	97.73	7.88	2.69	97.61	6.49	2.62	97.53	7.71	2.29	97.2	6.16	2.38	97.3
	Cars	−57.63	−16.23	64.56	−47.56	−10.07	70.74	−47.11	−9.44	71.37	−40.44	−10.45	70.35	−18.94	−0.06	80.76	−5.52	0.36	81.16	−13.29	0.44	81.23	−4.79	−0.01	80.75
	CIFAR100	−24.27	−0.03	79.61	−4.27	0.06	79.71	−3.24	−0.64	79.01	−2.24	1.59	81.25	6.78	3.86	83.5	7.67	4.72	84.37	2.89	0.82	80.46	5.78	5.26	84.87
	DTD	−19.41	−11.81	60.27	−10.8	−10.96	61.12	−14.26	−9.1	62.87	−12.45	−9.2	62.87	−2.93	0.48	72.71	−3.67	0.8	72.82	−0.96	0.05	72.29	−1.65	1.12	73.19
	EuroSAT	−21.56	−0.91	94.63	−20.07	−0.17	95.37	−15.2	−0.35	95.19	−22.02	−0.43	95.11	4.15	−0.13	95.41	−10.28	0.11	95.65	−1.93	0.24	95.78	3.37	0.3	95.83
	GTSRB	−9.32	−1.78	84.9	−6.33	0.86	87.62	−2.92	0.68	87.31	−4.92	1.68	88.3	4.72	3.14	89.81	4.17	4.15	90.75	7.17	3.24	89.87	7.09	3.86	90.64
	MNIST	34.62	−0.04	98.61	28.9	0.15	98.81	36.26	0.11	98.74	7.31	0.33	98.99	28.31	0.22	98.9	13.05	0.13	98.8	2.89	0.2	98.6	14.56	−0.1	98.56
	RESISC45	−36.56	−6.03	85.84	−22.29	−4.35	87.52	−20.76	−4.52	87.35	−23.3	−4.33	87.54	−1.51	−0.76	91.11	−2.51	−0.59	91.29	0.46	−0.25	91.62	0.4	0.17	92.05
SVHN → C100	SVHN	54.08	27.54	93.06	56.19	28.59	94.13	56.72	27.67	93.18	45.18	28.12	93.47	42.11	21.68	87.15	33.62	22.57	88.06	57.84	31.11	97.33	58.61	31.29	96.86
	CIFAR100	23.18	8.26	87.9	21.64	8.26	87.91	23.78	7.98	87.63	15.63	8.08	87.74	23.06	7.9	87.54	13.47	6.99	86.64	24.37	8.3	87.94	13.56	7.65	87.28
	Cars	−39.5	−7.77	73.01	−20.4	−8.31	72.5	−22.41	−8.41	72.4	−17.7	−8.16	72.64	−6.09	−0.14	80.69	−1.99	0.27	81.07	−11.19	−0.62	80.18	−0.99	−0.49	80.31
	CIFAR10	−10.04	0.52	95.43	−0.8	0.8	95.7	−2.31	0.92	95.83	0.52	1.12	96.05	2.7	1.93	96.85	3.01	1.84	96.75	−0.05	1.33	96.24	2.62	3.15	97.17
	DTD	−12.18	−9.1	62.98	−12.71	−8.56	63.51	−13.83	−9.36	62.61	−9.89	−8.3	63.78	−5.21	0.0	72.23	−3.19	0.27	72.29	−4.04	−1.54	70.69	−0.54	−0.32	72.76
	EuroSAT	−24.52	−0.52	95.02	−18.85	−0.48	95.06	−20.76	−0.52	95.02	−15.85	0.09	95.63	1.87	−0.13	95.41	−10.09	0.13	95.67	−6.89	−0.07	95.46	1.28	0.31	95.85
	GTSRB	−6.26	−0.03	86.65	−6.6	1.59	88.36	−5.4	1.31	87.94	2.42	3.44	90.06	4.96	2.79	89.46	4.92	3.62	90.22	2.67	3.43	90.06	5.49	4.33	91
	MNIST	21.79	0.19	98.84	35.77	0.22	98.88	38.24	0.25	98.88	11.51	0.36	99.02	25.84	0.18	98.86	15.59	0.25	98.92	6.95	0.43	98.83	15.78	0.34	99
	RESISC45	−25.98	−4.84	87.03	−23.4	−4.6	87.27	−20.44	−4.52	87.35	−23.06	−3.35	88.52	−6.05	−0.89	90.98	−4.52	−0.78	91.1	−8.84	−0.94	90.94	−3.28	−0.87	91

Table 7: **Δ_{ZS}, Δ_{LP} and $\mathcal{A}_{\text{LP}}$ for models fine-tuned on SVHN → CIFAR10 and SVHN → CIFAR100.** LP-init-LDIFS consistently outperforms other baselines even in the sequence setup.

Dataset		Fine-tuning baselines																							
Fine-tune	Eval	FLYP			FLYP-CE			ZS-init-CE			LP-init-CE (LP-FT)			ZS-init-L2SP			LP-init-L2SP			ZS-init-LDIFS (Ours)			LP-init-LDIFS (Ours)		
		$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{ZS}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$
EuroSAT	CIFAR10	−57.93	−23.21	65.15	−49.37	−23.02	65.46	−53.46	−23.74	65.01	−46.95	−22.55	66.13	−40.33	−14.04	74.4	−31.21	−8.9	79.55	−20.3	−8.21	80.15	**−14.1**	**−3.67**	**84.69**
	SVHN	−20.75	−19.91	47.77	−14.33	−24.64	43.08	−14.14	−22.23	45.54	−18.34	−16.54	51.14	−2.63	−12.81	54.91	−3.17	−6.92	60.73	**0.14**	−7.23	60.45	0.06	**−2.17**	**65.51**
	EuroSAT	56.46	**3.04**	**98**	55.29	3.09	97.76	**56.74**	2.83	97.52	3.74	2.81	97.80	55.78	1.56	96.46	−11.91	2.89	96.70	56.31	3.01	97.97	−1.6	3.02	97.98

Table 8: Δ_{ZS}, Δ_{LP} and $\mathcal{A}_{\text{LP}}$ for CLIP RN50 fully fine-tuned on EuroSAT.

Dataset		Fine-tuning baselines															
Fine-tune	Eval	FLYP		FLYP-CE		ZS-init-CE		LP-init-CE (LP-FT)		ZS-init-L2SP		LP-init-L2SP		ZS-init-LDIFS (Ours)		LP-init-LDIFS (Ours)	
		$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$	$\Delta_{\text{LP}}(\uparrow)$	$\mathcal{A}_{\text{LP}}(\uparrow)$
EuroSAT	CIFAR10	−9.44	83.43	−8.13	84.74	−6.01	86.86	−4.76	88.11	−2.32	90.55	−0.11	92.76	0.07	92.8	**0.16**	**93.03**
	SVHN	−8.92	54.24	−8.16	55.0	−6.63	56.53	−5.10	58.06	−1.87	61.29	−1.21	61.95	1.37	64.53	**2.68**	**65.84**
	EuroSAT	1.63	98.89	**1.65**	**98.91**	1.41	98.67	1.53	98.79	0.95	98.21	1.42	98.68	1.47	98.73	1.6	98.86

Table 9: Δ_{LP} and $\mathcal{A}_{\text{LP}}$ for FLAVA ViT-B/16 fully fine-tuned on EuroSAT.