

## A APPENDIX

### OVERVIEW

The appendix provides supplementary material to support and expand upon the main findings of our paper. Additionally, **we provide code in the supplementary material**. To ensure clarity the contents are organized as follows:

- **Evaluation Protocol:** We begin by providing comprehensive details on the evaluation protocol and datasets used throughout our experiments to ensure reproducibility.
- **Performance on ViT-L-14 Architecture:** We present a comparative performance analysis of AXIS and aTLAS using the larger ViT-L-14 architecture, demonstrating that the advantages of our method scale effectively to more powerful models.
- **Detailed Main Results:** We then provide extensive results with the ViT-B-32 architecture. These tables offer a granular performance breakdown, detailing per-target-task accuracy for different numbers of aggregated source task vectors and varying budgets of trainable parameters ( $N$ ).
- **In-depth Robustness Analyses:** We conduct a series of thorough evaluations to validate the robustness of our framework under challenging conditions. These include:
  - Resilience to 12 common image corruptions across five distinct severity levels.
  - Performance evaluation across different levels of training data availability for the target task.
  - Robustness against altered source parameters, including scenarios with noisy or heavily pruned task vectors.
- **Component Selection:** We present a detailed ablation study comparing our default component aggregation strategy (top components) against a range of alternative methods.
- **Impact of Final SVD:** We provide details of the role of the final SVD re-parameterization step in stabilizing the transfer basis across a couple of selection strategies.
- **Comparison with PEFT Methods:** To offer a wider perspective on performance, we compare our merge-based framework to selected PEFT methods, which operate under the assumption of having access only to the pre-trained weights.

#### A.1 EVALUATION PROTOCOL

To ensure a direct and fair comparison, we adopt the comprehensive benchmark, publicly released task vectors, and training protocols established by the authors of aTLAS. Their framework provides task vectors obtained by fine-tuning the pre-trained CLIP Radford et al. (2021) model on distinct image recognition datasets: Stanford Cars Krause et al. (2013), DTD Cimpoi et al. (2014), EuroSAT Helber et al. (2019), GTSRB Stallkamp et al. (2011), MNIST LeCun (1998), RESISC45 Cheng et al. (2017), SUN397 Xiao et al. (2016), SVHN Netzer et al. (2011), CIFAR10 Krizhevsky et al. (2009), CIFAR100 Krizhevsky et al. (2009), ImageNet Russakovsky et al. (2015), STL10 Coates et al. (2011), Food101 Bossard et al. (2014), Caltech101 Fei-Fei et al. (2006), Caltech256 Griffin et al. (2007), FGVCAircraft Maji et al. (2013), Flowers102 Nilsback & Zisserman (2008), Oxford Pets Parkhi et al. (2012), CUB200 Welinder et al. (2010), PascalVOC Everingham et al. (2015), Country211 Radford et al. (2021), and UCF101 Soomro et al. (2012). The original fine-tuning for these vectors was performed using the AdamW optimizer Loshchilov & Hutter (2017) with a learning rate of  $10^{-5}$ , a batch size of 128, and a weight decay of 0.1 for the ViT-B-32 architecture. Table 3 provides dataset details, their corresponding hyperparameters, and the fine-tuning accuracy achieved with full-finetuning.

During the target task adaptation stage, we fine-tune the merged model for each dataset independently, using the same hyperparameters as the aTLAS baseline (each adaptation runs for 10 epochs with a learning rate of  $10^{-1}$ ). The batch size is adjusted based on the model architecture: 64 for the ViT-B-32 model and 128 for the larger ViT-L-14 model. For the ViT-L-14 architecture, both methods originally use two steps of gradient accumulation. To ensure a controlled and reproducible evaluation provided by aTLAS, the source task vectors are aggregated incrementally in a

Dataset	Classes	Splits			Epochs	Fine-tuned accuracy (%)	
		train	val	test		ViT-B/32	ViT-L/14
Cars	196	7,330	814	8,041	35	78.26	91.67
DTD	47	3,384	376	1,880	76	78.94	84.73
EuroSAT	10	21,600	2,700	2,700	12	98.89	99.81
GTSRB	43	23,976	2,664	12,630	11	99.14	99.30
MNIST	10	55,000	5,000	10,000	5	99.65	99.77
RESISC45	45	17,010	1,890	6,300	15	95.94	97.14
SUN397	397	17,865	1,985	19,850	14	75.40	81.98
SVHN	10	68,257	5,000	26,032	4	97.38	97.97
CIFAR10	10	45,000	5,000	10,000	5	98.05	99.22
CIFAR100	100	45,000	5,000	10,000	6	89.09	93.01
ImageNet	1,000	1,276,167	5,000	50,000	10	76.41	85.52
STL10	10	4,500	500	8,000	4	98.55	99.62
Food101	101	70,750	5,000	25,250	15	88.68	95.37
Caltech101	101	6,941	694	1,736	10	94.41	94.82
Caltech256	257	22,037	2,448	6,122	8	92.60	97.17
FGVCAircraft	100	3,334	3,333	3,333	60	40.65	68.11
Flowers102	102	1,020	1,020	6,149	40	90.08	97.84
OxfordIIITPet	37	3,312	368	3,669	5	92.15	95.91
CUB200	200	5,395	599	5,794	20	73.56	86.35
PascalVOC	20	7,844	7,818	14,976	10	88.42	92.05
Country211	211	31,650	10,550	21,100	15	21.99	38.06
UCF101	101	7,639	1,898	3,783	20	85.01	92.55

Table 3: Comparison of full fine-tuning model accuracy per dataset

fixed, pre-defined sequence. The order of aggregation is as follows: Cars, DTD, EuroSAT, GTSRB, MNIST, RESISC45, SUN397, SVHN, CIFAR10, CIFAR100, ImageNet, STL10, Food101, Caltech101, Caltech256, FGVCAircraft, Flowers102, OxfordIIITPet, CUB200, PascalVOC, Country211, and UCF101. Each experimental run was conducted once with a single random seed across our comprehensive evaluation, which included 21 target tasks, multiple aggregation levels, and varying parameter budgets.

#### A.1.1 COMPUTATIONAL ENVIRONMENT

All experiments were conducted within a high-performance computing (HPC) cluster equipped with a heterogeneous GPU environment. The available resources included partitions with NVIDIA RTX 4090, NVIDIA V100, and NVIDIA A100 GPUs. The results reported in this paper, generated using the ViT-L-14 architecture, were obtained with nodes equipped with NVIDIA A100-SXM4-80GB GPUs. Our software stack was built upon the CUDA 12.2 toolkit with NVIDIA driver version 535.183.01.

#### A.2 PERFORMANCE ON ViT-L-14 ARCHITECTURE

To validate the scalability and effectiveness of our approach on larger models, we replicated our experiments using the ViT-L-14 architecture. The results demonstrate the advantages of the AXIS framework. The performance comparison for the  $N = 10\%$  and  $N = 20\%$  parameter budget is illustrated in Figure 12, where AXIS consistently outperforms aTLAS as the number of aggregated source tasks increases. Further analysis across different parameter budgets confirms these findings.

#### A.3 IN-DEPTH ROBUSTNESS ANALYSES

##### A.3.1 ROBUSTNESS TO INPUT PERTURBATIONS

To further probe the robustness capabilities of AXIS and aTLAS, we evaluate them against a set of 12 common image corruptions Hendrycks & Dietterich (2019). Each corruption type is applied to

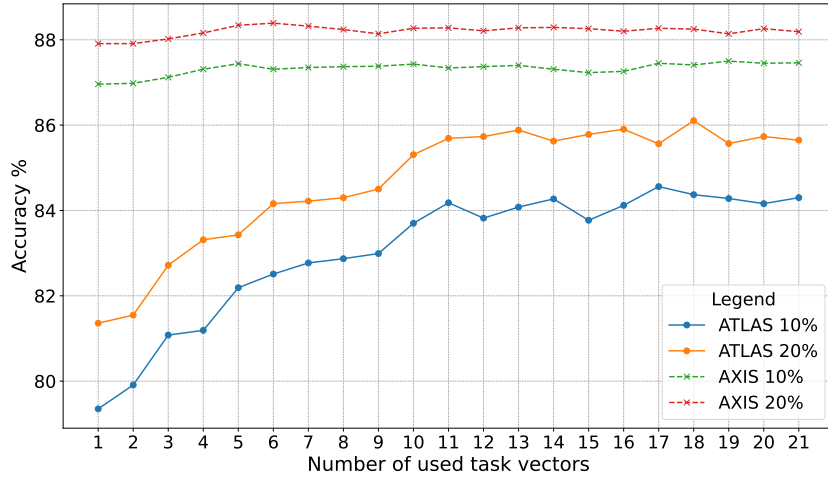


Figure 12: AXIS outperforms aTLAS on the ViT-L-14 architecture with  $N = 10\%$  and  $N = 20\%$  of trainable singular values. Each point is the mean accuracy across 21 independently evaluated target tasks. The plot illustrates the accuracy gain as the number of aggregated source tasks increases.

the test set of target task images at five distinct severity levels to simulate a range of degradations. As illustrated in Figure 13, our proposed method, AXIS, maintains a slightly average performance advantage (0.83 percentage points). This margin is particularly pronounced for corruption types where overall accuracy remains high, indicating better robustness in moderately challenging conditions. A detailed breakdown by severity level delineates this trend more clearly (see Figure 14). AXIS demonstrates greater resilience across the initial four perturbation levels, outperforming aTLAS by margins of 2.04 percentage points for the lowest corruption severity.

Furthermore, we extend our robustness evaluation to scenarios with partial input information, a challenge simulated using patch dropout. A detailed, step-by-step analysis, presented in Table 4, illustrates how the model’s resilience to input masking evolves as the incremental aggregation of each source task vector is performed. This granular breakdown demonstrates that the fusion of diverse knowledge sources enhances the model’s ability to perform predictions even when significant portions of the input are omitted.

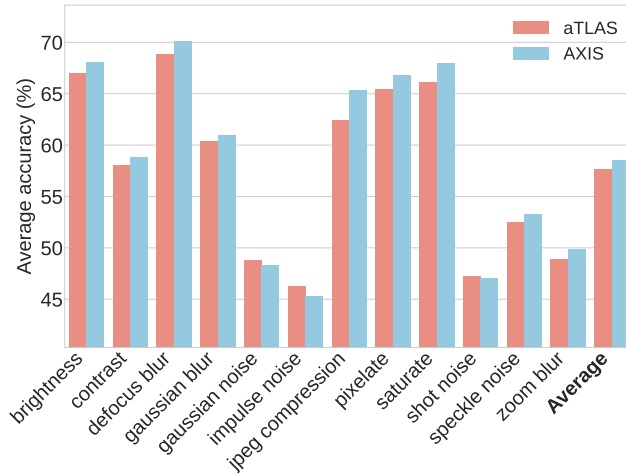


Figure 13: The accuracy across each type of corruption is evaluated for all severity levels ranging from 1 to 5 for all 21 target tasks.

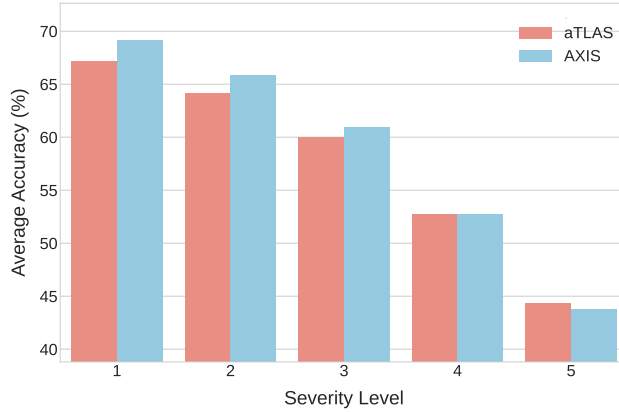


Figure 14: Severity levels average over all 12 image corruptions.

TV	Input Patch Dropout (%)															
1	77.31	75.53	72.78	67.51	63.61	56.86	52.11	44.66	39.79	28.74						
2	77.51 (+0.20)	75.80 (+0.27)	73.40 (+0.62)	69.04 (+1.53)	65.38 (+1.77)	59.54 (+2.68)	55.17 (+3.05)	48.36 (+3.69)	43.63 (+3.84)	32.67 (+3.93)						
3	77.52 (+0.21)	75.88 (+0.35)	73.49 (+0.71)	69.27 (+1.76)	65.72 (+2.12)	60.00 (+3.13)	55.87 (+3.75)	49.28 (+4.62)	44.84 (+5.05)	34.08 (+5.34)						
4	77.85 (+0.54)	76.12 (+0.60)	73.68 (+0.89)	69.43 (+1.91)	65.92 (+2.31)	59.97 (+3.10)	55.74 (+3.62)	49.12 (+4.45)	44.58 (+4.79)	33.49 (+4.75)						
5	77.81 (+0.50)	76.24 (+0.71)	73.98 (+1.20)	69.82 (+2.30)	66.51 (+2.90)	60.95 (+4.08)	56.87 (+4.76)	50.27 (+5.60)	45.65 (+5.86)	34.59 (+5.86)						
6	78.02 (+0.71)	76.41 (+0.88)	73.96 (+1.18)	69.78 (+2.27)	66.40 (+2.80)	60.53 (+3.66)	56.54 (+4.42)	50.10 (+5.43)	45.63 (+5.84)	34.64 (+5.90)						
7	78.08 (+0.76)	76.48 (+0.95)	74.23 (+1.45)	70.00 (+2.49)	66.82 (+3.21)	61.16 (+4.30)	57.01 (+4.90)	50.43 (+5.77)	45.74 (+5.95)	34.45 (+5.71)						
8	78.33 (+1.02)	76.67 (+1.14)	74.27 (+1.49)	70.27 (+2.76)	66.96 (+3.35)	61.21 (+4.35)	57.13 (+5.01)	50.71 (+6.04)	46.12 (+6.33)	34.98 (+6.24)						
9	78.41 (+1.10)	76.74 (+1.21)	74.42 (+1.64)	70.29 (+2.78)	66.88 (+3.27)	61.49 (+4.63)	57.63 (+5.52)	51.24 (+6.57)	47.02 (+7.23)	36.10 (+7.36)						
10	78.16 (+0.85)	76.60 (+1.07)	74.20 (+1.42)	69.85 (+2.33)	66.37 (+2.77)	60.55 (+3.69)	56.32 (+4.20)	49.77 (+5.11)	45.17 (+5.38)	34.25 (+5.51)						
11	78.40 (+1.09)	76.87 (+1.34)	74.44 (+1.66)	70.29 (+2.78)	66.81 (+3.20)	60.74 (+3.88)	56.21 (+4.09)	49.26 (+4.59)	44.47 (+4.68)	32.81 (+4.07)						
12	78.51 (+1.20)	76.90 (+1.38)	74.56 (+1.78)	70.34 (+2.83)	66.92 (+3.32)	61.11 (+4.24)	57.05 (+4.93)	50.31 (+5.65)	45.78 (+5.99)	33.99 (+5.26)						
13	78.37 (+1.06)	76.71 (+1.19)	74.34 (+1.55)	70.15 (+2.63)	66.78 (+3.17)	61.02 (+4.15)	56.85 (+4.74)	49.97 (+5.31)	45.32 (+5.53)	33.81 (+5.08)						
14	78.41 (+1.10)	76.83 (+1.30)	74.42 (+1.64)	70.16 (+2.65)	66.75 (+3.15)	60.87 (+4.00)	56.71 (+4.60)	49.65 (+4.98)	44.90 (+5.10)	33.05 (+4.31)						
15	78.34 (+1.02)	76.81 (+1.28)	74.50 (+1.71)	70.28 (+2.77)	66.82 (+3.21)	60.74 (+3.87)	56.32 (+4.20)	49.24 (+4.57)	44.64 (+4.85)	33.14 (+4.40)						
16	78.42 (+1.11)	76.85 (+1.32)	74.70 (+1.92)	70.45 (+2.94)	67.11 (+3.50)	61.37 (+4.50)	56.96 (+4.85)	50.16 (+5.50)	45.73 (+5.94)	34.51 (+5.78)						
17	78.41 (+1.09)	76.82 (+1.29)	74.57 (+1.79)	70.38 (+2.87)	67.06 (+3.45)	61.32 (+4.45)	57.10 (+4.98)	50.41 (+5.74)	45.93 (+6.14)	34.91 (+6.17)						
18	78.54 (+1.23)	76.94 (+1.41)	74.63 (+1.85)	70.53 (+3.01)	67.36 (+3.76)	61.77 (+4.91)	57.62 (+5.51)	50.92 (+6.26)	46.45 (+6.66)	34.92 (+6.19)						
19	78.58 (+1.27)	76.91 (+1.38)	74.61 (+1.83)	70.20 (+2.69)	66.87 (+3.26)	61.19 (+4.32)	56.97 (+4.86)	50.48 (+5.82)	46.05 (+6.26)	34.64 (+5.90)						
20	78.50 (+1.19)	76.75 (+1.22)	74.51 (+1.73)	70.14 (+2.63)	66.93 (+3.32)	61.19 (+4.33)	57.25 (+5.13)	50.58 (+5.91)	46.31 (+6.52)	35.05 (+6.31)						
21	<b>78.48</b> (+1.16)	<b>76.82</b> (+1.29)	<b>74.68</b> (+1.90)	<b>70.52</b> (+3.00)	<b>67.37</b> (+3.76)	<b>61.84</b> (+4.97)	<b>57.82</b> (+5.71)	<b>51.18</b> (+6.51)	<b>46.63</b> (+6.83)	<b>35.39</b> (+6.65)						

Table 4: Performance analysis of AXIS under increasing input masking. The table illustrates that aggregating more source task vectors (TV) enhances model robustness to input patch dropout. We report the mean accuracy (%) across all target tasks for dropout rates from 0% to 50%. Each row corresponds to a different number of aggregated sources, and values in parentheses show the improvement in percentage points (p.p.) over the first, single task vector baseline (first row).

### A.3.2 TRAINING DATA AVAILABILITY

To assess the data efficiency of our approach and its robustness in limited data scenarios, we investigate the performance of our method compared to aTLAS under varying levels of training data availability for the target task. For this experiment, we reduce the size of the target task’s training dataset, creating subsets with 5%, 10%, 25%, 50%, 75%, and 95% of the original samples. The results, illustrated in Figure 15, demonstrate that our method maintains a significant performance advantage over aTLAS across the broad majority of data availability levels.

### A.3.3 ROBUSTNESS AGAINST ALTERED SOURCE PARAMETERS

For a detailed analysis of the framework’s robustness, we refer to Table 16, which provides a comprehensive performance breakdown under two challenging scenarios: contamination by a single noisy source vector and aggregation of heavily pruned (95%) source vectors.

## A.4 COMPONENT SELECTION

To study our hypothesis that the most useful transferable knowledge is encapsulated within the principal singular components, we conducted a comprehensive ablation study. We evaluated the impact of different component selection and aggregation strategies on final model performance. The

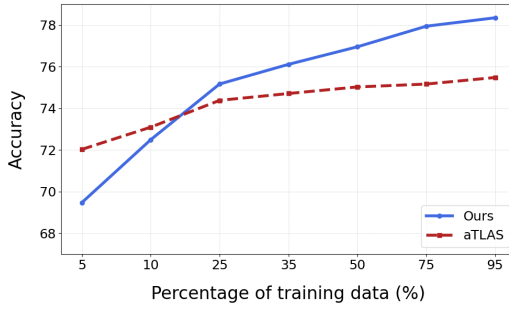


Figure 15: AXIS performs better with smaller amounts of training data in almost all cases.

goal was to ensure that our default approach, aggregating components with the highest singular values, is effective to other plausible alternatives, especially with the highest number of source task vectors. We compared the following seven strategies:

- **Top Components (our default):** As described in the main paper, we perform a global ranking of all singular components from all source tasks and select the top-K based on their singular values ( $\sigma_k$ ) to form the merged matrix  $\Delta_m = \sum_{k=1}^K u_k \sigma_k v_k^\top$ .
- **Bottom Components:** A control strategy where we select the K components with the lowest singular values from the global ranking.
- **Arbitrary Components:** A second control strategy where K components are arbitrarily selected from the global pool.
- **Average Top Components:** This baseline first distills each source task matrix  $\Delta_i$  into its top-K principal components. Next, all these resulting low-rank matrices are averaged into a single matrix. Finally, we perform a new SVD on this averaged matrix and select its top-K components to form the final  $\Delta_m$ .
- **Average Bottom Components:** The inverse of the "average top components" baseline, used as a control. First, each source task matrix is reduced to a low-rank approximation using only its own bottom-K singular components. Second, these resulting low-rank matrices are averaged, and a final selection of the bottom-K components is performed via SVD on this single, averaged matrix.
- **Equal Top Contribution:** This strategy ensures a balanced representation from all source tasks. Instead of a global ranking, it selects an equal number of the top singular components from each individual source task. If the total budget is K components and there are  $T - 1$  sources, we select the top  $K/(T - 1)$  components from each task. These are then pooled and summed to form  $\Delta_m$ .

The results, presented in Figure 17, demonstrate that the top components strategy slightly outperforms on average all other alternatives across a varying number of aggregated source tasks. For example, the top components strategy achieved an average score of 78.23 across all used task vectors, slightly edging out the equal top contribution approach, which averaged 78.19.

Additionally, we compare how different selection strategies for the top-ranking components affect accuracy when using the largest number of source task vectors, as illustrated in Figure 18. For this configuration, the top components strategy yielded the highest accuracy. These results are averaged across all target tasks. Additionally, we provided detailed results on the main aggregation strategies per target dataset in the Table 9.

#### A.5 IMPACT OF FINAL SVD

To empirically validate the importance of the final SVD re-parameterization, as discussed in the main text, we conduct a detailed ablation study. Table 5 presents a performance comparison of four different component aggregation strategies, each evaluated with and without the final SVD step.

Task Vectors	aTLAS			AXIS (ours)		
	intact	corrupted	pruned	intact	corrupted	pruned
3	71.22	61.59	68.25	77.52	77.56	77.85
4	71.86	61.41	69.43	77.85	77.70	77.99
5	72.34	60.78	70.16	77.81	77.77	78.13
6	72.95	60.38	69.50	78.02	77.76	78.28
7	73.58	60.77	71.19	78.08	77.66	78.30
8	73.86	60.38	71.42	78.33	77.82	78.28

Figure 16: Robustness to altered source task vectors. AXIS shows higher resilience to corruption and pruning compared to aTLAS.

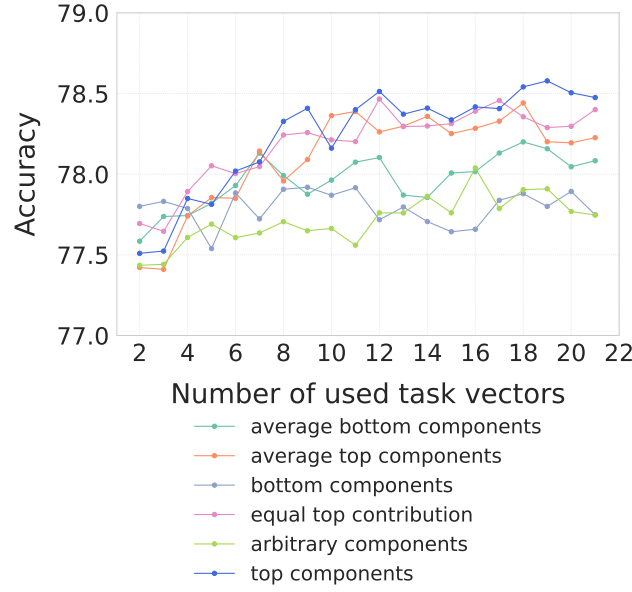


Figure 17: Performance comparison of seven different SVD component aggregation strategies  $K$  with constant  $N=10\%$ . The plot shows the average accuracy across all target tasks as the number of used source task vectors increases. Our default strategy, top components, yields the best performance with the largest number of sources.

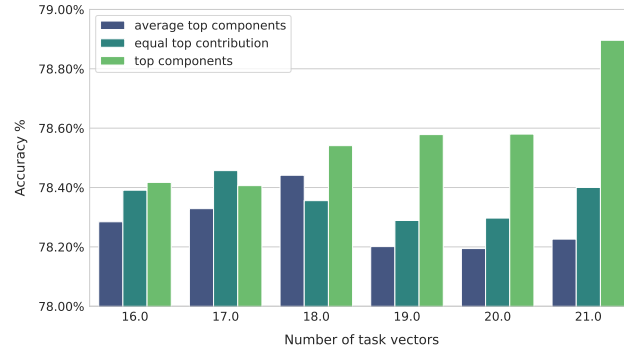


Figure 18: Detailed performance comparison of SVD component aggregation strategies, focusing on small variations within top components. While all strategies show comparable performance, the top components generally maintain a slight edge, particularly with a higher number of aggregated source tasks.

The omission of the final SVD step (denoted as ‘SVD X’) is particularly detrimental to the top components strategy, resulting in a significant performance drop (e.g., over eight percentage points when aggregating 9 task vectors). In contrast, strategies based on bottom or average components exhibit significantly higher resilience to this omission. We hypothesize that two related factors drive this phenomenon. First, the top components, representing high-magnitude task-specific knowledge, likely exhibit more substantial destructive interference when their non-orthogonal vectors are directly summed. Second, this instability may be amplified during the fine-tuning process. Without a shared orthogonal basis provided by the final SVD, the learnable parameters (a subset of singular values) may conflict with the frozen components, as their underlying vectors are not decorrelated. This could lead to an unstable optimization process where adjustments to learnable components negatively interfere with the knowledge stored in the frozen ones. The relative stability of the bottom components strategy suggests that the interference from low-magnitude components is negligible, making the final orthogonalization beneficial but not as critical.

Aggregated Task Vectors	Top		Bottom		Average top		Average bottom	
	SVD ✓	SVD ✗	SVD ✓	SVD ✗	SVD ✓	SVD ✗	SVD ✓	SVD ✗
1	77.31	77.35	77.63	77.57	77.42	77.42	77.62	77.25
2	77.51	77.38	77.80	77.65	77.42	77.41	77.58	77.23
3	77.52	76.36	77.83	77.80	77.41	77.37	77.74	77.29
4	77.85	76.49	77.79	77.61	77.74	77.75	77.74	77.33
5	77.81	76.56	77.54	77.75	77.86	77.83	77.82	77.14
6	78.02	76.39	77.88	77.85	77.85	77.95	77.93	77.35
7	78.08	76.40	77.72	77.85	78.14	78.20	78.13	77.51
8	78.33	71.16	77.91	77.84	77.96	77.98	77.99	77.53
9	78.41	69.85	77.92	77.85	78.09	78.13	77.88	77.64
10	78.16	71.58	77.87	77.84	78.36	78.24	77.96	77.50
11	78.40	78.52	77.92	77.84	78.39	78.42	78.07	77.49
12	78.51	78.39	77.72	77.80	78.26	78.28	78.10	77.34
13	78.37	78.49	77.80	77.77	78.30	78.38	77.87	77.68
14	78.41	78.37	77.71	77.72	78.36	78.20	77.86	77.54
15	78.34	78.53	77.64	77.66	78.25	78.21	78.01	77.52
16	78.42	78.57	77.66	77.76	78.28	78.29	78.02	77.36
17	78.41	78.51	77.84	77.77	78.33	78.28	78.13	77.56
18	78.54	78.60	77.88	77.76	78.44	78.30	78.20	77.70
19	78.58	78.50	77.80	77.65	78.20	78.31	78.16	77.55
20	78.50	78.45	77.89	77.79	78.19	78.18	78.05	77.64
21	78.48	78.49	77.75	77.78	78.23	78.27	78.08	77.37

Table 5: Performance comparison of different aggregation strategies with and without the final SVD step, across a varying number of aggregated task vectors and different component selection strategies.

#### A.6 DETAILED MAIN RESULTS

For a comprehensive and granular evaluation of our proposed framework, Tables 6–8 present a detailed, per-dataset comparison of AXIS and the aTLAS baseline.

STV	Method	CIFAR100	CIFAR10	CUB200	Caltech101	Caltech256	Cars	Country211
1	aTLAS (N=10%)	72.95	93.76	54.47	89.86	85.10	61.21	17.69
	aTLAS (N=20%)	73.62	94.15	55.38	91.65	85.53	62.12	17.92
	aTLAS (N=40%)	75.09	95.20	56.80	93.38	87.59	63.77	18.05
	AXIS (N=10%)	77.00	95.85	57.61	93.89	88.44	63.54	17.70
	AXIS (N=20%)	79.28	96.63	60.15	94.41	89.19	65.58	18.39
	AXIS (N=40%)	81.45	97.10	62.50	94.99	89.38	65.94	18.64
5	aTLAS (N=10%)	73.90	94.52	54.83	91.53	85.43	62.06	17.78
	aTLAS (N=20%)	74.77	95.17	55.94	92.68	87.59	62.53	18.02
	aTLAS (N=40%)	75.29	95.31	56.85	93.78	88.06	63.89	18.17
	AXIS (N=10%)	77.51	96.50	58.41	93.61	88.01	63.95	18.17
	AXIS (N=20%)	79.96	96.84	59.22	94.70	89.48	67.23	18.60
	AXIS (N=40%)	82.28	97.13	62.46	94.24	89.89	69.69	18.86
10	aTLAS (N=10%)	78.92	96.40	55.11	91.88	86.21	62.37	18.06
	aTLAS (N=20%)	79.68	96.58	55.78	93.72	86.82	62.90	18.24
	aTLAS (N=40%)	80.65	96.90	55.47	94.82	88.29	64.15	18.41
	AXIS (N=10%)	80.09	96.96	57.85	94.82	88.76	64.66	18.24
	AXIS (N=20%)	81.31	97.10	59.58	94.82	89.53	67.07	18.08
	AXIS (N=40%)	82.64	97.49	61.74	94.64	89.20	69.92	19.22
15	aTLAS (N=10%)	78.95	96.46	55.89	94.70	88.11	62.04	18.16
	aTLAS (N=20%)	79.81	96.81	57.08	95.22	89.19	64.15	18.30
	aTLAS (N=40%)	80.62	97.19	57.82	96.08	89.38	64.88	18.51
	AXIS (N=10%)	80.14	96.85	58.68	94.64	88.65	65.43	18.31
	AXIS (N=20%)	81.55	97.25	60.94	95.56	89.89	66.86	18.48
	AXIS (N=40%)	82.83	97.38	63.00	95.28	90.33	69.99	19.24
21	aTLAS (N=10%)	78.91	96.53	55.85	94.53	88.81	63.29	18.07
	aTLAS (N=20%)	79.94	96.79	57.46	94.64	89.43	64.21	18.36
	aTLAS (N=40%)	80.84	97.14	58.01	95.28	89.89	65.09	18.32
	AXIS (N=10%)	80.11	96.93	58.46	94.99	88.76	65.09	18.48
	AXIS (N=20%)	81.69	97.13	61.10	94.64	89.95	66.88	18.58
	AXIS (N=40%)	82.96	97.39	62.70	95.45	90.75	70.77	19.42

Table 6: Detailed results per target dataset for various numbers of source task vectors (STV). Part 1 of 3.

STV	Method	DTD	EuroSAT	FGVCAircraft	Flowers102	Food101	GTSRB	MNIST
1	aTLAS (N=10%)	48.78	88.81	22.62	67.39	85.11	54.90	82.44
	aTLAS (N=20%)	51.49	90.85	23.64	67.96	85.09	59.20	84.84
	aTLAS (N=40%)	56.97	95.04	24.75	70.25	85.73	78.45	93.38
	AXIS (N=10%)	67.02	97.30	29.70	77.49	85.81	89.57	97.36
	AXIS (N=20%)	70.80	97.70	30.66	81.15	86.28	93.20	98.46
	AXIS (N=40%)	74.15	98.30	19.65	81.20	86.93	94.22	98.76
5	aTLAS (N=10%)	53.03	94.11	22.86	68.56	85.27	66.85	89.08
	aTLAS (N=20%)	54.04	94.48	24.15	68.26	85.41	71.35	91.97
	aTLAS (N=40%)	58.67	95.44	24.83	69.58	85.86	79.96	93.44
	AXIS (N=10%)	65.69	97.41	30.48	77.22	86.05	90.74	97.78
	AXIS (N=20%)	70.96	97.63	33.75	80.09	86.62	93.45	98.57
	AXIS (N=40%)	73.09	98.22	16.83	82.84	87.02	94.51	98.81
10	aTLAS (N=10%)	55.96	95.59	24.18	69.02	85.27	77.00	95.42
	aTLAS (N=20%)	59.57	95.93	24.54	69.60	85.71	83.70	96.44
	aTLAS (N=40%)	64.26	96.93	26.70	72.30	85.94	88.06	97.25
	AXIS (N=10%)	68.14	98.00	31.95	76.65	86.15	90.02	98.02
	AXIS (N=20%)	70.85	98.33	29.70	79.10	86.49	93.61	98.54
	AXIS (N=40%)	71.91	98.19	19.20	77.82	87.07	94.73	98.96
15	aTLAS (N=10%)	56.44	95.15	24.93	70.22	85.60	78.31	96.15
	aTLAS (N=20%)	60.21	96.11	25.86	73.61	85.99	83.08	96.94
	aTLAS (N=40%)	62.71	96.81	28.14	74.48	86.17	87.39	97.06
	AXIS (N=10%)	67.82	97.78	31.05	77.25	86.18	91.00	98.20
	AXIS (N=20%)	70.59	98.19	34.92	82.09	86.61	93.67	98.70
	AXIS (N=40%)	71.38	98.26	39.15	83.67	87.11	94.76	98.89
21	aTLAS (N=10%)	56.44	95.07	25.62	71.23	85.72	78.02	95.98
	aTLAS (N=20%)	60.37	96.26	26.37	72.09	85.91	83.45	96.94
	aTLAS (N=40%)	63.24	96.96	26.25	75.09	86.29	88.38	97.58
	AXIS (N=10%)	67.98	97.81	30.75	77.87	86.32	91.06	98.11
	AXIS (N=20%)	70.64	98.22	34.50	82.31	86.57	93.46	98.64
	AXIS (N=40%)	72.18	98.52	38.97	83.74	87.15	94.43	98.96

Table 7: Detailed results per target dataset for various numbers of source task vectors (STV). Part 2 of 3.



STV	Method	OxfordIIITPet	PascalVOC	RESISC45	STL10	SUN397	SVHN	UCF101
1	aTLAS (N=10%)	90.19	82.99	71.19	97.99	64.42	62.10	65.05
	aTLAS (N=20%)	90.73	84.21	72.14	98.16	64.95	67.11	65.61
	aTLAS (N=40%)	90.71	86.51	80.40	98.49	66.16	86.49	68.94
	AXIS (N=10%)	89.92	85.77	87.51	97.65	66.80	86.63	71.00
	AXIS (N=20%)	89.86	86.77	89.95	97.80	68.48	89.76	74.91
	AXIS (N=40%)	90.24	86.53	91.84	97.08	70.05	91.35	77.98
5	aTLAS (N=10%)	90.62	85.49	74.56	97.91	64.85	83.23	65.95
	aTLAS (N=20%)	91.31	86.16	77.16	98.35	65.43	84.09	67.94
	aTLAS (N=40%)	91.99	86.72	80.43	98.34	66.29	86.34	68.86
	AXIS (N=10%)	90.60	86.71	87.90	97.74	67.27	90.87	71.45
	AXIS (N=20%)	90.19	87.09	90.41	97.65	68.69	92.18	76.37
	AXIS (N=40%)	90.27	86.99	91.90	97.26	69.75	92.87	78.03
10	aTLAS (N=10%)	91.77	86.19	79.13	98.24	66.33	85.66	67.57
	aTLAS (N=20%)	91.61	86.63	82.16	98.21	66.76	87.45	69.36
	aTLAS (N=40%)	91.50	87.11	84.87	98.24	67.04	89.06	71.66
	AXIS (N=10%)	90.11	86.50	88.38	97.73	67.28	87.92	73.17
	AXIS (N=20%)	90.32	87.05	90.48	97.59	68.91	90.65	77.37
	AXIS (N=40%)	89.53	86.75	92.75	96.86	70.41	92.61	78.09
15	aTLAS (N=10%)	91.63	86.87	78.79	98.50	66.43	85.62	68.68
	aTLAS (N=20%)	92.78	87.39	82.40	98.70	67.48	87.66	70.90
	aTLAS (N=40%)	92.18	87.62	84.97	98.53	67.82	89.14	72.51
	AXIS (N=10%)	91.09	86.92	87.94	98.13	67.61	88.24	73.17
	AXIS (N=20%)	91.03	87.64	90.54	97.89	68.70	89.97	75.87
	AXIS (N=40%)	90.22	87.18	92.22	97.68	70.56	92.91	77.64
21	aTLAS (N=10%)	92.23	87.11	80.52	98.36	66.63	86.83	70.05
	aTLAS (N=20%)	92.61	87.56	81.25	98.55	66.99	87.69	71.35
	aTLAS (N=40%)	92.91	88.15	84.40	98.55	67.88	89.14	73.09
	AXIS (N=10%)	91.25	87.25	88.25	98.05	67.62	88.56	74.28
	AXIS (N=20%)	90.81	87.46	90.86	97.95	68.96	90.38	77.35
	AXIS (N=40%)	90.71	86.97	91.97	97.30	70.29	92.64	79.96

Table 8: Detailed results per target dataset for various numbers of source task vectors (STV). Part 3 of 3.

TV	Strategy	CIFAR100	CIFAR10	CUB200	Catsch101	Catsch256	Cats	Country211	DTD	EuroSAT	FGVCAircraft	Flowers102	Food101	GTSRB	MNIST	OxfordIIITPet	PascalVOC	RESISC45	STL10	SUN397	SVHN	UCF101
1	bottom components	76.58	95.95	58.15	94.64	88.52	64.07	17.59	67.87	97.78	29.61	79.62	85.20	87.87	97.31	91.09	86.78	88.73	98.20	66.90	84.79	72.98
	top components	77.00	95.85	57.61	93.89	88.44	63.54	17.70	67.02	97.30	29.70	77.49	85.81	89.57	97.36	89.92	85.77	87.51	97.65	66.80	86.63	71.00
2	bottom components	77.27	95.96	58.42	94.70	88.52	63.97	17.82	65.48	97.74	29.73	79.31	85.68	88.58	97.67	91.44	87.20	88.16	98.15	67.44	86.38	74.17
	arbitrary components	77.14	95.97	57.99	94.64	87.90	63.56	17.77	65.37	97.70	29.52	77.67	85.69	88.79	97.55	91.52	87.17	87.14	98.08	66.64	85.68	72.64
	top components	77.81	96.18	57.44	93.84	87.70	63.51	17.75	67.45	96.89	29.43	75.18	85.88	90.40	97.83	89.94	86.61	88.11	97.96	66.90	87.81	73.06
3	bottom components	77.51	95.75	58.78	95.10	88.32	64.63	17.75	66.17	97.67	29.85	80.52	85.71	88.27	97.63	91.14	86.89	88.00	98.06	67.20	85.68	73.80
	arbitrary components	77.46	96.12	57.90	94.41	87.85	63.79	17.88	66.17	97.93	29.49	78.06	85.54	88.36	97.57	90.76	86.36	87.73	98.10	66.43	86.51	71.85
	top components	77.37	96.08	57.59	93.84	87.99	64.21	17.84	66.17	97.37	30.03	77.07	85.79	89.72	97.69	90.02	86.28	87.48	97.71	67.60	87.28	72.85
4	bottom components	77.36	96.11	58.35	94.47	88.37	64.18	17.91	67.02	97.63	30.30	80.19	85.77	88.61	97.75	91.11	86.87	87.37	98.21	67.11	85.56	73.25
	arbitrary components	77.17	96.13	58.32	94.70	87.99	63.96	17.81	66.22	97.11	30.39	79.35	85.56	87.78	97.69	91.01	86.75	87.19	98.21	67.06	86.28	73.09
	top components	78.32	96.01	57.89	93.15	88.32	64.23	17.91	66.86	97.85	29.82	77.72	85.88	90.32	98.00	90.11	86.79	87.41	97.84	67.35	89.26	73.80
5	bottom components	77.62	95.83	57.70	94.59	88.48	64.08	17.78	65.90	97.41	29.49	78.86	85.74	88.95	97.50	90.73	86.73	88.02	98.31	66.86	84.98	72.75
	arbitrary components	77.04	96.19	58.01	94.12	87.98	64.07	17.73	66.70	97.74	30.18	77.82	85.75	89.49	97.81	90.49	87.05	88.02	98.21	66.63	87.38	73.09
	top components	77.51	96.50	58.41	93.61	88.01	63.95	18.17	65.69	97.41	30.48	77.22	86.05	90.74	97.78	90.60	86.71	87.90	97.74	67.27	90.87	71.45
6	bottom components	78.01	95.96	58.25	94.82	88.73	64.28	18.22	66.81	97.70	30.42	79.48	85.89	88.87	97.70	91.28	87.17	88.02	98.19	66.94	86.08	72.75
	arbitrary components	77.54	96.17	57.99	94.53	88.08	64.21	17.82	65.32	97.89	30.45	78.48	85.77	88.43	97.50	90.84	86.69	87.92	98.01	67.09	87.58	71.42
	top components	77.63	96.11	58.46	94.35	88.68	65.03	18.45	66.91	97.56	30.09	76.94	85.91	90.73	98.02	90.38	86.31	88.19	97.91	67.90	90.84	72.01
7	bottom components	78.04	95.93	58.44	94.87	88.52	64.51	17.72	65.37	97.59	30.03	78.37	85.94	88.57	97.39	91.28	87.07	88.02	98.08	66.88	86.09	73.43
	arbitrary components	77.92	95.83	57.94	95.28	87.83	64.35	17.99	65.96	97.63	29.82	78.37	85.90	88.73	97.81	90.65	86.97	87.79	98.28	66.92	85.92	73.38
	top components	78.12	96.20	57.99	94.41	88.22	64.72	17.82	67.61	97.81	30.51	77.54	86.15	91.43	98.36	89.97	86.53	88.22	97.80	67.43	90.80	71.95
8	bottom components	78.16	95.97	58.13	94.53	88.58	64.59	17.96	65.80	97.81	30.36	80.24	85.74	88.50	97.35	91.47	86.91	87.95	98.41	67.41	86.21	73.94
	arbitrary components	77.56	96.07	57.34	93.95	88.01	63.87	18.00	66.44	97.67	30.21	78.63	85.69	88.73	97.61	91.41	86.73	87.62	98.00	67.24	86.44	74.60
	top components	79.05	96.45	58.42	93.84	88.91	64.64	18.04	66.65	97.59	30.75	79.13	86.15	91.44	98.39	90.73	86.88	88.48	97.80	67.43	90.96	73.14
9	bottom components	78.29	96.12	58.30	93.95	88.39	64.51	17.92	66.65	97.74	30.90	80.00	85.79	88.16	97.70	91.50	87.15	87.90	98.26	67.14	86.12	73.80
	arbitrary components	77.95	96.13	56.94	94.64	88.50	64.11	17.85	67.45	96.96	29.43	78.50	85.86	88.49	97.85	90.57	87.03	87.84	98.20	67.04	86.20	73.09
	top components	79.42	96.83	58.47	95.74	88.40	64.57	18.32	67.50	97.59	30.93	78.08	86.17	92.24	98.50	90.32	86.87	88.46	97.59	67.11	91.00	72.46
10	bottom components	77.86	96.12	57.49	95.10	88.66	64.36	18.03	66.01	98.04	30.39	80.37	85.95	88.28	97.59	91.50	87.13	87.56	98.11	67.37	85.85	73.49
	arbitrary components	77.27	96.08	57.99	94.64	88.32	63.98	17.91	65.37	97.44	30.54	79.51	85.96	88.23	97.83	90.98	87.06	87.71	98.11	67.31	85.98	72.69
	top components	80.09	96.96	57.85	94.82	88.76	64.66	18.24	68.14	98.00	31.95	76.65	86.15	90.02	98.02	90.11	86.50	88.38	97.73	67.28	87.92	73.17
11	bottom components	77.84	95.81	57.80	93.84	88.75	64.48	18.06	67.34	97.81	30.78	80.00	85.95	87.68	97.80	91.47	87.00	88.03	98.29	67.39	86.34	73.78
	arbitrary components	78.16	96.12	57.13	93.72	88.66	64.02	17.82	65.43	97.93	29.64	79.05	86.02	89.25	97.73	90.81	86.97	86.83	97.99	67.12	86.79	71.58
	top components	80.14	96.97	59.11	95.39	88.48	64.51	18.39	66.97	97.96	30.45	78.55	86.15	90.39	98.17	90.98	87.37	87.92	97.78	67.47	88.99	74.25
12	bottom components	77.83	95.69	58.06	94.24	88.52	64.06	17.99	66.91	97.81	29.94	79.31	85.82	88.84	97.76	90.52	87.09	87.89	98.18	67.48	86.09	72.03
	arbitrary components	78.18	96.29	57.58	93.89	88.44	64.22	17.75	67.66	97.85	30.93	78.18	86.01	87.66	97.72	90.92	86.89	87.43	98.08	67.06	86.60	73.62
	top components	80.11	96.92	58.99	95.45	88.61	64.54	18.34	68.40	98.15	31.65	78.44	86.19	90.32	98.28	90.71	87.07	88.60	98.04	67.44	88.71	73.80
13	bottom components	77.82	95.89	58.53	94.18	88.86	64.12	18.20	67.55	97.93	29.70	78.73	85.90	87.64	97.57	91.50	87.07	87.84	98.31	67.42	86.25	72.72
	arbitrary components	78.41	95.96	58.06	95.05	88.26	64.37	18.09	66.49	97.78	29.19	79.61	85.82	88.60	97.76	90.52	86.92	87.29	98.20	67.38	86.35	72.88
	top components	80.05	96.83	58.85	94.70	88.81	65.10	18.38	67.39	98.19	30.84	76.81	86.17	90.28	98.18	91.36	87.26	88.54	98.05	67.55	88.84	73.62
14	bottom components	77.75	95.82	58.16	94.76	88.91	64.26	18.04	67.07	97.52	29.37	78.96	85.84	88.27	97.67	90.98	86.62	87.92	98.39	67.28	85.74	72.51
	arbitrary components	78.21	96.00	57.63	93.95	88.32	64.57	18.09	67.29	97.59	30.69	79.92	86.02	88.14	97.88	91.01	87.08	87.78	98.08	66.98	86.53	73.38
	top components	80.10	96.83	58.60	95.33	88.78	64.66	18.31	67.29	98.04	30.93	77.46	86.23	90.68	98.22	91.09	87.06	88.73	98.25	67.56	88.73	73.94
15	bottom components	78.01	95.79	58.01	93.95	88.96	64.23	17.93	67.13	97.19	29.43	78.63	85.92	88.47	97.64	90.65	87.01	87.78	98.43	67.27	85.55	72.51
	arbitrary components	77.94	96.24	57.27	94.53	88.71	64.28	17.79	65.90	97.85	30.33	78.66	85.97	88.90	97.92	90.73	87.02	88.24	98.21	66.95	86.53	72.98
	top components	80.14	96.85	58.68	94.64	88.65	65.43	18.31	67.82	97.78	31.05	77.25	86.18	91.00	98.20	91.09	86.92	87.94	98.13	67.61	88.24	73.17
16	bottom components	77.94	95.90	57.73	94.30	88.48	64.66	18.21	65.32	98.00	29.88	79.74	85.93	88.27	97.56	91.06	87.13	87.65	98.28	67.39	85.97	72.32
	arbitrary components	78.40	96.26	57.70	94.70	88.29	64.59	18.22	67.98	97.78	30.63	80.09	86.06	88.70	97.85	91.11	86.76	88.16	98.31	67.05	87.45	72.72
	top components	80.13	96.78	59.06	94.47	88.88	65.30	18.28	68.09	97.74	30.63	78.37	86.23	90.58	98.03	90.79	87.16	87.76	97.91	67.62	88.64	74.31
17	bottom components	77.64	95.87	57.85	94.99	88.30	64.40	17.89	66.49	97.70	30.45	80.47	85.99	88.18	97.78	90.73	87.35	88.03	98.35	67.60	85.91	72.64
	arbitrary components	78.15	96.27	57.90	94.47	88.14	64.61	17.97	65.74	97.59	30.06	79.35	86.03	88.47	97.70	91.25	87.23	87.78	98.21	66.98	86.42	73.20
	top components	80.17	96.91	58.77	95.45	88.89	65.03	18.27	66.76	97.89	30.12	78.96	86.23	90.82	98.10	91.20	86.91	87.75	97.95	67.67	88.38	74.31
18	bottom components	77.55	96.01	58.47	94.18	88.91	64.71	17.88	66.81	98.04	29.37	80.89	85.93	89.04	97.51	90.65	87.09	87.68	98.39	67.50	86.62	72.22
	arbitrary components	78.34	96.14	57.85	94.07	88.37	65.10	18.05	66.76	97.44	29.04	80.68	86.10	88.92	97.77	91.41	87.23	87.27	98.36	67.20	86.65	73.22
	top components	80.00	96.86	58.70	94.87	89.10	65.10	18.22	68.78	97.74	31.95	77.74	86.17	90.67	98.19	91.20	86.90	88.37	97.89	67.58	88.29	75.05
19	bottom components	7																				