Training-Free Pretrained Model Merging
——*Supplementary Material*——

	Avg Acc	Prime	Odd
M. A	71.71	98.81	44.60
M. B	68.41	38.13	98.68
Avg	63.99	58.61	69.37
Rebasin	71.71	66.64	76.77
A. Align	90.79	91.71	89.87
Ours$_{A.}$	**90.93**	**91.72**	**90.14**
Zipit	92.95	93.65	92.24
W. Zip	89.35	87.90	90.79
Ours$_{Z.}$	**94.62**	**94.51**	**94.72**

Table S1. The per-task accuracy on multi-task MNIST Dataset.

S1. Details for Figure 1

For Figure 1, the experiments are conducted on ResNet-50. We first train it on CIFAR10 to get the first parent, and then retrain only the 5-th convolution layer (64 units) to obtain the second parent. Then we average the two sets of units in the 5-th convolution layer pairwise, getting $64 \times 64 = 4,096$ merged units. The "best merged" unit is the one whose similarity to the parents is maximized among the $4,096$ units. As each merged unit has two parents, "low bound" here means the smaller similarity value between the merged unit and its two parents. We use Pearson correlation to measure similarity.

S2. Additional Results

Original models and ensemble methods. In Table S3, Table S4 and Table S2, we provide the results of original models and ensemble methods for the experiments in Section 4.1 as the reference.

Models used in Section 4.3. The architecture of the model is a four-layer MLP, and each hidden layer has 1024 units. Each model is trained as CLIP image encoder. The per-task accuracy is shown in Table S1. We provide the average results of 5 different random seeds.

More advanced ViTs. Here we provide the results on DINO [2] and Swin-Transformer [3] in Table S7. It can seen that the proposed method consistently outperforms the two competitors, again validate the superiority of our method.

Method	Joint Acc	Avg Acc	T. A	T. B
Model A	41.86	45.22	77.15	13.28
Model B	40.81	45.14	13.30	76.98
Ensemble	51.75	77.06	77.15	76.98

Table S2. Results of two original models and ensemble method for Table 3

Statistical significance. Here we provide some stds of our methods and the two competitors in Table S5 and Table S6, which proves the significance of the improvements.

S3. Convergence of MuDSC

Algorithm 1 adopts a well-proved iterative algorithm [1], where each iteration increases the similarity until it converges.

S4. Complexity of MuDSC

As solving Eq. 1 dominates the computation of Alg. 1, here we simply discuss the complexity of solving Eq. 1. The activation-based methods are single-round methods as they can solve Eq. 1 in just one round, while weight-based methods and the proposed MuDSC are multi-round methods as they solve Eq. 1 in an iterative manner. Tab. S8 provides some results. It can be seen that MuDSCs needs less rounds than prior weight-based methods if two models are trained from scratch.

References

[1] Samuel Ainsworth, Jonathan Hayase, and Siddhartha Srinivasa. Git re-basin: Merging models modulo permutation symmetries. In *The Eleventh International Conference on Learning Representations*, 2023. 1

[2] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021. 1

[3] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hi-

Model	Resnet20								Resnet20GN			
Dataset	CIFAR100(50+50)				CIFAR10(5+5)				CIFAR100(50+50)			
Method	Joint	Avg	T. A	T. B	Joint	Avg	T. A	T. B	Joint	Avg	T. A	T. B
Model A	41.48	53.44	82.71	24.18	48.59	70.98	96.67	45.29	38.70	49.16	77.11	21.22
Model B	41.30	53.20	23.99	82.41	48.58	72.31	47.51	97.12	38.64	49.00	20.87	77.13
Ensemble	69.51	82.56	82.71	82.41	84.12	96.89	96.67	97.12	63.18	77.12	77.10	77.13

Table S3. Results of original models and ensemble methods for Table 1

Model	Resnet26				Resnet50GN				ViT			
Method	Joint	Avg	T. A	T. B	Joint	Avg	T. A	T. B	Joint	Avg	T. A	T. B
Model A	42.89	54.31	84.72	23.89	45.02	56.99	89.24	24.75	47.57	58.10	93.05	23.14
Model B	43.05	54.41	23.46	85.36	45.09	57.32	25.31	89.33	47.17	58.28	23.71	92.86
Ensemble	71.43	85.04	84.72	85.36	76.88	89.28	89.23	89.32	82.69	92.95	93.05	92.86

Table S4. Results of original models and ensemble methods for Table 2

Model	Resnet20						Resnet20GN		
Dataset	CIFAR100(50+50)			CIFAR10(5+5)			CIFAR100(50+50)		
Method	Joint	T. A	T. B	Joint	T. A	T. B	Joint	T. A	T. B
Rebasin	$41.33_{\pm 1.52}$	$57.31_{\pm 1.23}$	$56.58_{\pm 0.28}$	$60.61_{\pm 0.14}$	$88.46_{\pm 0.18}$	$88.68_{\pm 0.69}$	$13.85_{\pm 0.14}$	$22.99_{\pm 0.36}$	$21.37_{\pm 0.42}$
A. Align	$44.33_{\pm 0.13}$	$61.61_{\pm 0.17}$	$60.66_{\pm 1.46}$	$\mathbf{61.71}_{\pm 0.13}$	$88.63_{\pm 0.06}$	$\mathbf{89.78}_{\pm 0.52}$	$29.37_{\pm 1.07}$	$41.05_{\pm 1.44}$	$43.05_{\pm 0.85}$
MuDSC$_{Align}$	$\mathbf{45.50}_{\pm 0.38}$	$\mathbf{63.06}_{\pm 0.46}$	$\mathbf{62.56}_{\pm 0.48}$	$60.84_{\pm 0.14}$	$\mathbf{89.04}_{\pm 0.25}$	$89.63_{\pm 0.21}$	$\mathbf{31.84}_{\pm 0.60}$	$\mathbf{45.34}_{\pm 1.17}$	$\mathbf{45.29}_{\pm 0.79}$
Zipit	$54.69_{\pm 0.15}$	$67.11_{\pm 0.77}$	$66.44_{\pm 0.68}$	$82.44_{\pm 0.76}$	$94.22_{\pm 0.14}$	$95.00_{\pm 0.95}$	$29.93_{\pm 1.09}$	$39.99_{\pm 0.73}$	$42.41_{\pm 0.86}$
W.Zip	$55.16_{\pm 0.20}$	$68.58_{\pm 0.29}$	$66.71_{\pm 0.08}$	$82.85_{\pm 0.13}$	$94.42_{\pm 0.15}$	$94.99_{\pm 0.11}$	$14.28_{\pm 1.07}$	$19.17_{\pm 0.78}$	$22.72_{\pm 0.88}$
MuDSC$_{Zip}$	$\mathbf{56.01}_{\pm 0.25}$	$\mathbf{68.80}_{\pm 0.05}$	$\mathbf{67.47}_{\pm 0.33}$	$\mathbf{83.09}_{\pm 0.13}$	$\mathbf{94.56}_{\pm 0.24}$	$\mathbf{95.21}_{\pm 0.39}$	$\mathbf{30.05}_{\pm 0.39}$	$\mathbf{40.39}_{\pm 0.70}$	$\mathbf{42.65}_{\pm 0.56}$

Table S5. Results of MuDSC in Table 1 including std.

Model	Resnet26			Resnet50GN			ViT		
Method	Joint	T. A	T. B	Joint	T. A	T. B	Joint	T. A	T. B
Rebasin	$61.39_{\pm 0.31}$	$74.48_{\pm 0.36}$	$75.10_{\pm 0.18}$	$74.52_{\pm 0.12}$	$85.06_{\pm 0.32}$	$84.50_{\pm 0.13}$	$\mathbf{70.16}_{\pm 0.15}$	$84.32_{\pm 0.05}$	$84.32_{\pm 0.02}$
A. Align	$61.91_{\pm 0.44}$	$75.03_{\pm 0.43}$	$75.79_{\pm 0.34}$	$74.44_{\pm 0.13}$	$84.99_{\pm 0.04}$	$84.56_{\pm 0.01}$	$69.99_{\pm 0.16}$	$84.20_{\pm 0.07}$	$84.24_{\pm 0.12}$
MuDSC$_{Align}$	$\mathbf{62.84}_{\pm 0.50}$	$\mathbf{75.87}_{\pm 0.38}$	$\mathbf{76.40}_{\pm 0.35}$	$\mathbf{74.66}_{\pm 0.09}$	$\mathbf{85.25}_{\pm 0.07}$	$\mathbf{84.58}_{\pm 0.03}$	$70.09_{\pm 0.03}$	$\mathbf{84.38}_{\pm 0.13}$	$\mathbf{84.40}_{\pm 0.04}$
Zipit	$60.23_{\pm 0.70}$	$73.20_{\pm 0.89}$	$74.17_{\pm 0.71}$	$72.05_{\pm 0.52}$	$83.06_{\pm 0.29}$	$82.92_{\pm 0.12}$	$68.57_{\pm 0.16}$	$82.79_{\pm 0.12}$	$83.30_{\pm 0.18}$
W.Zip	$61.28_{\pm 0.06}$	$74.42_{\pm 0.23}$	$74.96_{\pm 0.08}$	$74.52_{\pm 0.12}$	$85.06_{\pm 0.32}$	$84.50_{\pm 0.13}$	$\mathbf{70.16}_{\pm 0.15}$	$84.32_{\pm 0.05}$	$84.32_{\pm 0.02}$
MuDSC$_{Zip}$	$\mathbf{61.58}_{\pm 0.27}$	$\mathbf{74.61}_{\pm 0.24}$	$\mathbf{75.41}_{\pm 0.38}$	$\mathbf{74.71}_{\pm 0.01}$	$\mathbf{85.14}_{\pm 0.01}$	$\mathbf{84.62}_{\pm 0.03}$	$70.10_{\pm 0.03}$	$\mathbf{84.41}_{\pm 0.08}$	$\mathbf{84.36}_{\pm 0.05}$

Table S6. Results of MuDSC in Table 2 including std.

Model	DINO-S			Swin-T		
Method	Joint	T. A	T. B	Joint	T. A	T. B
Rebasin	$66.22_{\pm 0.25}$	$81.07_{\pm 0.10}$	$78.37_{\pm 0.43}$	$75.41_{\pm 0.08}$	$87.46_{\pm 0.16}$	$84.99_{\pm 0.29}$
A. Align	$61.28_{\pm 1.73}$	$76.92_{\pm 1.33}$	$74.46_{\pm 1.61}$	$67.86_{\pm 0.56}$	$81.42_{\pm 0.57}$	$79.12_{\pm 0.47}$
MuDSC A.	$\mathbf{66.25}_{\pm 0.25}$	$\mathbf{81.24}_{\pm 0.10}$	$\mathbf{79.20}_{\pm 0.5}$	$\mathbf{75.76}_{\pm 0.12}$	$\mathbf{87.78}_{\pm 0.16}$	$\mathbf{85.56}_{\pm 0.32}$
Zipit	$59.73_{\pm 0.22}$	$79.90_{\pm 0.15}$	$74.24_{\pm 0.56}$	$63.39_{\pm 0.81}$	$75.22_{\pm 1.22}$	$73.09_{\pm 0.75}$
W. Zip	$66.21_{\pm 0.25}$	$81.11_{\pm 0.10}$	$78.37_{\pm 0.44}$	$75.41_{\pm 0.08}$	$87.46_{\pm 0.16}$	$84.99_{\pm 0.29}$
MuDSC Z.	$\mathbf{66.66}_{\pm 0.29}$	$\mathbf{81.27}_{\pm 0.09}$	$\mathbf{79.35}_{\pm 0.48}$	$\mathbf{75.73}_{\pm 0.10}$	$\mathbf{87.74}_{\pm 0.15}$	$\mathbf{85.52}_{\pm 0.33}$

Table S7. CIFAR100 Results on DINO and Swin-Transformer.

	Align Time/Round	# Rounds			Zip Time/Round	# Rounds		
		A. Align	Rebasin	MuDSC A.		Zipit	W. Align	MuDSC Z.
Resnet20	0.12sec	1	10	5	2.76sec	1	6	5
Resnet26(pre)	0.42sec	1	4	3	16.12sec	1	3	3
ViT-S(pre)	0.97sec	1	2	3	1.02min	1	3	3

Table S8. Time to solve Eq. 1 and the rounds to converge.

erarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 1