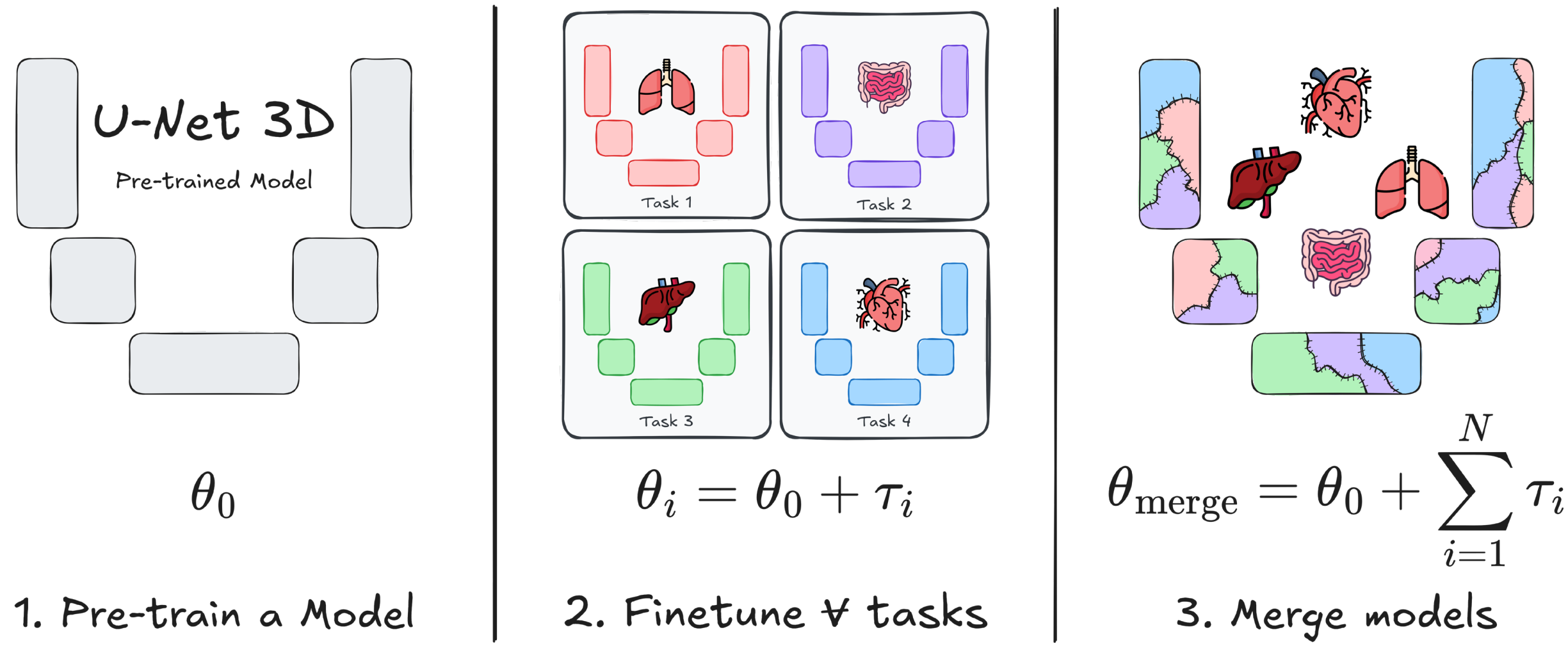




### 1. Model Merging

Model merging is a technique that combines the weights of multiple neural network models into a single unified model that inherits strengths from each component.



Starting from a pretrained model, we can finetune it multiple times for each different task. We can then merge all these models into a single one, able to perform all the tasks in a single forward pass.

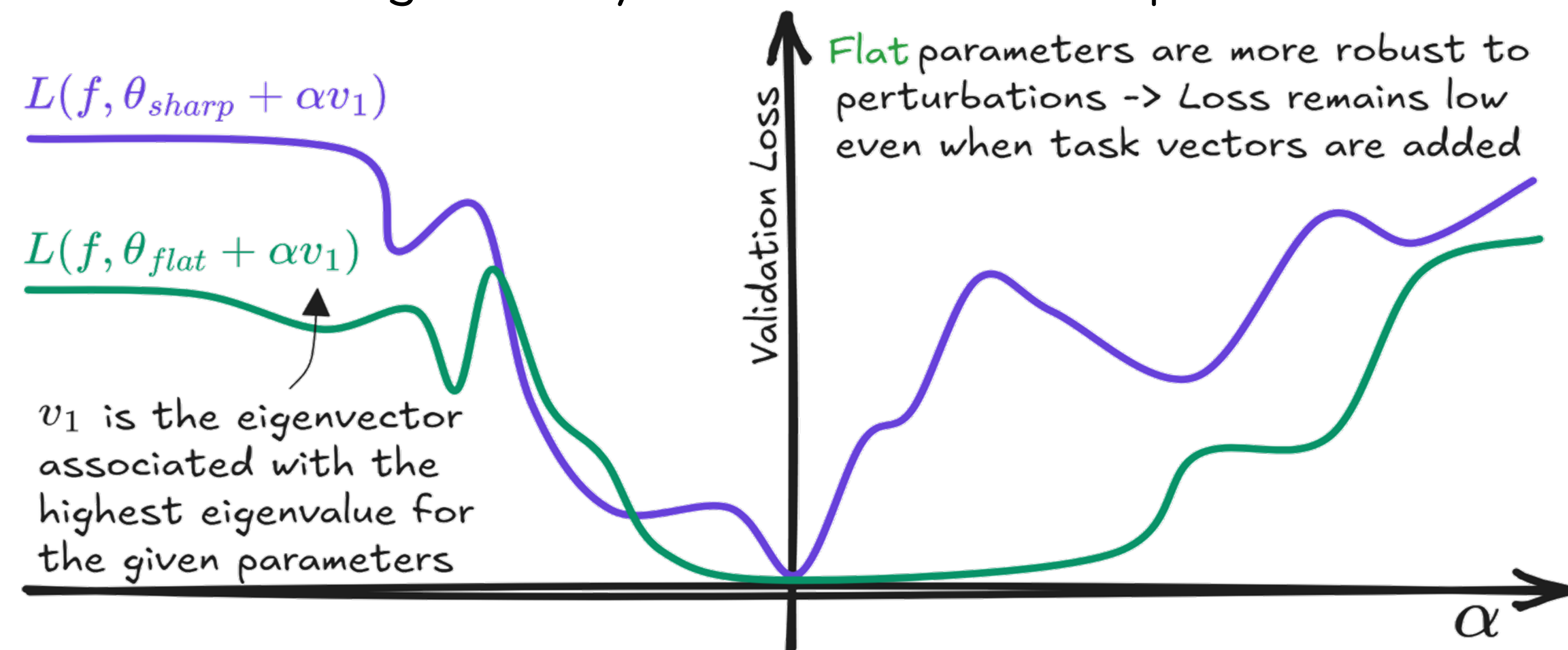
### 2. Role of Pretraining

Research has focused only on post-training phase, about task vectors, by proposing different merging methods and advanced algorithms to reduce task-wise interference.

Can we do something during the pretraining to ease the merging process?

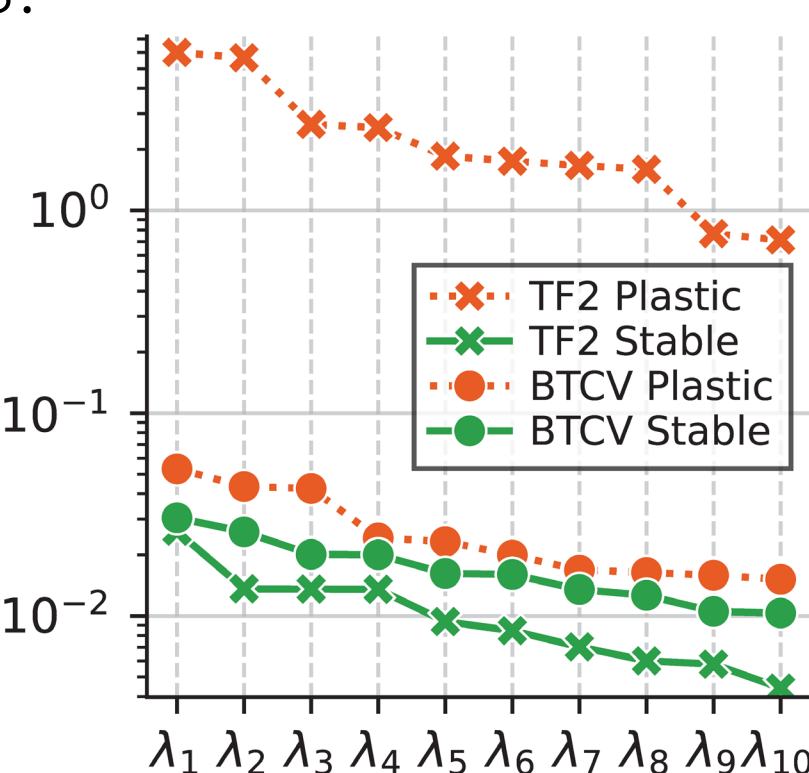
### 3. Biasing towards Wider Minima

The entire spectrum of eigenvalues of the Hessian matrix is related to the geometry of the loss landscape.



A larger maximum eigenvalue suggests that the loss landscape is steeper along at least one dimension, which corresponds to a sharper minimum. Smaller eigenvalues suggest wider minima because the surface of the loss changes less drastically in those directions.

Regime	Dataset	BS	DO	LR	Dice↑	∑λᵢ↓	λ₁↓
Stable	Cui	4	0.5	10 <sup>-3</sup>	34.93	0.57	0.02
Plastic	Cui	8	0.0	10 <sup>-4</sup>	42.68	40.71	6.00
Stable	AMOS	4	0.5	10 <sup>-3</sup>	43.76	2.30	0.03
Plastic	AMOS	8	0.0	10 <sup>-4</sup>	46.87	58.46	0.05



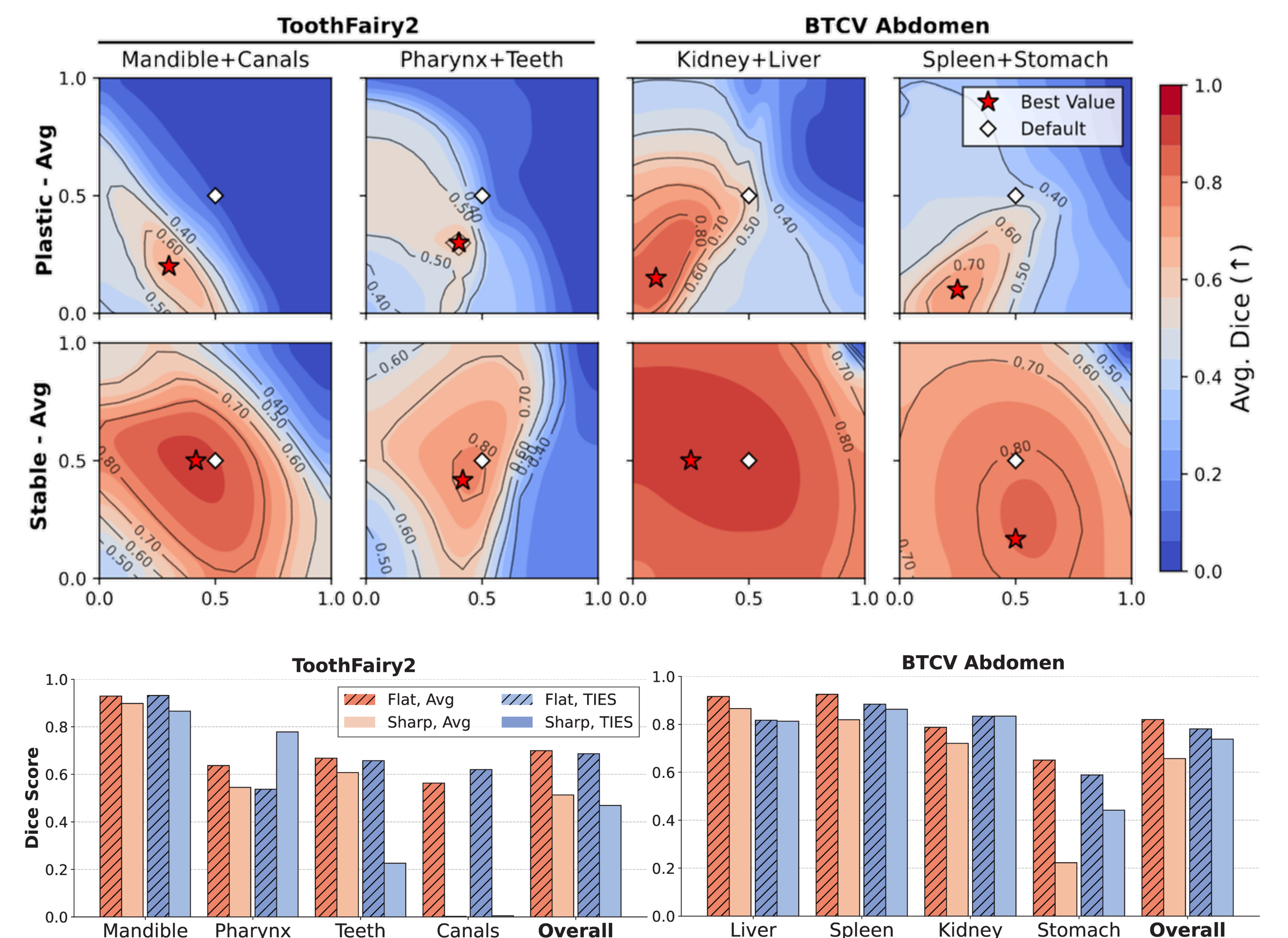
### 4. Datasets

We chose four datasets, two for pretraining and two for finetuning different tasks. Two of them are about the abdominal region (AMOS and BTCV Abdomen), two are about the maxillofacial area (Cui and ToothFairy2).

Dataset	Modality	Volumes	Structs	Shape
AMOS [12] (pre-training)	CT	240	15	148 × 533 × 560
BTCV Abdomen [15]		30	13	125 × 512 × 512
Cui [6] (pre-training)	CBCT	151	42	322 × 402 × 402
ToothFairy2 [3]		480	42	169 × 356 × 375

### 5. Wide Minima Improve Merging

Models finetuned from a stable pretraining phase show better performance and greater stability when merged.



### 6. Results

Results show that merging two tasks yields results similar to performing a joint training of the two tasks simultaneously and that we can merge up to four tasks without degrading performance significantly.

Dataset	w	Merging Strategy	Spl. Kid.	Spl. Liv.	Spl. Sto.	Kid. Liv.	Kid. Sto.	Liv. Sto.	Avg.
BTCV Abdomen	Default ◇	Average [10]	91.41	92.18	80.61	90.85	77.18	80.91	85.52
		TIES [26]	82.80	90.76	76.83	88.69	58.56	76.69	79.05
	Best ★	Average [10]	92.64	92.22	82.09	91.01	78.97	81.80	86.45
		TIES [26]	92.42	91.88	81.55	91.07	77.72	81.04	85.95
-	-	Joint	91.40	93.34	78.38	92.31	91.79	88.86	89.35
Dataset	w	Merging Strategy	Mand. IACs	Mand. Teeth	Mand. Phar.	IACs Teeth	IACs Phar.	Teeth Phar.	Avg.
ToothFairy2	Default ◇	Average [10]	89.54	82.55	87.70	56.08	57.08	81.27	75.70
		TIES [26]	88.70	79.89	88.96	58.51	63.44	73.72	75.54
	Best ★	Average [10]	89.67	82.78	91.89	68.16	75.19	81.27	81.49
		TIES [26]	89.24	82.33	91.97	67.94	68.91	81.07	80.24
-	-	Joint	98.75	97.33	98.26	83.04	97.10	93.61	94.68