

MAPEX: Modality-Aware Pruning of Experts for Remote Sensing Foundation Models

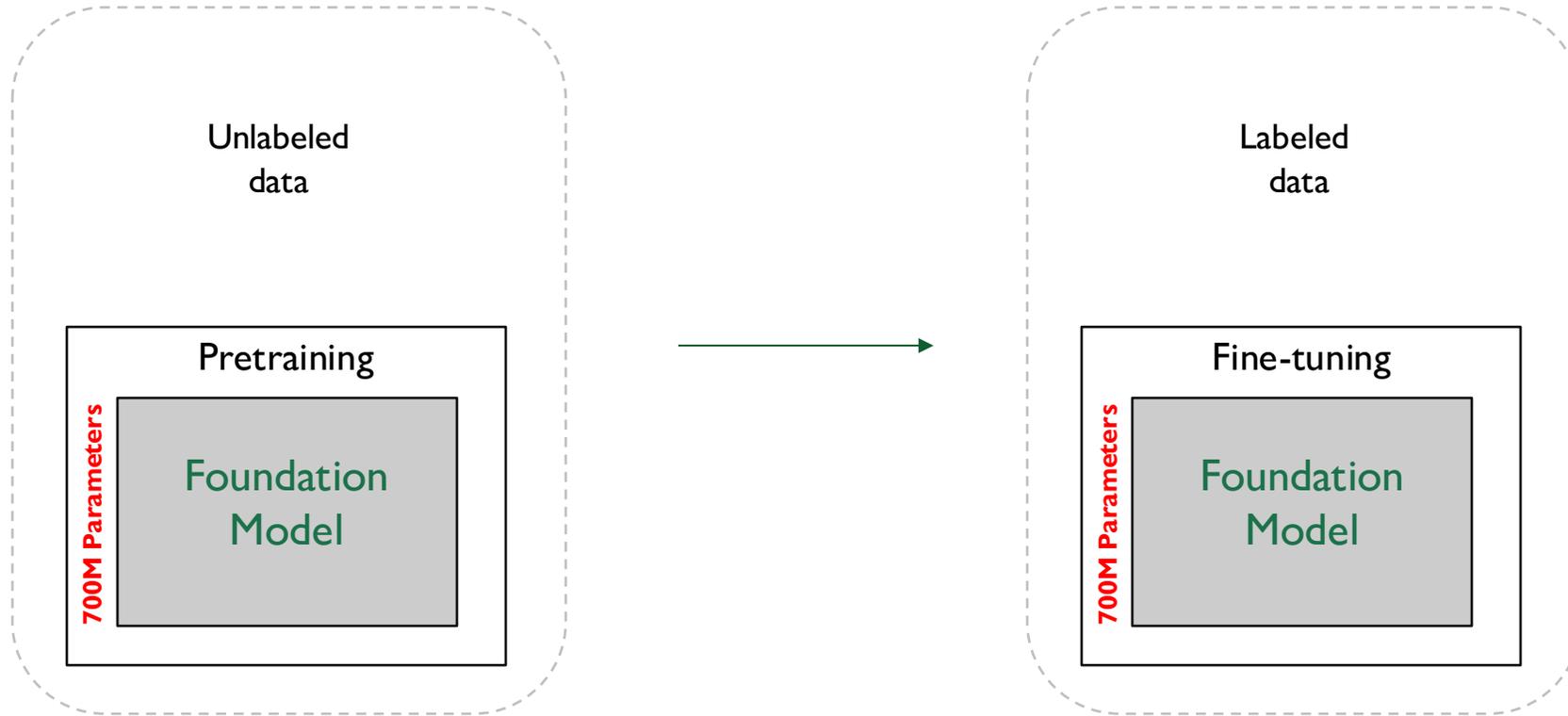
Joëlle Hanna*, Linus Scheibenreif*, Damian Borth

AIML Lab, University of St. Gallen, Switzerland

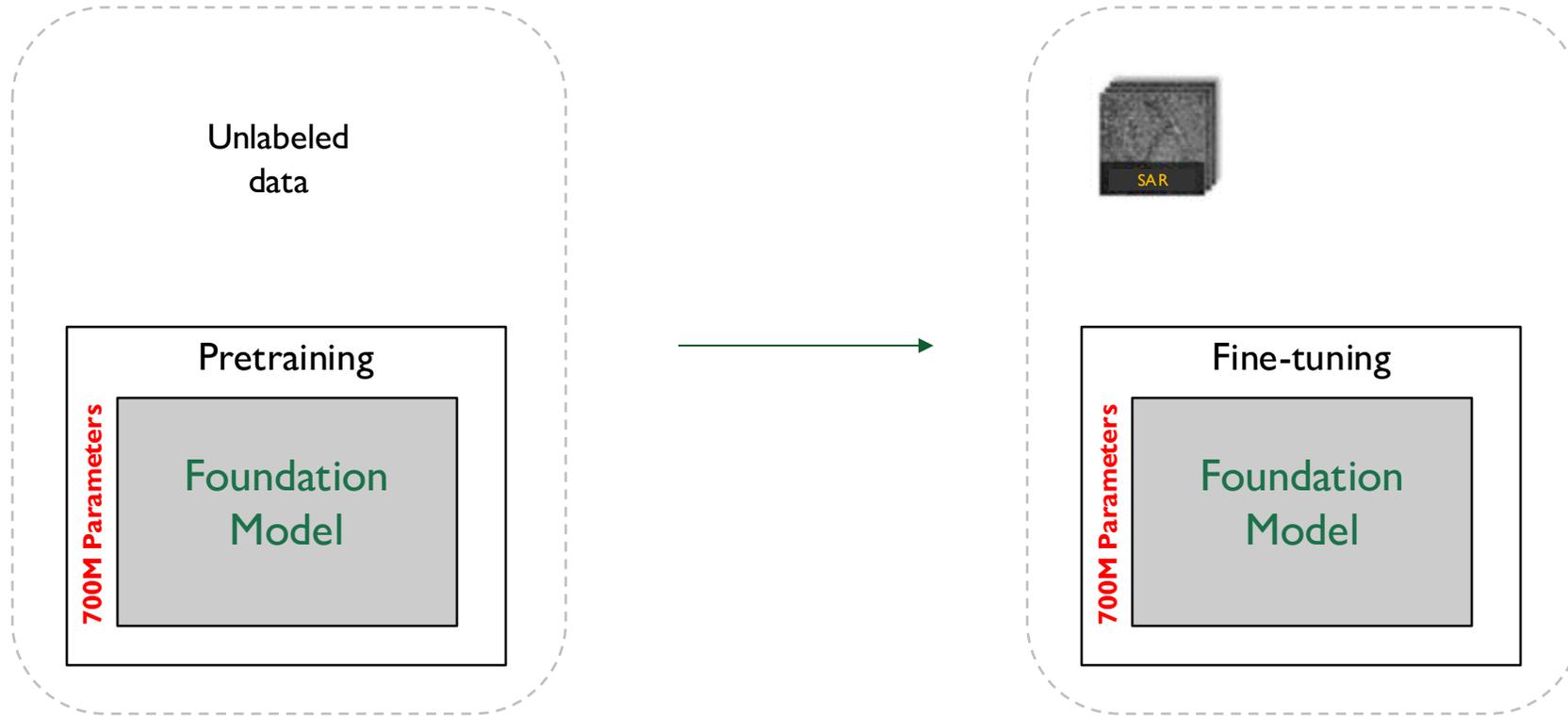
ESA-NASA International Workshop on AI Foundation Model for EO
Frascati, Italy

*equal contribution

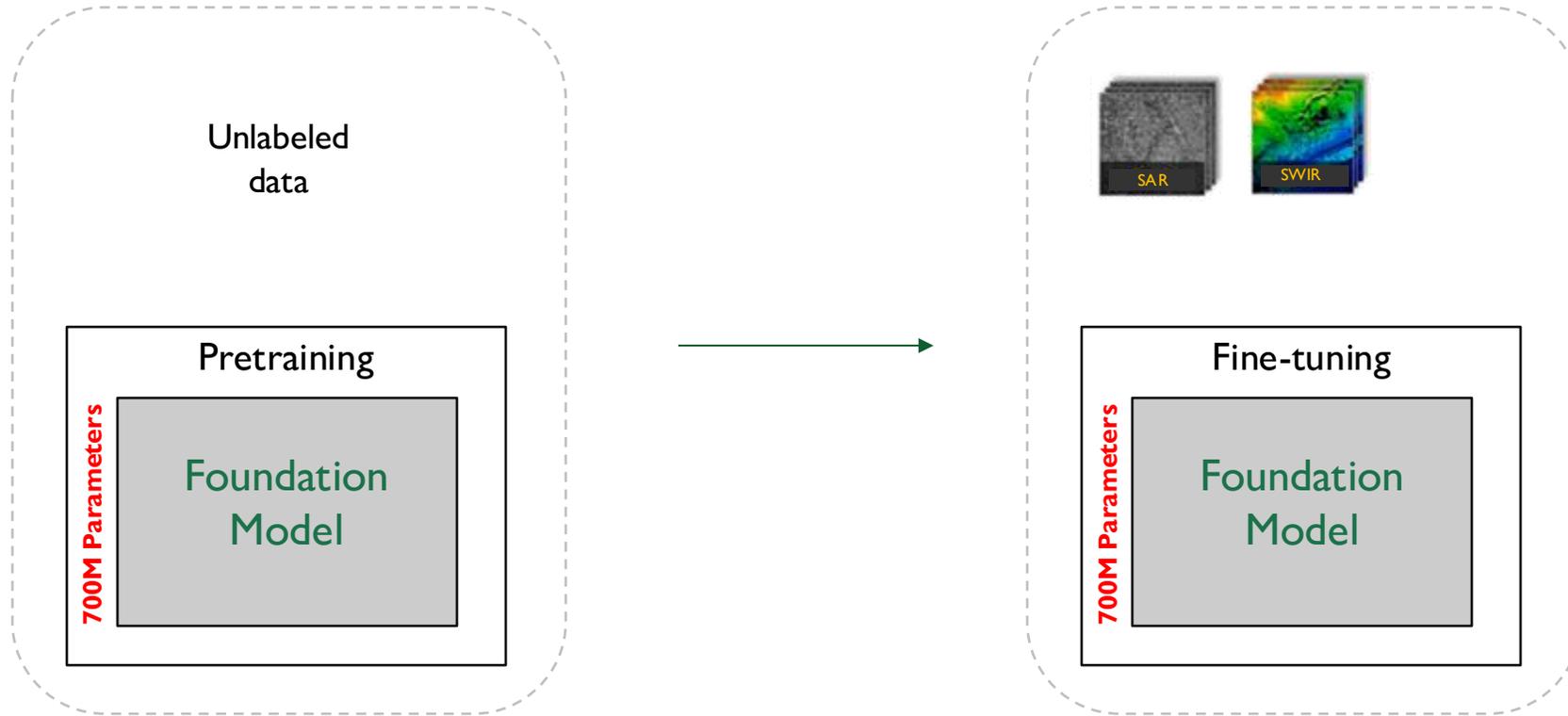
The Modality Mismatch Problem



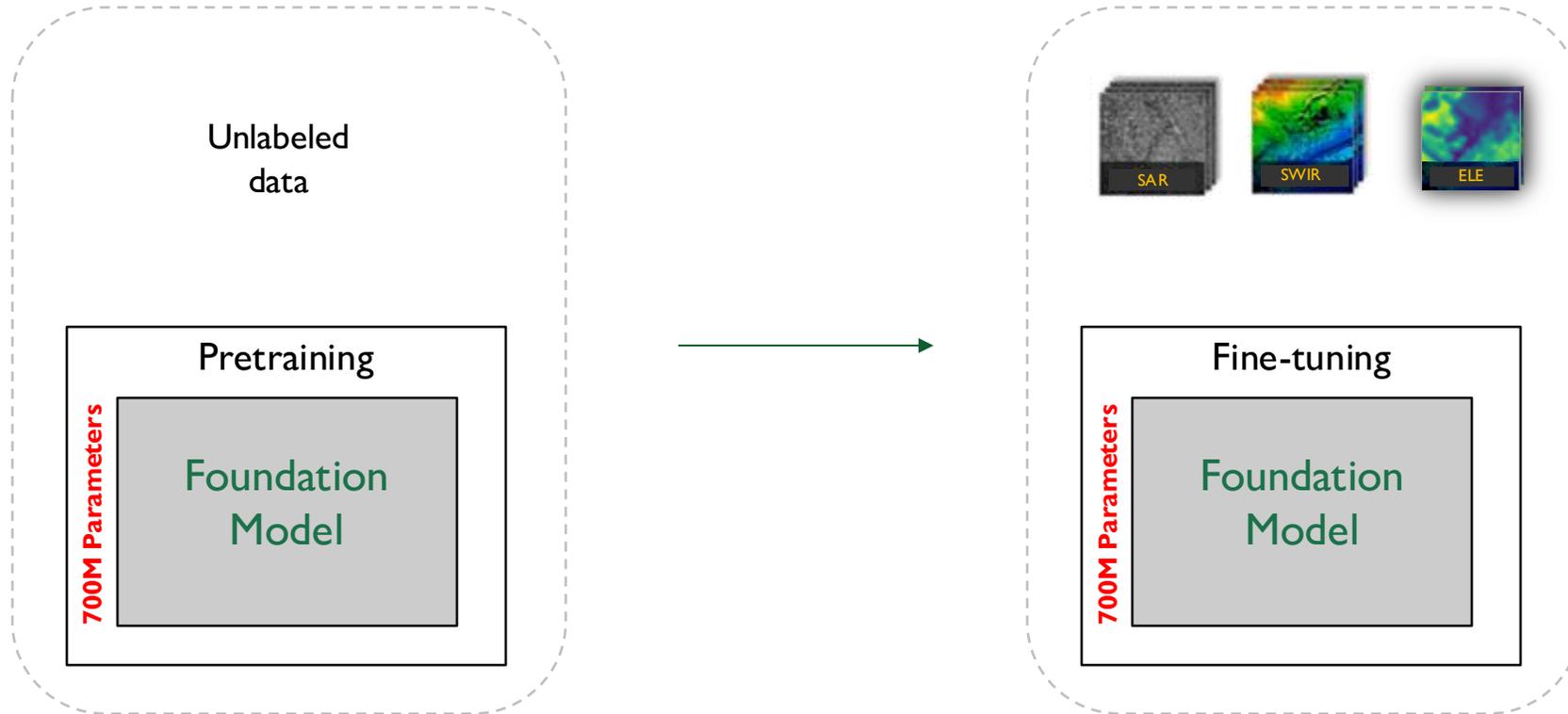
The Modality Mismatch Problem



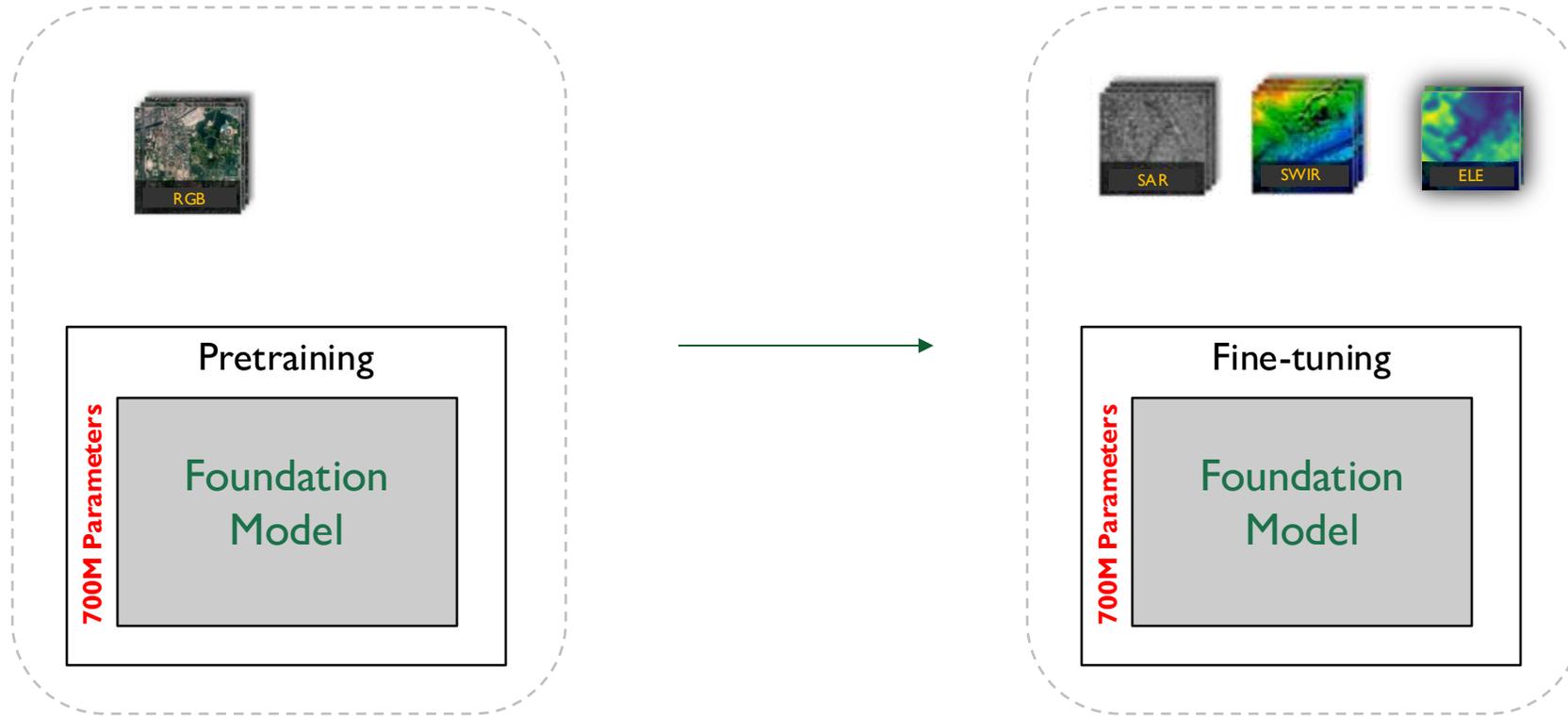
The Modality Mismatch Problem



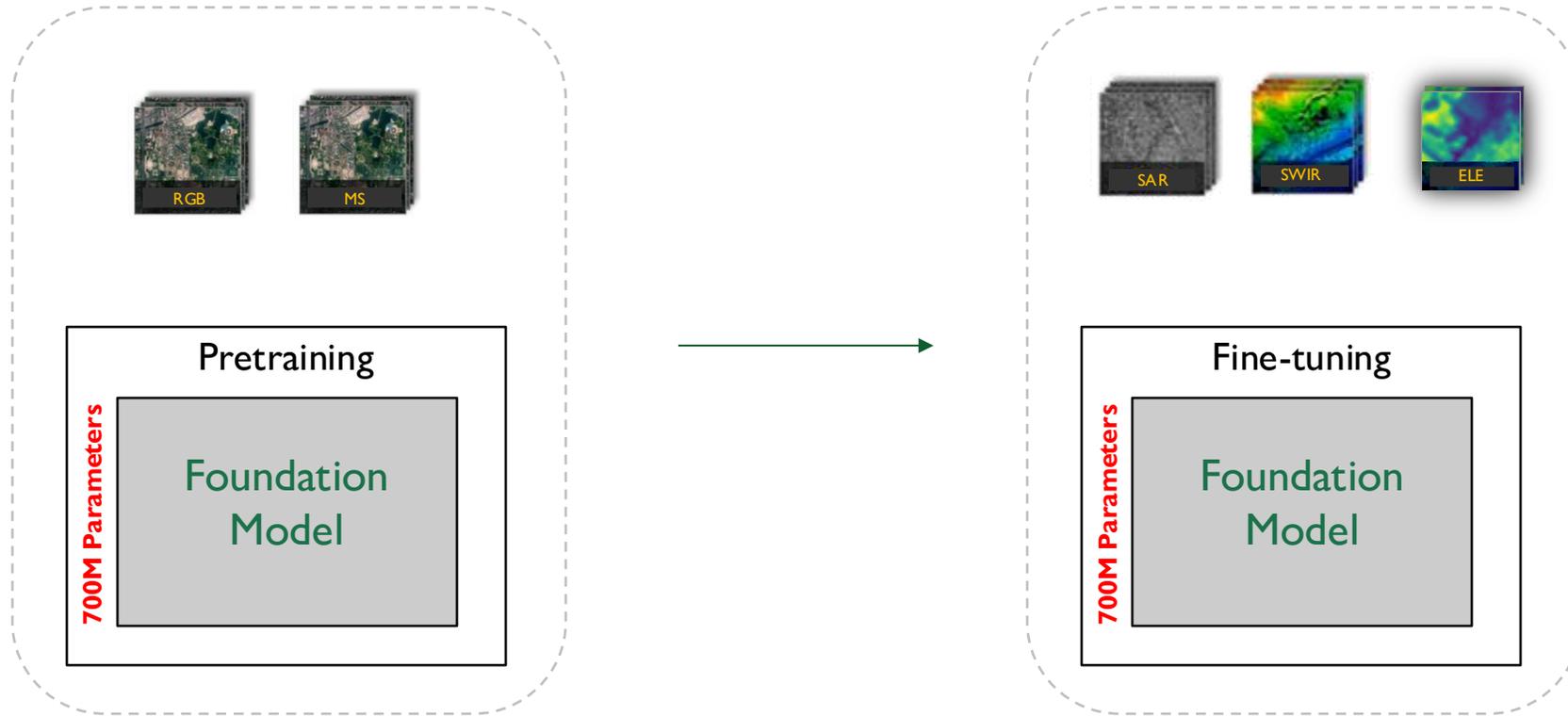
The Modality Mismatch Problem



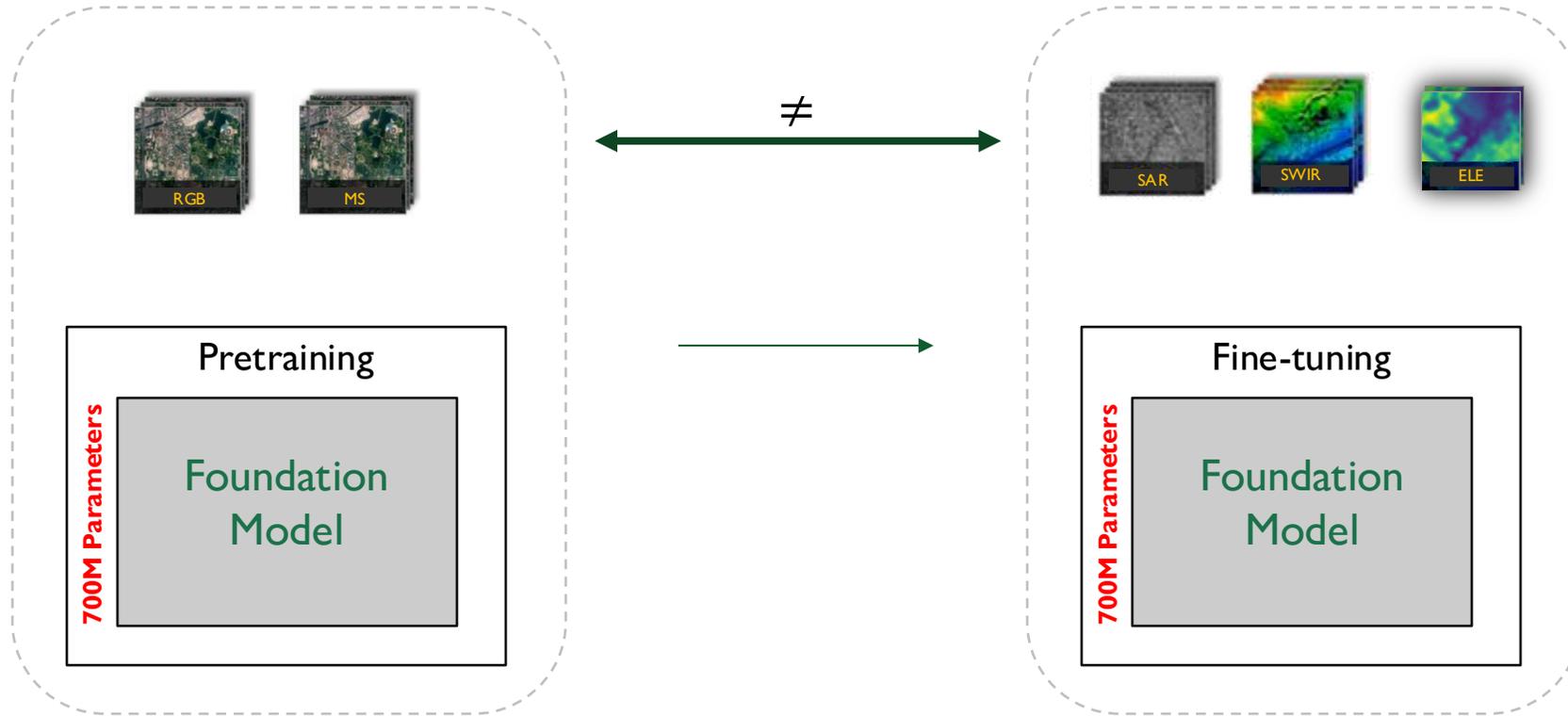
The Modality Mismatch Problem



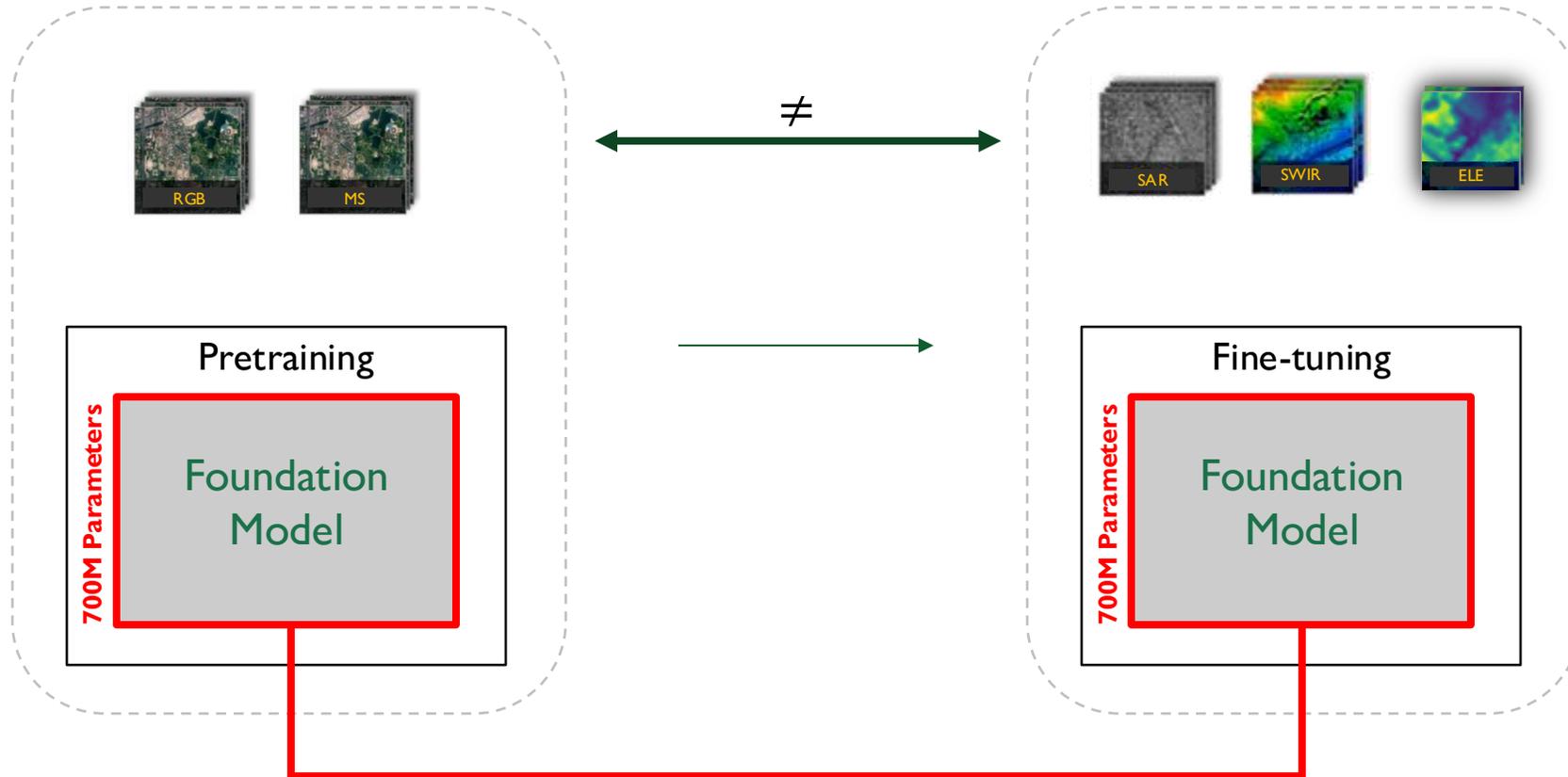
The Modality Mismatch Problem



The Modality Mismatch Problem

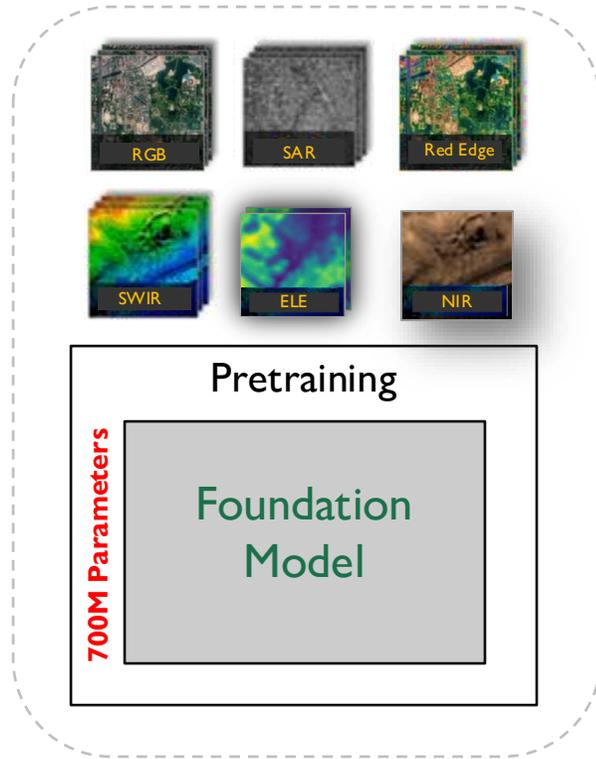


The Modality Mismatch Problem

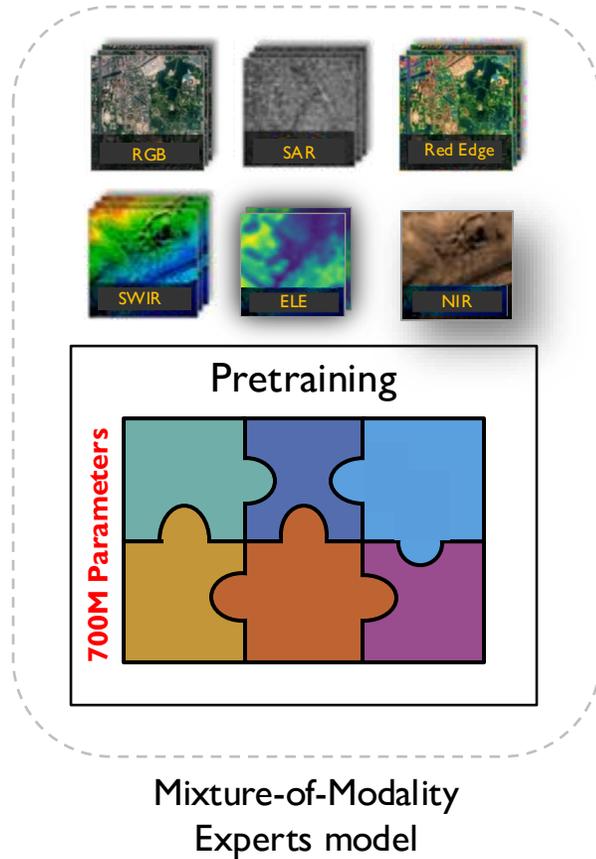


Pretraining leads to large models \neq downstream applications require lightweight, task-specific ones

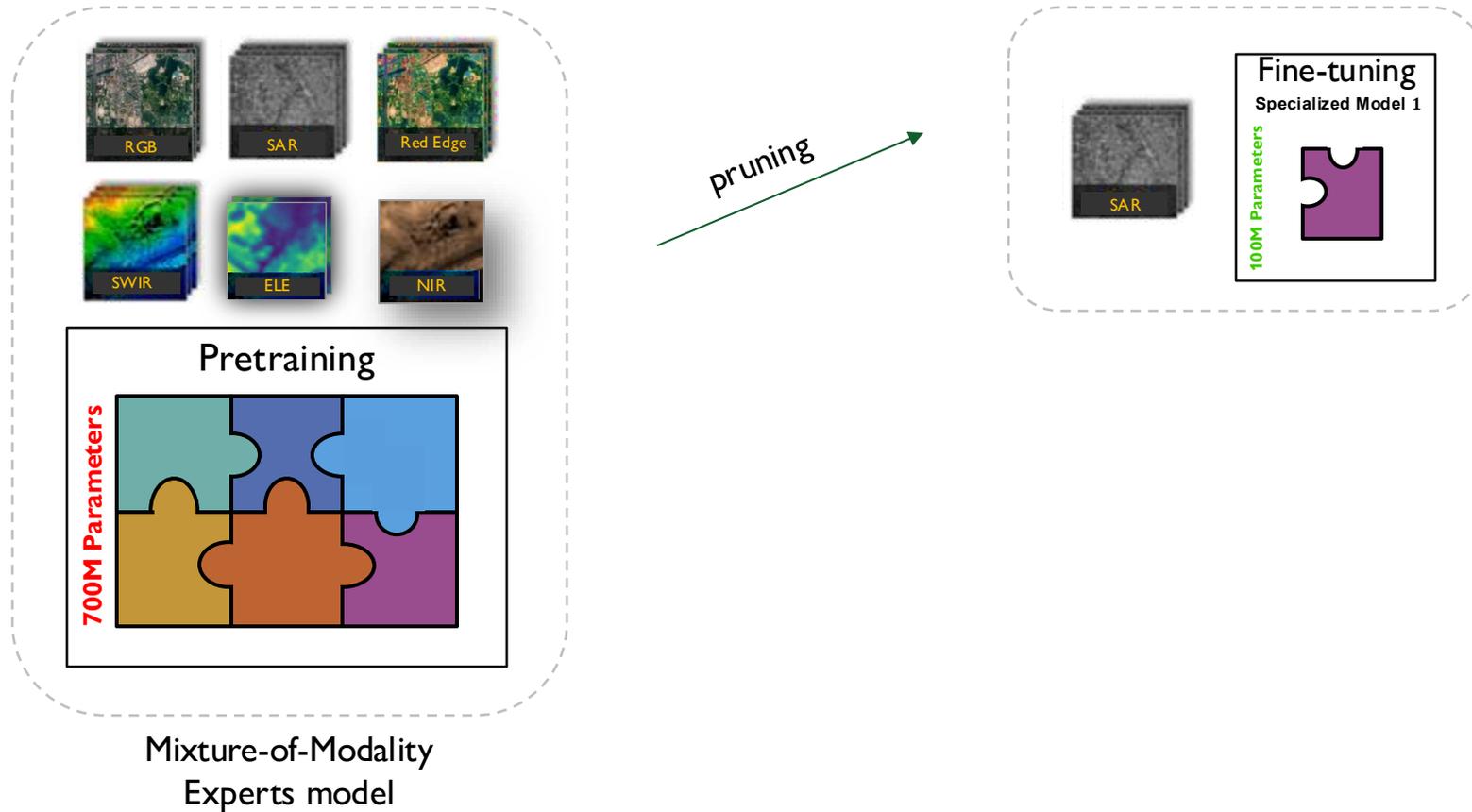
From Foundation to Lightweight Specialization



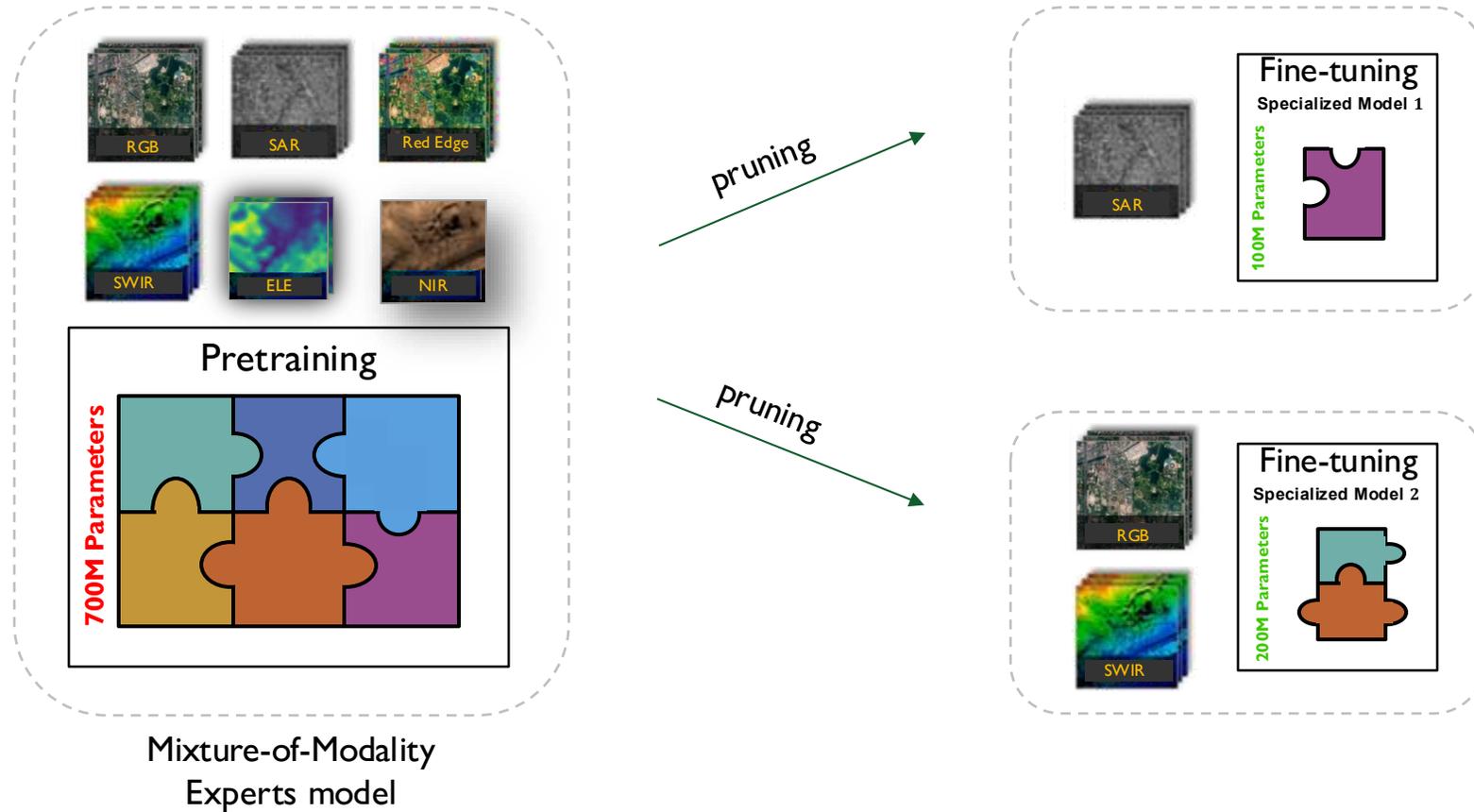
From Foundation to Lightweight Specialization



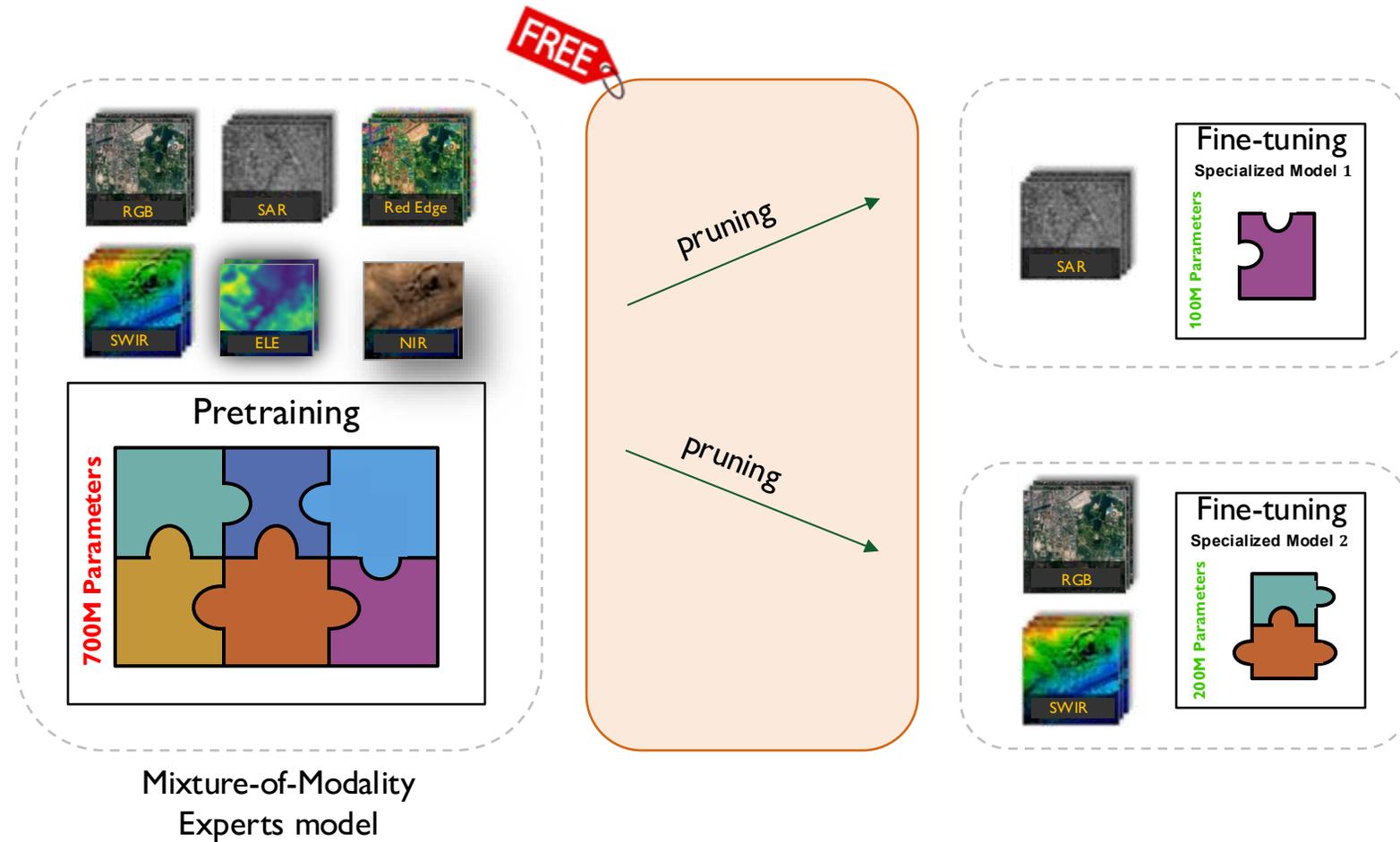
From Foundation to Lightweight Specialization



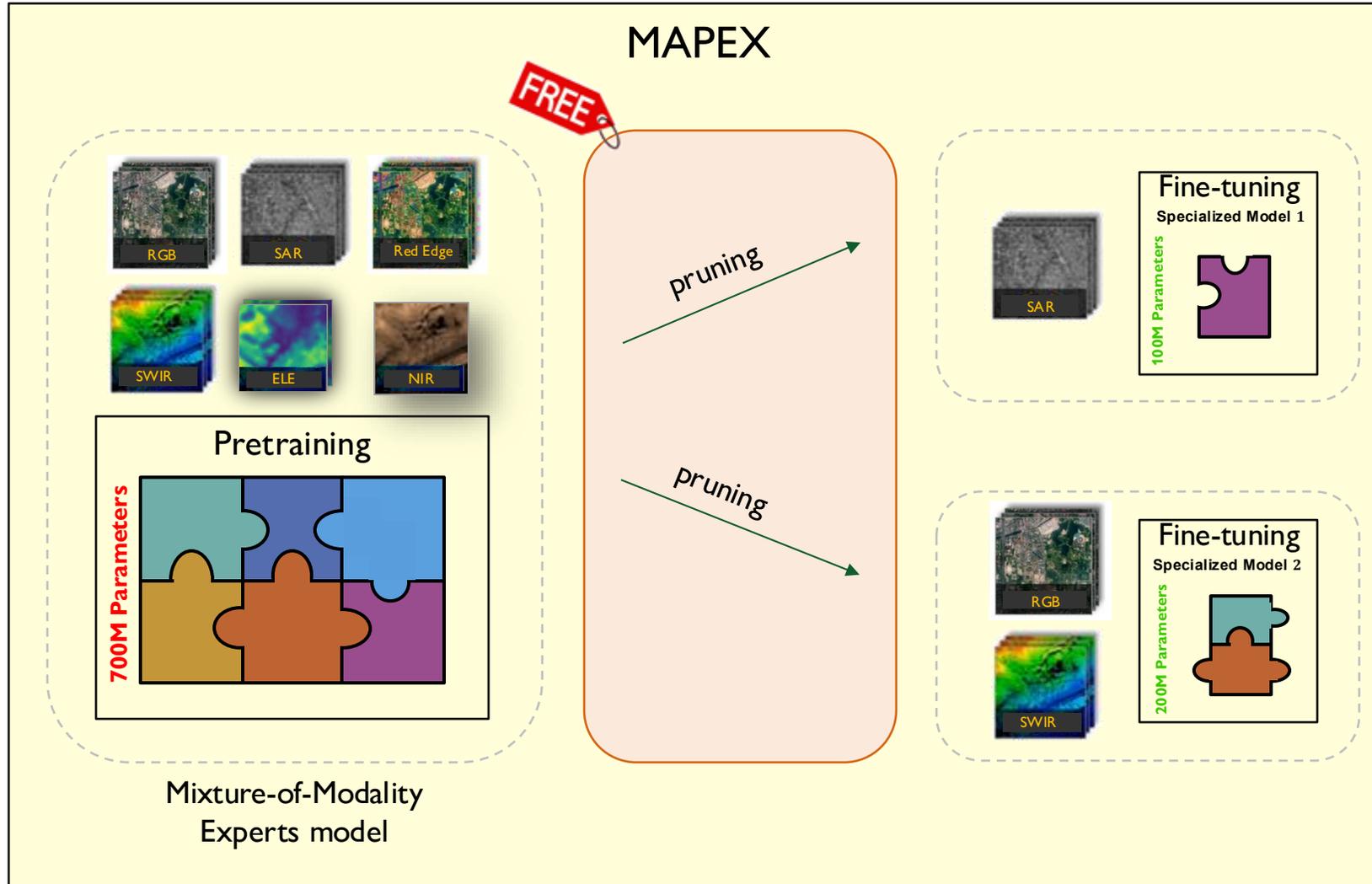
From Foundation to Lightweight Specialization

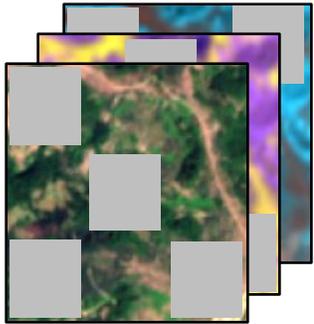


From Foundation to Lightweight Specialization

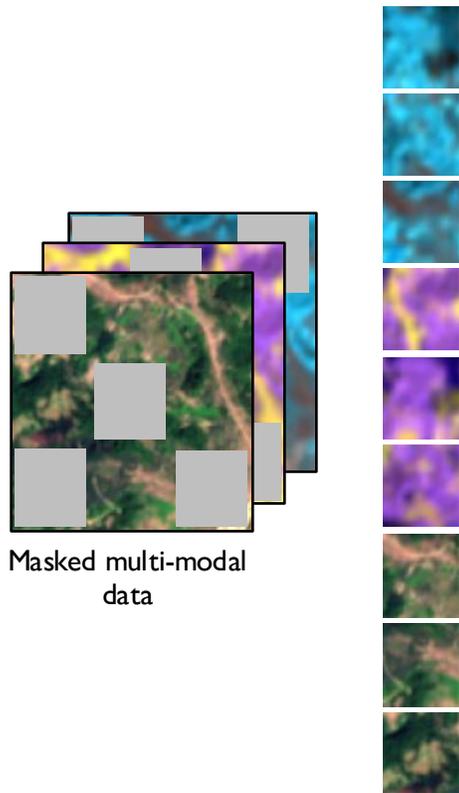


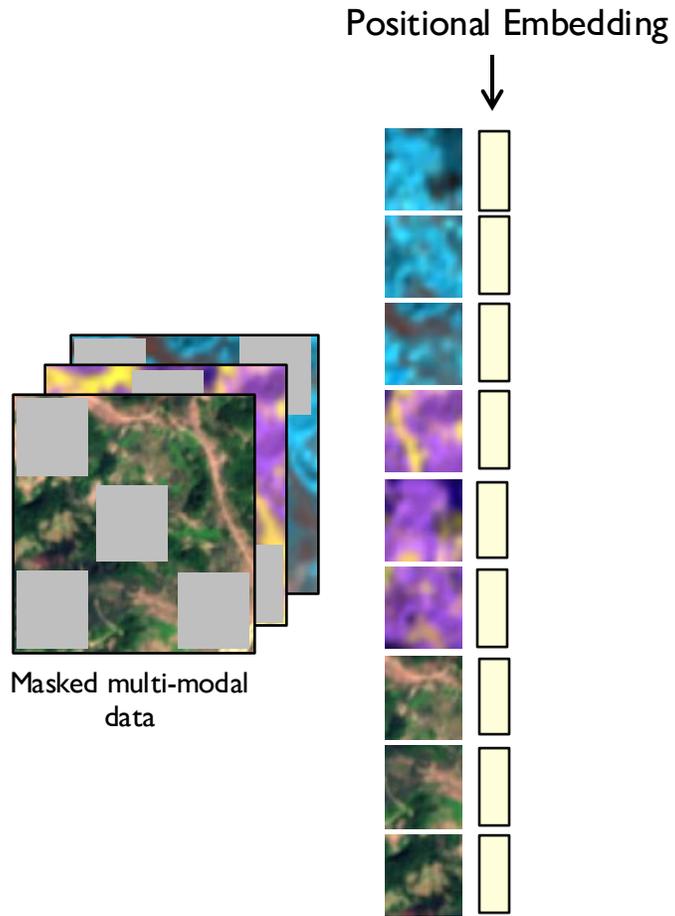
From Foundation to Lightweight Specialization

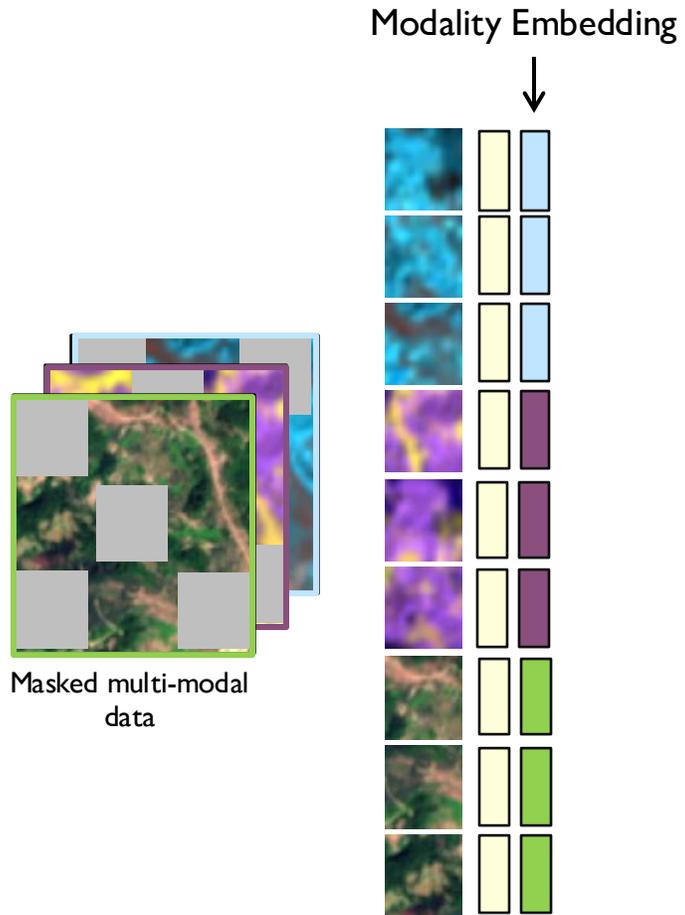


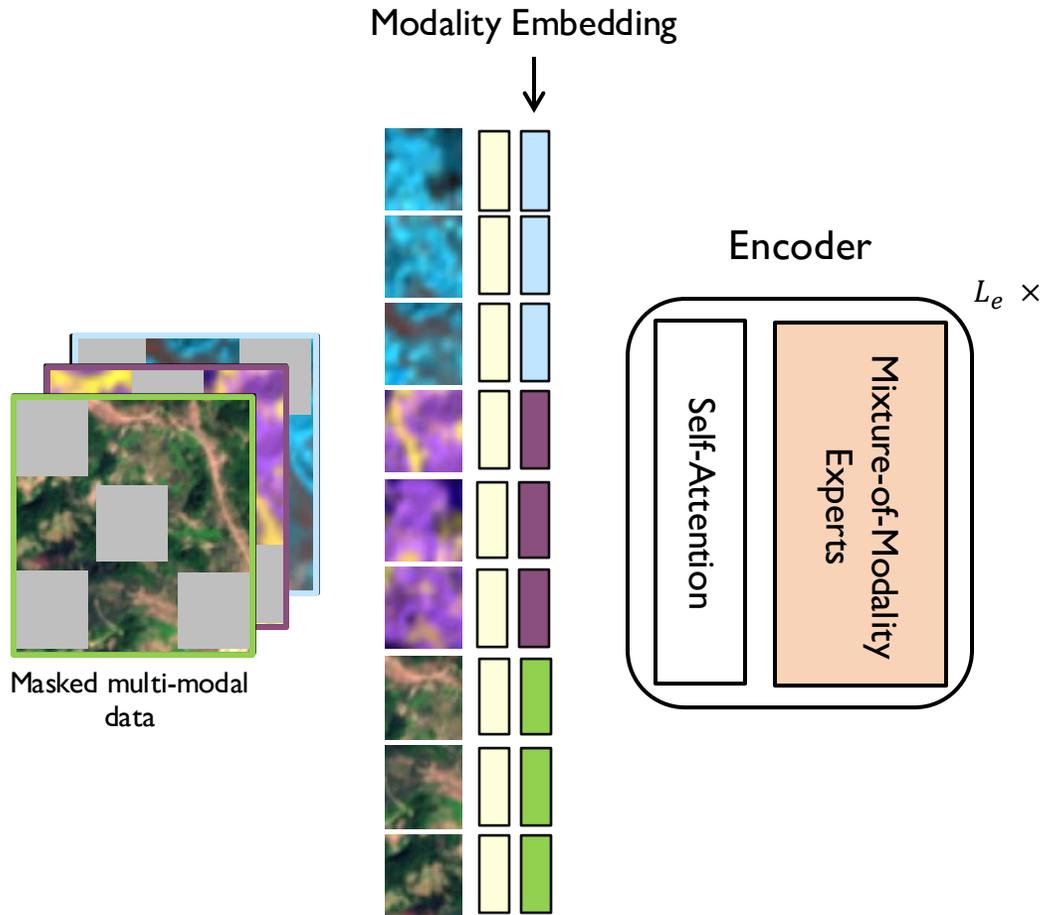


Masked multi-modal
data

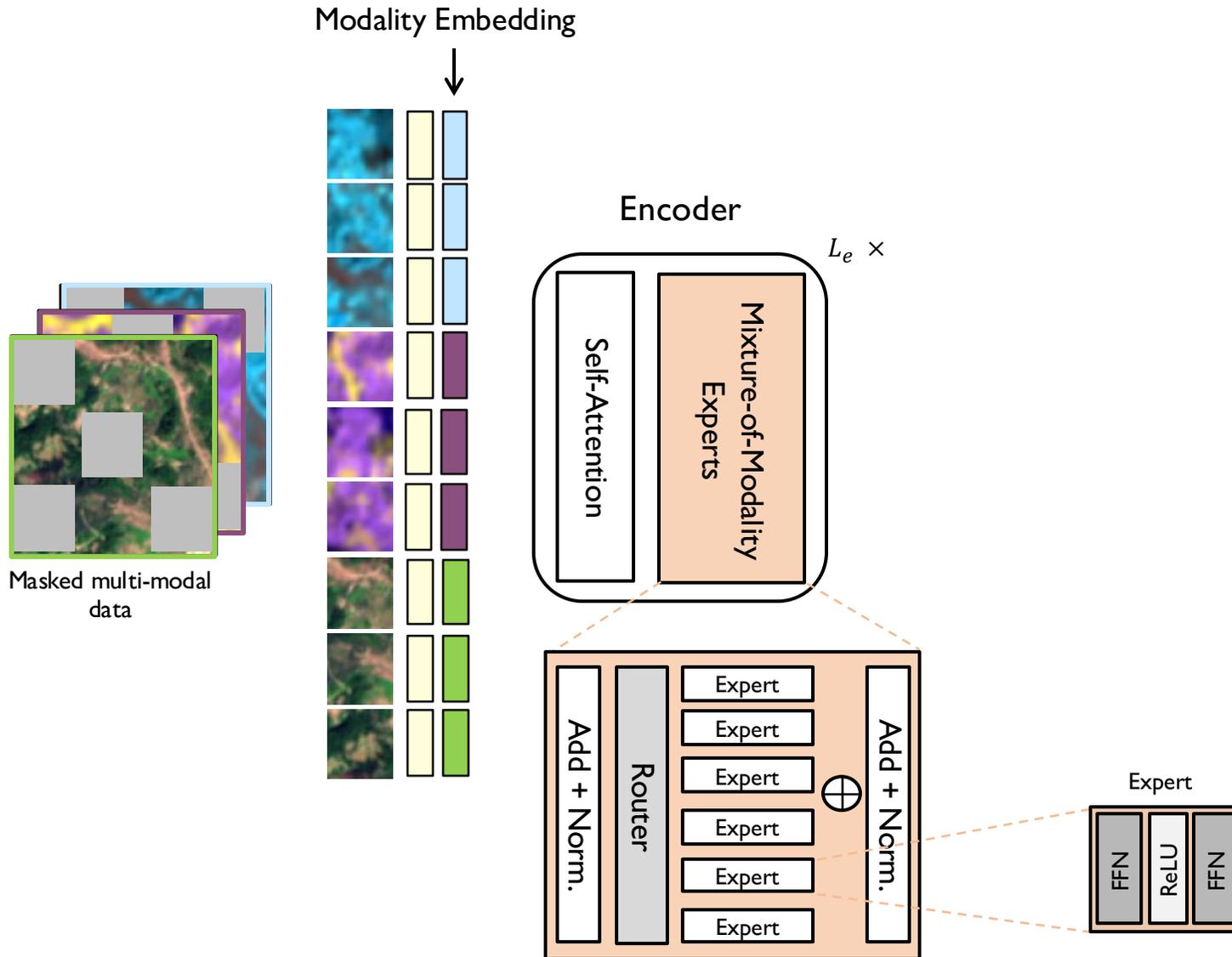




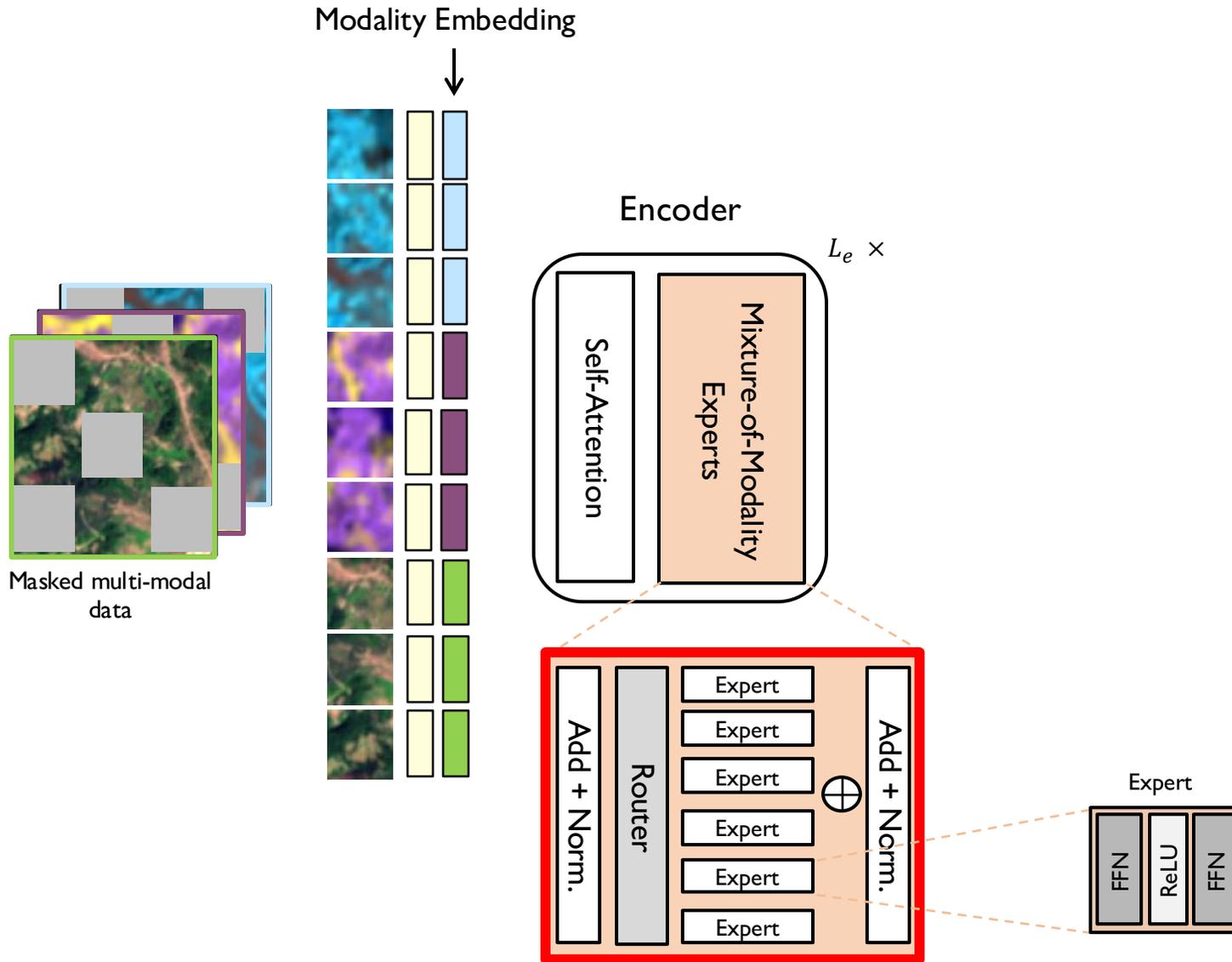




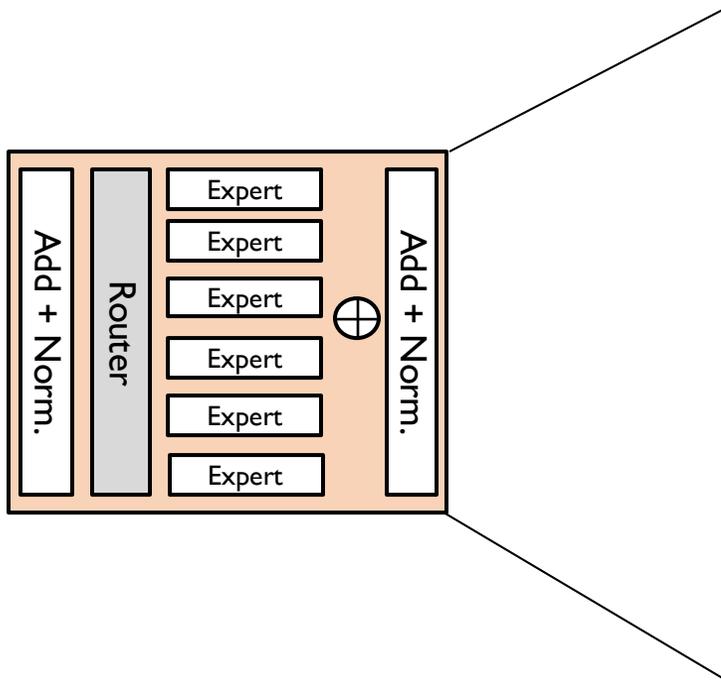
MAPEX - Mixture of Modality Experts



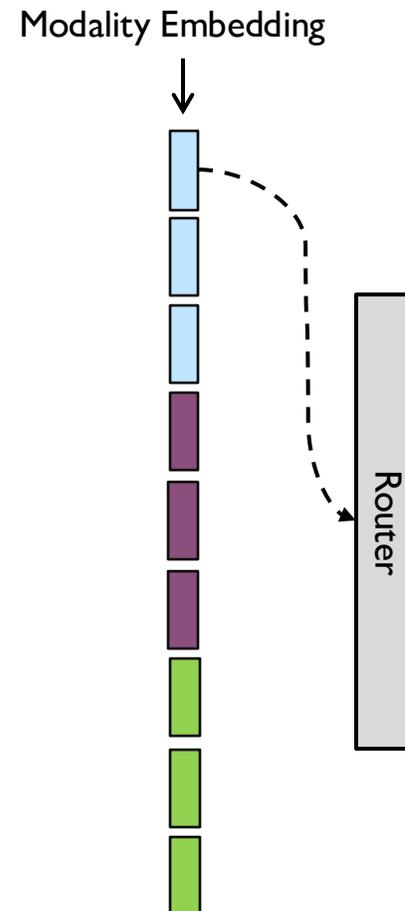
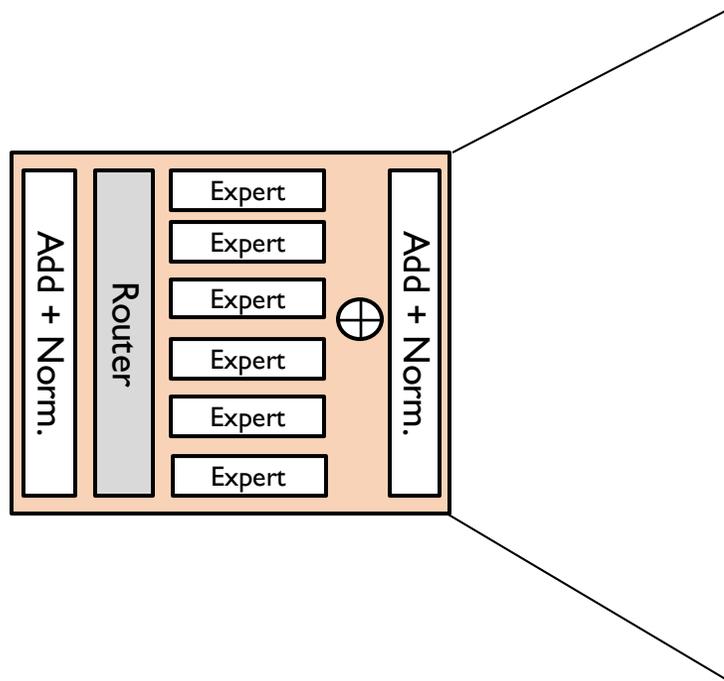
MAPEX - Mixture of Modality Experts



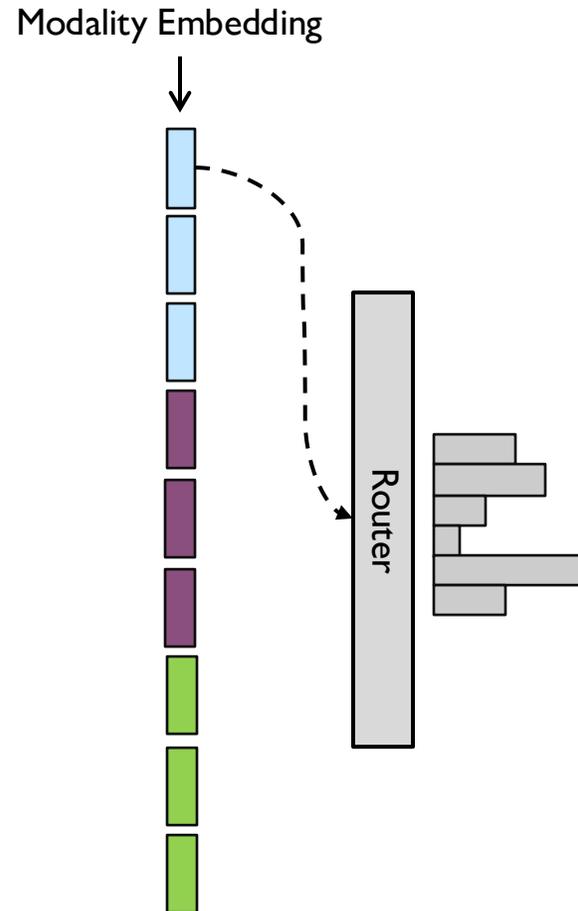
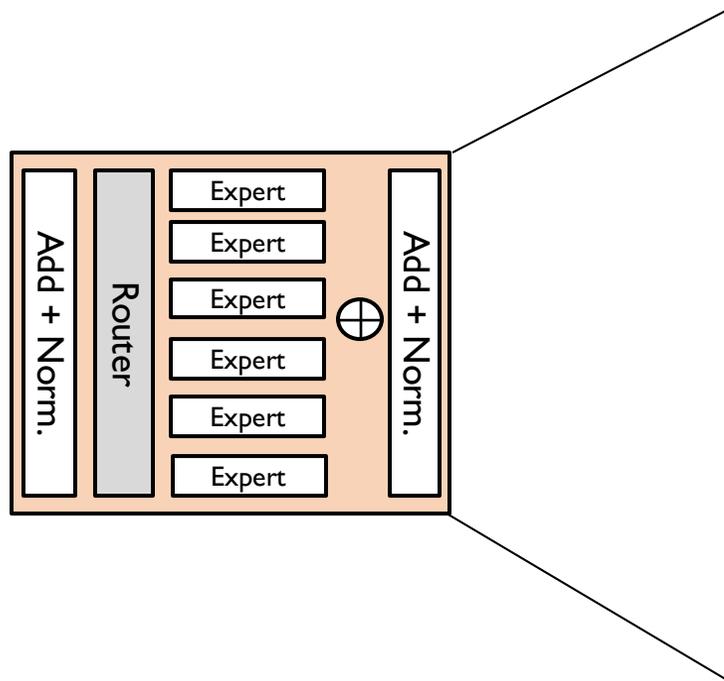
MAPEX - Modality-based Routing



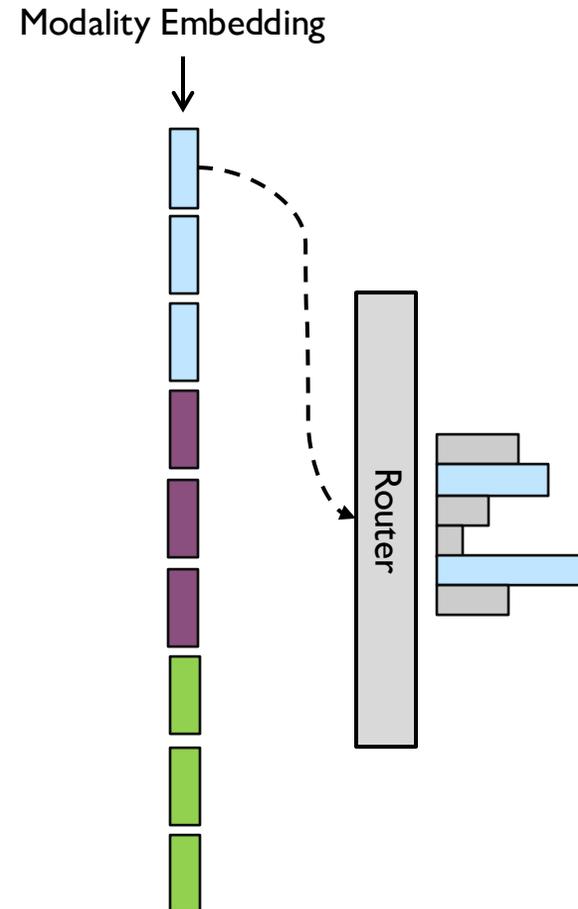
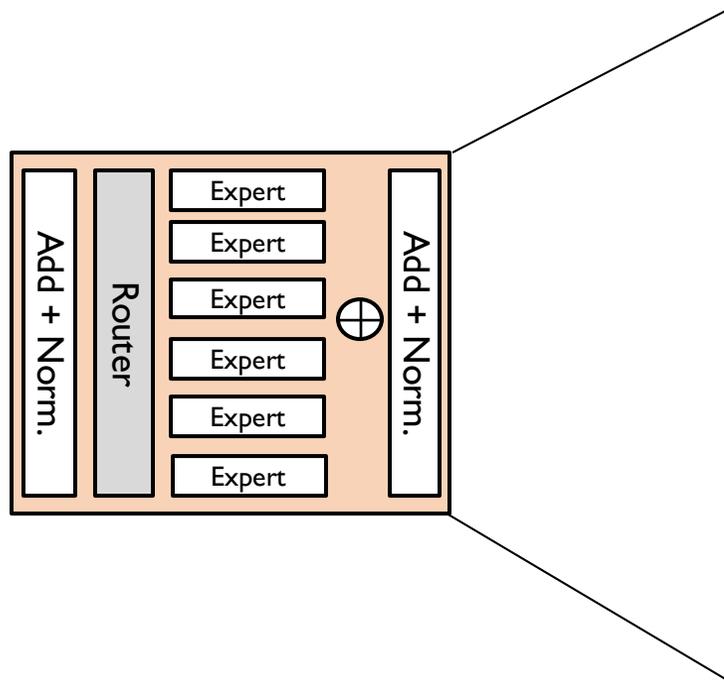
MAPEX - Modality-based Routing



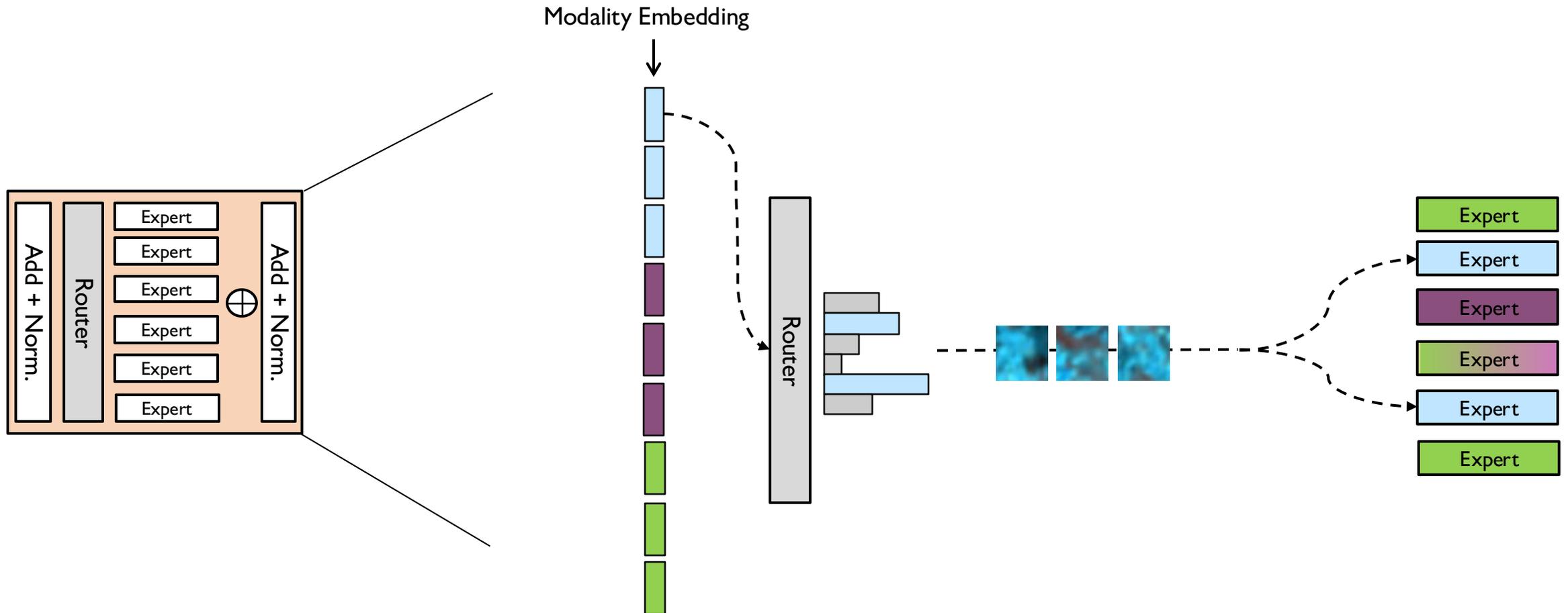
MAPEX - Modality-based Routing



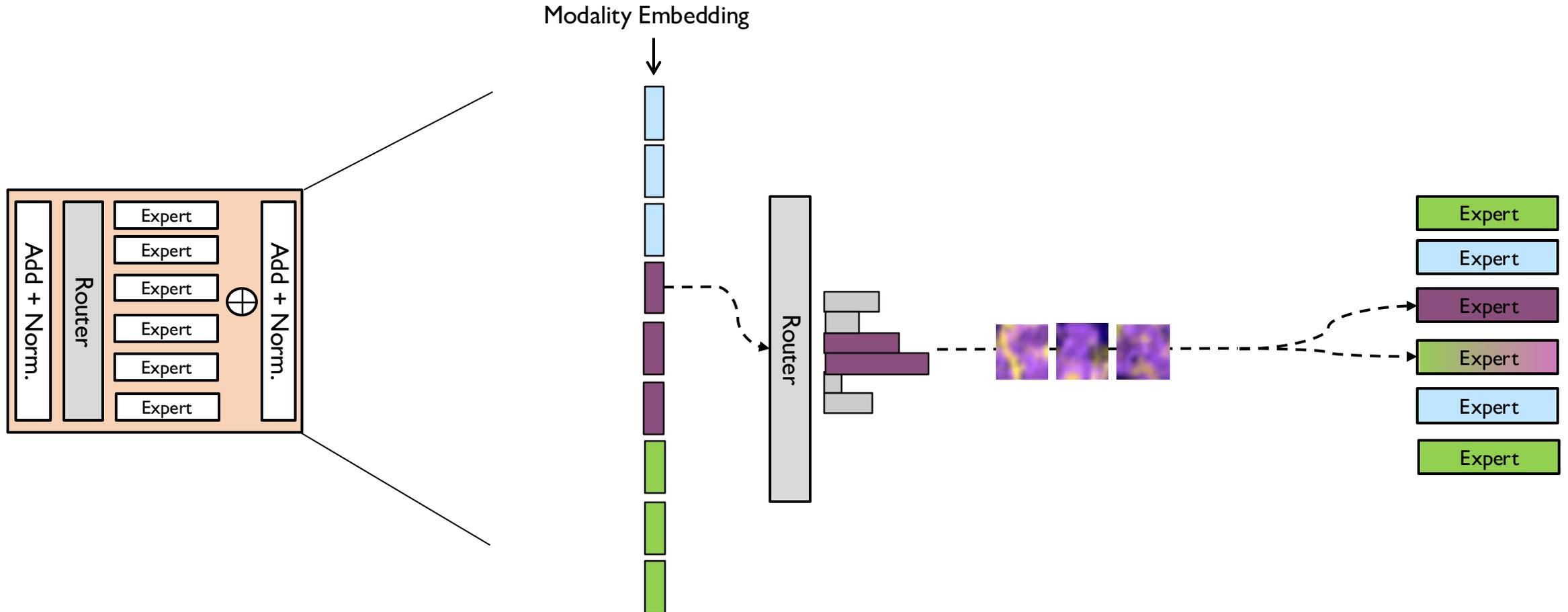
MAPEX - Modality-based Routing



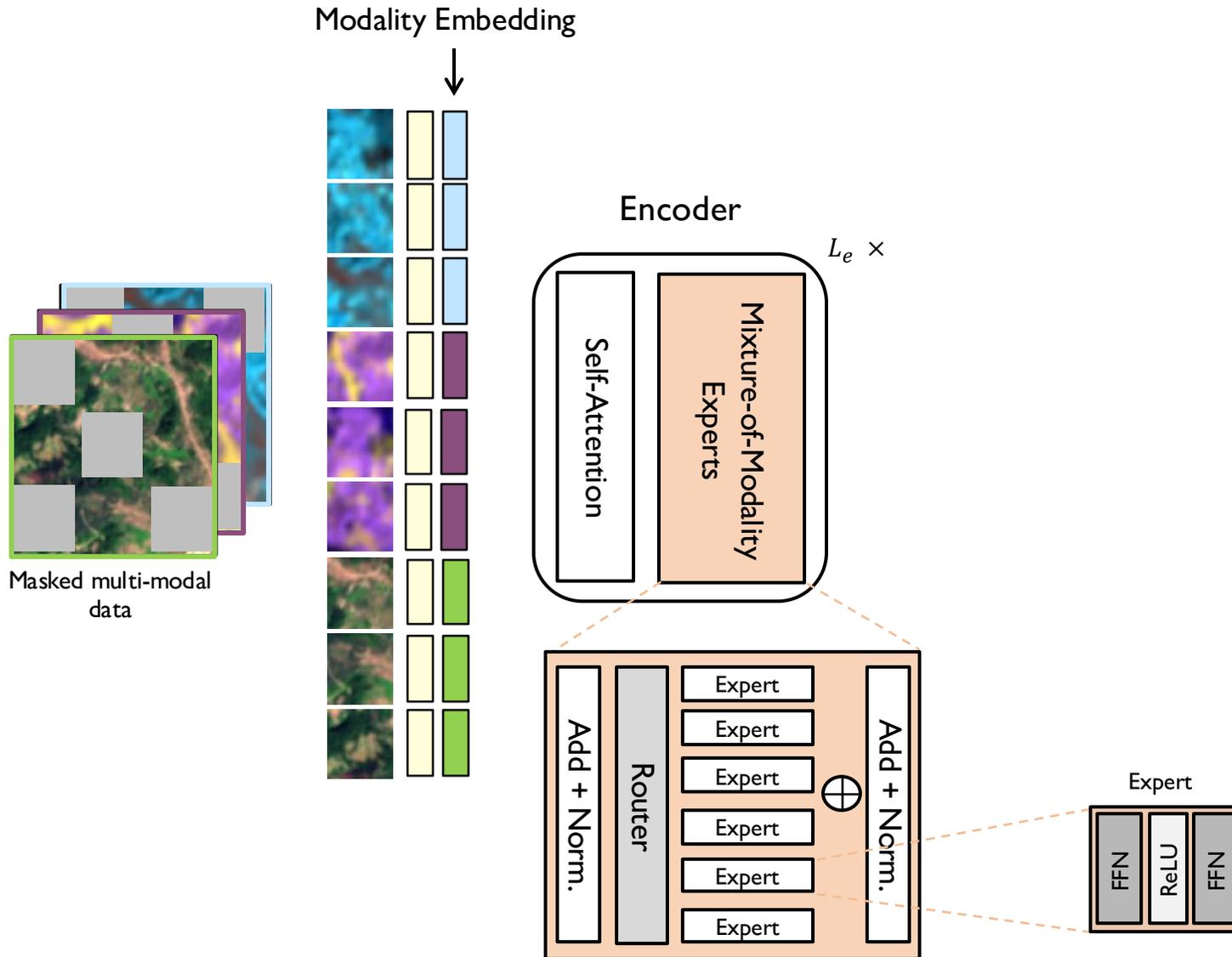
MAPEX - Modality-based Routing



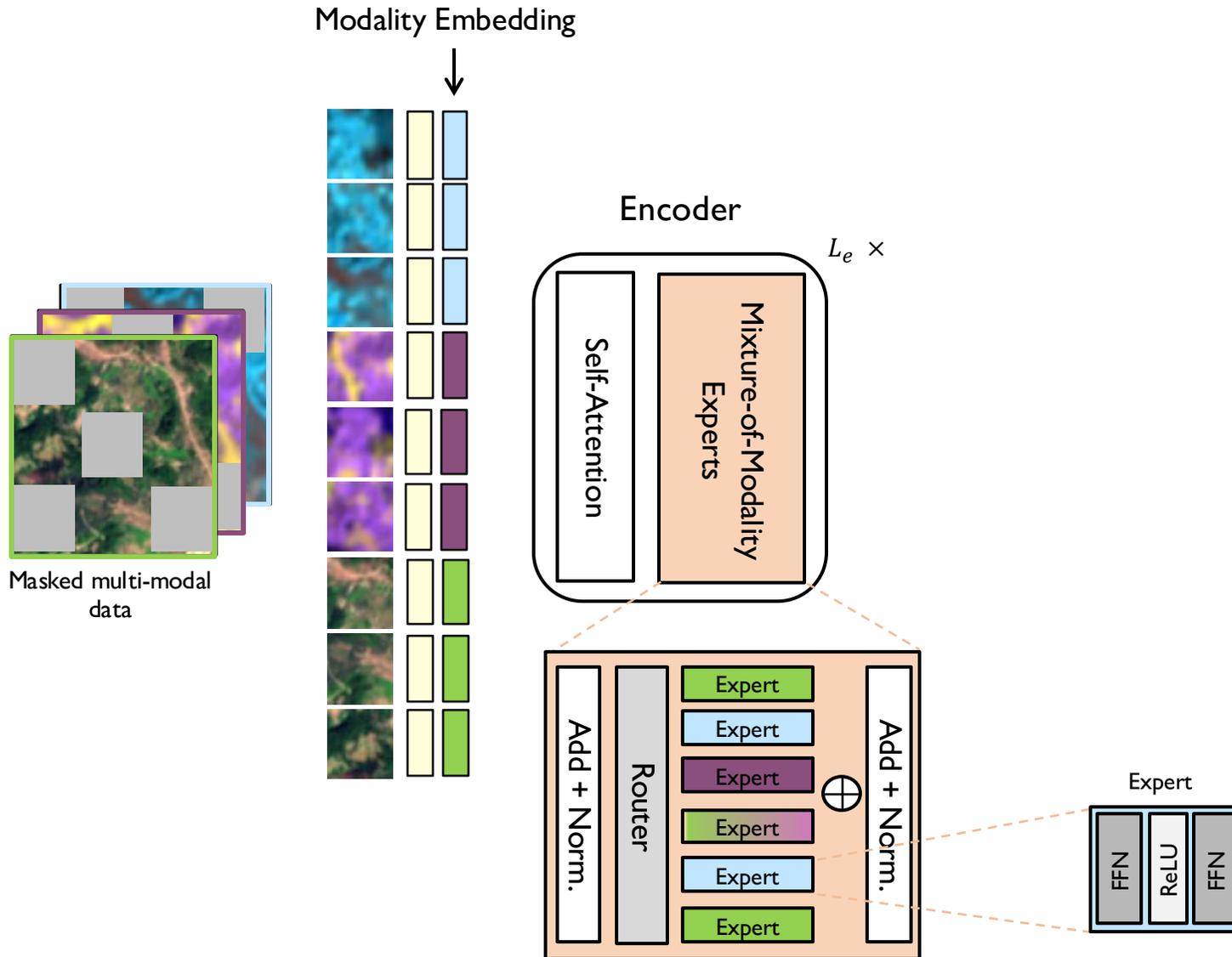
MAPEX - Modality-based Routing

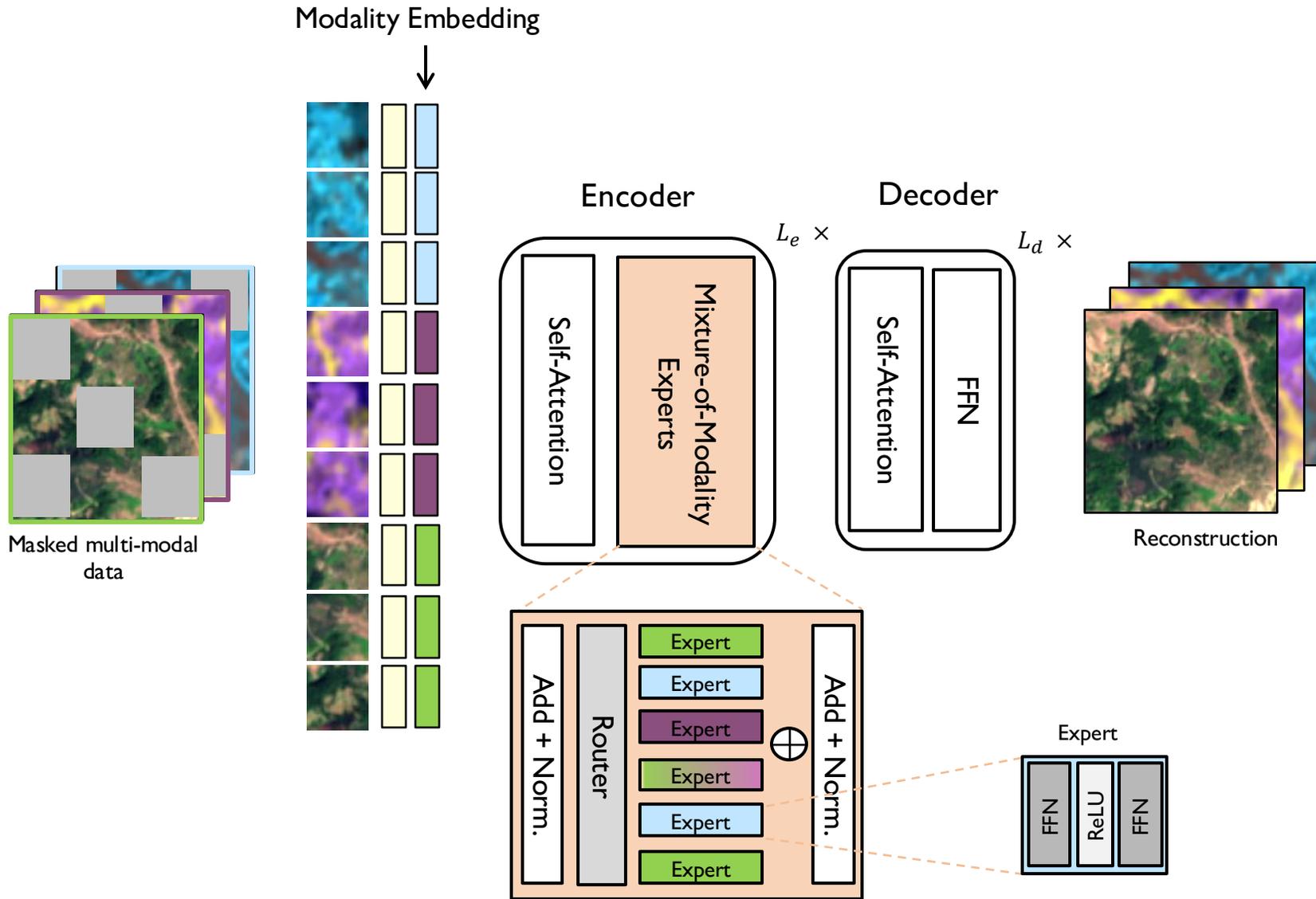


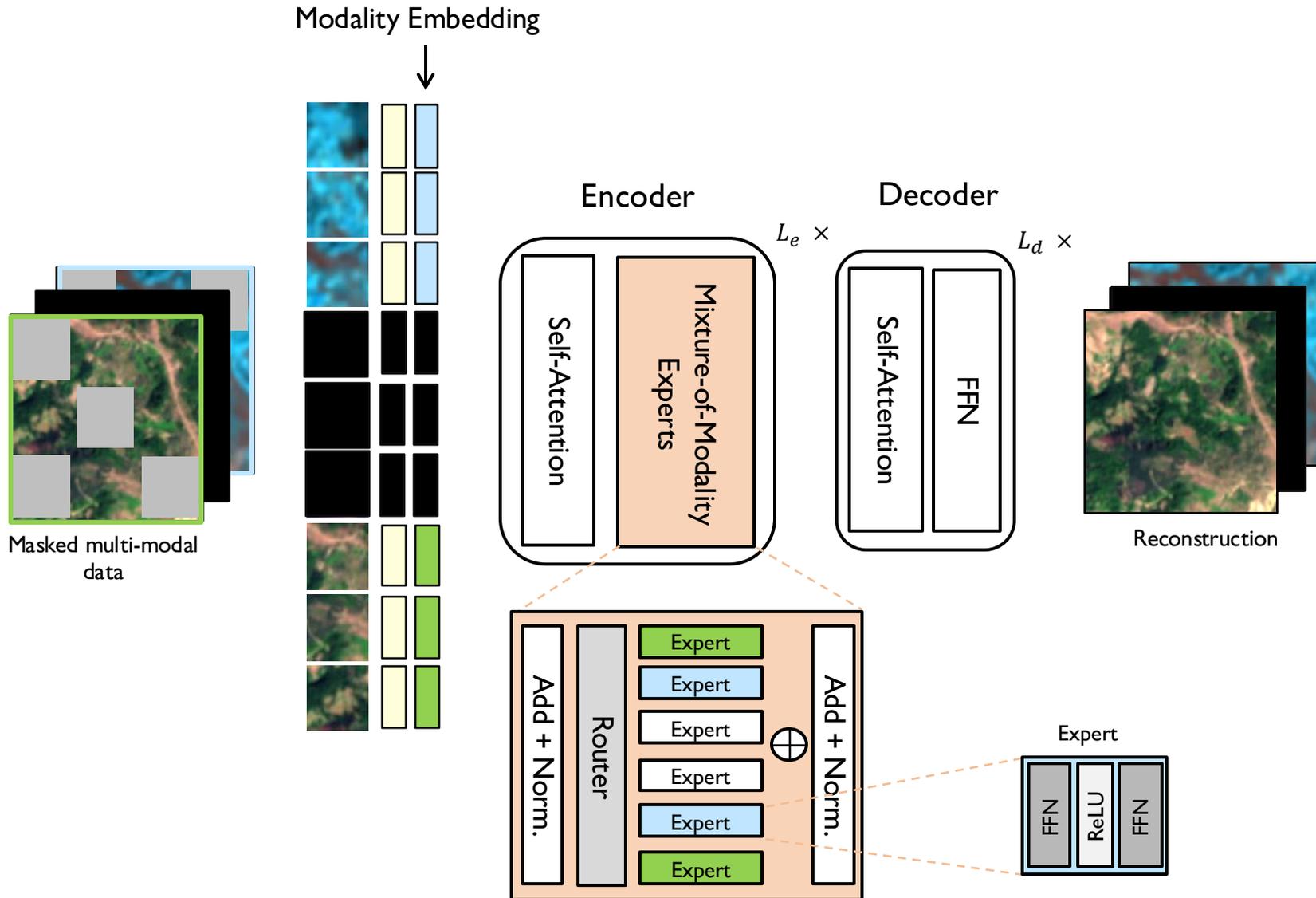
MAPEX - Mixture of Modality Experts



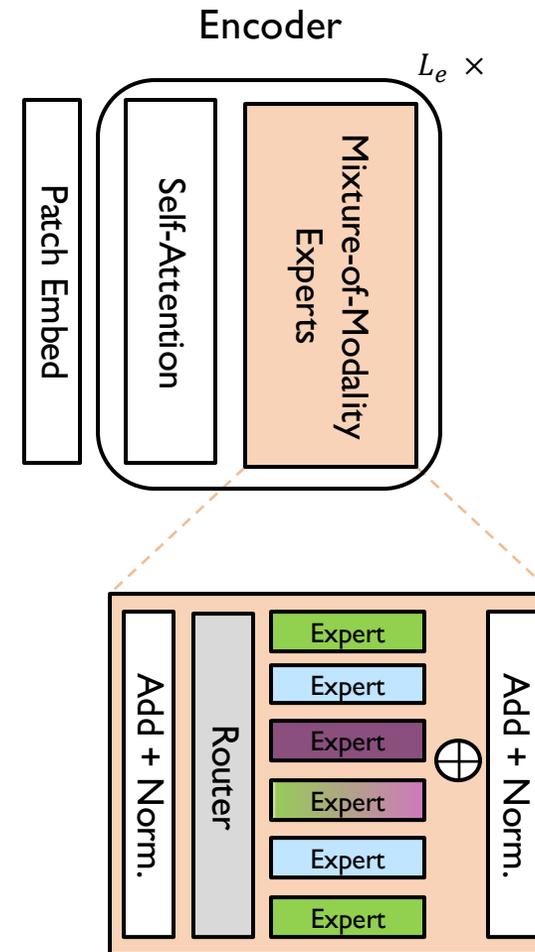
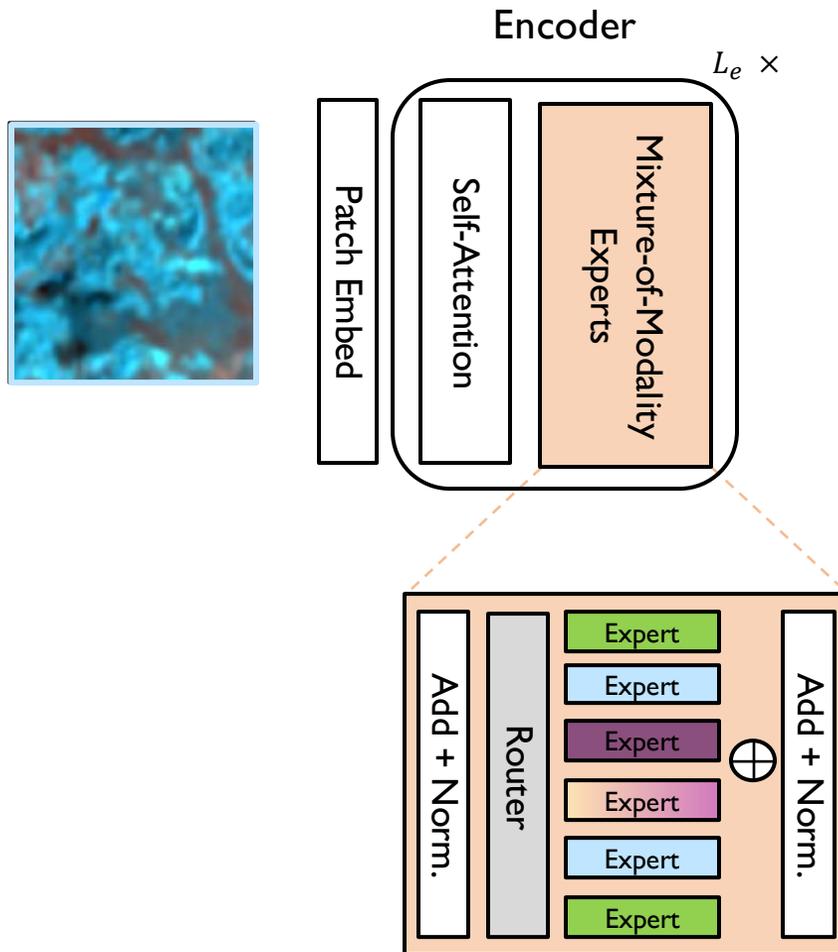
MAPEX - Mixture of Modality Experts



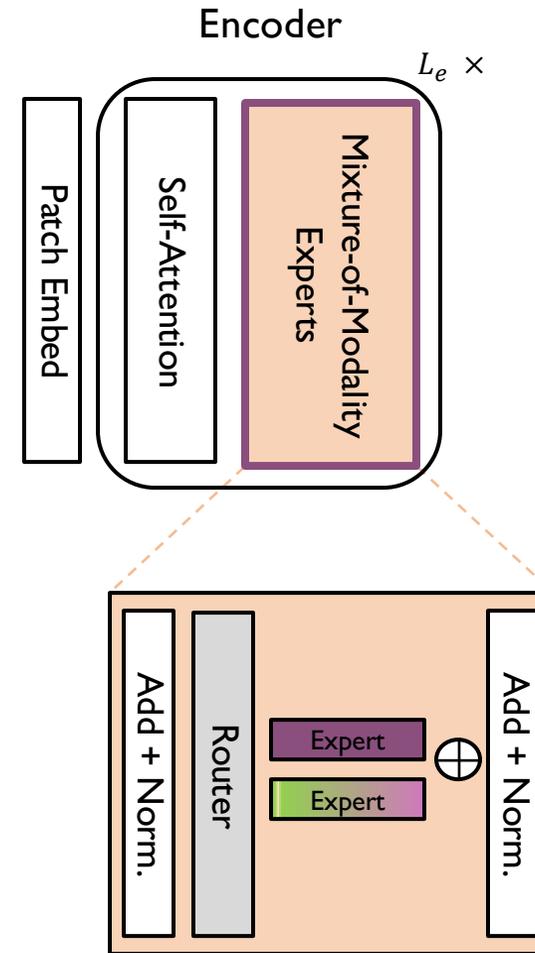
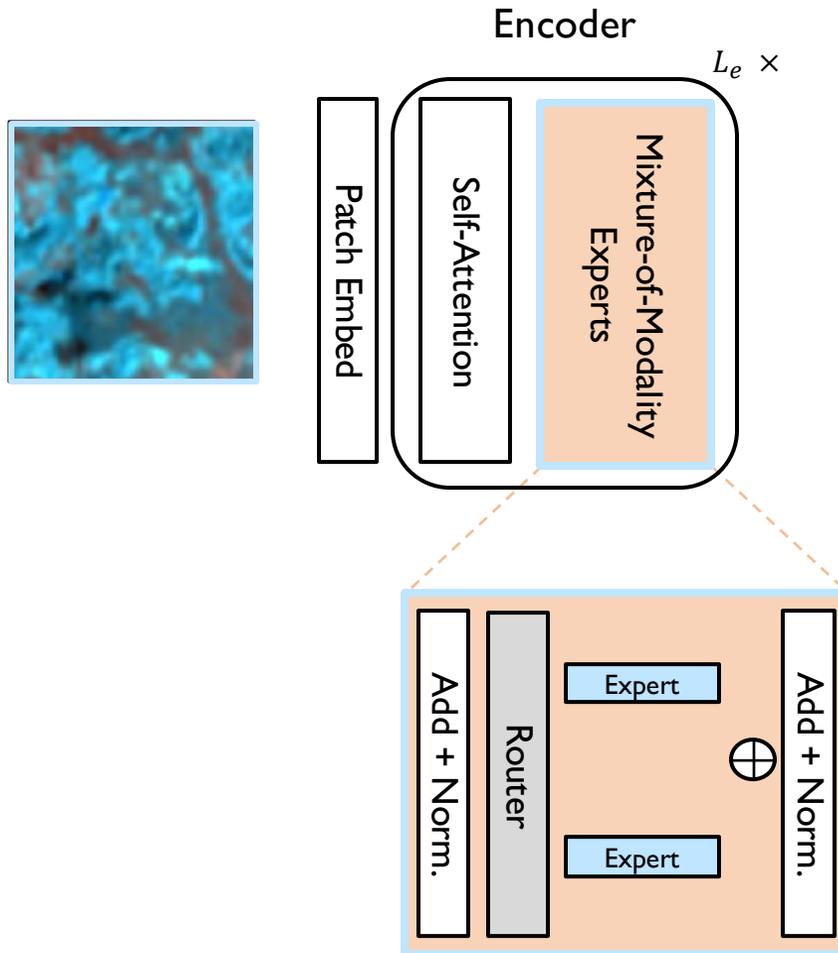




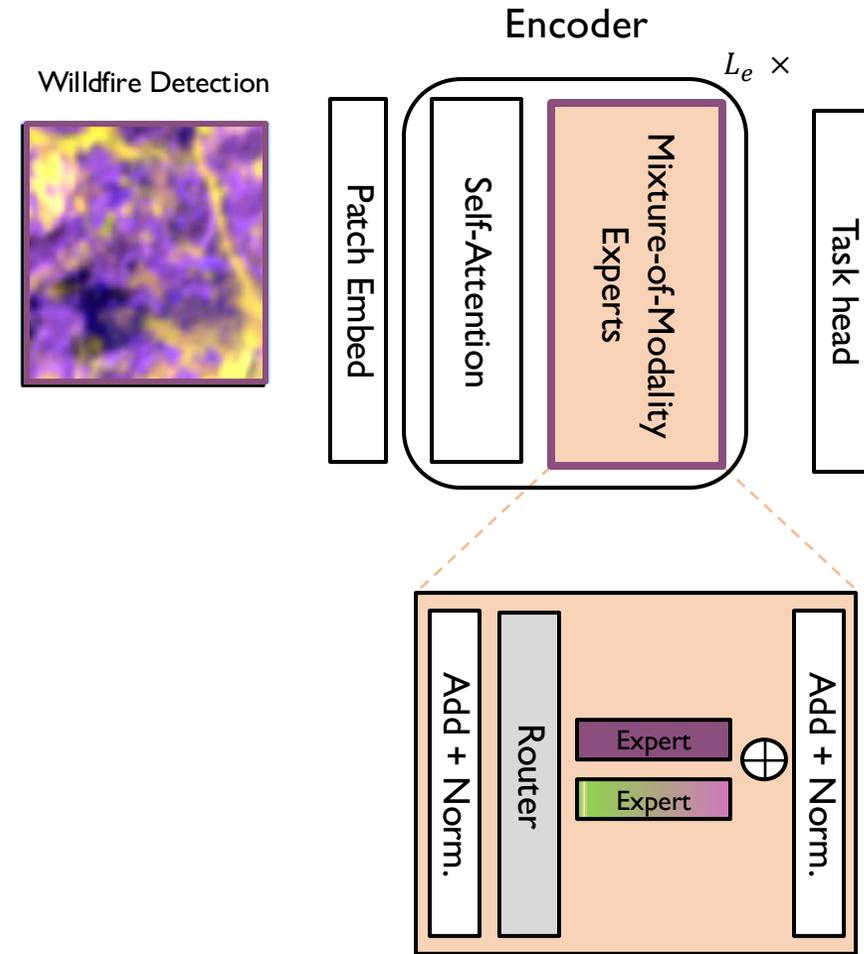
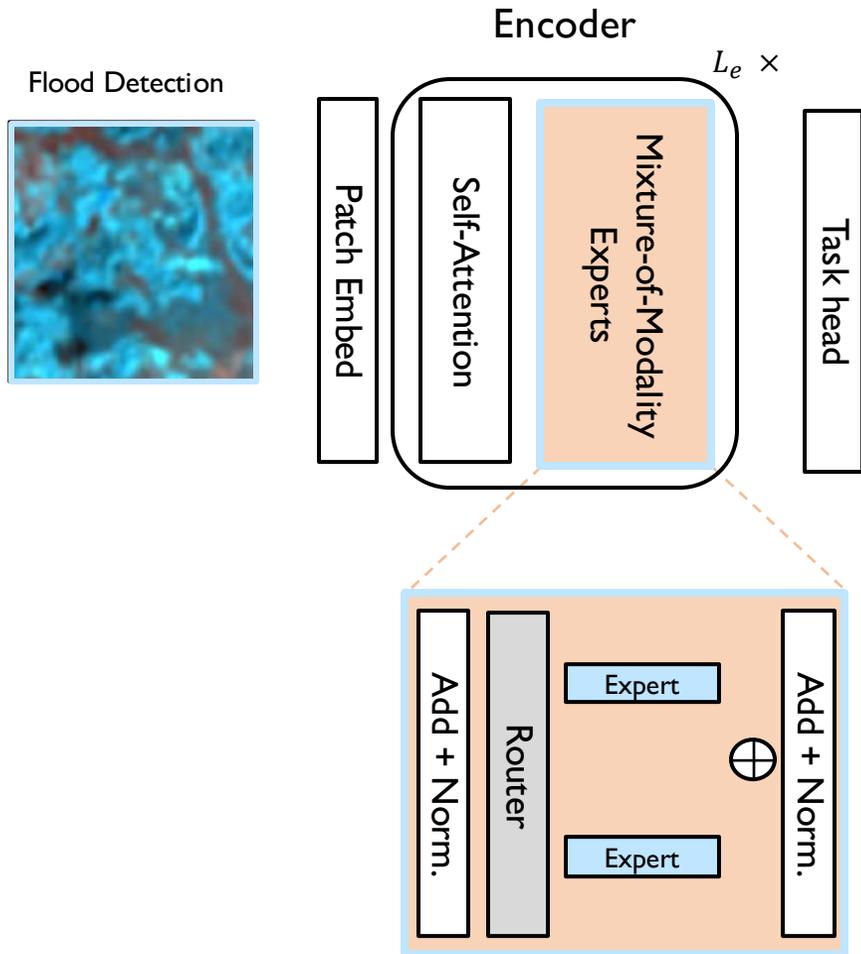
MAPEX - Modality Aware Pruning



MAPEX - Modality Aware Pruning

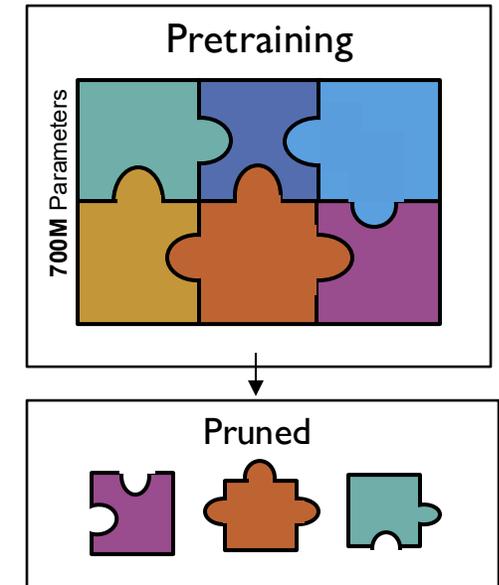


MAPEX - Modality Aware Pruning



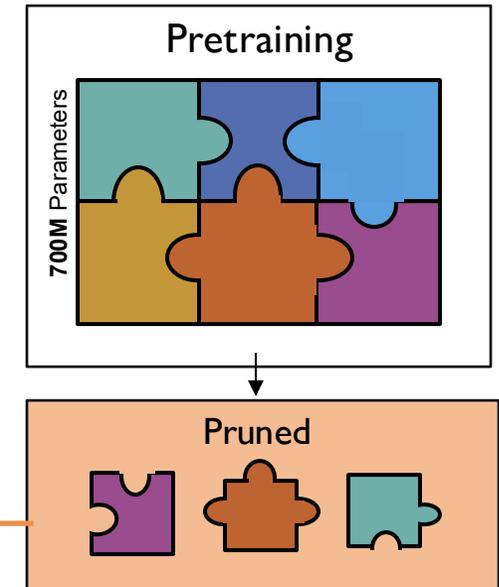
Land cover classification

Modality	Supervised FS	SatMAE [10]		Scale-MAE [36]		MAPEX (Ours)	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
RGB	71.2	64.5	70.8	75.1	79.9	67.2	75.6
Red Edge	73.3	65.5	69.7	74.7	80.3	66.7	77.8
SWIR	76.5	68.3	74.5	76.2	83.7	72.4	79.8
NIR	56.1	50.0	56.9	49.9	57.2	56.8	59.4
SAR	69.7	62.2	67.2	65.3	73.1	66.1	73.6
ELE	51.8	46.5	47.1	46.6	51.4	48.2	58.8



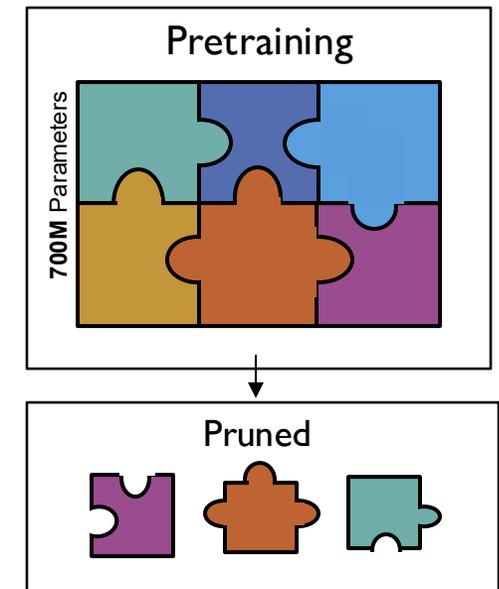
Land cover classification

Modality	Supervised FS	SatMAE [10]		Scale-MAE [36]		MAPEX (Ours)	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
RGB	71.2	64.5	70.8	75.1	79.9	67.2	75.6
Red Edge	73.3	65.5	69.7	74.7	80.3	66.7	77.8
SWIR	76.5	68.3	74.5	76.2	83.7	72.4	79.8
NIR	56.1	50.0	56.9	49.9	57.2	56.8	59.4
SAR	69.7	62.2	67.2	65.3	73.1	66.1	73.6
ELE	51.8	46.5	47.1	46.6	51.4	48.2	58.8



Land cover classification

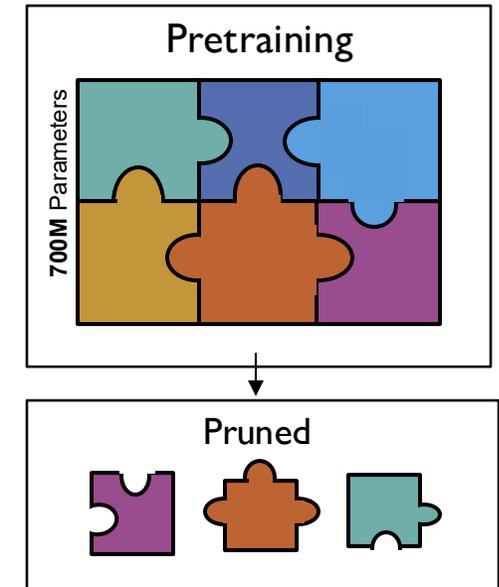
Modality	Supervised FS	SatMAE [10]		Scale-MAE [36]		MAPEX (Ours)	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
RGB	71.2	64.5	70.8	75.1	79.9	67.2	75.6
Red Edge	73.3	65.5	69.7	74.7	80.3	66.7	77.8
SWIR	76.5	68.3	74.5	76.2	83.7	72.4	79.8
NIR	56.1	50.0	56.9	49.9	57.2	56.8	59.4
SAR	69.7	62.2	67.2	65.3	73.1	66.1	73.6
ELE	51.8	46.5	47.1	46.6	51.4	48.2	58.8



The proposed pretraining approach consistently improves performance - by 5% to 13% depending on the modality - across all specialized models.

Land cover classification

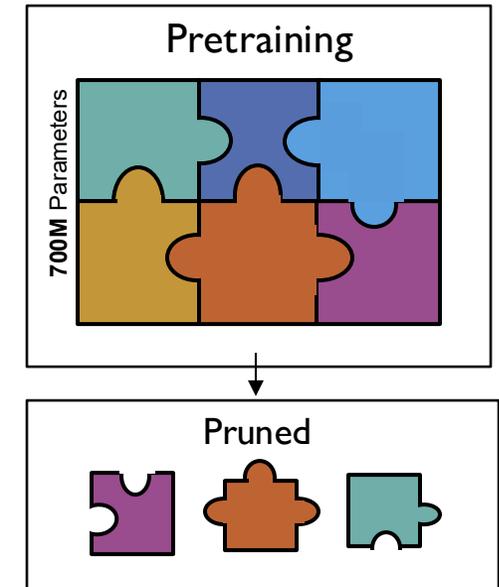
Modality	Supervised FS	SatMAE [10]		Scale-MAE [36]		MAPEX (Ours)	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
RGB	71.2	64.5	70.8	75.1	79.9	67.2	75.6
Red Edge	73.3	65.5	69.7	74.7	80.3	66.7	77.8
SWIR	76.5	68.3	74.5	76.2	83.7	72.4	79.8
NIR	56.1	50.0	56.9	49.9	57.2	56.8	59.4
SAR	69.7	62.2	67.2	65.3	73.1	66.1	73.6
ELE	51.8	46.5	47.1	46.6	51.4	48.2	58.8



For some modalities, Scale-MAE performs better than our method, but it is 2.5 times larger. Our smaller models achieve performance close to that of much bigger models.

Land cover classification

Modality	Supervised FS	SatMAE [10]		Scale-MAE [36]		MAPEX (Ours)	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
RGB	71.2	64.5	70.8	75.1	79.9	67.2	75.6
Red Edge	73.3	65.5	69.7	74.7	80.3	66.7	77.8
SWIR	76.5	68.3	74.5	76.2	83.7	72.4	79.8
NIR	56.1	50.0	56.9	49.9	57.2	56.8	59.4
SAR	69.7	62.2	67.2	65.3	73.1	66.1	73.6
ELE	51.8	46.5	47.1	46.6	51.4	48.2	58.8



For other modalities that are different from those seen by Scale-MAE or SatMAE during pretraining, our method consistently outperforms them. This highlights the problem of modality mismatch. With our approach, we can include new modalities during pretraining and still end up with smaller models that perform well on downstream tasks.

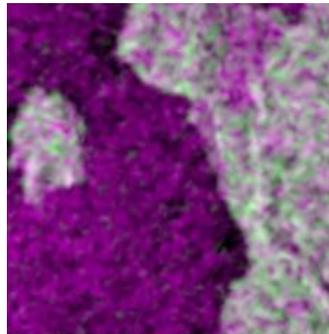
Flood Detection

Modality	FS	SatMAE		Scale-MAE		MAPEX	
		k -NN	FT	k -NN	FT	k -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SAR	78.1	77.4	80.3	83.0	82.6	80.1	83.2

Sentinel-2 RGB



Sentinel-1 VV VH



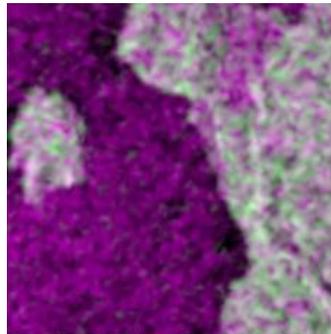
Flood Detection

Modality	FS	SatMAE		Scale-MAE		MAPEX	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SAR	78.1	77.4	80.3	83.0	82.6	80.1	83.2

Sentinel-2 RGB



Sentinel-1 VV VH



Flood Detection

Modality	FS	SatMAE		Scale-MAE		MAPEX	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SAR	78.1	77.4	80.3	83.0	82.6	80.1	83.2

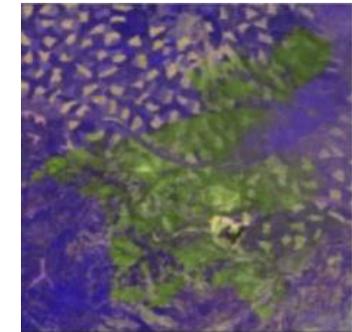
Wildfire Detection

Modality	FS	SatMAE		Scale-MAE		MAPEX	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SWIR	84.7	84.2	88.6	86.6	90.8	87.3	90.5

Sentinel-2 RGB



Sentinel-2 SWIR



Flood Detection

Modality	FS	SatMAE		Scale-MAE		MAPEX	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SAR	78.1	77.4	80.3	83.0	82.6	80.1	83.2

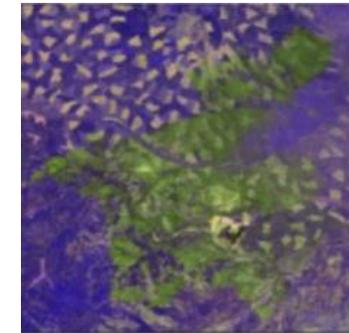
Wildfire Detection

Modality	FS	SatMAE		Scale-MAE		MAPEX	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SWIR	84.7	84.2	88.6	86.6	90.8	87.3	90.5

Sentinel-2 RGB



Sentinel-2 SWIR



Flood Detection

Modality	FS	SatMAE		Scale-MAE		MAPEX	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SAR	78.1	77.4	80.3	83.0	82.6	80.1	83.2

Wildfire Detection

Modality	FS	SatMAE		Scale-MAE		MAPEX	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SWIR	84.7	84.2	88.6	86.6	90.8	87.3	90.5

Land cover Segmentation

Modality	U-Net [38]	SatMAE (<i>vit_base</i>)	MAPEX
RGB	56.3	53.1	56.9

Flood Detection

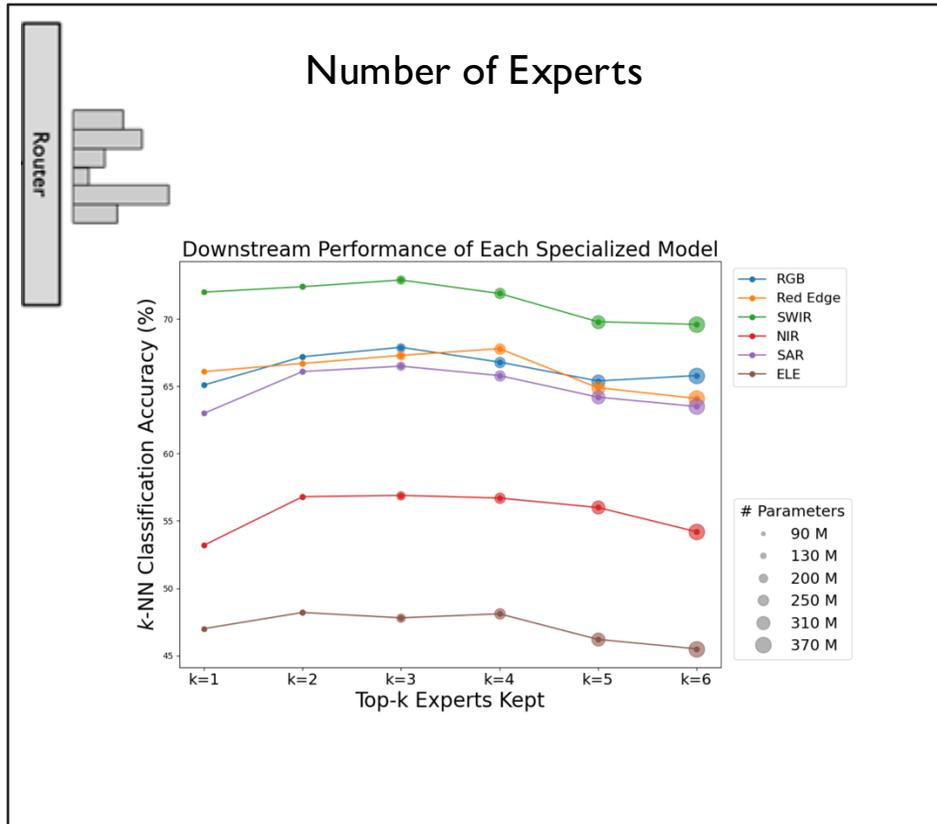
Modality	FS	SatMAE		Scale-MAE		MAPEX	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SAR	78.1	77.4	80.3	83.0	82.6	80.1	83.2

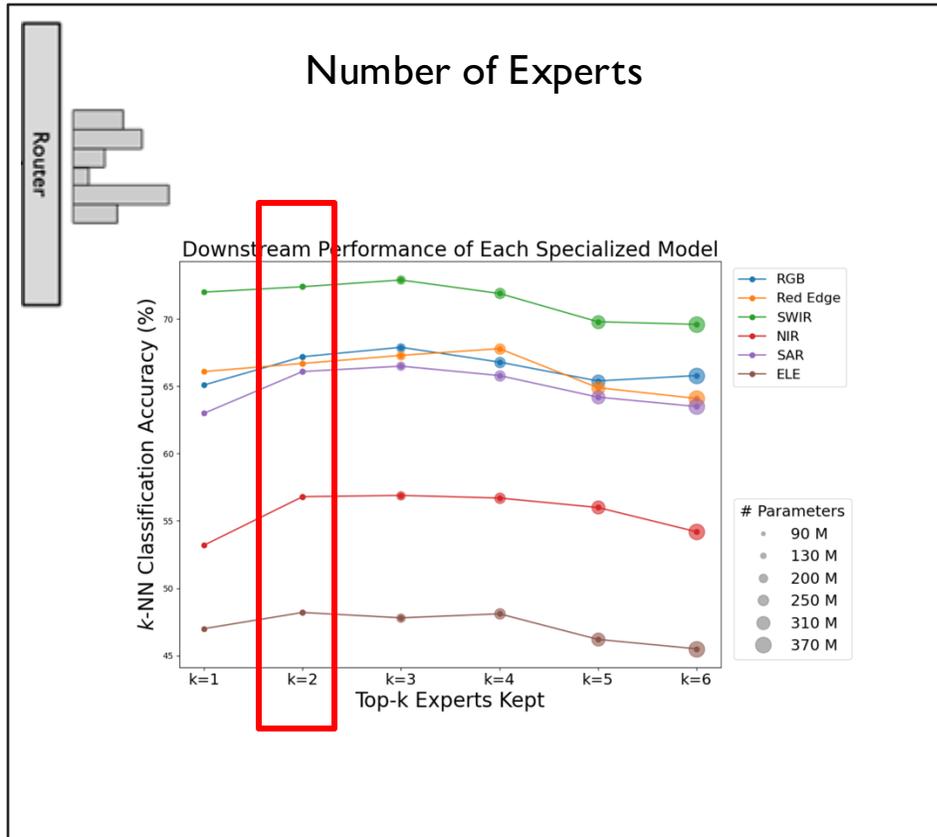
Wildfire Detection

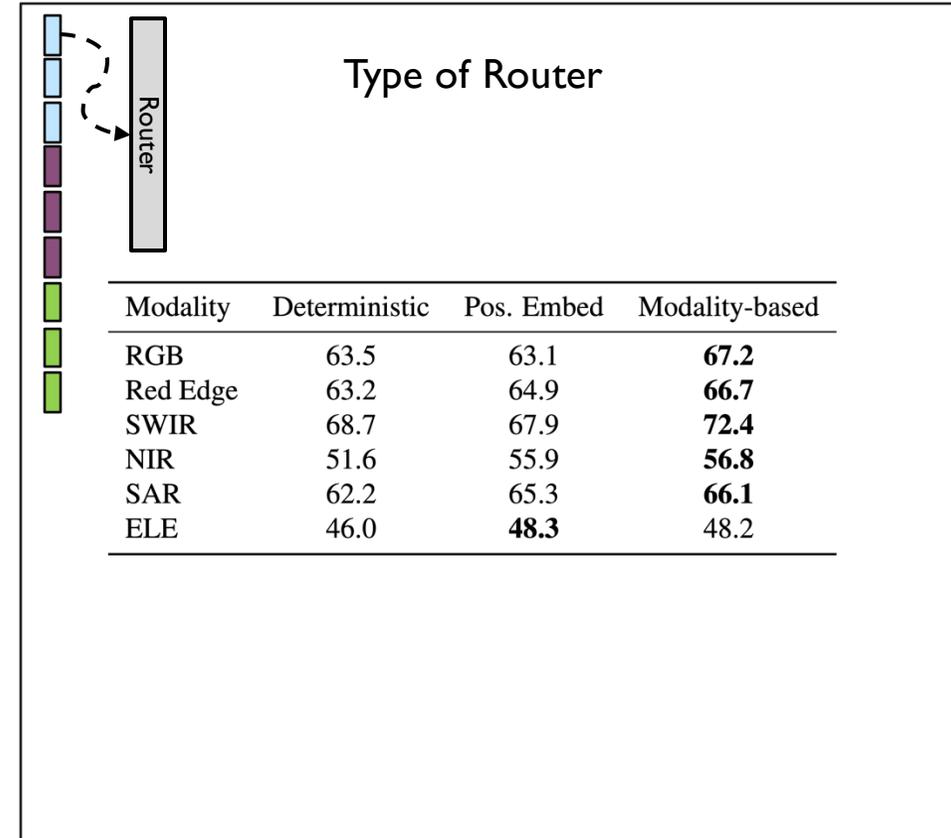
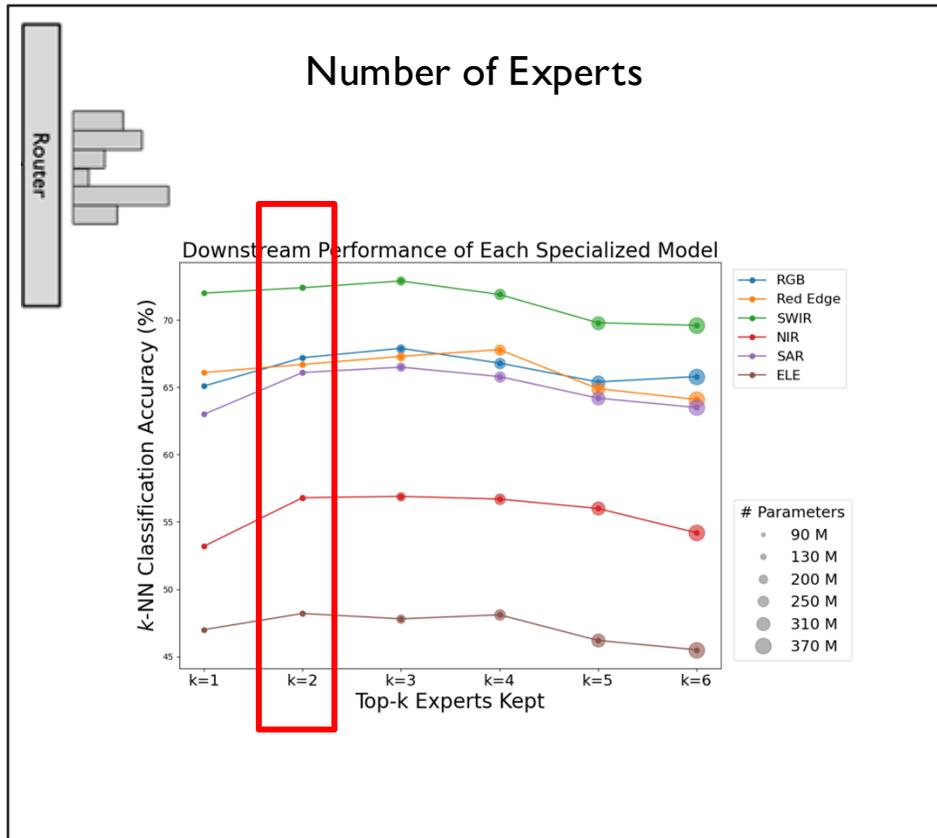
Modality	FS	SatMAE		Scale-MAE		MAPEX	
		<i>k</i> -NN	FT	<i>k</i> -NN	FT	<i>k</i> -NN	FT
Architecture	<i>MAPEX</i>	<i>vit_base</i>		<i>vit_large</i>		<i>MAPEX</i>	
# params	130M	90M		310M		130M	
SWIR	84.7	84.2	88.6	86.6	90.8	87.3	90.5

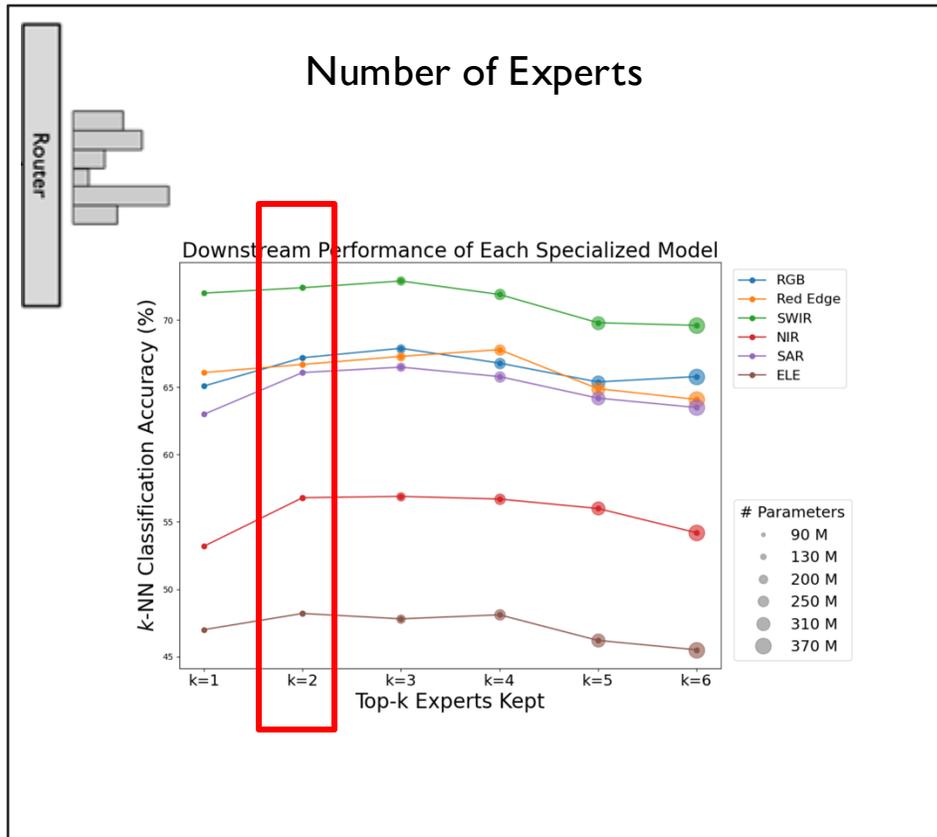
Land cover Segmentation

Modality	U-Net [38]	SatMAE (<i>vit_base</i>)	MAPEX
RGB	56.3	53.1	56.9

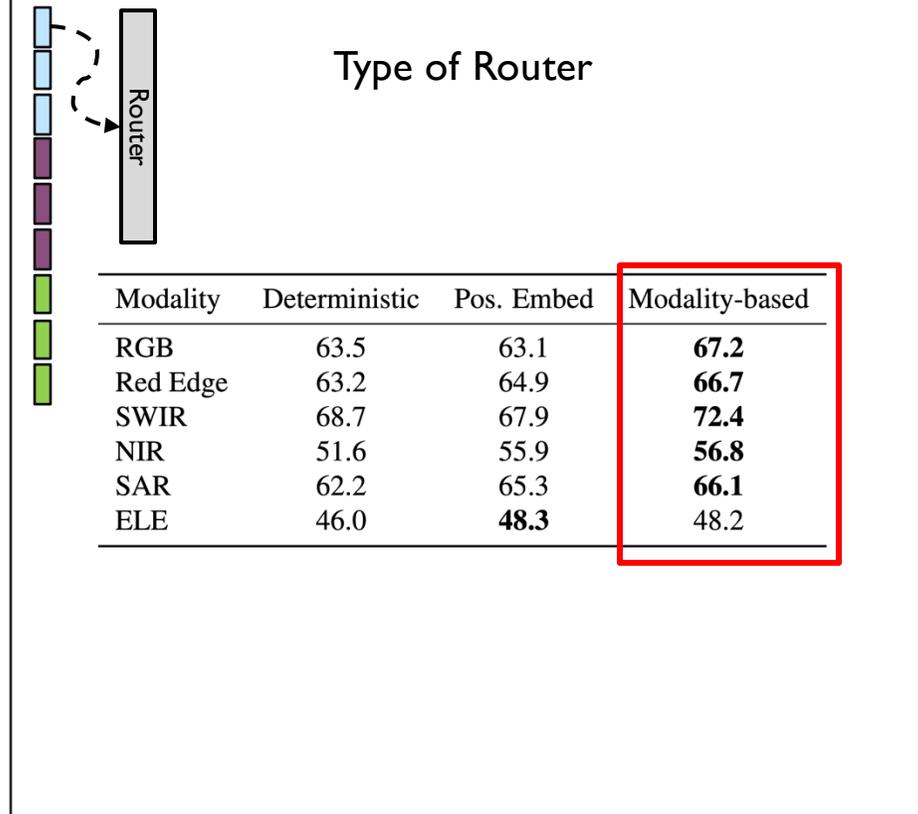








Type of Router



Modality	Deterministic	Pos. Embed	Modality-based
RGB	63.5	63.1	67.2
Red Edge	63.2	64.9	66.7
SWIR	68.7	67.9	72.4
NIR	51.6	55.9	56.8
SAR	62.2	65.3	66.1
ELE	46.0	48.3	48.2

Takeaways

From Foundation to Lightweight Specialization

MAPEX:

- combines the *best of both worlds*:
 - Multi-modal pretraining
 - Task-specific specialization

From Foundation to Lightweight Specialization

MAPEX:

- combines the *best of both worlds*:
 - Multi-modal pretraining
 - Task-specific specialization
- leads to smaller and (out-)performant models

From Foundation to Lightweight Specialization

MAPEX:

- combines the *best of both worlds*:
 - Multi-modal pretraining
 - Task-specific specialization
- leads to smaller and (out-)performant models
- enables flexible use of different input modalities

Thank you!