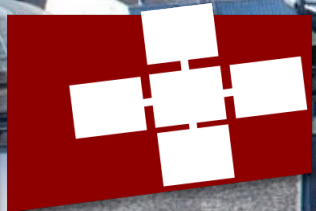


NIKOLI DRYDEN & TORSTEN HOEFLER

Spatial Mixture-of-Experts

NEURIPS 2022



This project received funding from the European Research Council under grant agreement MAELSTROM.



Data Has Spatial Structure

Data Has Spatial Structure

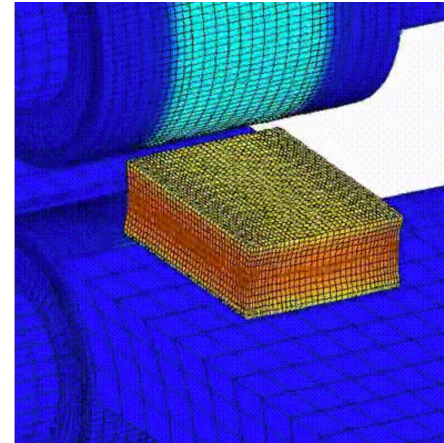
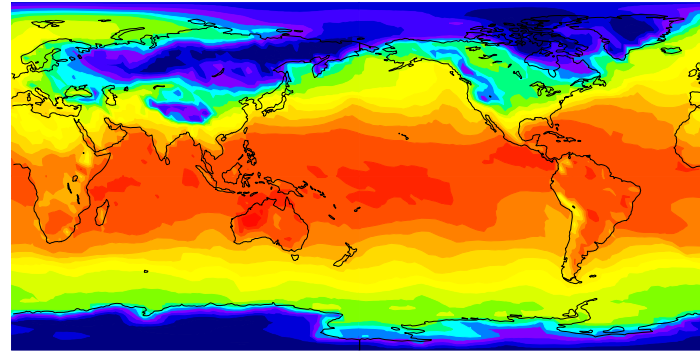
Weather

Geometric meshes

Registered data

Approximate alignment

And more...



Data Has Spatial Structure

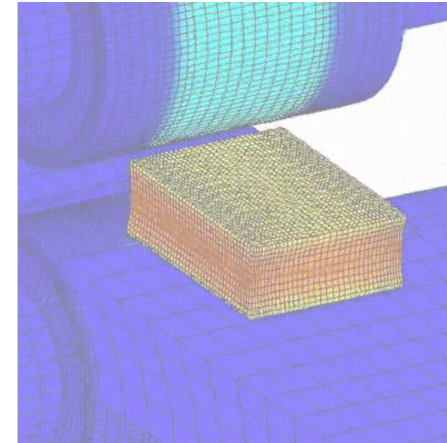
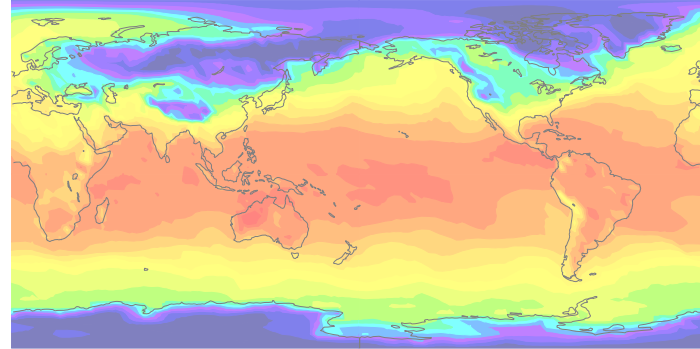
Weather

Geometric meshes

Registered data

Approximate alignment

And more...



Data Has Spatial Structure

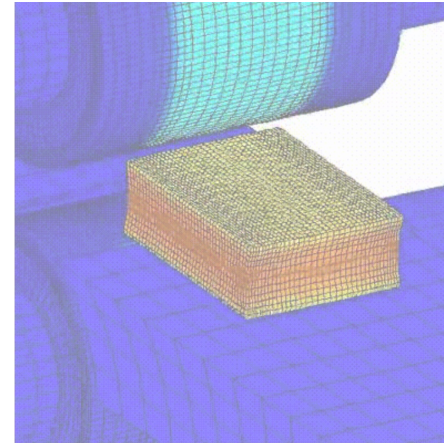
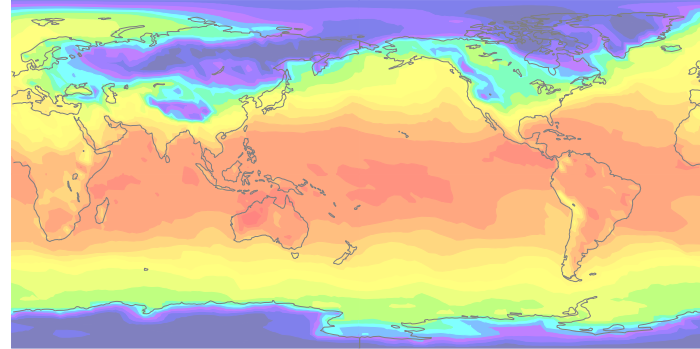
Weather

Geometric meshes

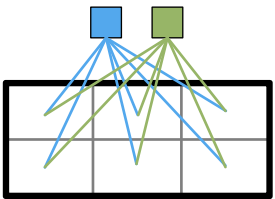
Registered data

Approximate alignment

And more...



Convolution



Data Has Spatial Structure

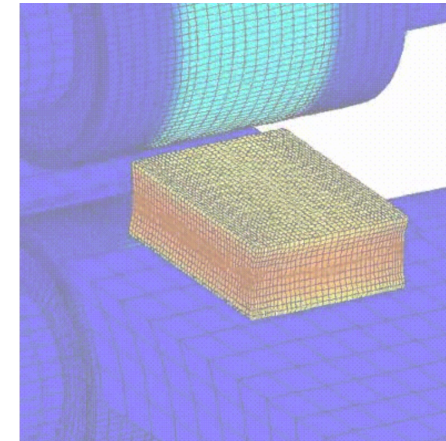
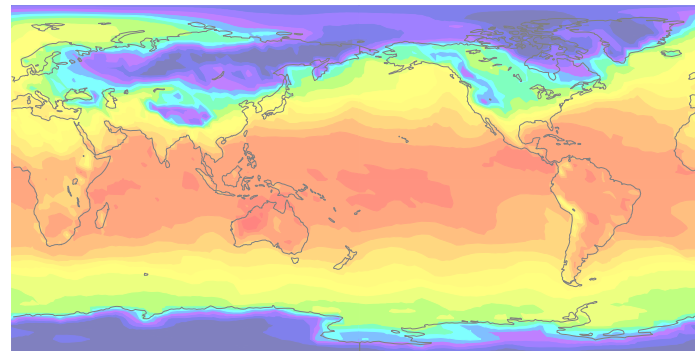
Weather

Geometric meshes

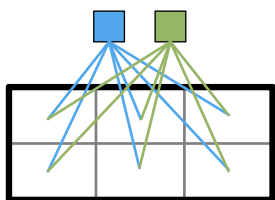
Registered data

Approximate alignment

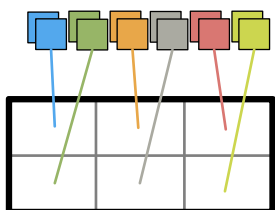
And more...



Convolution



Locally-connected



Data Has Spatial Structure

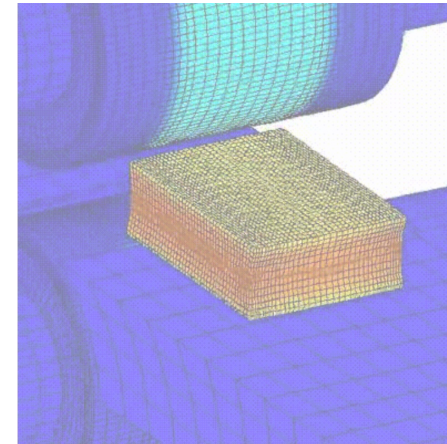
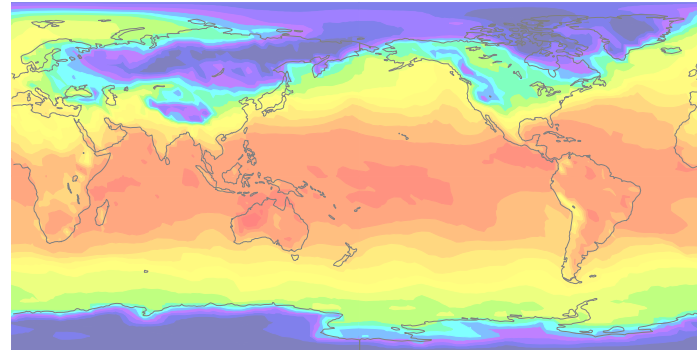
Weather

Geometric meshes

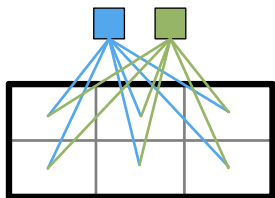
Registered data

Approximate alignment

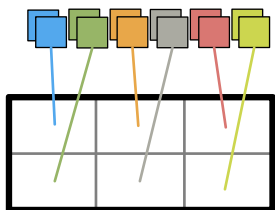
And more...



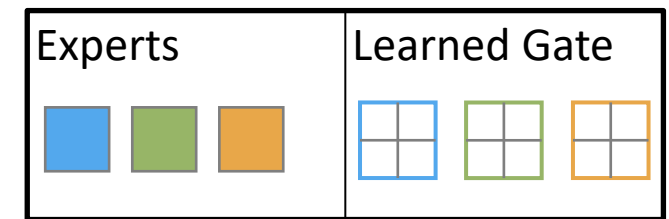
Convolution



Locally-connected



Spatial Mixture-of-Experts



Data Has Spatial Structure

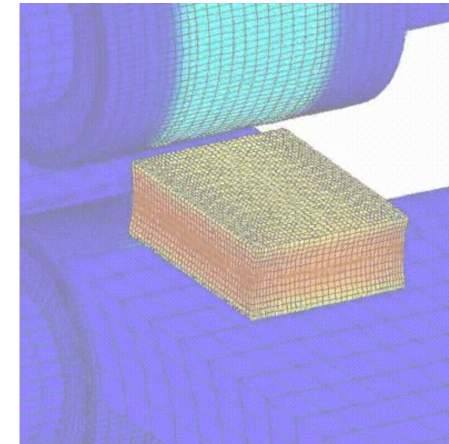
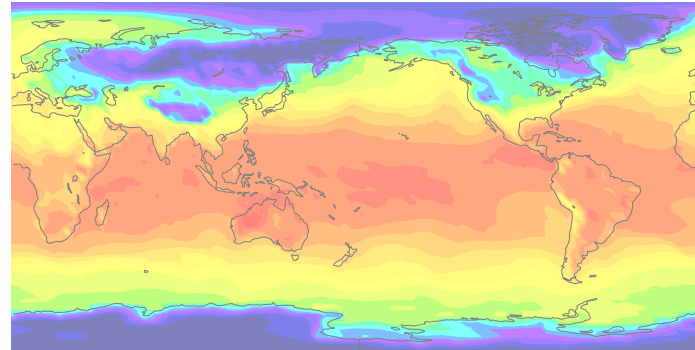
Weather

Geometric meshes

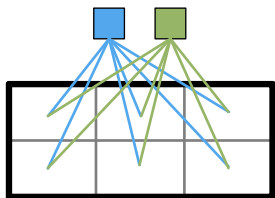
Registered data

Approximate alignment

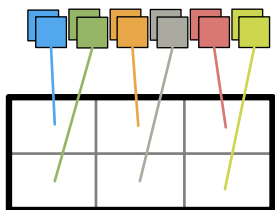
And more...



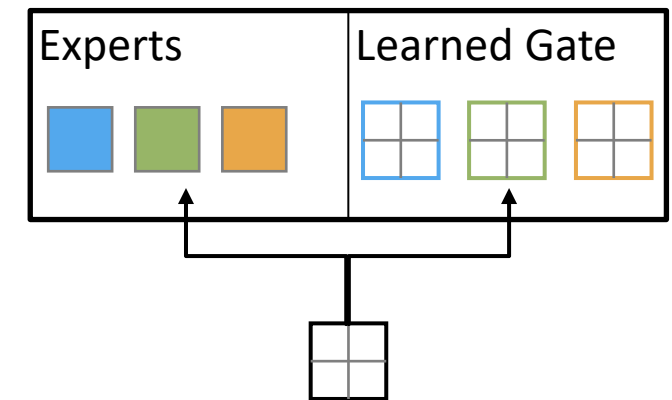
Convolution



Locally-connected



Spatial
Mixture-of-Experts



Data Has Spatial Structure

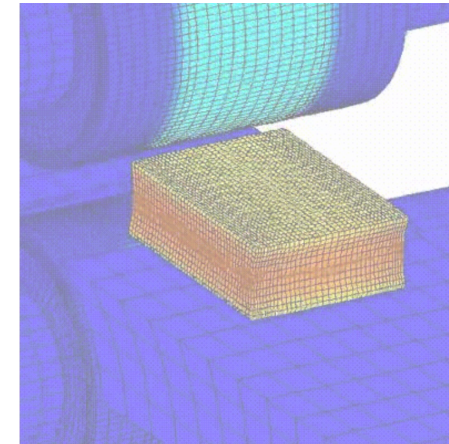
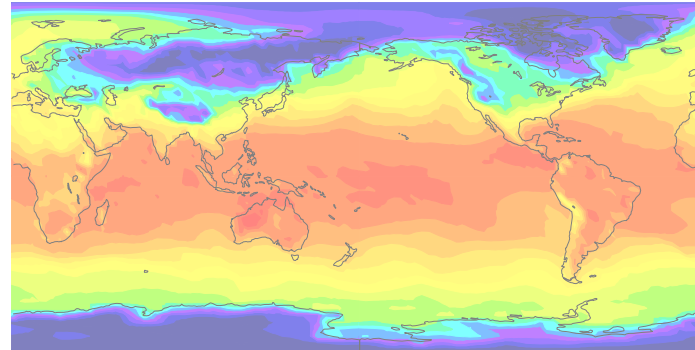
Weather

Geometric meshes

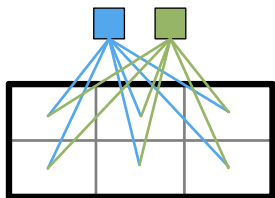
Registered data

Approximate alignment

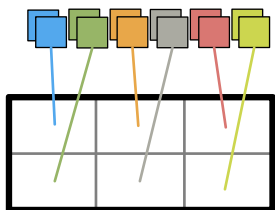
And more...



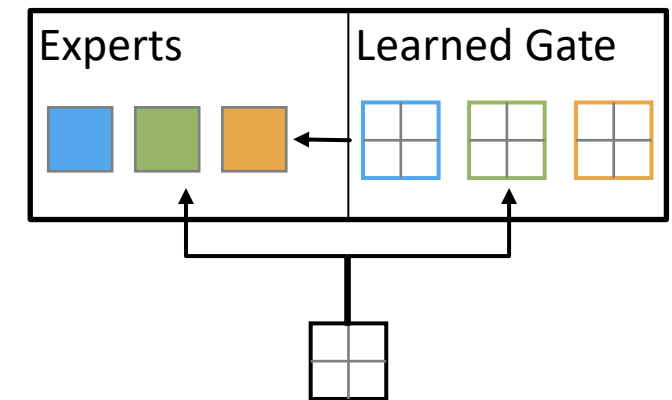
Convolution



Locally-connected



Spatial
Mixture-of-Experts



Data Has Spatial Structure

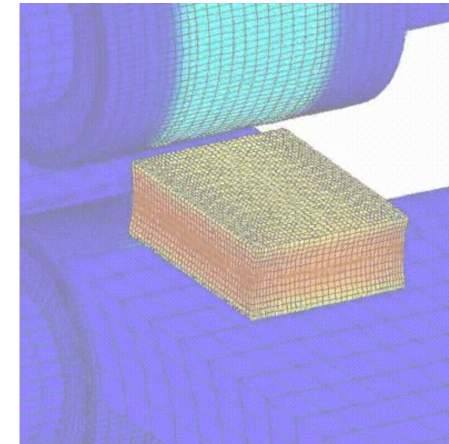
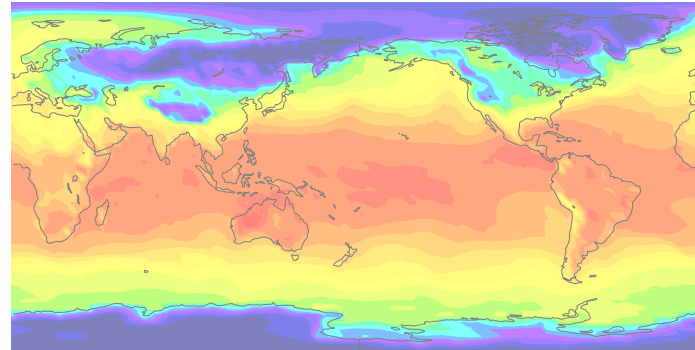
Weather

Geometric meshes

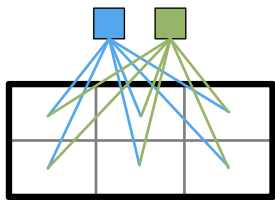
Registered data

Approximate alignment

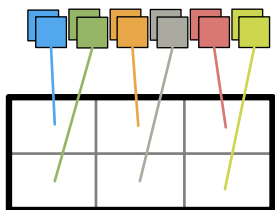
And more...



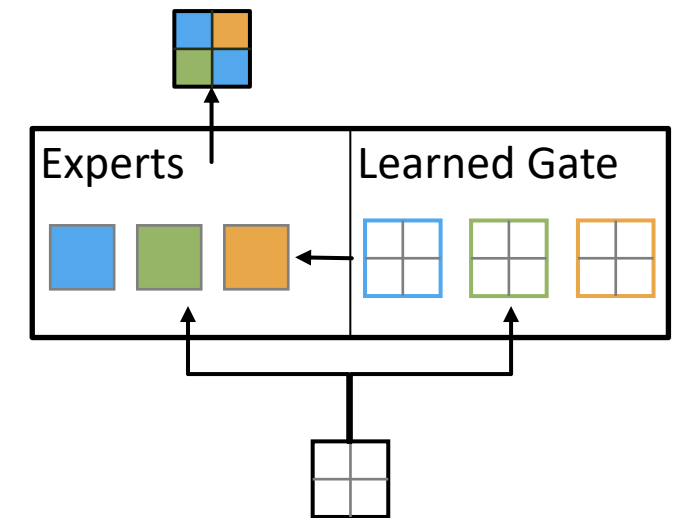
Convolution



Locally-connected



Spatial Mixture-of-Experts



Data Has Spatial Structure

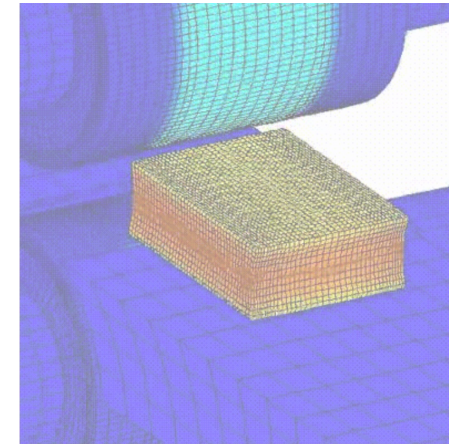
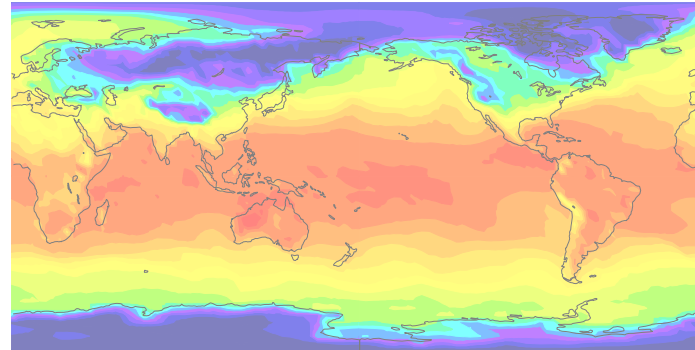
Weather

Geometric meshes

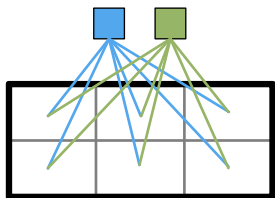
Registered data

Approximate alignment

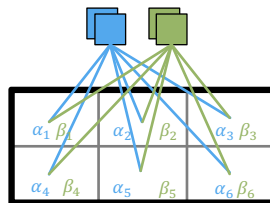
And more...



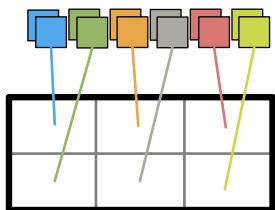
Convolution



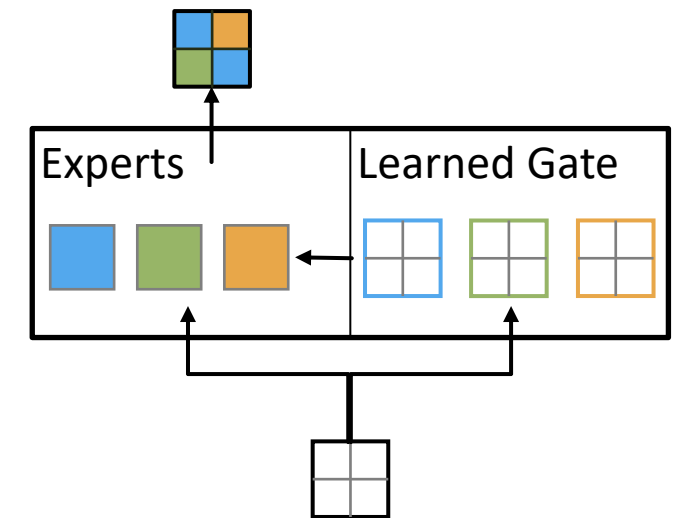
Low-rank locally-connected



Locally-connected



Spatial Mixture-of-Experts



Data Has Spatial Structure

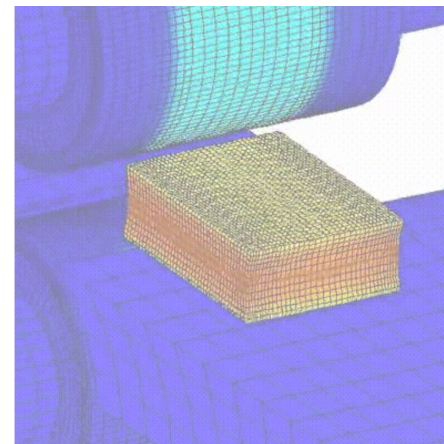
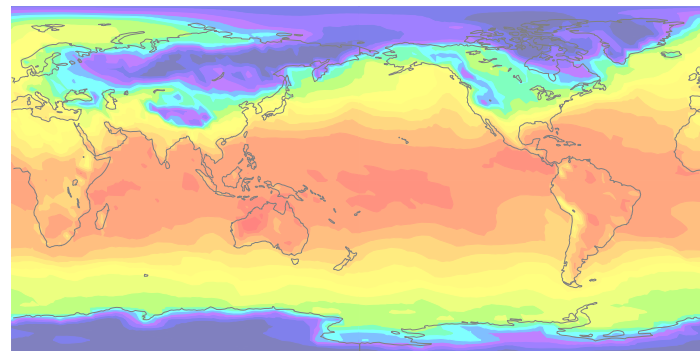
Weather

Geometric meshes

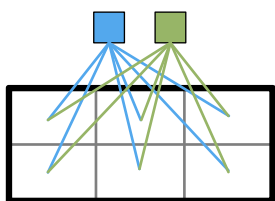
Registered data

Approximate alignment

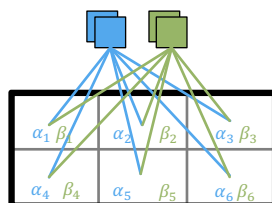
And more...



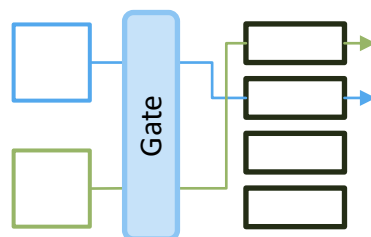
Convolution



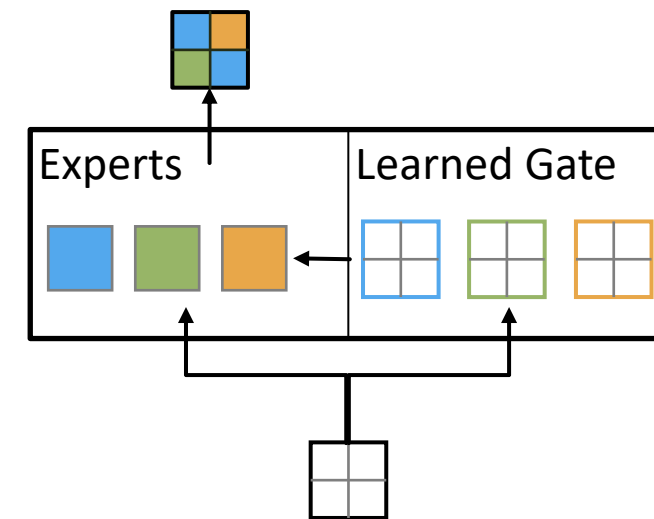
Low-rank locally-connected



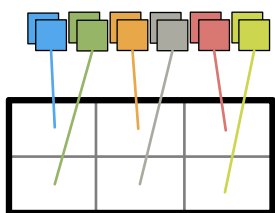
Mixture-of-Experts



Spatial
Mixture-of-Experts



Locally-connected

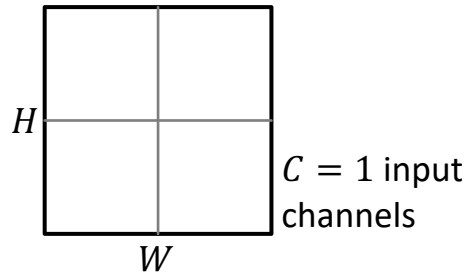


The icon depicts a stylized landscape with a mountain range, a river, and a circular target or map overlay, representing spatial data analysis.

Spatial Mixture-of-Experts



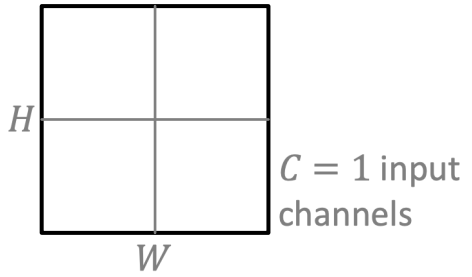
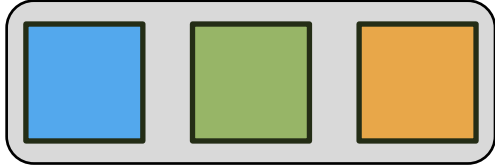
Spatial Mixture-of-Experts





Spatial Mixture-of-Experts

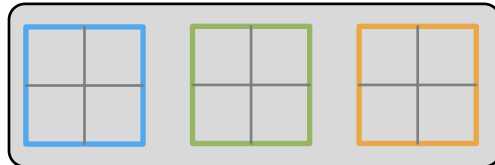
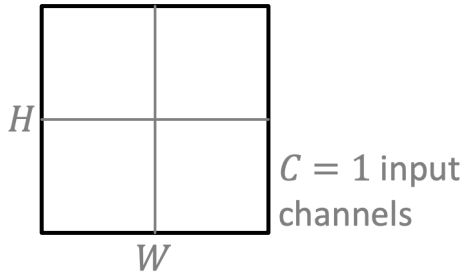
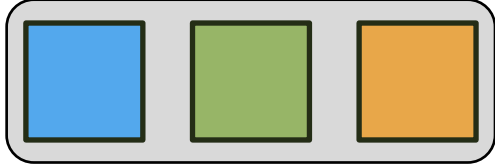
Experts \mathcal{E}





Spatial Mixture-of-Experts

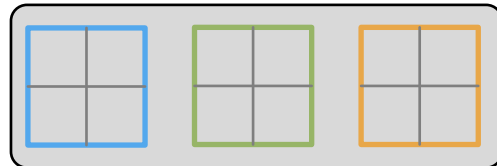
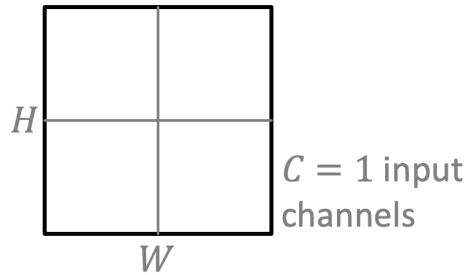
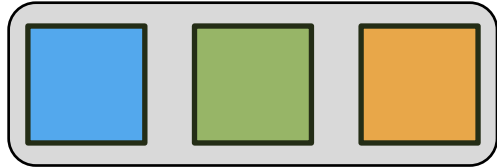
Experts \mathcal{E}





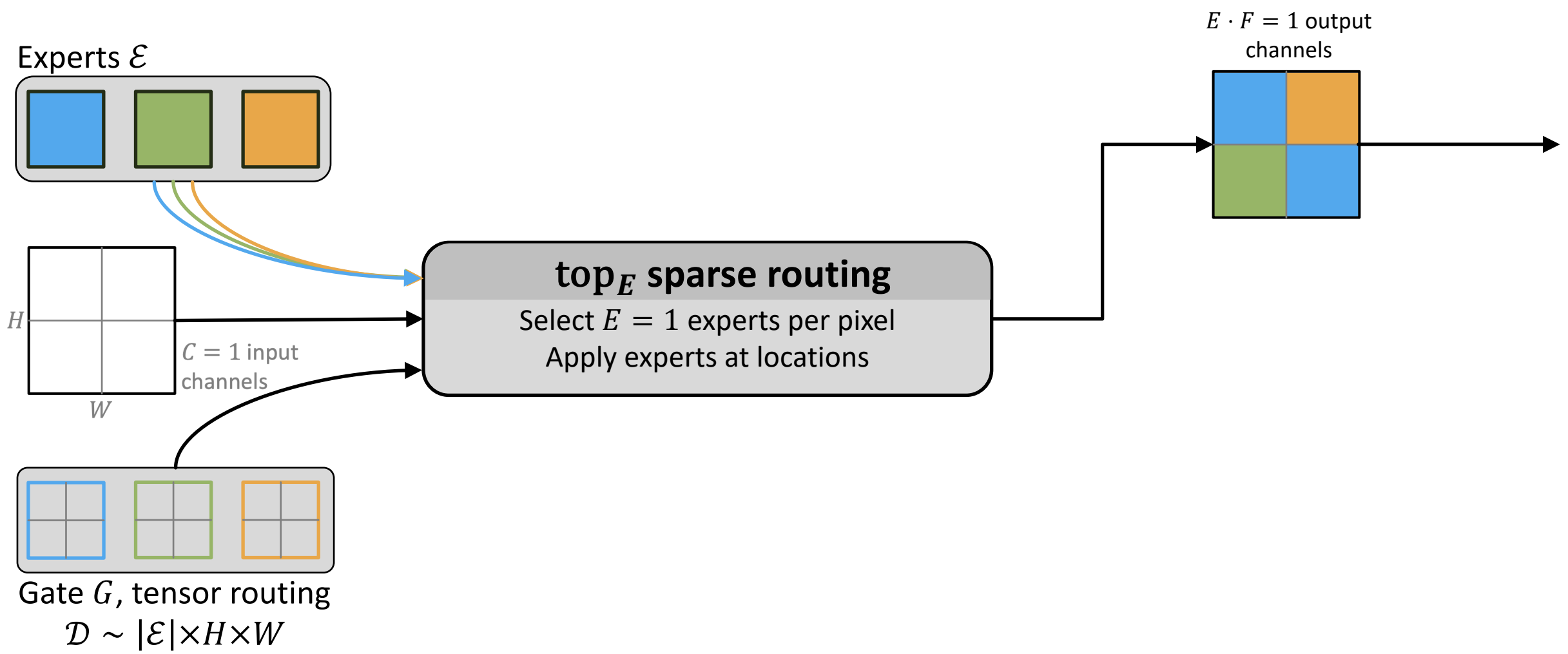
Spatial Mixture-of-Experts

Experts \mathcal{E}

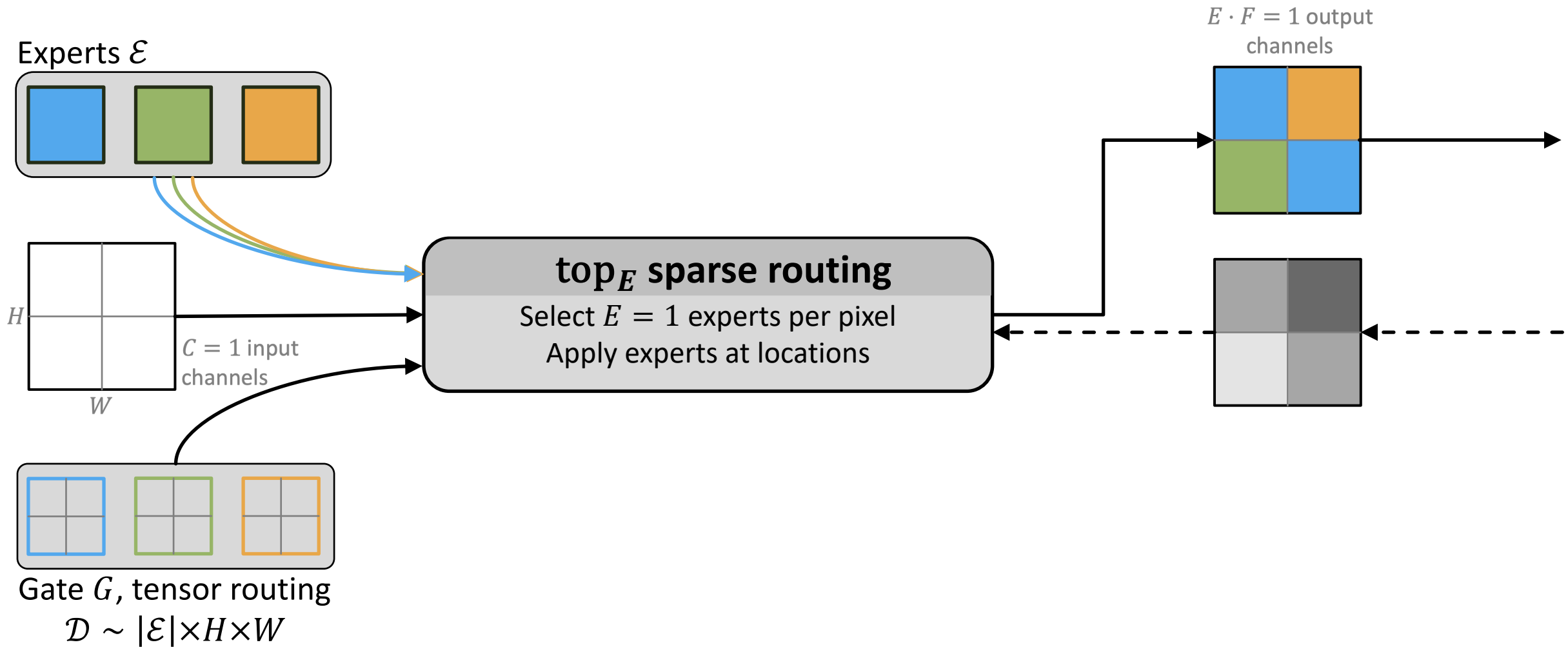


Gate G , tensor routing
 $\mathcal{D} \sim |\mathcal{E}| \times H \times W$

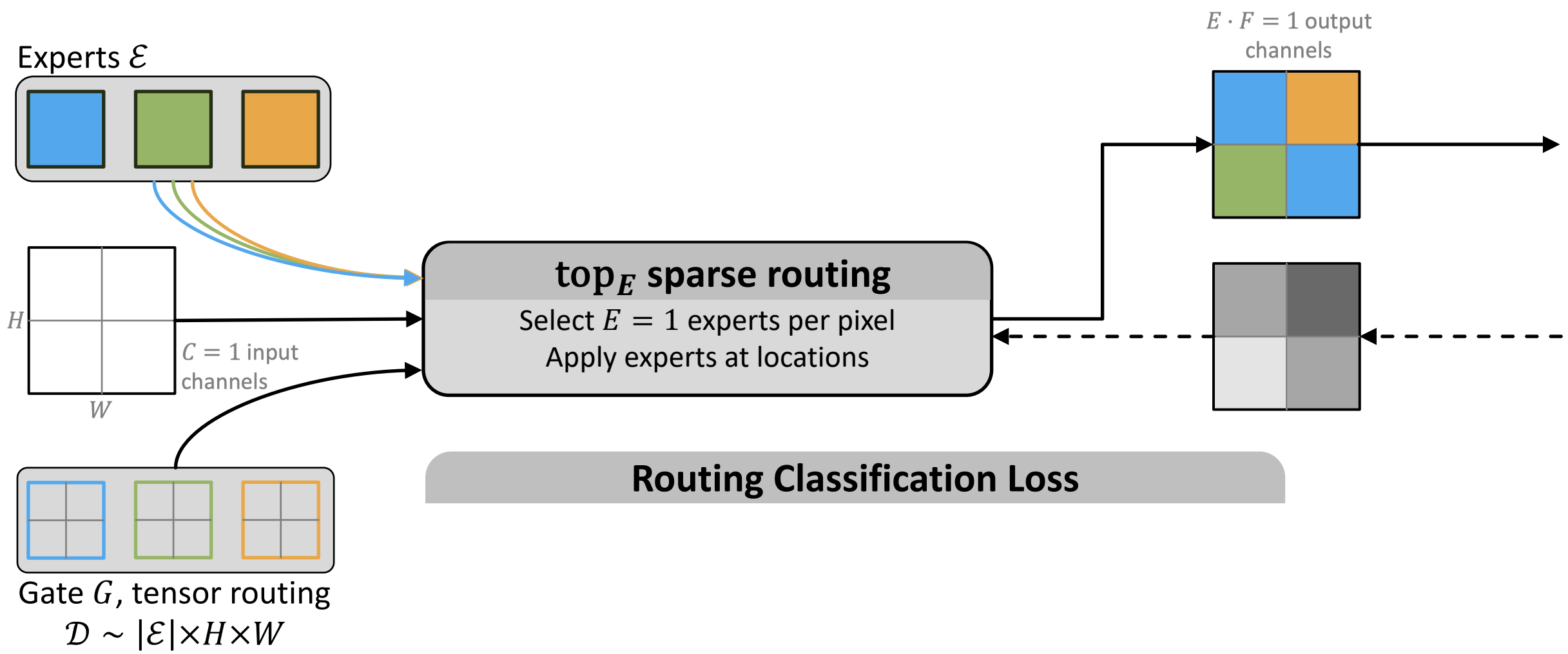
Spatial Mixture-of-Experts



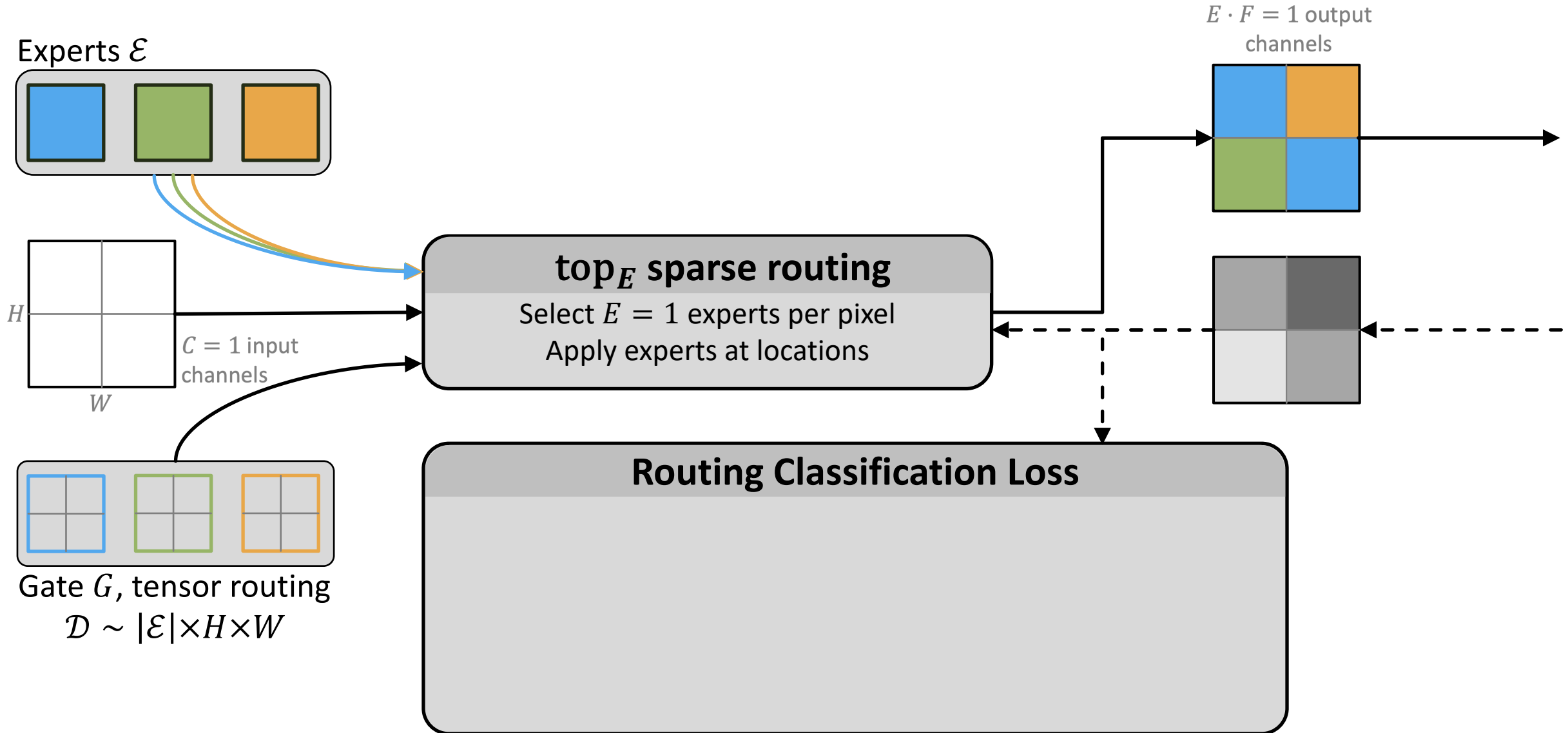
Spatial Mixture-of-Experts



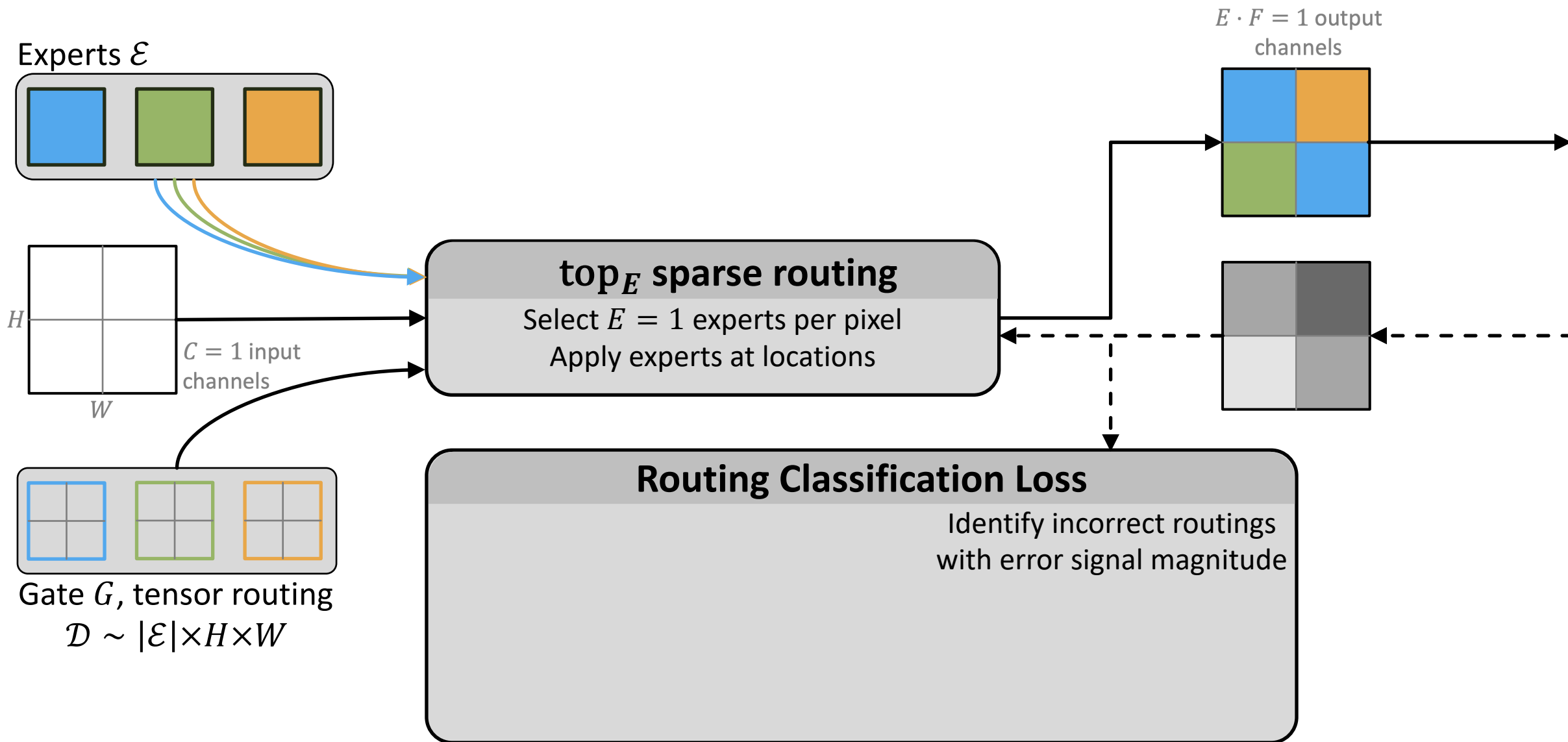
Spatial Mixture-of-Experts



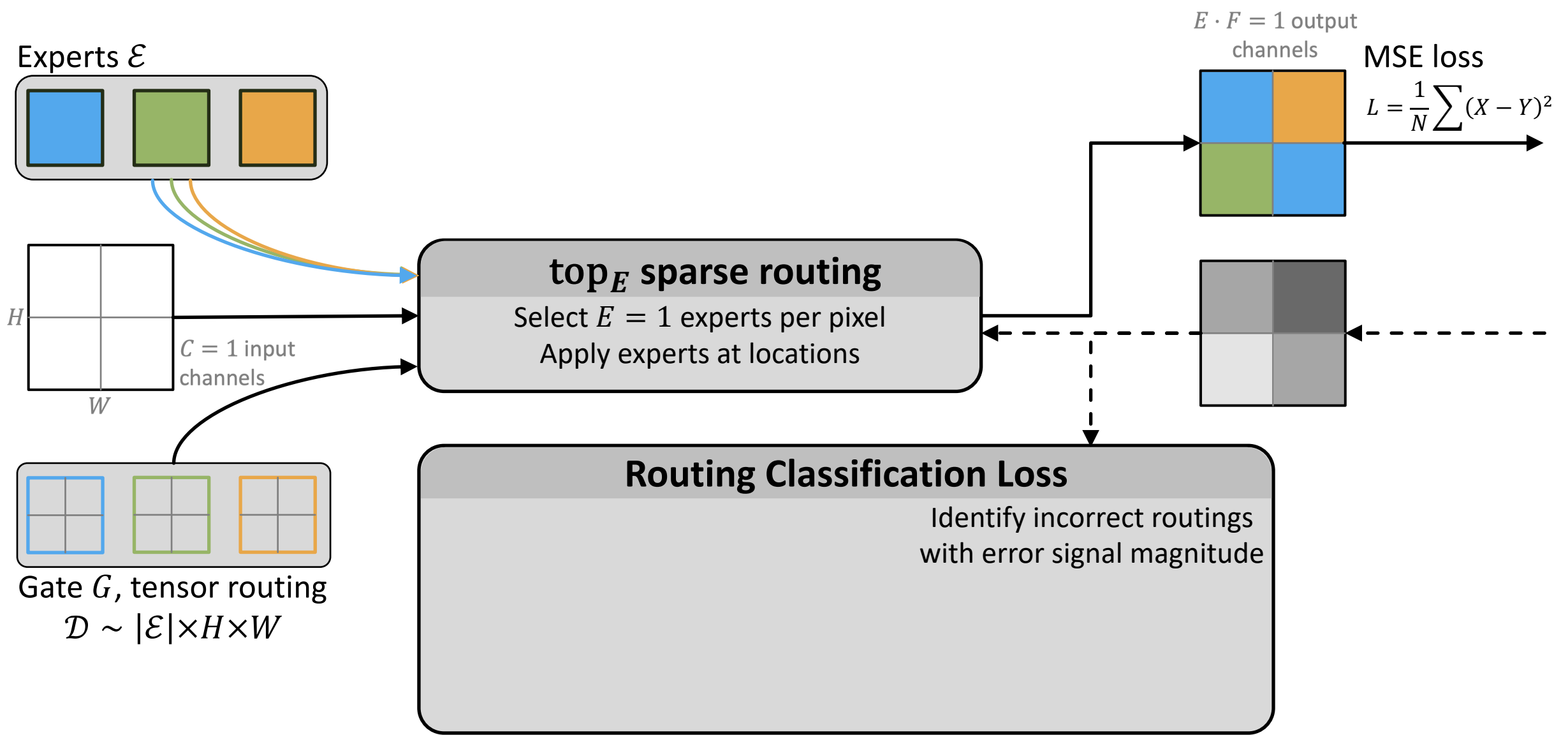
Spatial Mixture-of-Experts



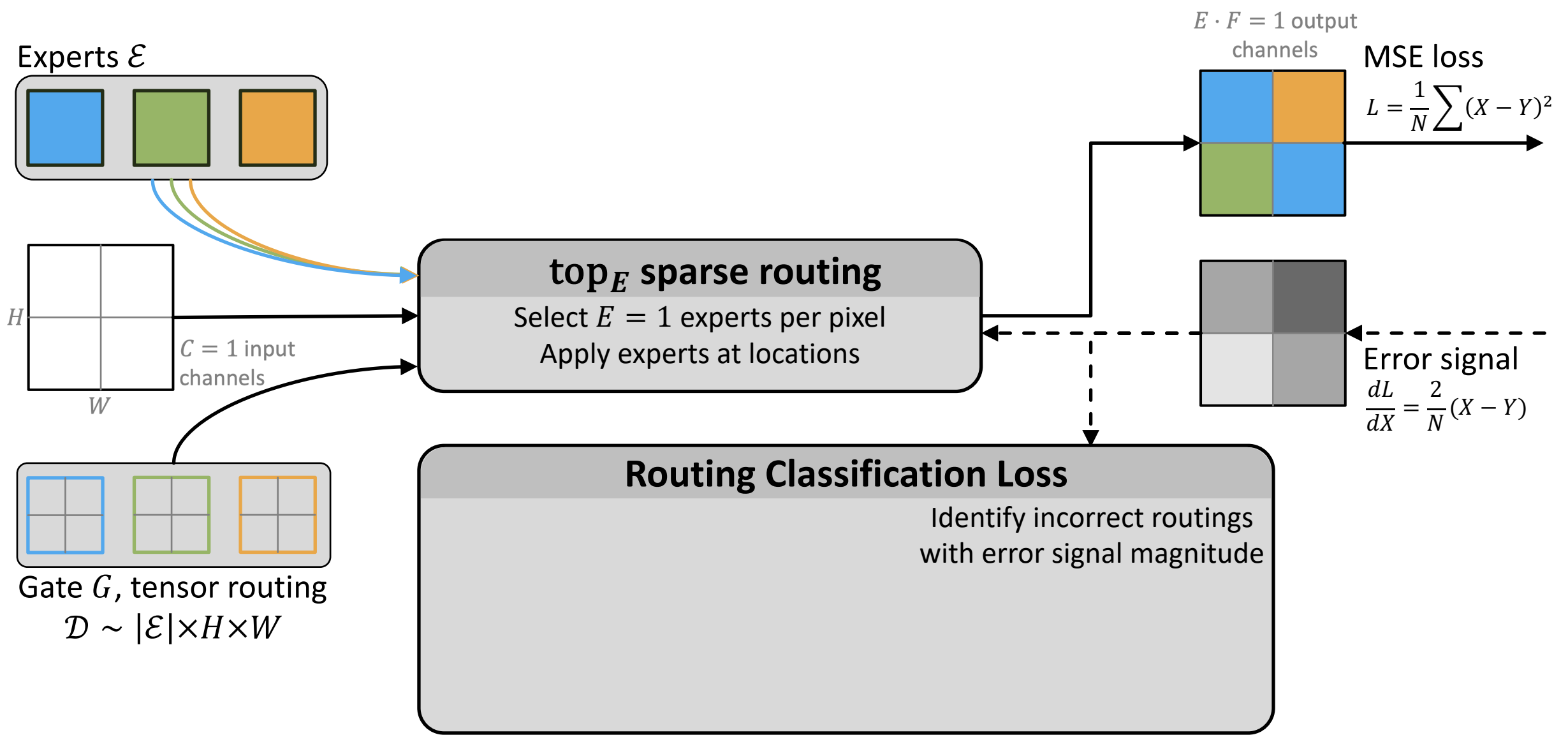
Spatial Mixture-of-Experts



Spatial Mixture-of-Experts

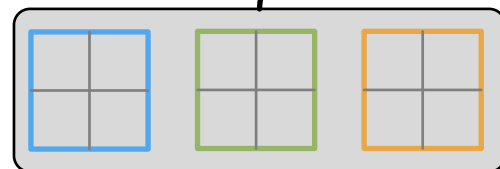
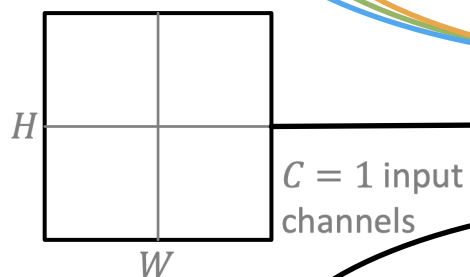
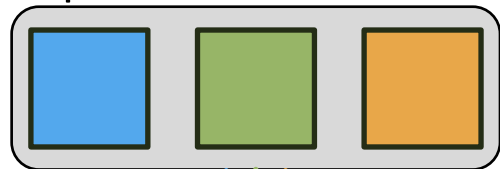


Spatial Mixture-of-Experts



Spatial Mixture-of-Experts

Experts \mathcal{E}



Gate G , tensor routing
 $\mathcal{D} \sim |\mathcal{E}| \times H \times W$

top_E sparse routing
 Select $E = 1$ experts per pixel
 Apply experts at locations

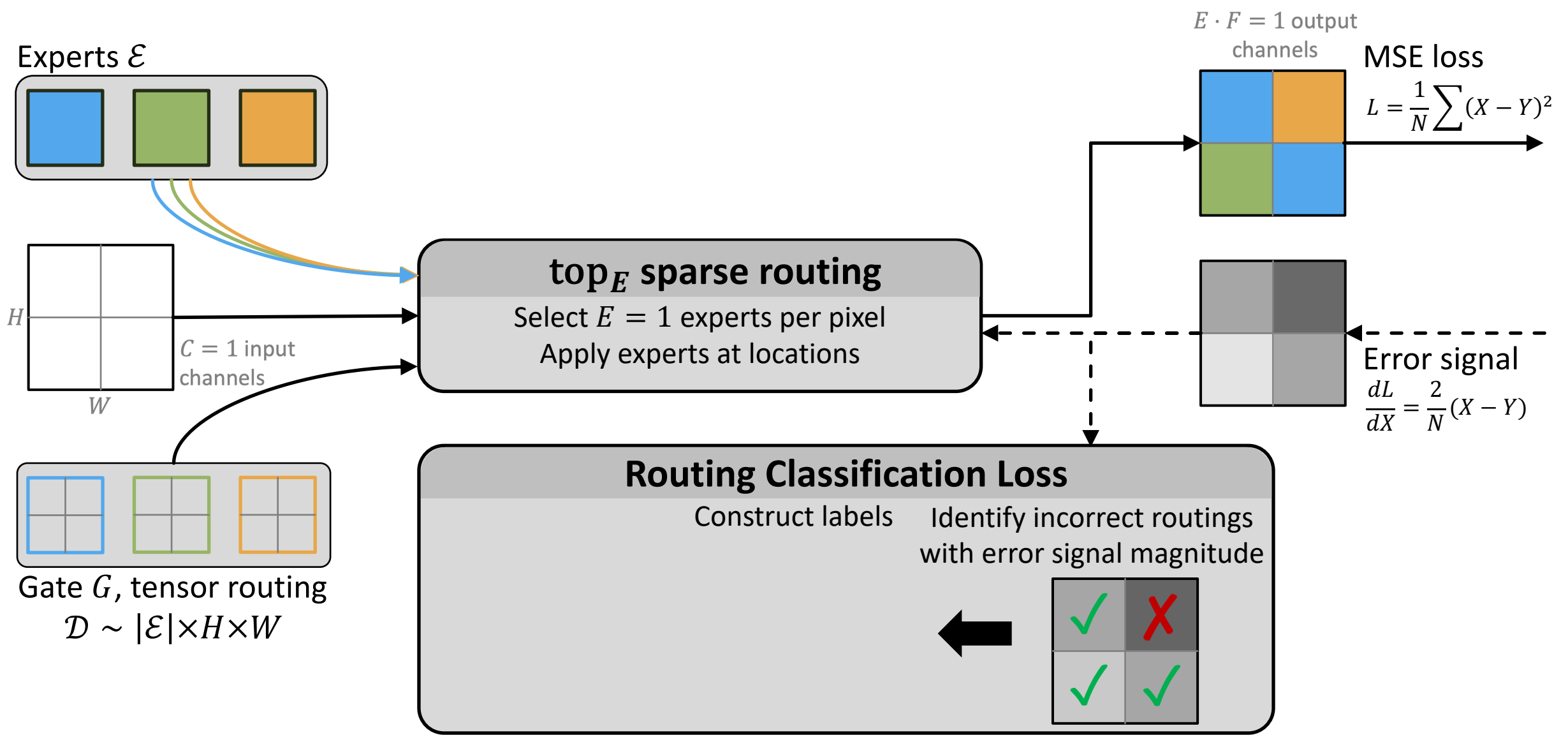
Routing Classification Loss
 Identify incorrect routings with error signal magnitude

$E \cdot F = 1$ output channels

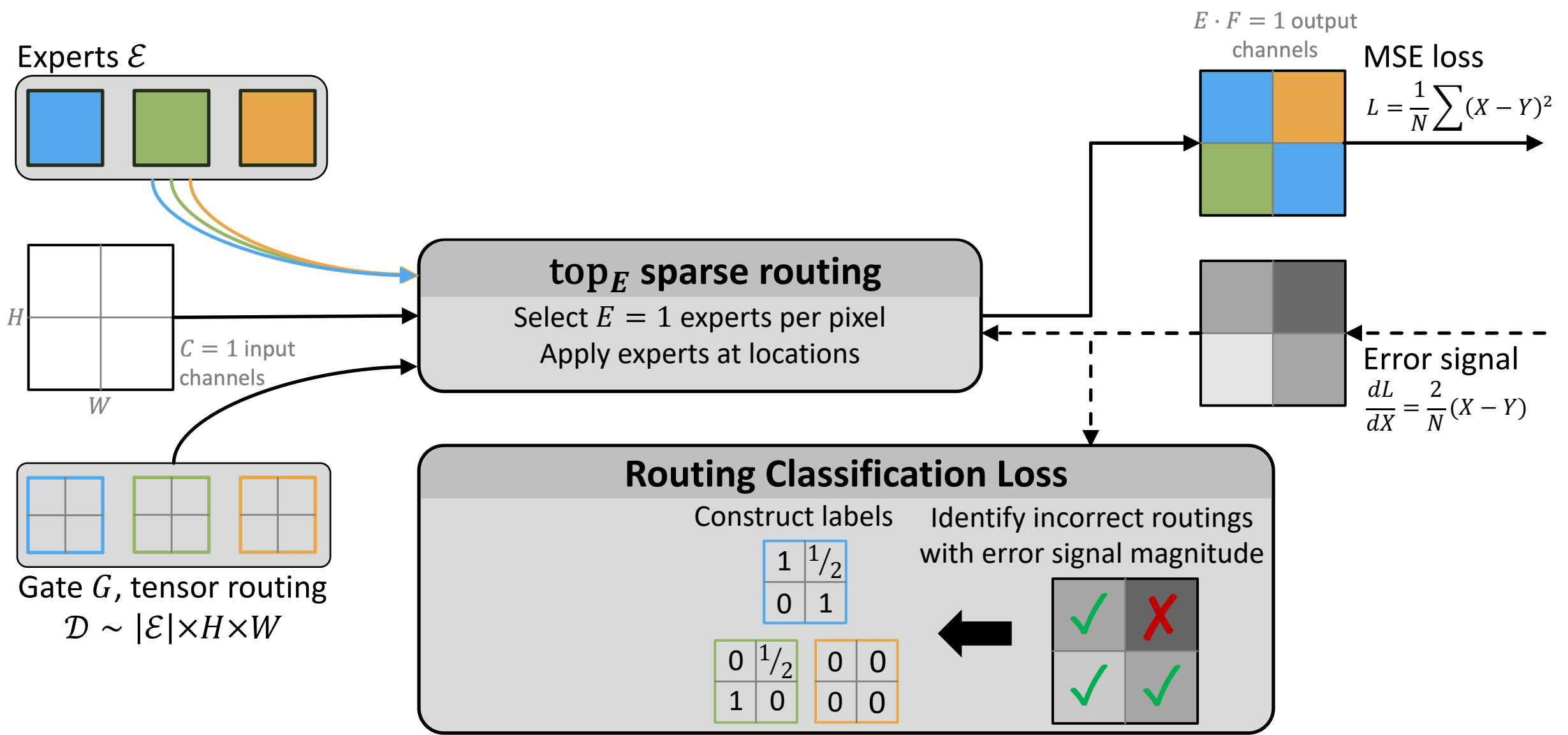
MSE loss
 $L = \frac{1}{N} \sum (X - Y)^2$

Error signal
 $\frac{dL}{dX} = \frac{2}{N} (X - Y)$

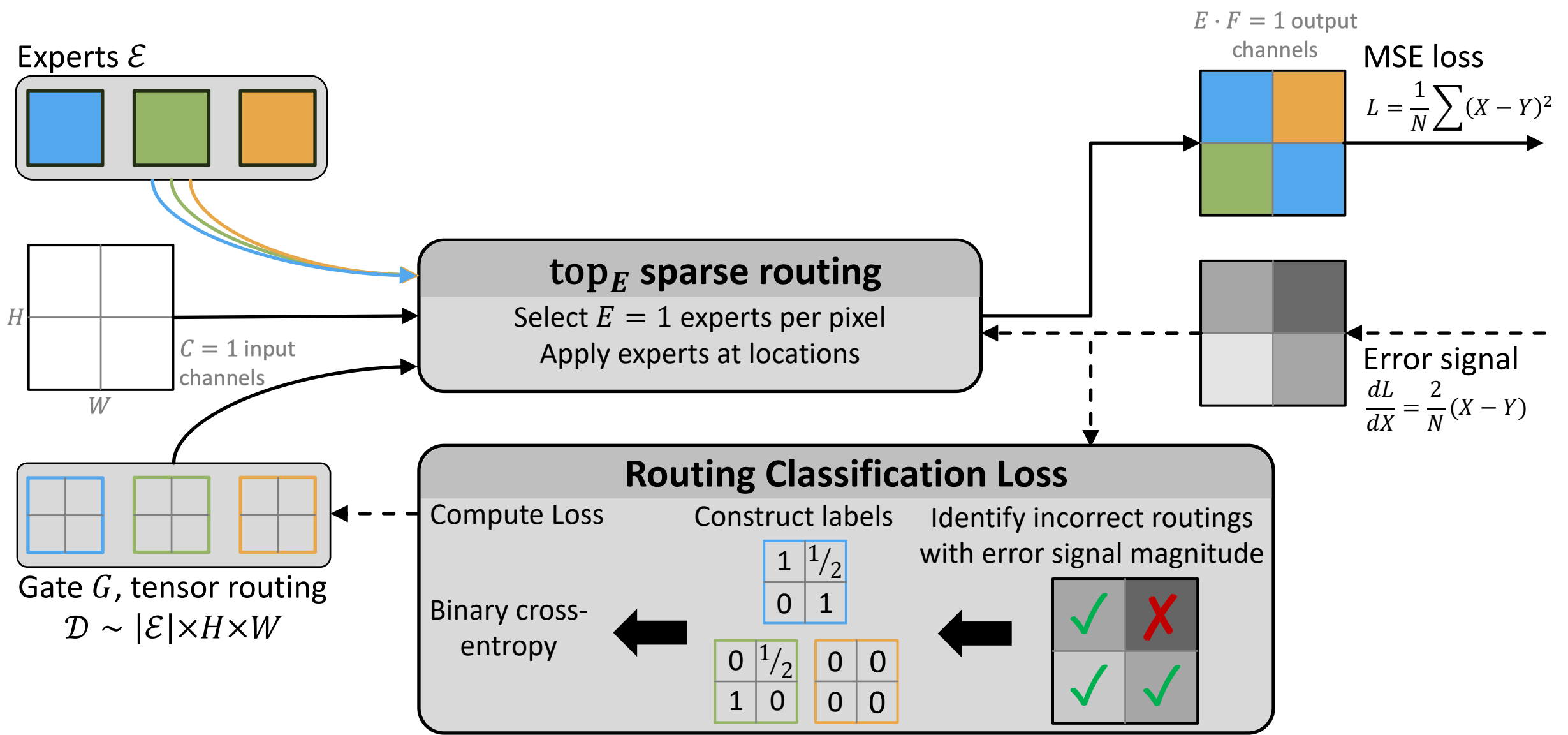
Spatial Mixture-of-Experts



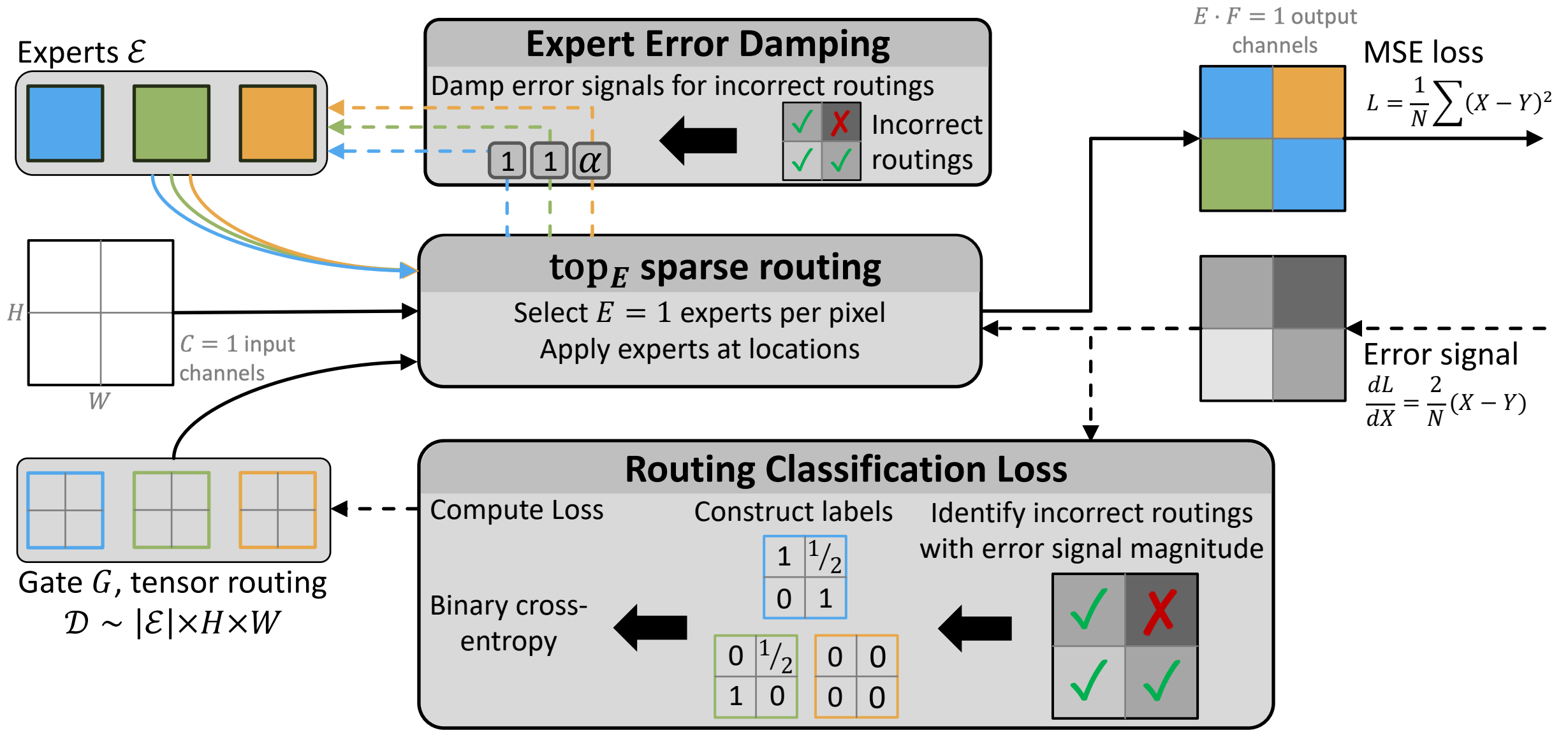
Spatial Mixture-of-Experts



Spatial Mixture-of-Experts



Spatial Mixture-of-Experts





Weather

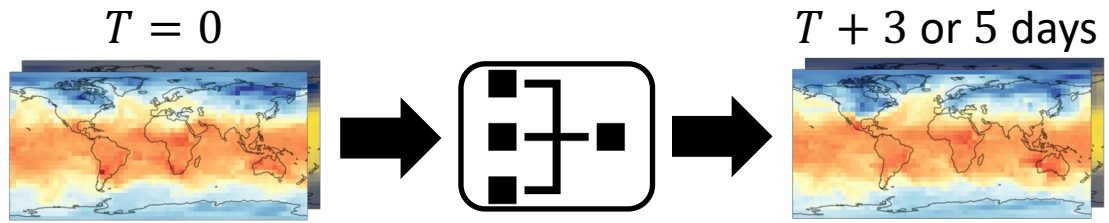
Weather

Medium-range weather prediction **WeatherBench**



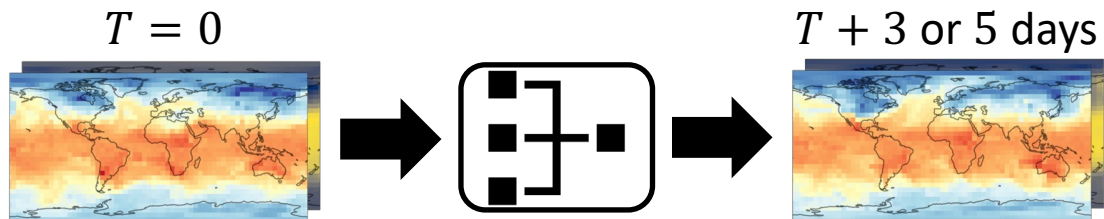
Weather

Medium-range weather prediction **WeatherBench**



Weather

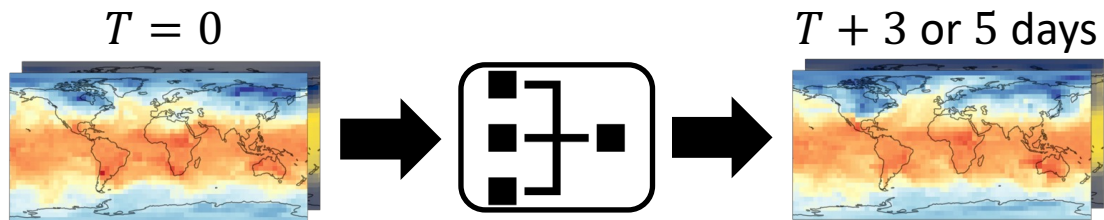
Medium-range weather prediction **WeatherBench**



Model	Z500 [m^2s^{-2}]		T850 [K]	
	3 days	5 days	3 days	5 days
Rasp & Thuerey	316 \pm 2.4	563 \pm 3.1	1.80 \pm 0.02	2.84 \pm 0.03
↪ 2x wide	310 \pm 2.0	555 \pm 2.8	1.76 \pm 0.03	2.78 \pm 0.01
LRLCN	290 \pm 1.4	549 \pm 1.9	1.73 \pm 0.03	2.79 \pm 0.01
ViT	438 \pm 2.8	638 \pm 3.1	2.24 \pm 0.04	2.88 \pm 0.03
SMoE	270\pm2.0	525\pm2.0	1.66\pm0.02	2.60\pm0.01

Weather

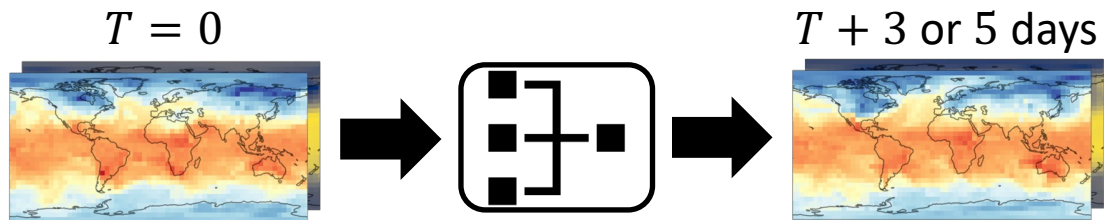
Medium-range weather prediction



Model	Z500 [m^2s^{-2}]		T850 [K]	
	3 days	5 days	3 days	5 days
Rasp & Thuerey	316 \pm 2.4	563 \pm 3.1	1.80 \pm 0.02	2.84 \pm 0.03
➔ 2x wide	310 \pm 2.0	555 \pm 2.8	1.76 \pm 0.03	2.78 \pm 0.01
LRLCN	290 \pm 1.4	549 \pm 1.9	1.73 \pm 0.03	2.79 \pm 0.01
ViT	438 \pm 2.8	638 \pm 3.1	2.24 \pm 0.04	2.88 \pm 0.03
SMoE	270\pm2.0	525\pm2.0	1.66\pm0.02	2.60\pm0.01
R&T (pretrained)	267 \pm 1.8	500 \pm 2.4	1.66 \pm 0.03	2.43 \pm 0.02
SMoE (pretrained)	253 \pm 2.1	488 \pm 1.7	1.57 \pm 0.02	2.34 \pm 0.02
➔ + extra ERA5	232 \pm 1.5	440 \pm 1.2	1.46 \pm 0.02	2.19 \pm 0.01
➔ + 1.4°	198\pm1.8	382\pm2.0	1.42\pm0.00	2.06\pm0.02

Weather

Medium-range weather prediction **WeatherBench**

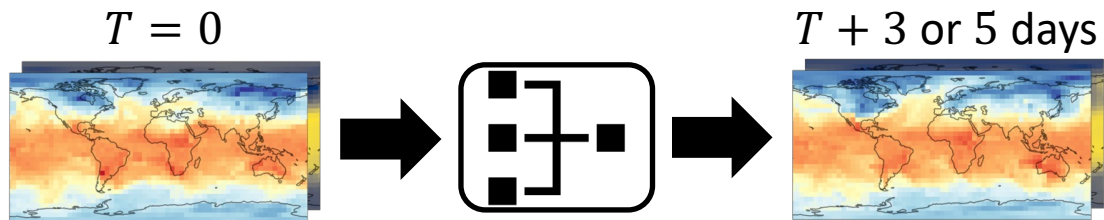


Model	Z500 [m^2s^{-2}]		T850 [K]	
	3 days	5 days	3 days	5 days
Rasp & Thuerey	316 \pm 2.4	563 \pm 3.1	1.80 \pm 0.02	2.84 \pm 0.03
↪ 2x wide	310 \pm 2.0	555 \pm 2.8	1.76 \pm 0.03	2.78 \pm 0.01
LRLCN	290 \pm 1.4	549 \pm 1.9	1.73 \pm 0.03	2.79 \pm 0.01
ViT	438 \pm 2.8	638 \pm 3.1	2.24 \pm 0.04	2.88 \pm 0.03
SMoE	270\pm2.0	525\pm2.0	1.66\pm0.02	2.60\pm0.01
R&T (pretrained)	267 \pm 1.8	500 \pm 2.4	1.66 \pm 0.03	2.43 \pm 0.02
SMoE (pretrained)	253 \pm 2.1	488 \pm 1.7	1.57 \pm 0.02	2.34 \pm 0.02
↪ + extra ERA5	232 \pm 1.5	440 \pm 1.2	1.46 \pm 0.02	2.19 \pm 0.01
↪ + 1.4°	198\pm1.8	382\pm2.0	1.42\pm0.00	2.06\pm0.02

Ensemble post-processing ENS-10

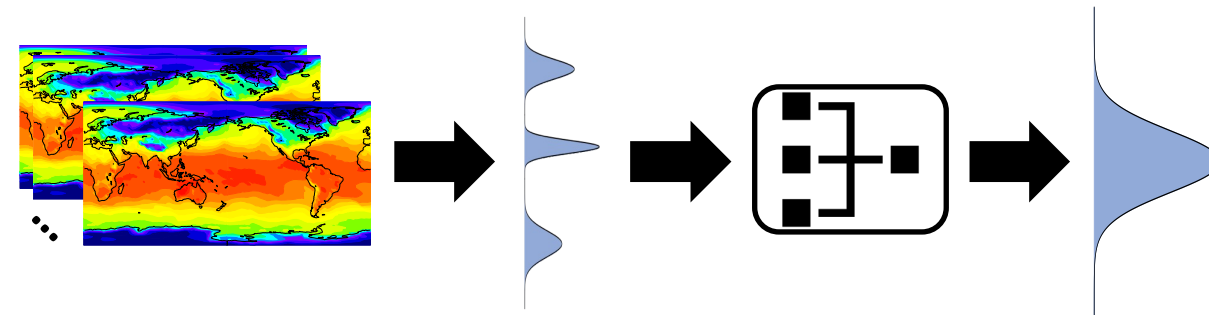
Weather

Medium-range weather prediction **WeatherBench**



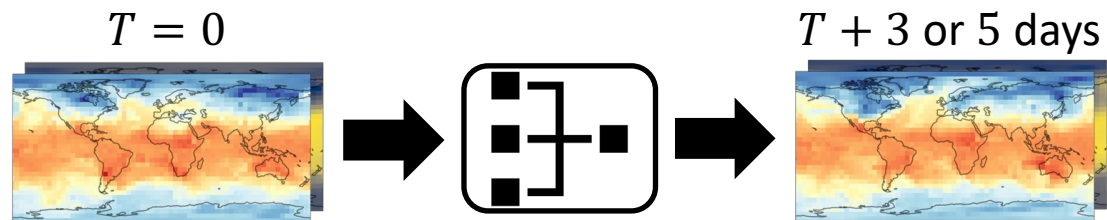
Model	Z500 [m^2s^{-2}]		T850 [K]	
	3 days	5 days	3 days	5 days
Rasp & Thuerey	316 \pm 2.4	563 \pm 3.1	1.80 \pm 0.02	2.84 \pm 0.03
↪ 2x wide	310 \pm 2.0	555 \pm 2.8	1.76 \pm 0.03	2.78 \pm 0.01
LRLCN	290 \pm 1.4	549 \pm 1.9	1.73 \pm 0.03	2.79 \pm 0.01
ViT	438 \pm 2.8	638 \pm 3.1	2.24 \pm 0.04	2.88 \pm 0.03
SMoE	270\pm2.0	525\pm2.0	1.66\pm0.02	2.60\pm0.01
R&T (pretrained)	267 \pm 1.8	500 \pm 2.4	1.66 \pm 0.03	2.43 \pm 0.02
SMoE (pretrained)	253 \pm 2.1	488 \pm 1.7	1.57 \pm 0.02	2.34 \pm 0.02
↪ + extra ERA5	232 \pm 1.5	440 \pm 1.2	1.46 \pm 0.02	2.19 \pm 0.01
↪ + 1.4°	198\pm1.8	382\pm2.0	1.42\pm0.00	2.06\pm0.02

Ensemble post-processing **ENS-10**



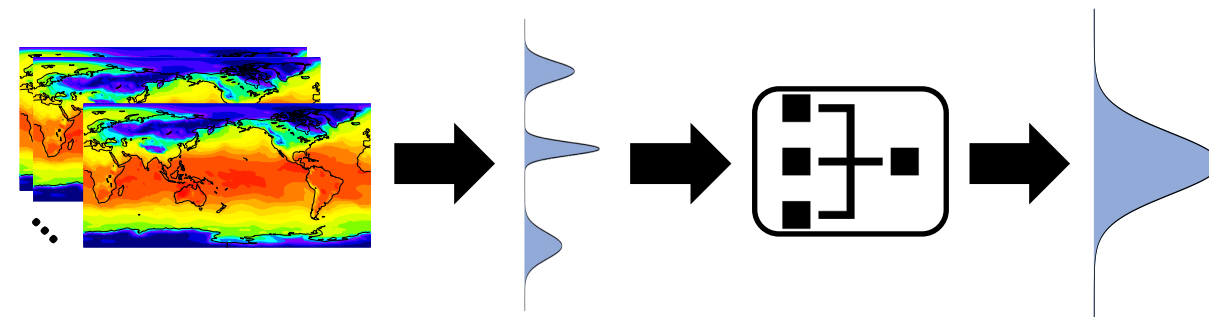
Weather

Medium-range weather prediction **WeatherBench**



Model	Z500 [m^2s^{-2}]		T850 [K]	
	3 days	5 days	3 days	5 days
Rasp & Thuerey	316 \pm 2.4	563 \pm 3.1	1.80 \pm 0.02	2.84 \pm 0.03
↪ 2x wide	310 \pm 2.0	555 \pm 2.8	1.76 \pm 0.03	2.78 \pm 0.01
LRLCN	290 \pm 1.4	549 \pm 1.9	1.73 \pm 0.03	2.79 \pm 0.01
ViT	438 \pm 2.8	638 \pm 3.1	2.24 \pm 0.04	2.88 \pm 0.03
SMoE	270\pm2.0	525\pm2.0	1.66\pm0.02	2.60\pm0.01
R&T (pretrained)	267 \pm 1.8	500 \pm 2.4	1.66 \pm 0.03	2.43 \pm 0.02
SMoE (pretrained)	253 \pm 2.1	488 \pm 1.7	1.57 \pm 0.02	2.34 \pm 0.02
↪ + extra ERA5	232 \pm 1.5	440 \pm 1.2	1.46 \pm 0.02	2.19 \pm 0.01
↪ + 1.4°	198\pm1.8	382\pm2.0	1.42\pm0.00	2.06\pm0.02

Ensemble post-processing **ENS-10**



Model	Z500 [m^2s^{-2}]		T850 [K]		T2M [K]	
	5 ens	10 ens	5 ens	10 ens	5 ens	10 ens
CRPS						
EMOS	79.12 \pm 0.12	78.80 \pm 0.21	0.721 \pm 0.01	0.706 \pm 0.04	0.720 \pm 0.00	0.711 \pm 0.03
U-Net	76.54 \pm 0.20	76.18 \pm 0.12	0.685 \pm 0.00	0.670 \pm 0.01	0.657 \pm 0.01	0.644 \pm 0.01
SMoE	68.94\pm0.14	67.43\pm0.12	0.612\pm0.01	0.590\pm0.02	0.601\pm0.02	0.594\pm0.02
ECCRP						
EMOS	29.21 \pm 0.18	29.02 \pm 0.13	0.247 \pm 0.00	0.245 \pm 0.02	0.244 \pm 0.00	0.241 \pm 0.02
U-Net	27.78 \pm 0.11	27.55 \pm 0.19	0.230 \pm 0.01	0.229 \pm 0.01	0.225 \pm 0.00	0.220 \pm 0.01
SMoE	23.79\pm0.20	23.10\pm0.16	0.207\pm0.03	0.197\pm0.03	0.199\pm0.01	0.190\pm0.02