



On Calibrating Semantic Segmentation Models: Analyses and An Algorithm

Dongdong Wang¹, Boqing Gong², and Liqiang Wang¹

¹University of Central Florida ²Google

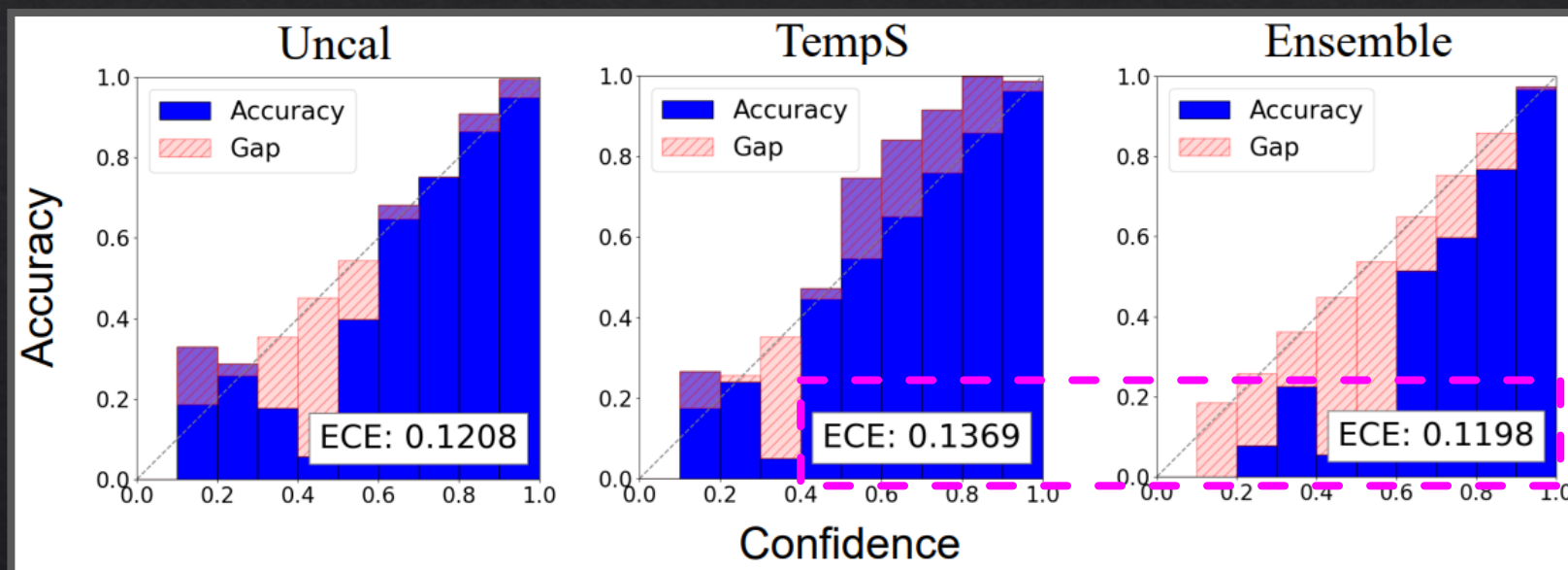
THU-PM-291

Presentation Roadmap (4-*n*s)

- *Motivation***n**
- *Investigation***n**
- *Innovation***n**
- *Observation***n**

- **Motivation**

- Evaluate/Study/Improve semantic segmentation model calibration.



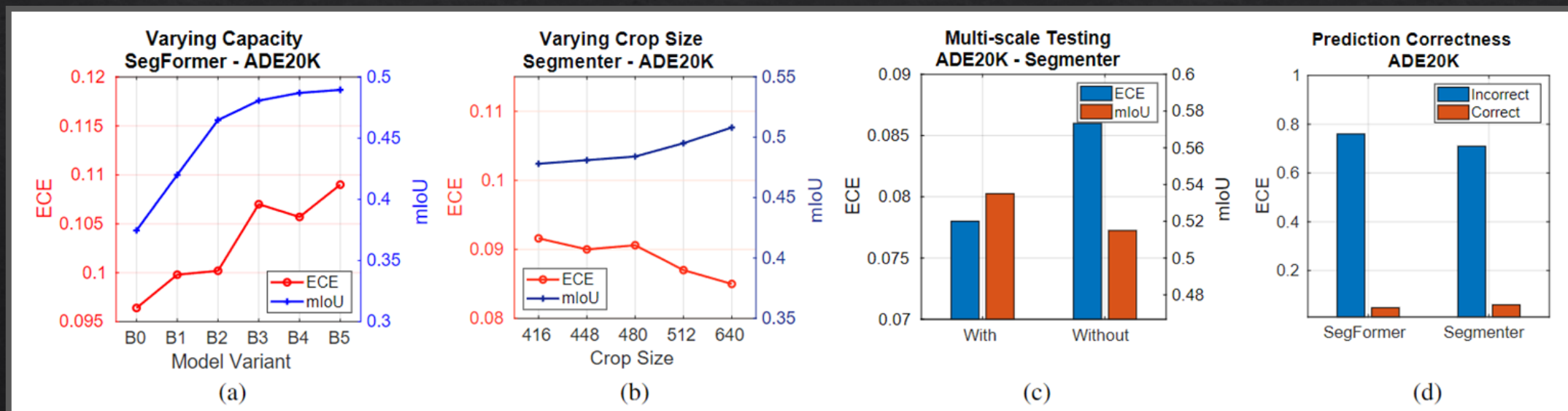
Segmenter on COCO-164K

Presentation Roadmap (4-*n*s)

- *Motivation*
- *Investigation***n**
- *Innovation***n**
- *Observation***n**

- **Investigation**

- Assess five SOTA models on ADE20K from four perspectives.



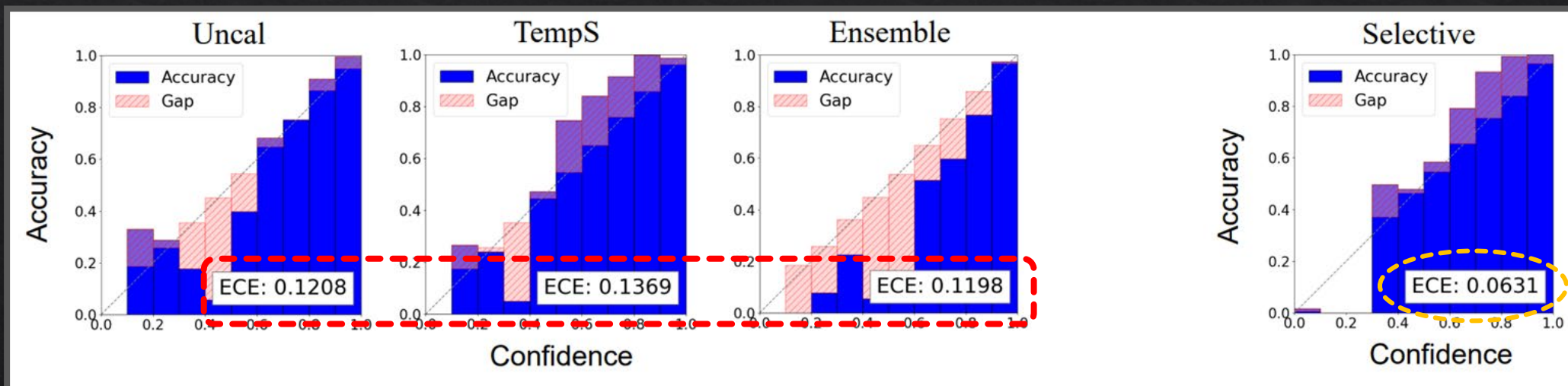
Presentation Roadmap (4-*n*s)

- *Motivation*
- *Investigation*
- *Innovation***n**
- *Observation***n**

- **Innovation**

- Propose a selective scaling to improve segmentation calibration.

$$\begin{cases} \hat{p} = \sigma_{SM}(z/T_1) & \text{if } \hat{y} \neq y \\ \hat{p} = \sigma_{SM}(z/T_2) & \text{if } \hat{y} = y \end{cases}$$



Presentation Roadmap (4-*n*s)

- *Motivation*
- *Investigation*
- *Innovation*
- *Observation***n**

• Observation

- Experiment with both in-domain and shift-domain data to justify the consistent superiority over existing calibrators.

In-domain Calibration

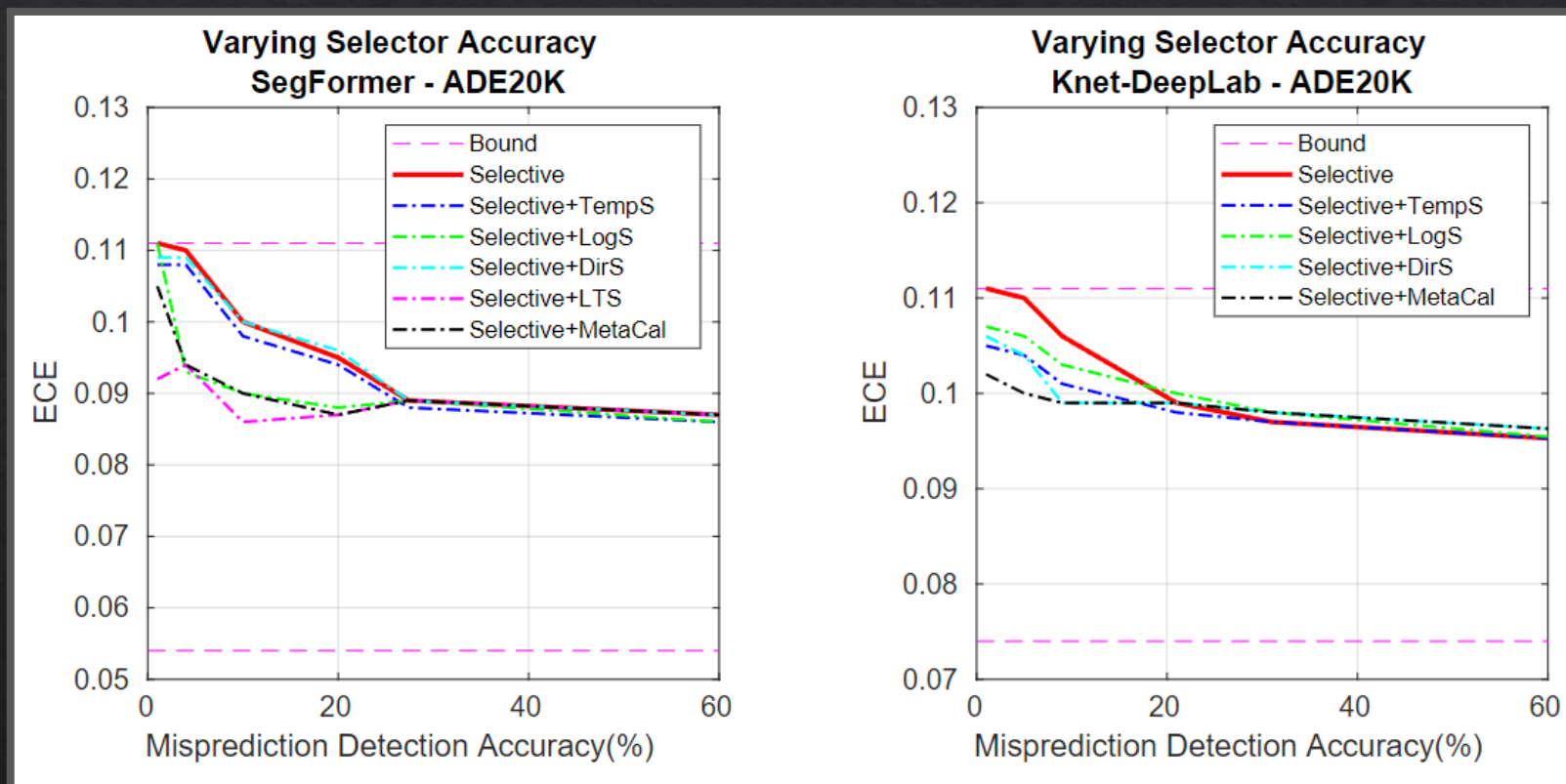
Dataset	Model	mIoU	Uncal	TempS	LogS	DirS	Meta-Cal*	LTS	Ens.	Selective
ADE20K	SegFormer-B5 [49]	49.13	0.111	0.109	0.110	0.110	0.103	0.105	0.109	0.086
ADE20K	Segmenter-L [40]	51.65	0.087	0.086	0.086	0.087	0.081	0.094	0.086	0.069
ADE20K	Knet-DeepLab [51]	45.06	0.111	0.105	0.107	0.106	0.102	0.118	0.110	0.095
ADE20K	Knet-SWIN-L [51]	52.46	0.098	0.094	0.093	0.097	0.089	0.134	0.096	0.078
ADE20K	ConvNeXt-L [24]	53.16	0.097	0.092	0.094	0.091	0.088	0.133	0.094	0.082
COCO-164K	SegFormer-B5 [49]	45.78	0.151	0.149	0.141	0.151	0.132	0.151	0.149	0.113
COCO-164K	Segmenter-L [40]	47.09	0.152	0.149	0.149	0.151	0.130	0.155	0.150	0.109
COCO-164K	Knet-DeepLab [51]	37.24	0.170	0.170	0.168	0.171	0.149	0.172	0.169	0.093
COCO-164K	Knet-SWIN-L [51]	46.49	0.161	0.159	0.161	0.160	0.142	0.162	0.160	0.102
COCO-164K	ConvNeXt-L [24]	46.48	0.160	0.157	0.158	0.159	0.141	0.162	0.159	0.108
BDD100K	SegFormer-B5 [49]	65.08	0.064	0.055	0.054	0.053	0.049	0.069	0.059	0.040
BDD100K	Segmenter-L [40]	61.33	0.055	0.045	0.043	0.042	0.037	0.071	0.052	0.031
BDD100K	Knet-DeepLab [51]	62.89	0.060	0.049	0.047	0.048	0.041	0.063	0.057	0.035
BDD100K	Knet-SWIN-L [51]	67.59	0.065	0.055	0.054	0.054	0.049	0.067	0.063	0.040
BDD100K	ConvNeXt-L [24]	67.26	0.064	0.054	0.053	0.056	0.049	0.065	0.063	0.038

Shift-domain Calibration

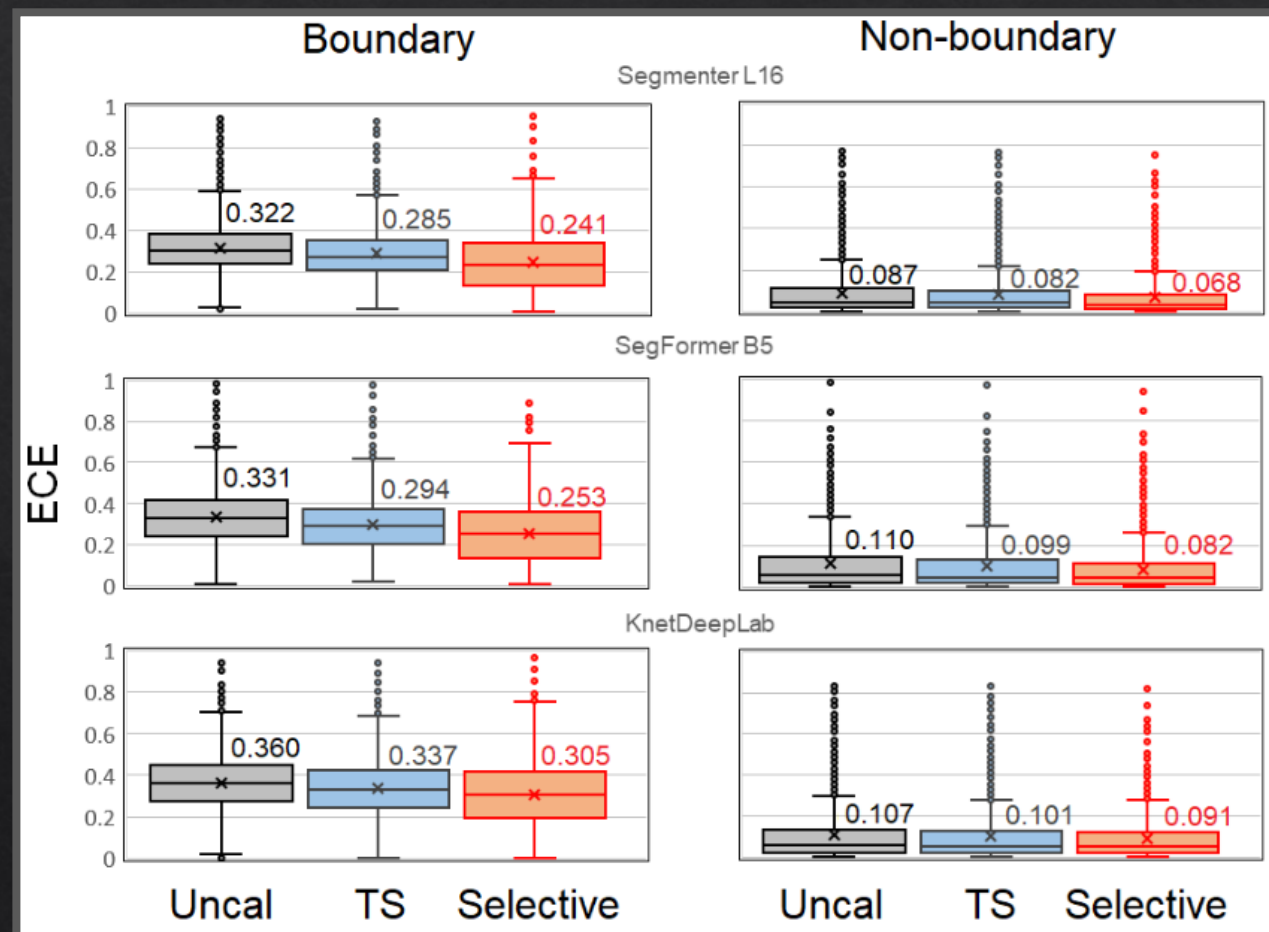
InD	SD	Model	mIoU	Uncal	TempS	LogS	DirS	Meta-Cal*	LTS	Ens.	Selective
ADE20K	COCO-164K	SegFormer-B5 [49]	8.20	0.467	0.466	0.468	0.469	0.415	0.420	0.465	0.331
ADE20K	COCO-164K	Segmenter-L [40]	9.60	0.395	0.394	0.395	0.394	0.322	0.331	0.394	0.259
ADE20K	COCO-164K	Knet-DeepLab [51]	7.02	0.447	0.445	0.446	0.446	0.403	0.439	0.444	0.351
ADE20K	COCO-164K	Knet-SWIN-L [51]	9.06	0.464	0.462	0.463	0.463	0.379	0.465	0.462	0.327
ADE20K	COCO-164K	ConvNeXt-L [24]	8.99	0.461	0.458	0.459	0.460	0.377	0.460	0.459	0.320
BDD100K	CityScapes	SegFormer-B5 [49]	63.05	0.087	0.083	0.082	0.084	0.076	0.085	0.087	0.069
BDD100K	CityScapes	Segmenter-L [40]	60.26	0.062	0.058	0.059	0.057	0.055	0.067	0.060	0.048
BDD100K	CityScapes	Knet-DeepLab [51]	61.27	0.073	0.070	0.069	0.071	0.063	0.073	0.071	0.056
BDD100K	CityScapes	Knet-SWIN-L [51]	67.46	0.080	0.078	0.077	0.077	0.071	0.079	0.079	0.063
BDD100K	CityScapes	ConvNeXt-L [24]	67.18	0.081	0.079	0.079	0.077	0.070	0.080	0.079	0.065
DAVIS-train	DAVIS-test	SegFormer-B4 [49]	89.33	0.033	0.031	0.031	0.031	0.055	0.098	0.034	0.024
DAVIS-train	DAVIS-test	Segmenter-B [40]	81.35	0.080	0.077	0.076	0.075	0.103	0.121	0.074	0.049
DAVIS-train	DAVIS-test	Knet-DeepLab [51]	83.12	0.076	0.072	0.073	0.073	0.098	0.107	0.075	0.047
DAVIS-train	DAVIS-test	Knet-SWIN-B [51]	89.22	0.045	0.041	0.041	0.040	0.065	0.101	0.044	0.029
DAVIS-train	DAVIS-test	ConvNeXt-B [24]	89.00	0.044	0.040	0.041	0.042	0.067	0.100	0.044	0.031
SN-7-SP-train	SN-7-SP-test	SegFormer-B4 [49]	57.23	0.655	0.636	0.625	0.634	0.711	0.701	0.652	0.532
SN-7-SP-train	SN-7-SP-test	Segmenter-B [40]	51.98	0.766	0.751	0.748	0.740	0.796	0.789	0.764	0.687
SN-7-SP-train	SN-7-SP-test	Knet-DeepLab [51]	55.58	0.704	0.688	0.679	0.681	0.768	0.720	0.696	0.655
SN-7-SP-train	SN-7-SP-test	Knet-SWIN-B [51]	59.42	0.609	0.576	0.559	0.563	0.631	0.618	0.604	0.504
SN-7-SP-train	SN-7-SP-test	ConvNeXt-B [24]	58.69	0.611	0.570	0.564	0.566	0.629	0.633	0.611	0.517
SN-7-TS-train	SN-7-TS-test	SegFormer-B4 [49]	54.39	0.700	0.673	0.669	0.673	0.712	0.723	0.695	0.582
SN-7-TS-train	SN-7-TS-test	Segmenter-B [40]	50.13	0.769	0.745	0.740	0.736	0.785	0.777	0.758	0.624
SN-7-TS-train	SN-7-TS-test	Knet-DeepLab [51]	56.73	0.651	0.622	0.621	0.619	0.689	0.675	0.644	0.576
SN-7-TS-train	SN-7-TS-test	Knet-SWIN-B [51]	62.42	0.619	0.587	0.579	0.584	0.630	0.621	0.609	0.551
SN-7-TS-train	SN-7-TS-test	ConvNeXt-B [24]	62.14	0.623	0.578	0.571	0.573	0.647	0.638	0.620	0.549
BraTS-train	BraTS-test	SegFormer-B4 [49]	45.81	0.155	0.137	0.135	0.135	0.129	0.150	0.134	0.120
BraTS-train	BraTS-test	Segmenter-B [40]	44.57	0.154	0.149	0.146	0.147	0.135	0.142	0.150	0.124
BraTS-train	BraTS-test	Knet-DeepLab [51]	46.05	0.188	0.179	0.176	0.176	0.158	0.187	0.184	0.148
BraTS-train	BraTS-test	Knet-SWIN-B [51]	47.99	0.155	0.146	0.143	0.144	0.132	0.157	0.152	0.123
BraTS-train	BraTS-test	ConvNeXt-B [24]	48.40	0.169	0.159	0.154	0.155	0.136	0.168	0.165	0.121
SYNTHIA	CityScapes	SegFormer-B5 [49]	33.12	0.286	0.264	0.262	0.263	0.230	0.288	0.284	0.218
SYNTHIA	CityScapes	Segmenter-L [40]	32.33	0.254	0.235	0.232	0.230	0.217	0.280	0.252	0.203
SYNTHIA	CityScapes	Knet-DeepLab [51]	30.19	0.311	0.298	0.296	0.296	0.272	0.305	0.309	0.241
SYNTHIA	CityScapes	Knet-SWIN-L [51]	34.31	0.281	0.273	0.270	0.270	0.239	0.288	0.270	0.226
SYNTHIA	CityScapes	ConvNeXt-L [24]	34.79	0.287	0.267	0.263	0.264	0.230	0.299	0.281	0.211

- *Ablation Study*

- Selector Accuracy



- *Ablation Study*
 - Pixel Location





On Calibrating Semantic Segmentation Models: Analyses and An Algorithm

Dongdong Wang¹, Boqing Gong², and Liqiang Wang¹

¹University of Central Florida ²Google

THU-PM-291