Predictive Inference with Feature Conformal Prediction

Jiaye Teng[1,3,4,*], Chuan Wen[1,3,4,*], Dinghuai Zhang[2,*], Yoshua Bengio[2], Yang Gao[1,3,4], Yang Yuan[1,3,4,†]

[1]IIIS, Tsinghua University [2]Mila - Quebec AI Institute [3]Shanghai Artificial Intelligence Laboratory [4]Shanghai Qizhi Institute
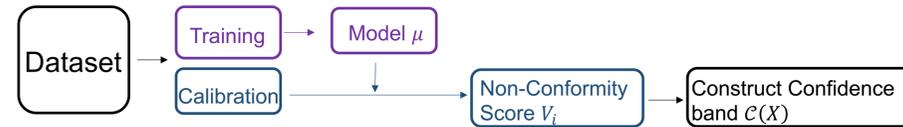
Conformal Prediction

Predictive Inference: Construct a confidence band $\mathcal{C}(X')$ to cover the true value Y' with high probability.

$$\mathbb{P}\big(Y' \in \mathcal{C}_{1-\alpha}(X')\big) \geq 1 - \alpha.$$

Conformal Prediction: A post-hoc and distribution-free method to construct the confidence band.

- Key intuition: use a holdout calibration fold to adjust the band.
- Just like what validation set do in generalization.
- Theoretical guarantee with some mild assumptions

Limitations in Conformal Prediction:

- Training independent: results hold for any model μ
- → FCP exploits semantic information of deep representation
- Non-personalized: band has same length $|\mathcal{C}(X)|$ individually
- → FCP returns different band length for each sample
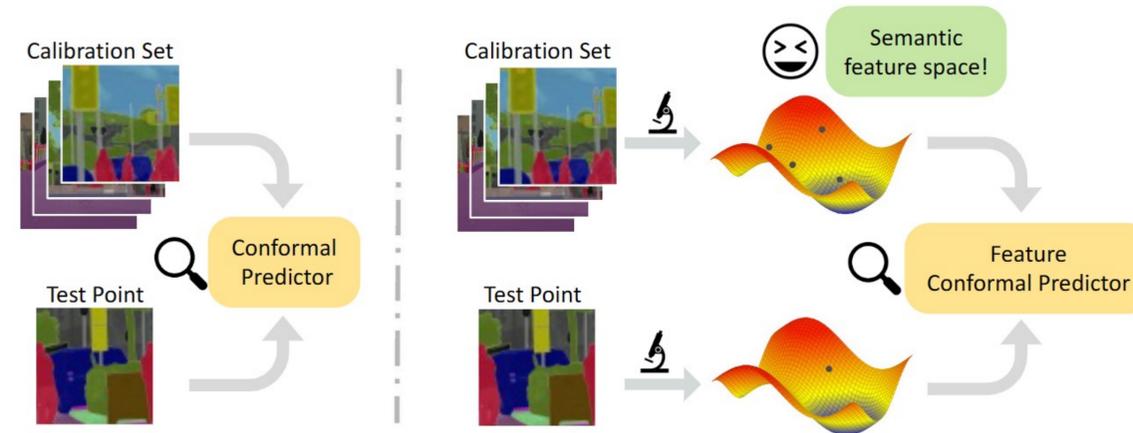
Feature Conformal Prediction
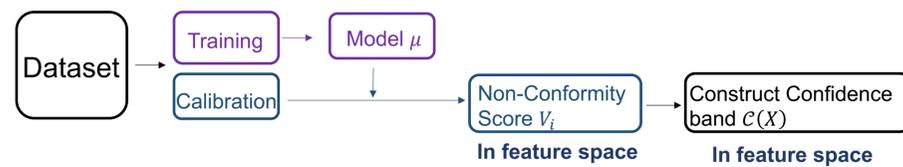
- Requires a new non-conformity score in feature space

Intuition: find the nearest feature that corrects the prediction

- Requires to transfer the band to the output space

Approach 1: detect whether a point falls in the band in the feature space

Approach 2: estimate the band in output space and then calculate its length

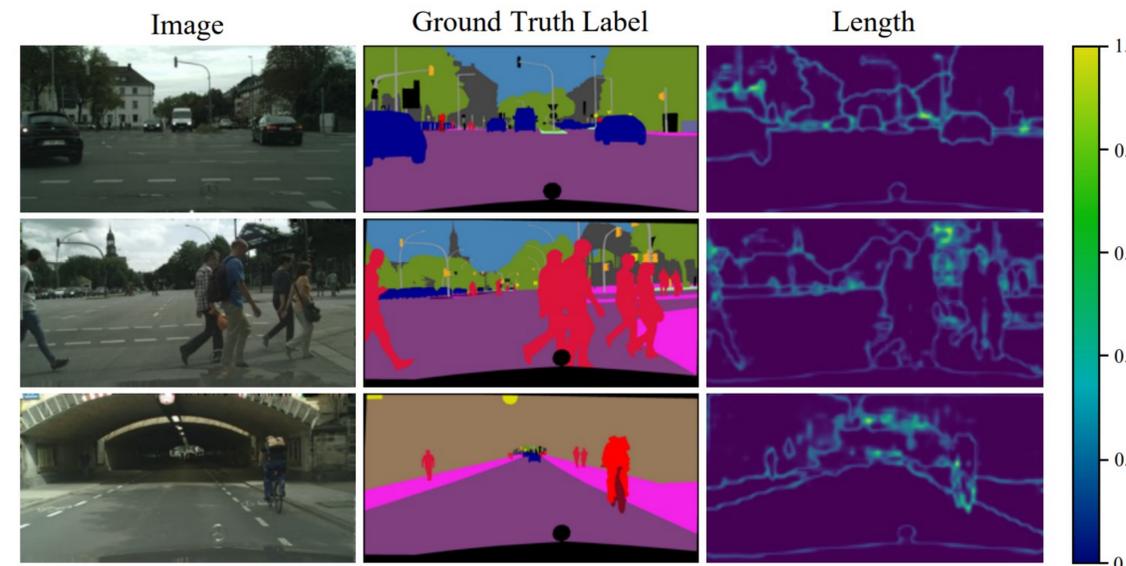## From Output Space to Feature Space

Conformal Prediction   **Feature Conformal Prediction**

Deploy Conformal Prediction In Feature Space
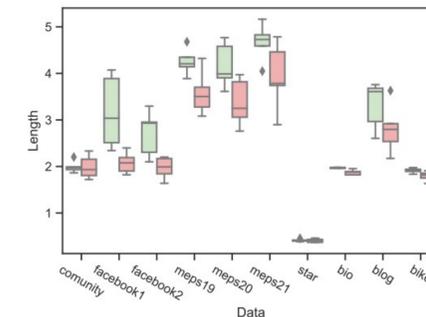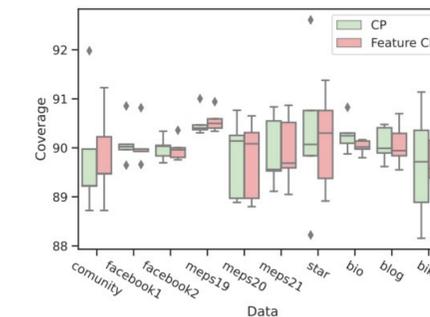
FCP in image segmentation

- Brighter region means uncertain (larger band)
- FCP puts more uncertainty on the boundaries for each image
- Vanilla CP put the uncertainty equally for each image
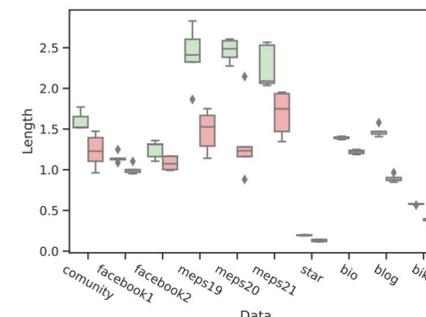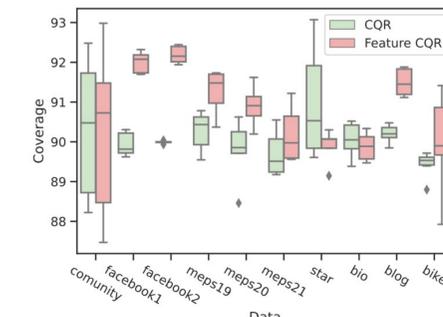

Experiment: further validation

FCP v.s. CP

- FCP can attain $1 - \alpha$ empirical coverage
- FCP has shorter band length

FCP for other CP variants

- FCP can be deployed into other variants of CP (e.g., CQR)
- FCP has better band length

FCP exploits training information

- FCP cannot outperform CP if the network is untrained
- Semantic information of deep representation is important

METHOD	VANILLA CP		FEATURE CP	
DATASET	COVERAGE	LENGTH	COVERAGE	LENGTH
COMMUNITY	90.28 ±1.70	**4.85**±0.22	90.68±1.33	4.92±0.77
FACEBOOK1	90.15 ±0.15	3.42 ±0.25	90.16 ±0.12	**3.20** ±0.50
FACEBOOK2	90.17 ±0.11	3.51 ±0.26	90.12 ±0.14	**3.34** ±0.39
MEPS19	90.81 ±0.46	**4.02** ±0.16	90.86 ±0.30	4.22 ±0.48
MEPS20	90.10 ±0.60	4.10 ±0.28	90.28 ±0.46	**4.02** ±0.41
MEPS21	89.78 ±0.44	4.08 ±0.16	89.85 ±0.58	**3.81** ±0.32
STAR	90.07 ±0.77	**2.23** ±0.18	89.47 ±1.84	2.24 ±0.40
BIO	90.06 ±0.10	**4.25** ±0.11	90.11 ±0.07	4.44 ±0.74
BLOG	90.13 ±0.34	**2.41** ±0.15	90.16 ±0.26	2.58 ±0.49
BIKE	89.53 ±0.78	4.65 ±0.15	89.61 ±0.86	**4.13** ±0.38