

Asymmetric Duos: Sidekicks Improve Uncertainty

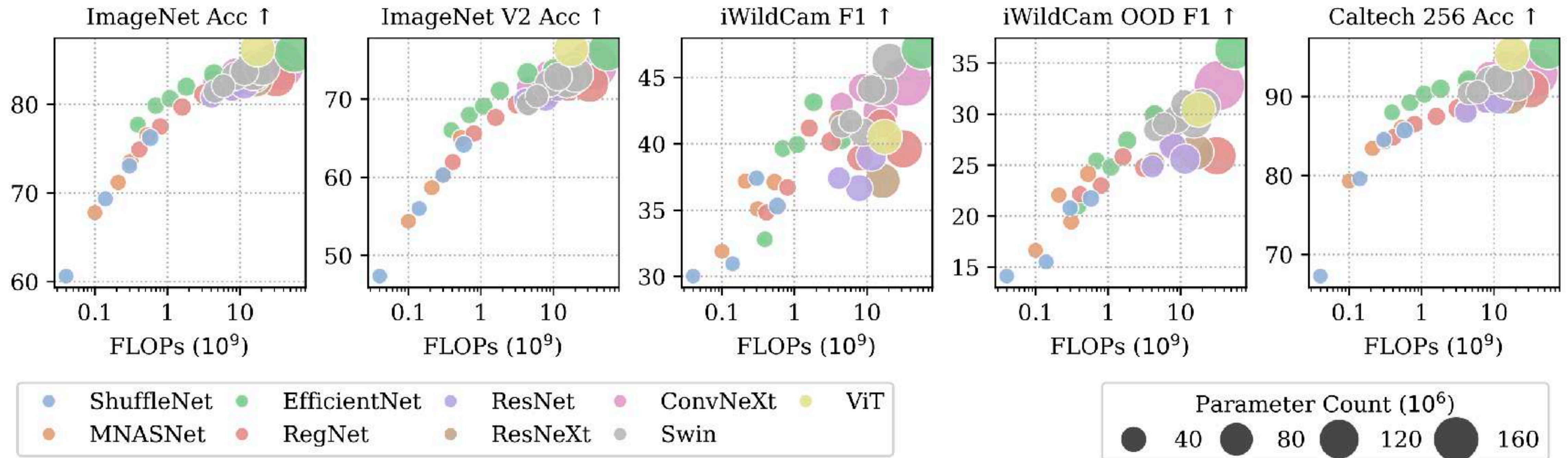
Pairing Models of different capabilities

by Tim Zhou, Evan Shelhamer, Geoff Pleiss

Published as a Spotlight at NeurIPS 2025

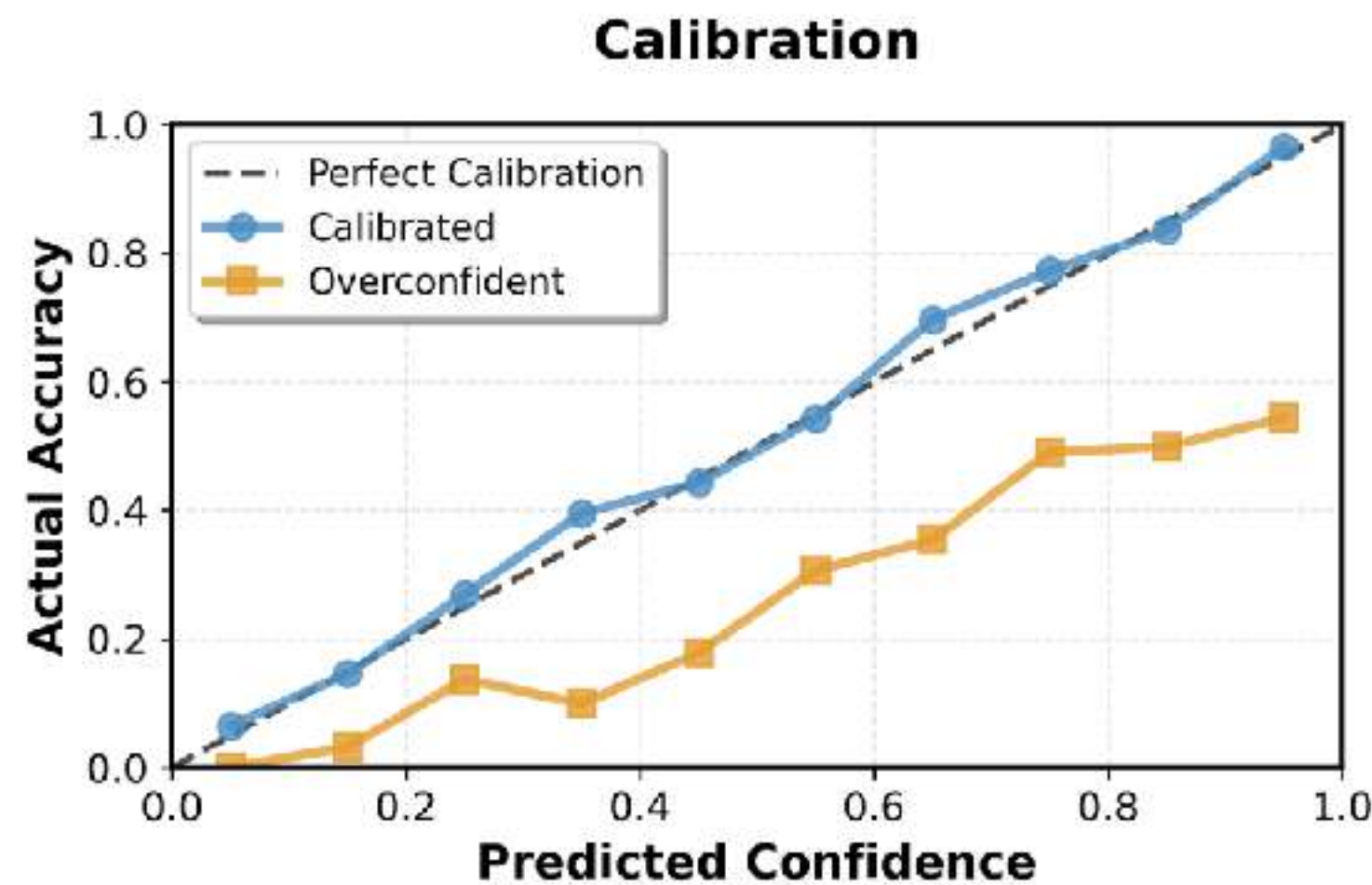


Bigger Models are More Accurate Models



In modern deep learning, the bigger the network the better the predictions, but bigger models require more computation to train and to deploy.

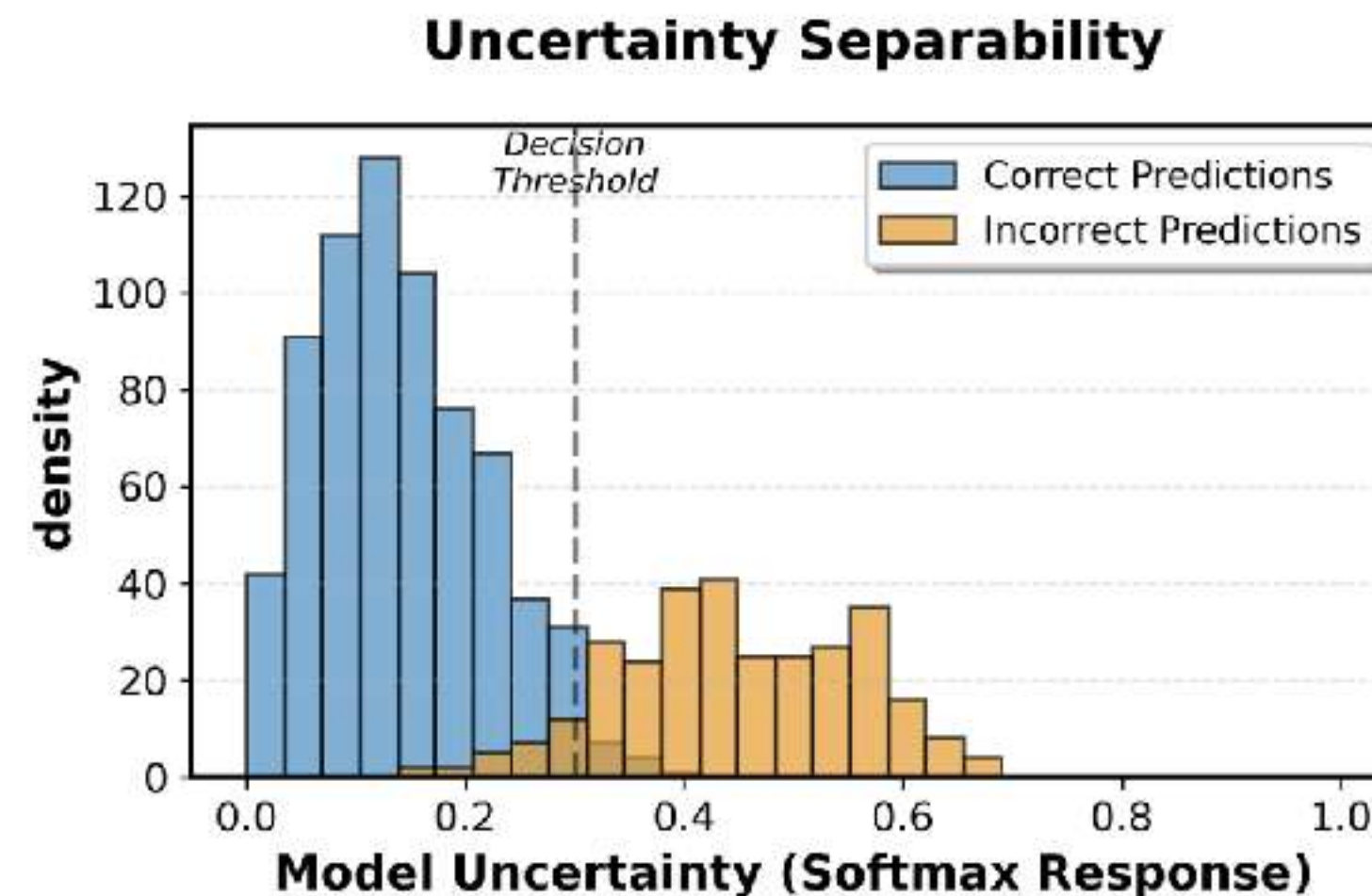
Measuring Uncertainty in Deep Learning



Calibration:

Does predicted uncertainty match actual accuracy?

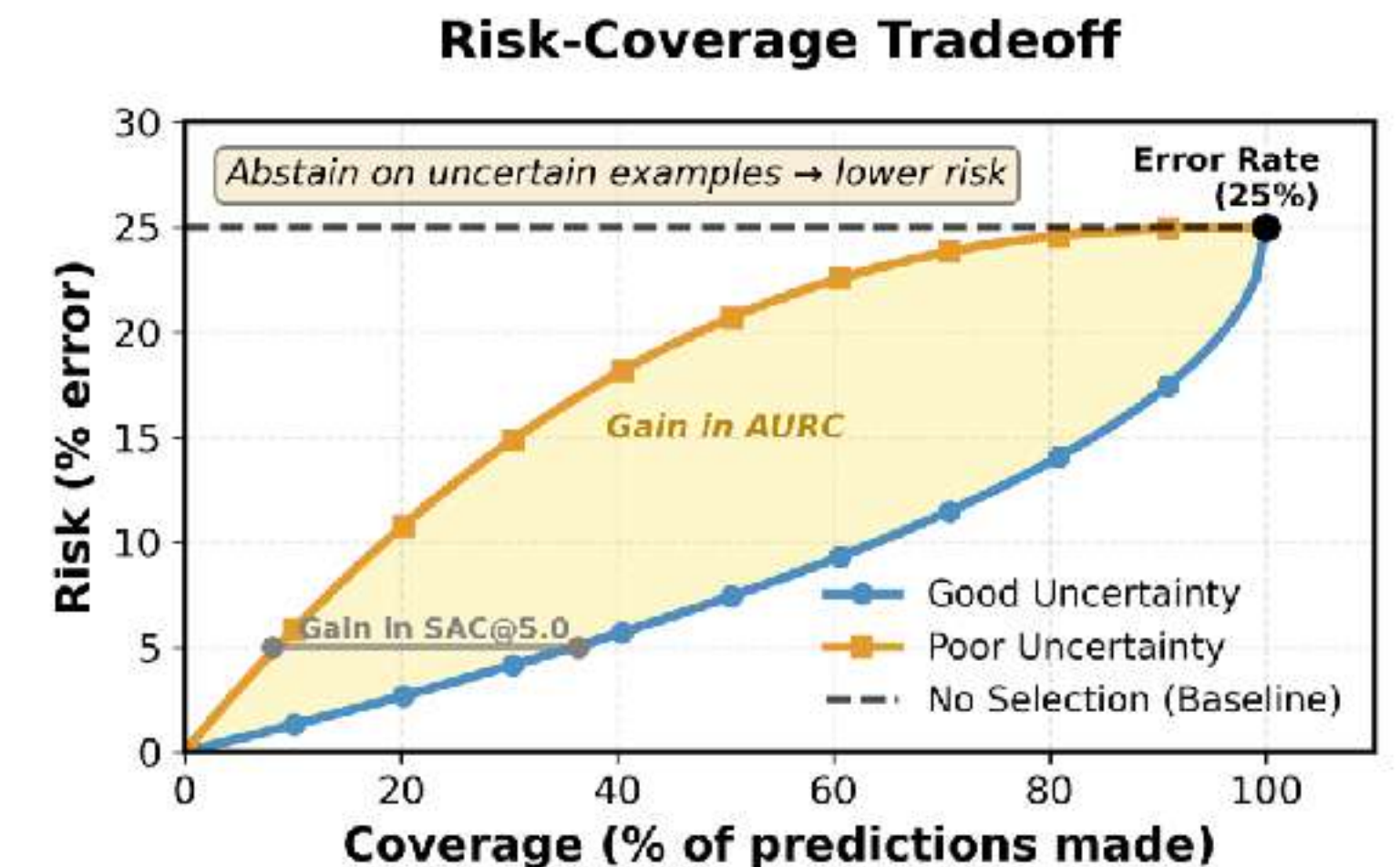
When a model is 80% confident, it should be correct ~80% of the time.



Separability:

Can uncertainty distinguish correct from incorrect predictions?

Mistakes should have higher uncertainty than correct predictions.



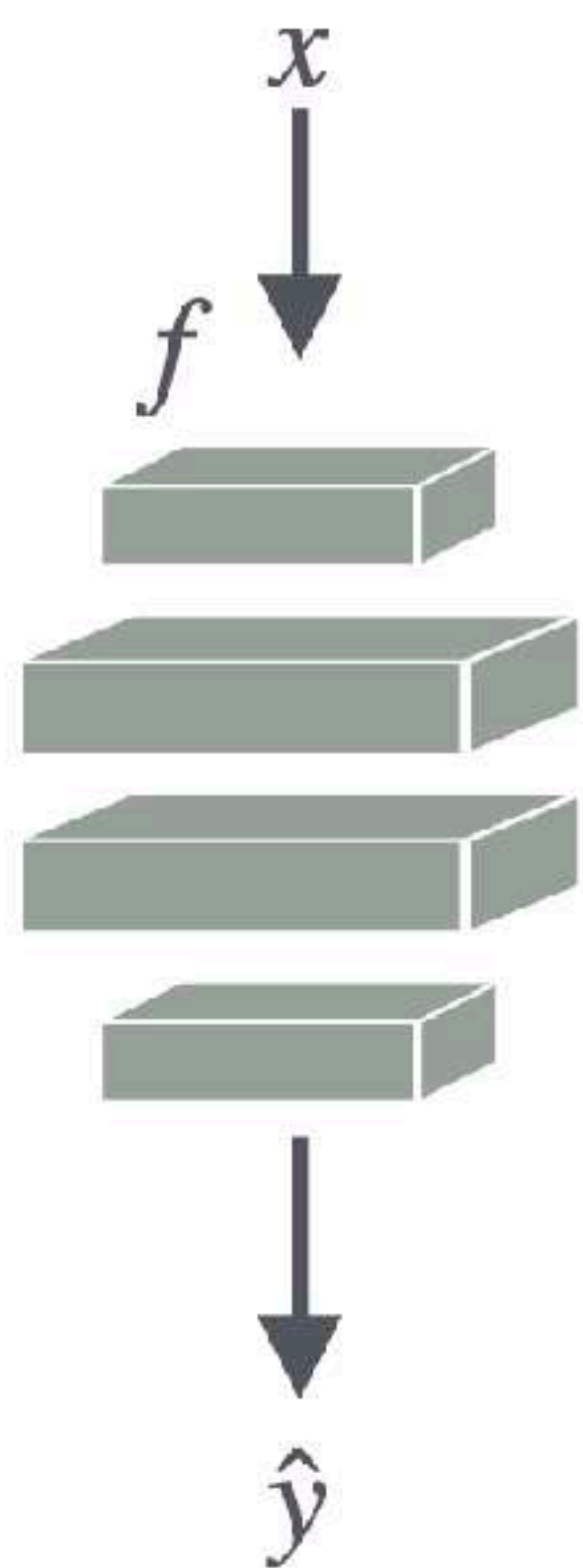
Synergy with Predictive Power:

Can we predict uncertainty without sacrificing predictive power?

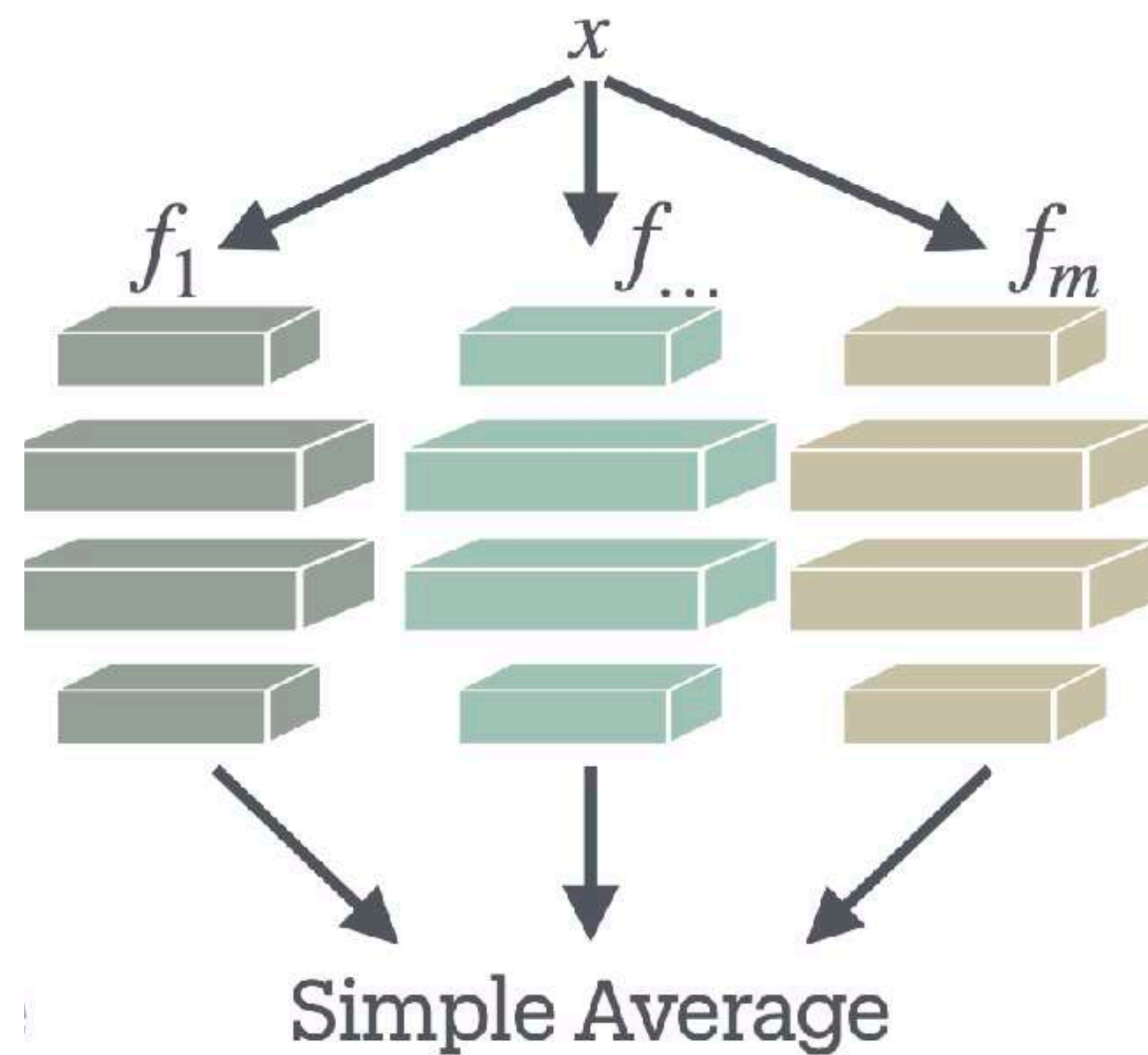
Separability is most useful when the model is already accurate.

Deep Ensembles work, but are costly

Single Model
1x compute

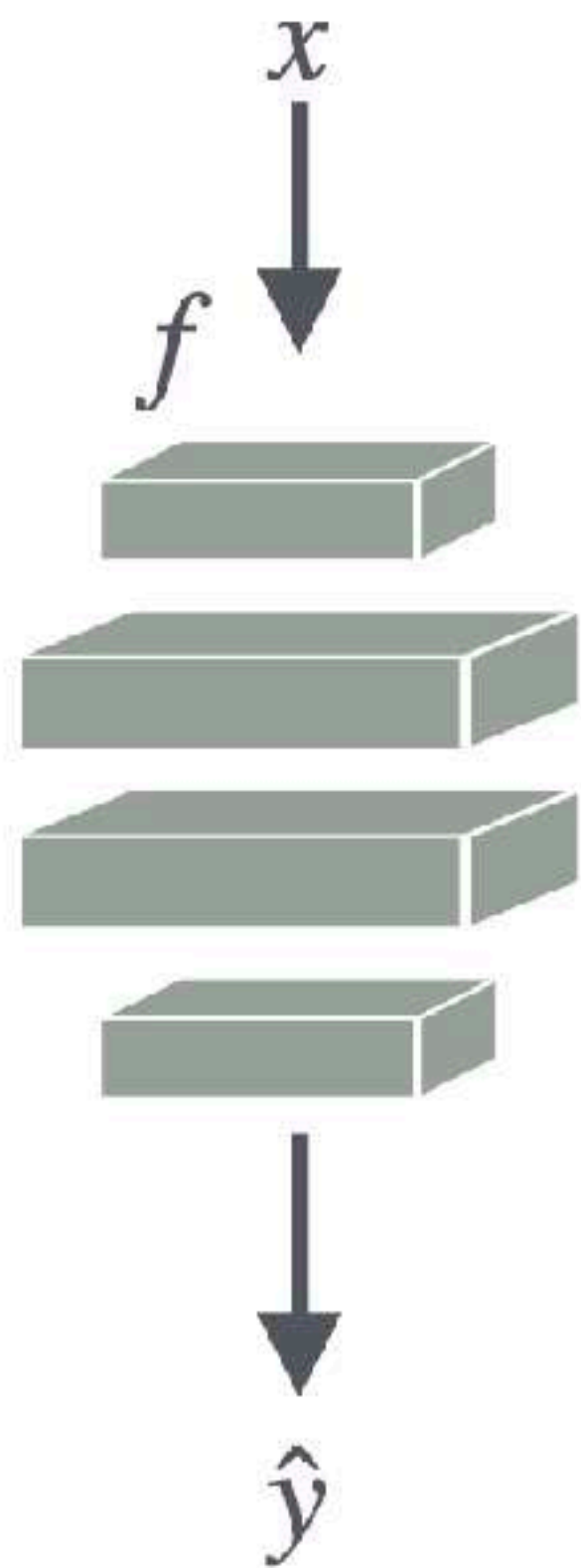


Deep Ensemble
 m x compute

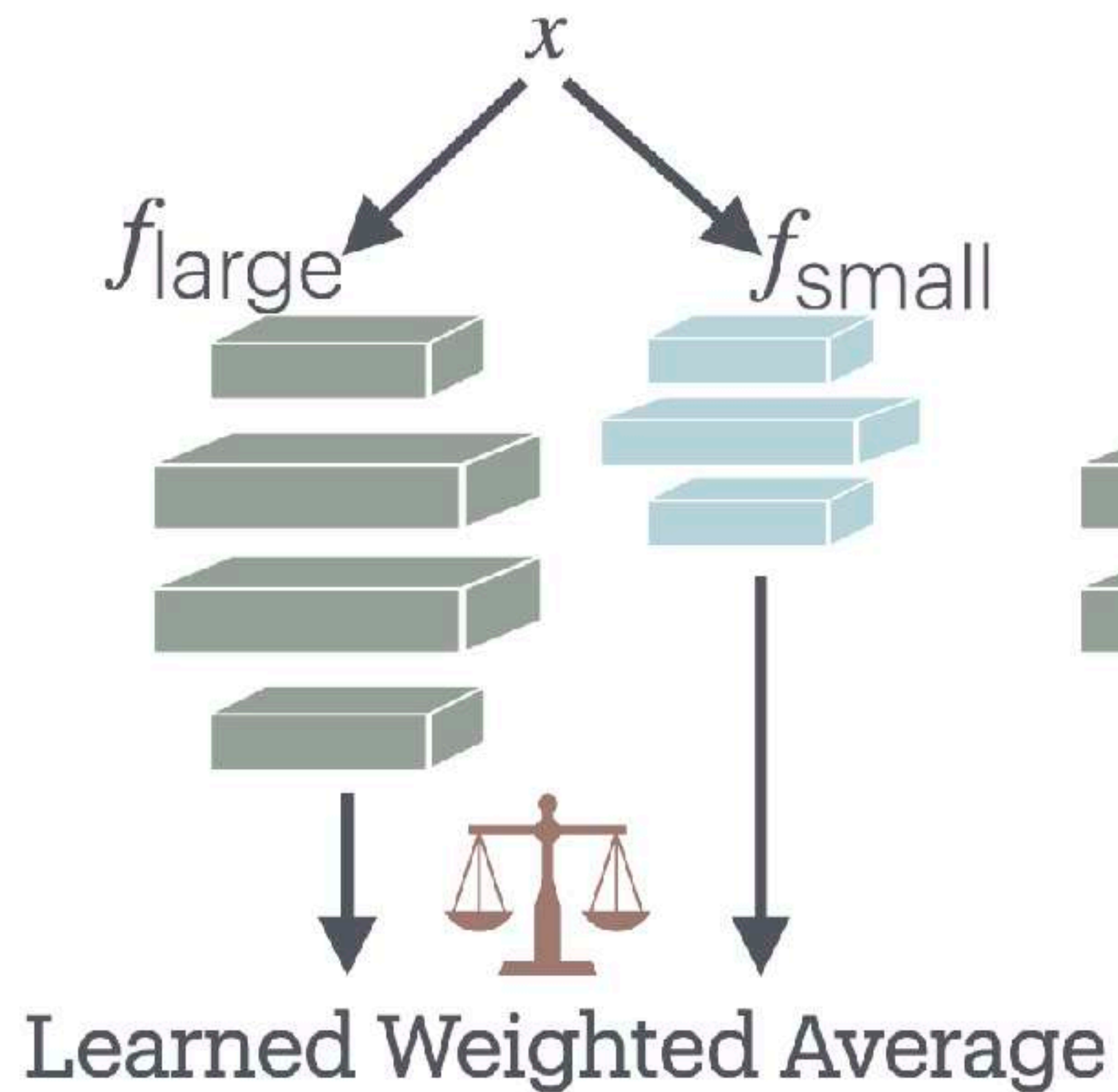


Asymmetric Duos pair a large model with a small model

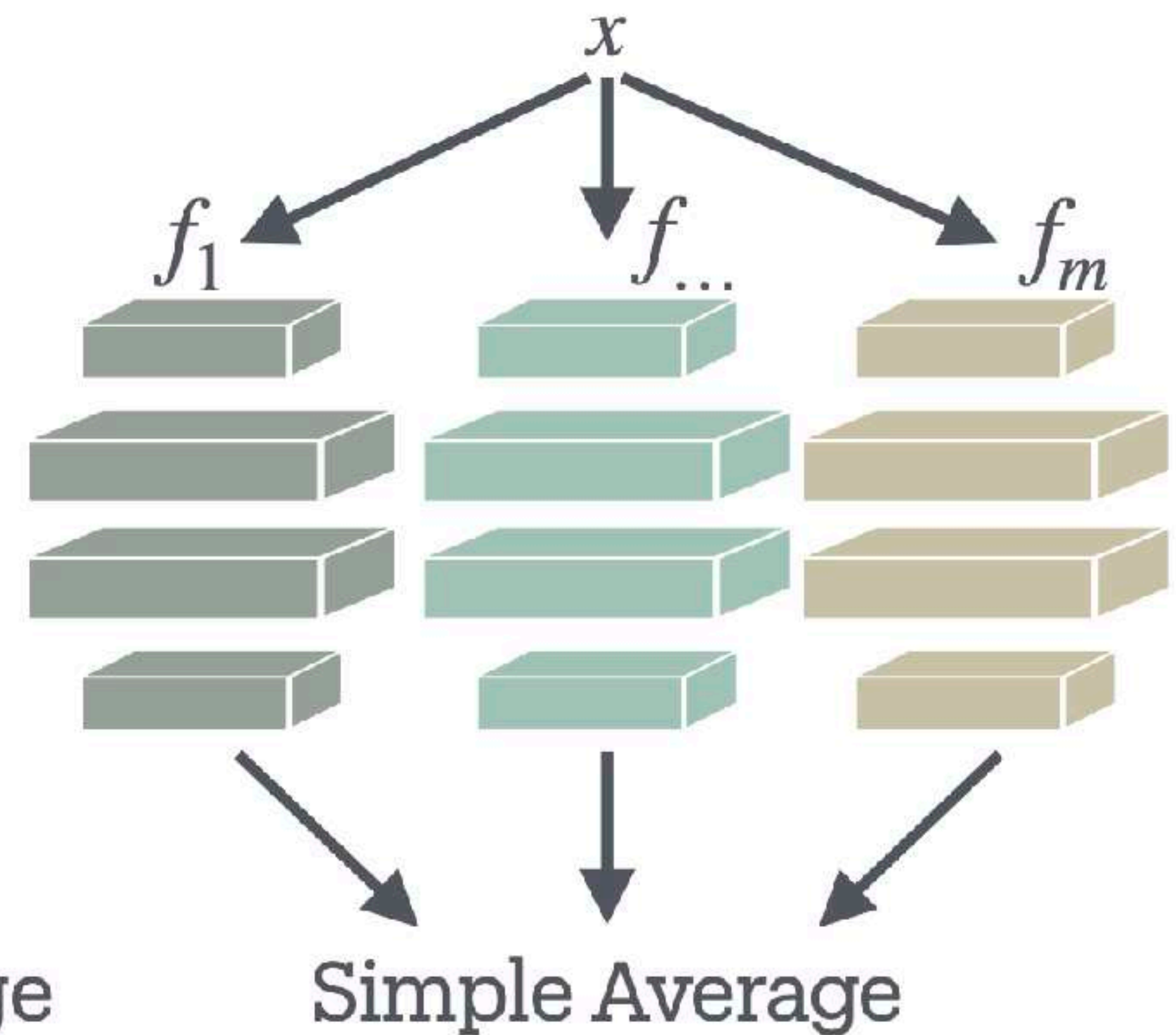
Single Model
1x compute



Asymmetric Duo (**Ours**)
 $(1 + \epsilon)$ x compute



Deep Ensemble
 m x compute

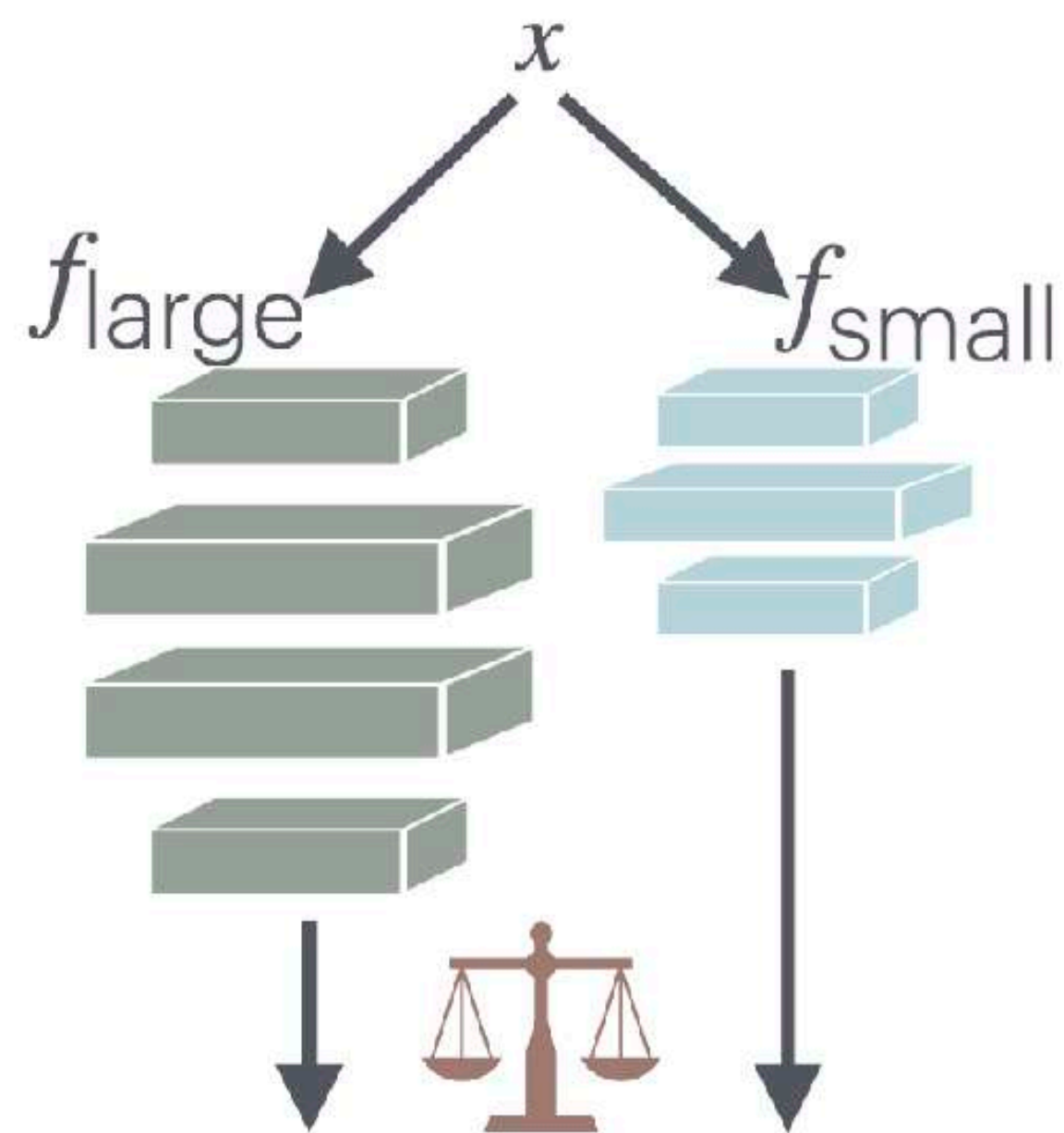


f_{large} and f_{small} are fine-tuned independently, no joint training needed.

Aggregating Asymmetric Duos

Asymmetric Duo (Ours)

$(1 + \epsilon)$ x compute



Learned Weighted Average

$$f_{Duo}(X) = \overbrace{f_{large}(X)}^{\text{logits}} \cdot \overbrace{T_{large}}^{\text{temp.}} + \overbrace{f_{small}(X)}^{\text{logits}} \cdot \overbrace{T_{small}}^{\text{temp.}}$$

$$\hat{Y}_{Duo}(X) = \arg \max_i [f_{Duo}(X)]_i \quad \text{unc}(f_{Duo}(X)) = 1 - \left[\sigma(f_{Duo}(X)) \right]_{\hat{Y}_{Duo}(X)}$$

T_{small}, T_{large} are tuned on the val. set used for hyperparameter tuning.

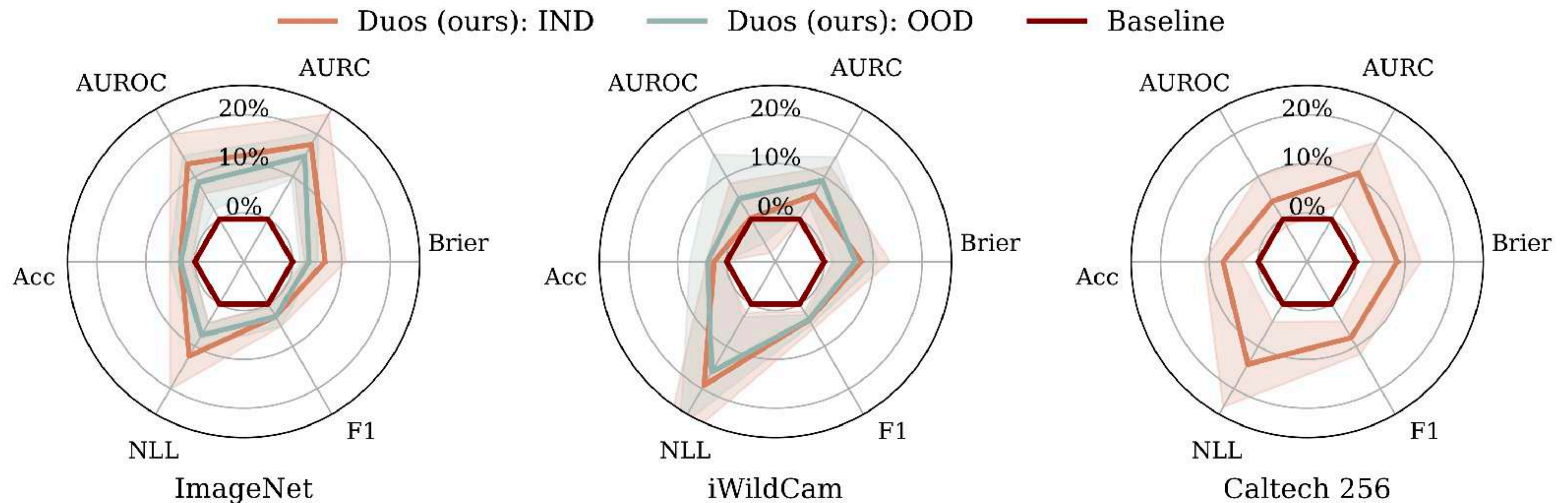
$T_{small} \approx 0 \implies$ small model doesn't help; revert to large model preds.

$T_{small} > 0 \implies$ small model improves preds. of large model

Do a lot with a little

Asymmetric Duos only need ~10–20% additional FLOPs to improve on f_{large} across metrics for

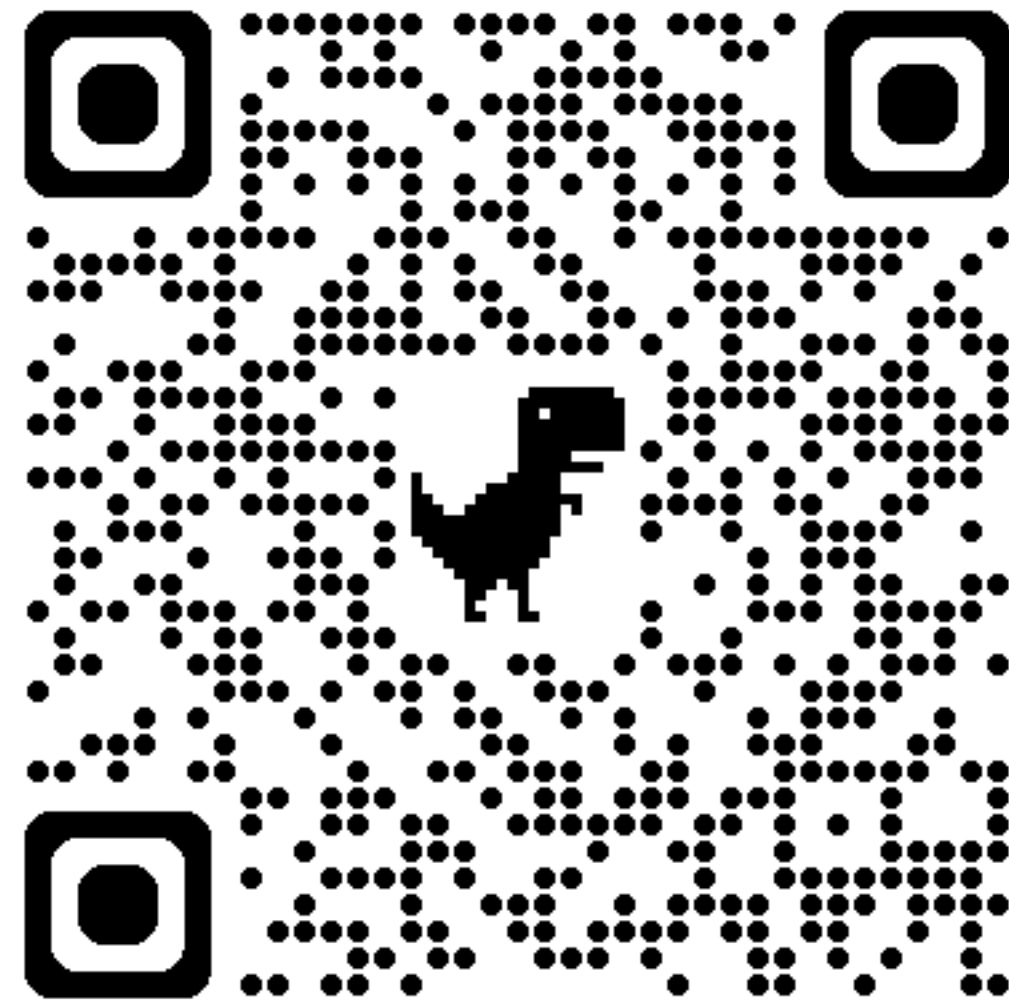
- predictive power (Acc, F1),
- uncertainty quantification (ECE, NLL, Brier),
- uncertainty separability (AUROC)
- selective classification (AURC, SAC).



Asymmetric Duos: Sidekicks Improve Uncertainty

read our paper!

by Tim Zhou, Evan Shelhamer, Geoff Pleiss



Published as a Spotlight at NeurIPS 2025

