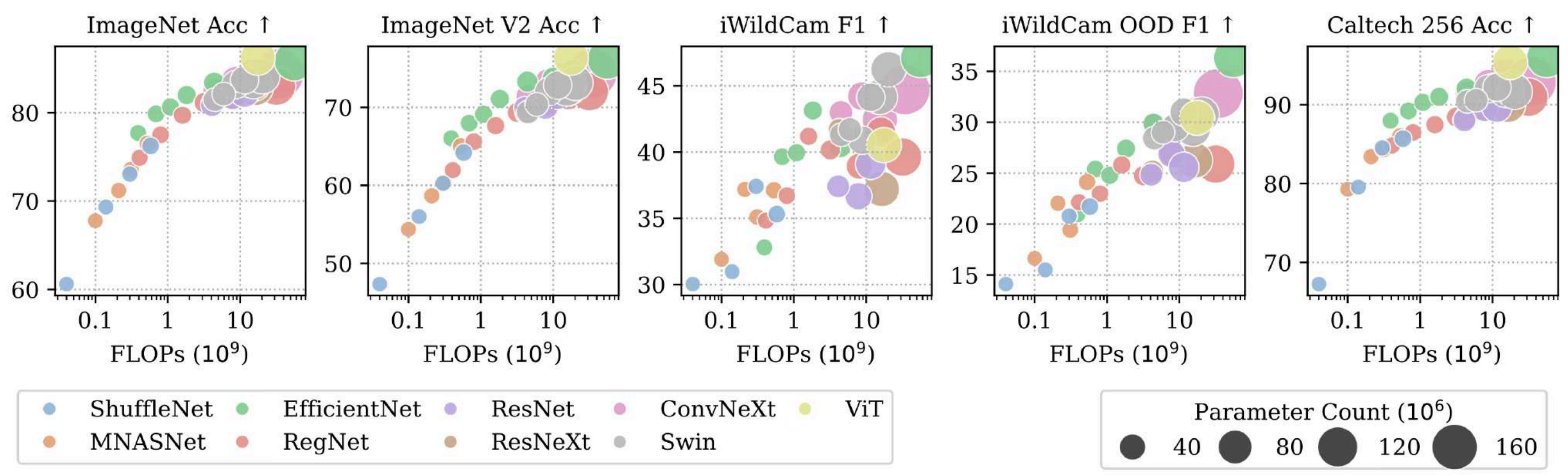
Asymmetric Duos: Sidekicks Improve Uncertainty

Tim G. Zhou, Evan Shelhamer, Geoff Pleiss

Bigger models are more accurate but costlier, can we measure uncertainty with little computation?

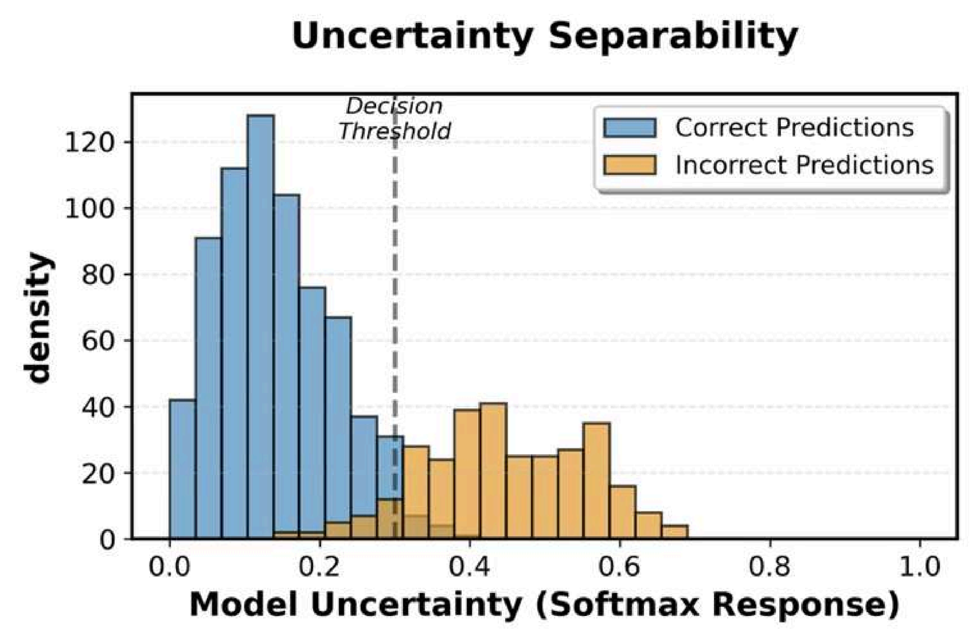


Measuring Uncertainty for Deep Learning

In high-stakes applications, knowing when a model is uncertain is as important as the prediction itself. Meaningful uncertainty predictions unlock smarter downstream decision making.

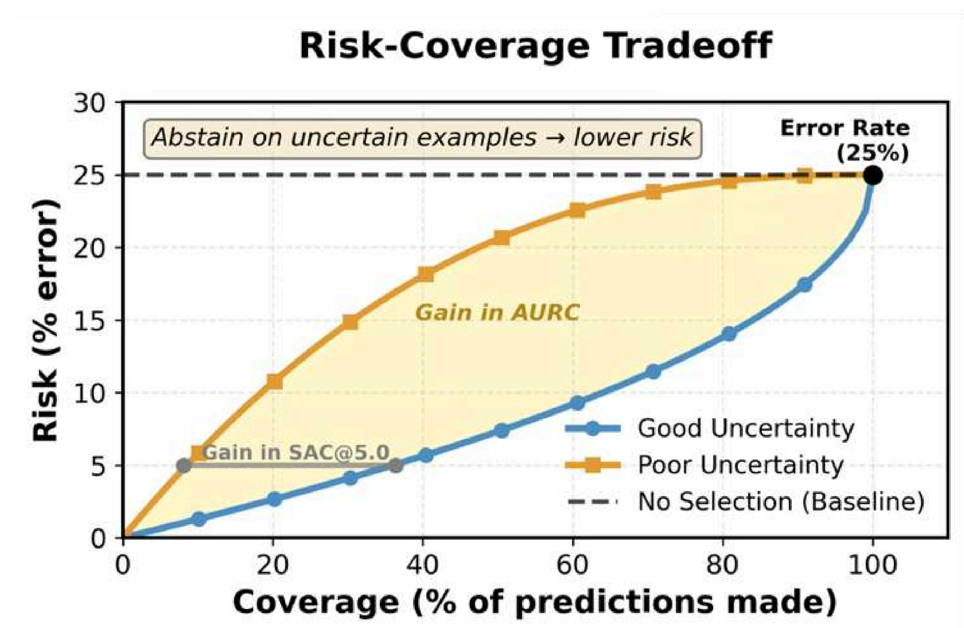
Separability

Mistakes should have higher uncertainty than correct predictions.



Synergy with Predictive Power

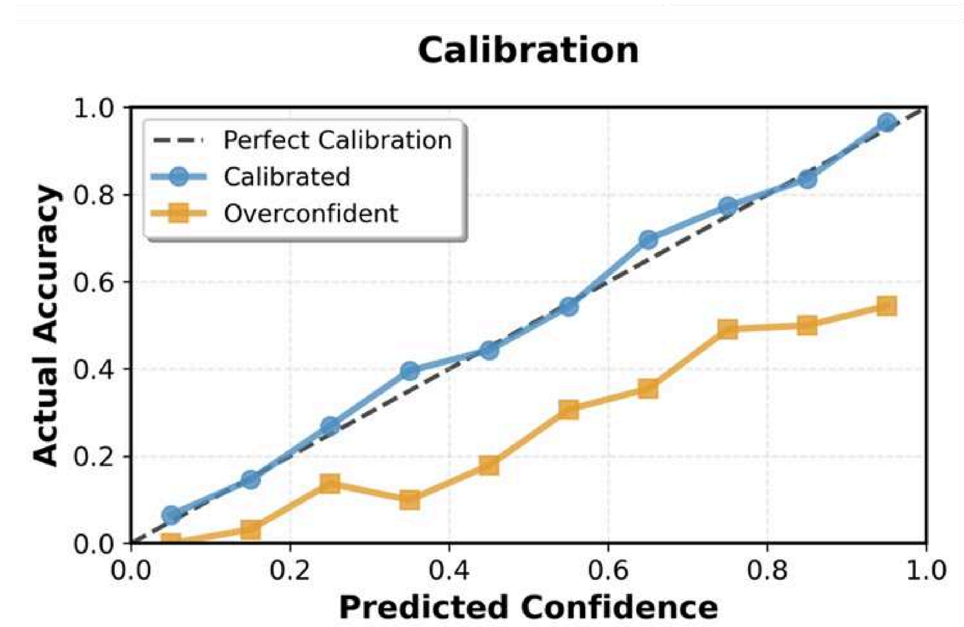
Separability is most useful when the model is already accurate. This synergy enables selective classification—abstaining on uncertain examples while maintaining high accuracy on the rest.



Calibration

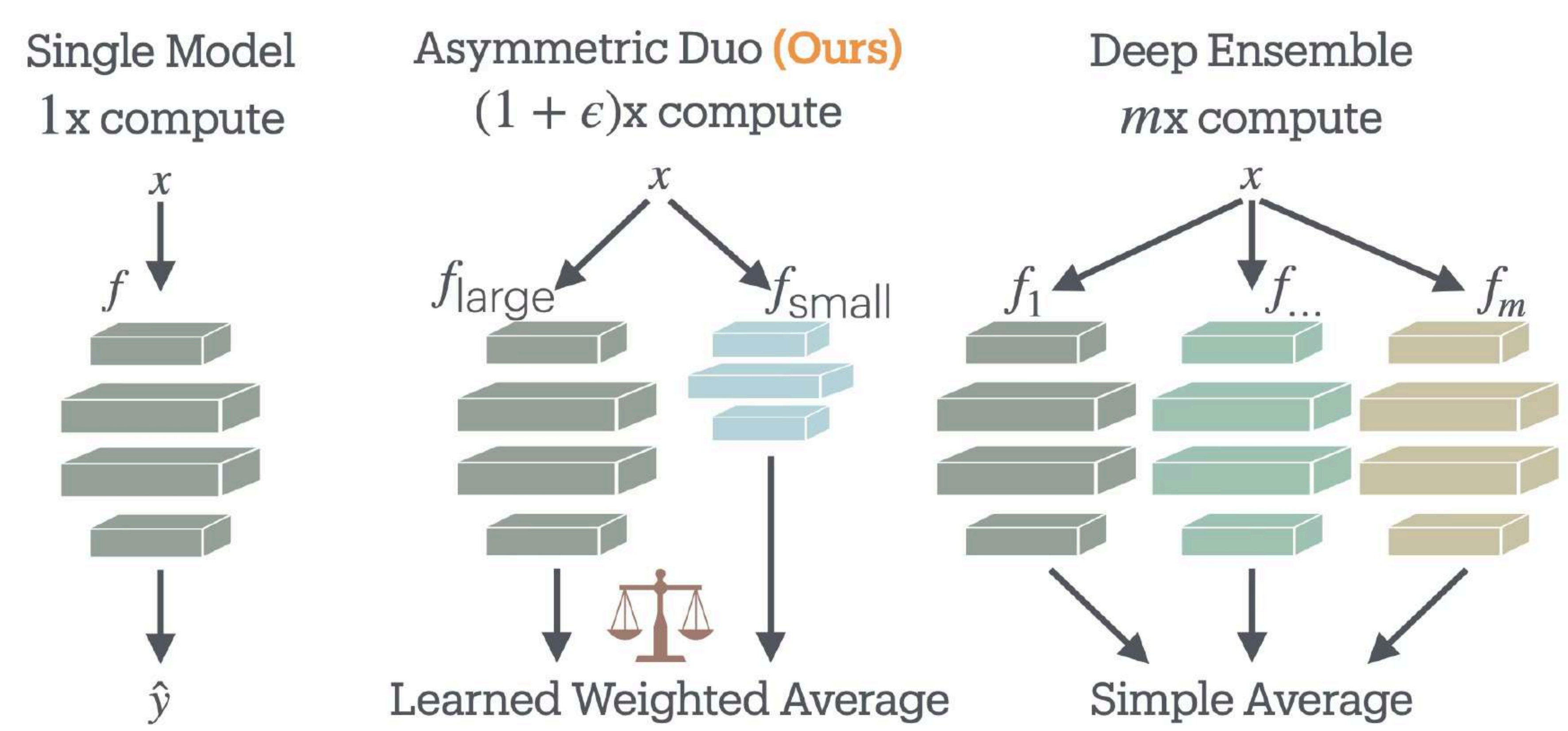
Does predicted uncertainty match actual accuracy?

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



Overview of Asymmetric Duos

Asymmetric Duos improve the uncertainty of predictions by pairing a large model with a smaller but less accurate "sidekick".



Large and small models are fine-tuned independently, no joint training.

Aggregating Predictions for Duos

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

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

Temperatures are tuned on the val. set used for hyperparameter tuning.
Tsmall=0 → small model doesn't help; revert to large model preds.
Tsmall>0 → small model improves predictions of large model

Measuring Computational Cost of Asymmetric Duos

We measure asymmetry as the relative increase in forward FLOPs per pass over using f_{large} alone.

$$\text{FLOPs Balance} = \%\,\text{Computation Increase} = \frac{\text{FLOPs}(f_{small})}{\text{FLOPs}(f_{large})}$$

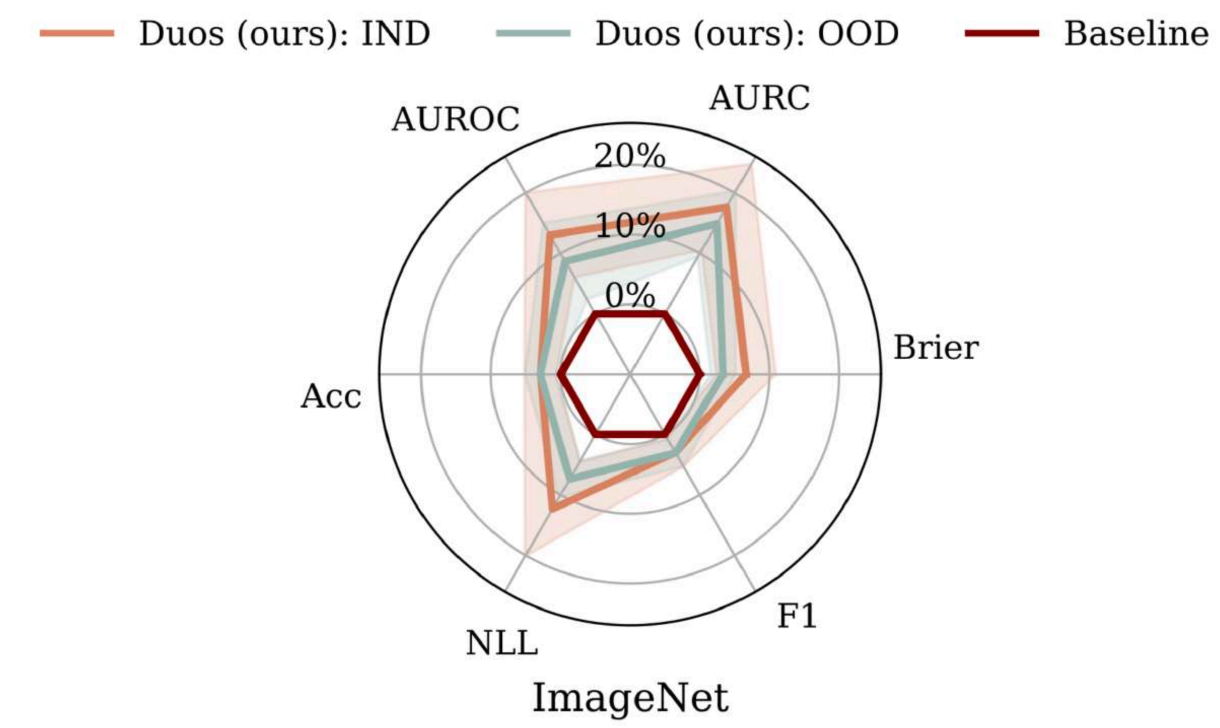
Duo Variants:

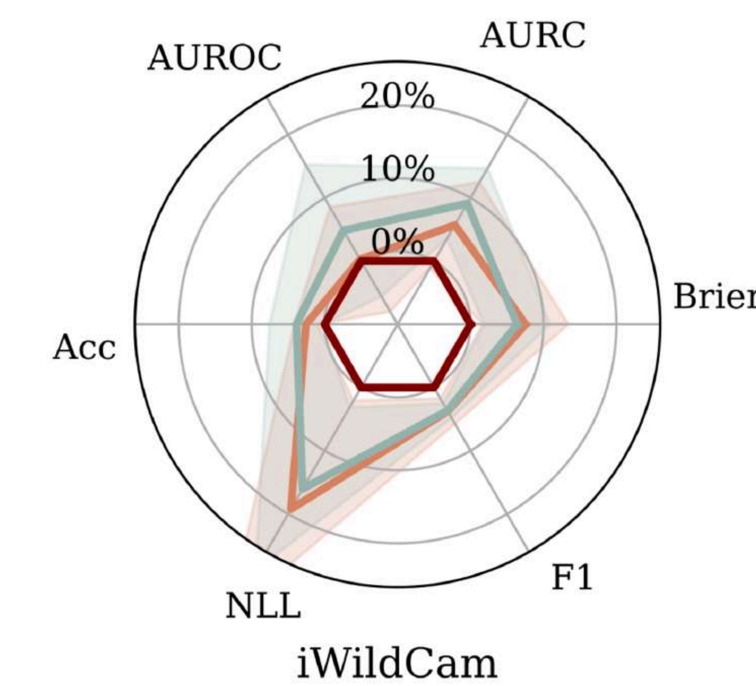
Unweighted Duos Equally weight the predictions of both models, disregarding their asymmetry, i.e. $T_{small} = T_{large} = 0.5$

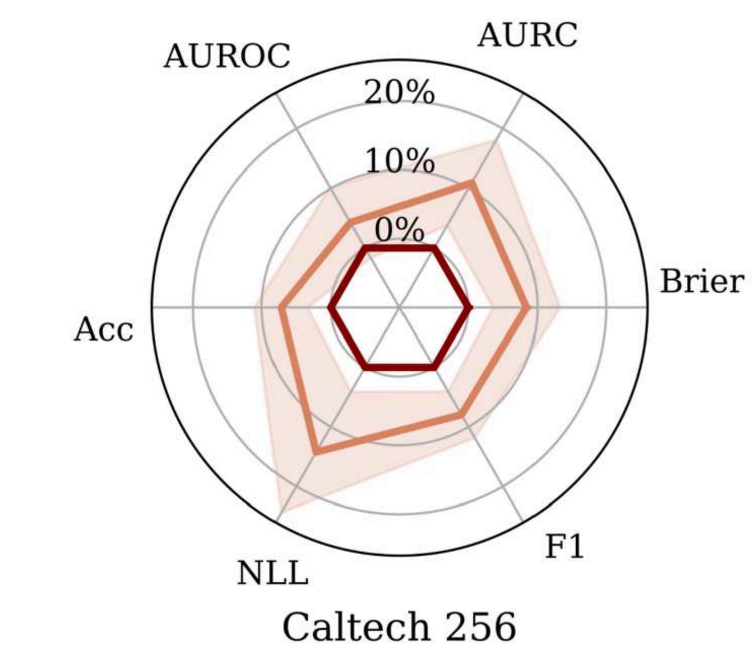
UQ-Only Duos relies on f_{large} for prediction and employs the learned temperature-weighted uncertainty quantification (UQ) only.

Fractional Additional Computation Improves Predictions

Asymmetric Duos only need ~10–20% additional FLOPs to improve on the base model across metrics for predictive power (Acc, F1), uncertainty quantification (ECE, NLL, Brier), uncertainty separability (AUROC), and selective classification (AURC, SAC).







All Duos evaluated here have FLOPs balance between 10% to 20%
Plotted values are proportional improvements w.r.t metric upper bound