Asymmetric Duos: Sidekicks Improve Uncertainty

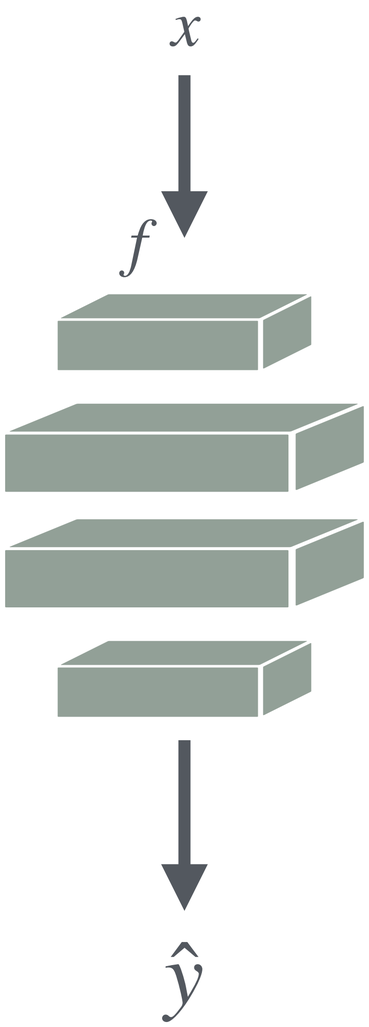
Tim G. Zhou, Evan Shelhamer, Geoff Pleiss

University of British Columbia, Vector Institute

NeurIPS 2025

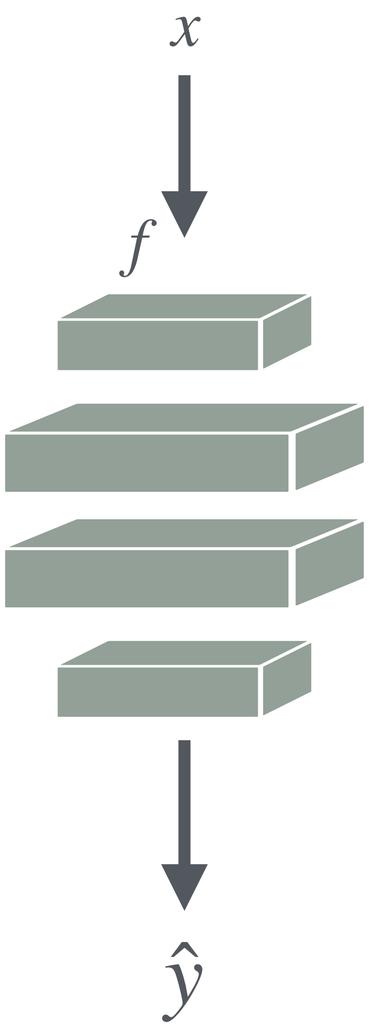
VECTOR INSTITUTE

Single Model



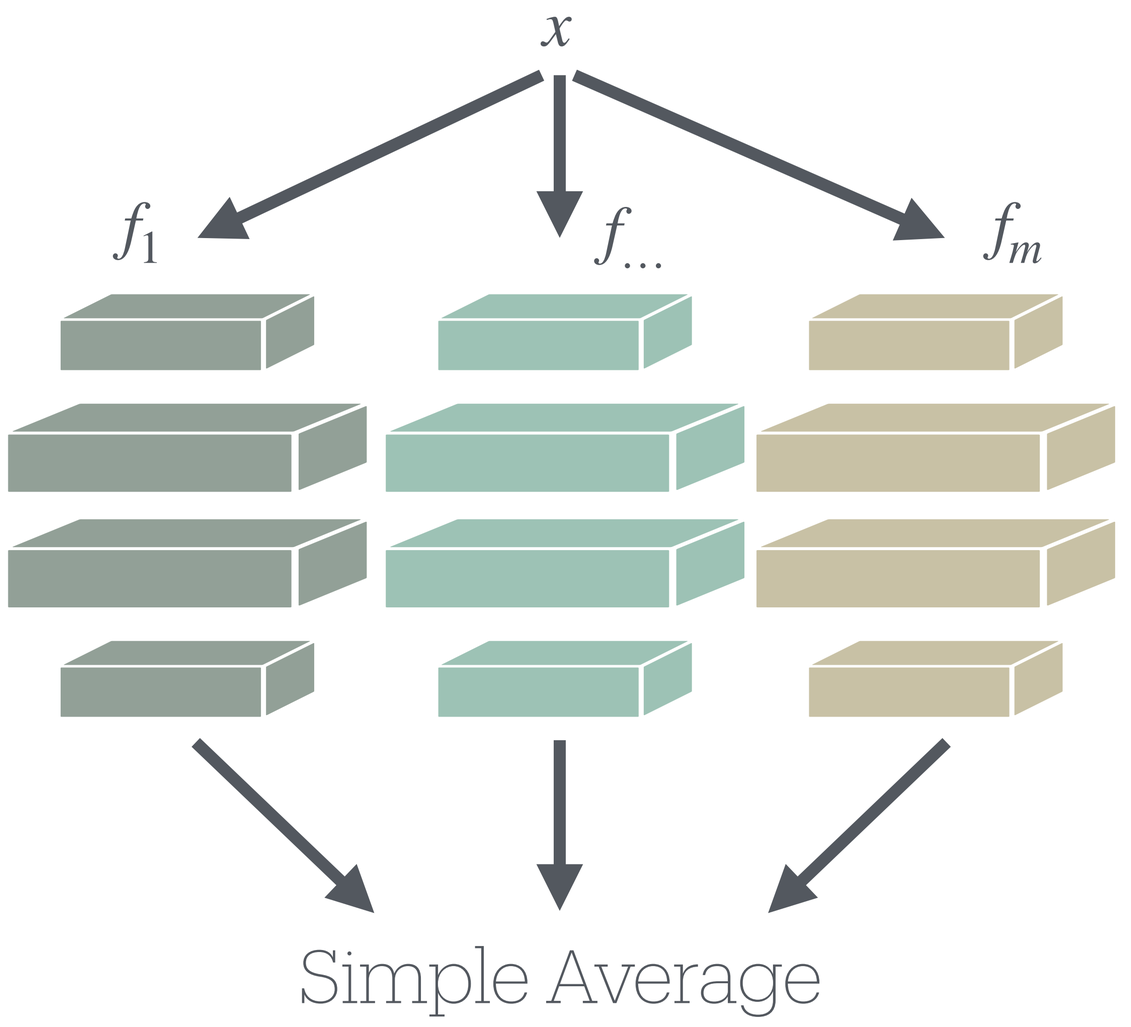
Single Model

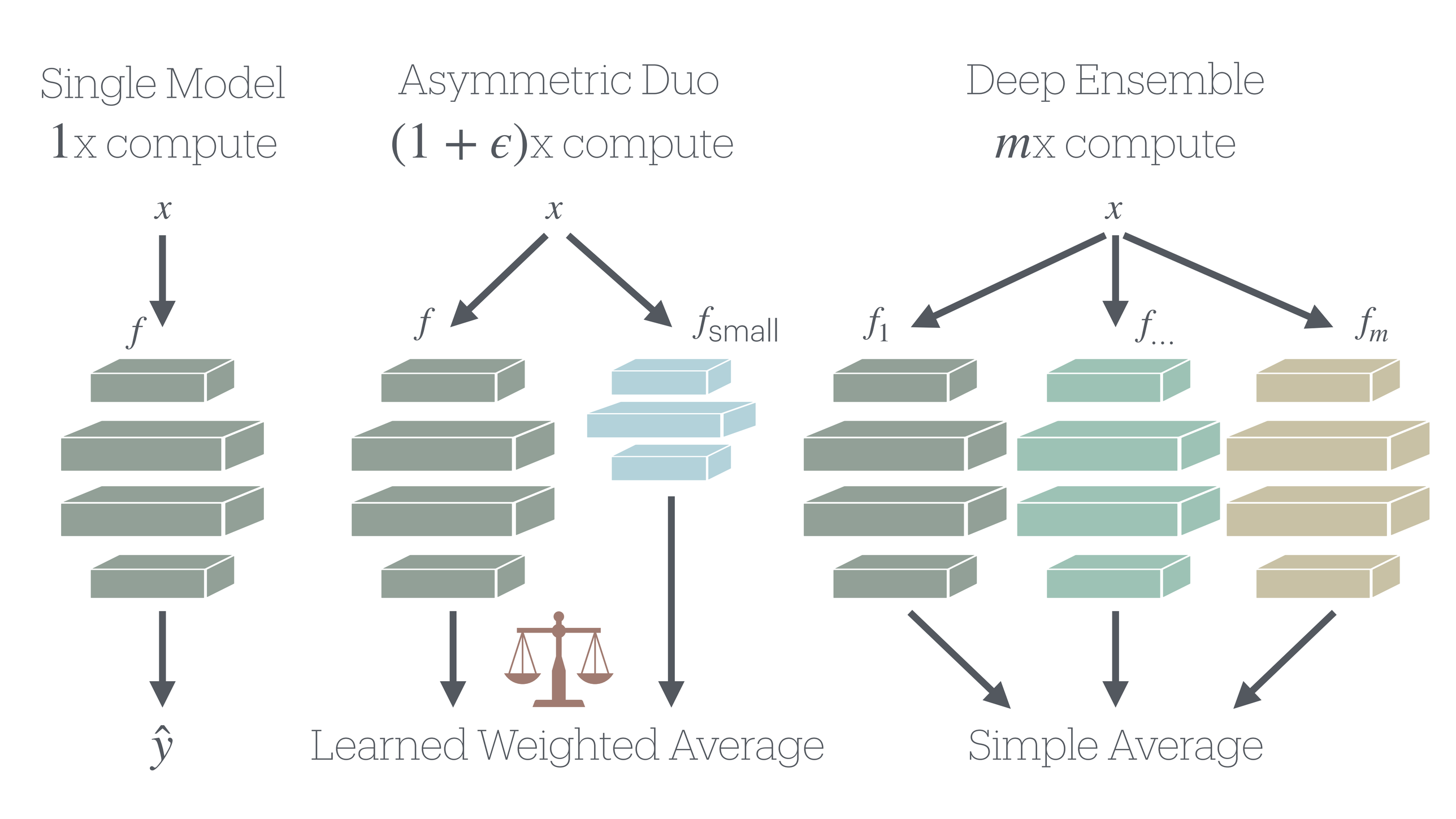
1x compute



Deep Ensemble

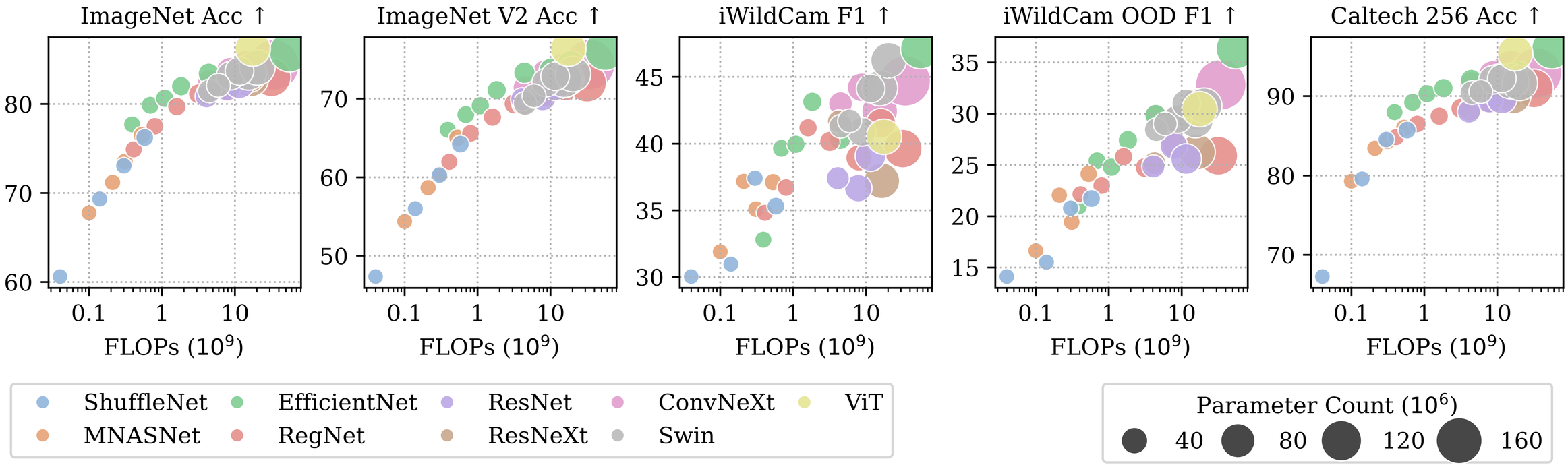
mx compute




Single Model
1x compute
Asymmetric Duo
$(1+\epsilon)$x compute
Deep Ensemble
mx compute
x
f
$\hat{y}$
f_{small}
Learned Weighted Average
f_1
$f_{...}$
f_m
Simple Average

Risk and Promises of Duo Asymmetry

Model size correlates with Accuracy

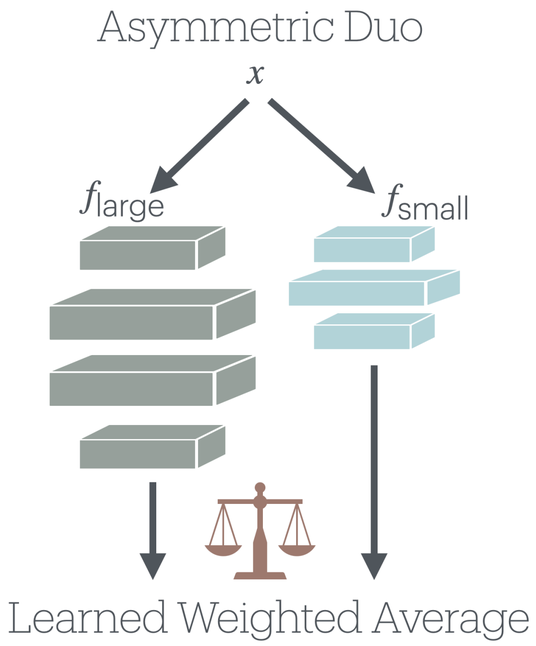


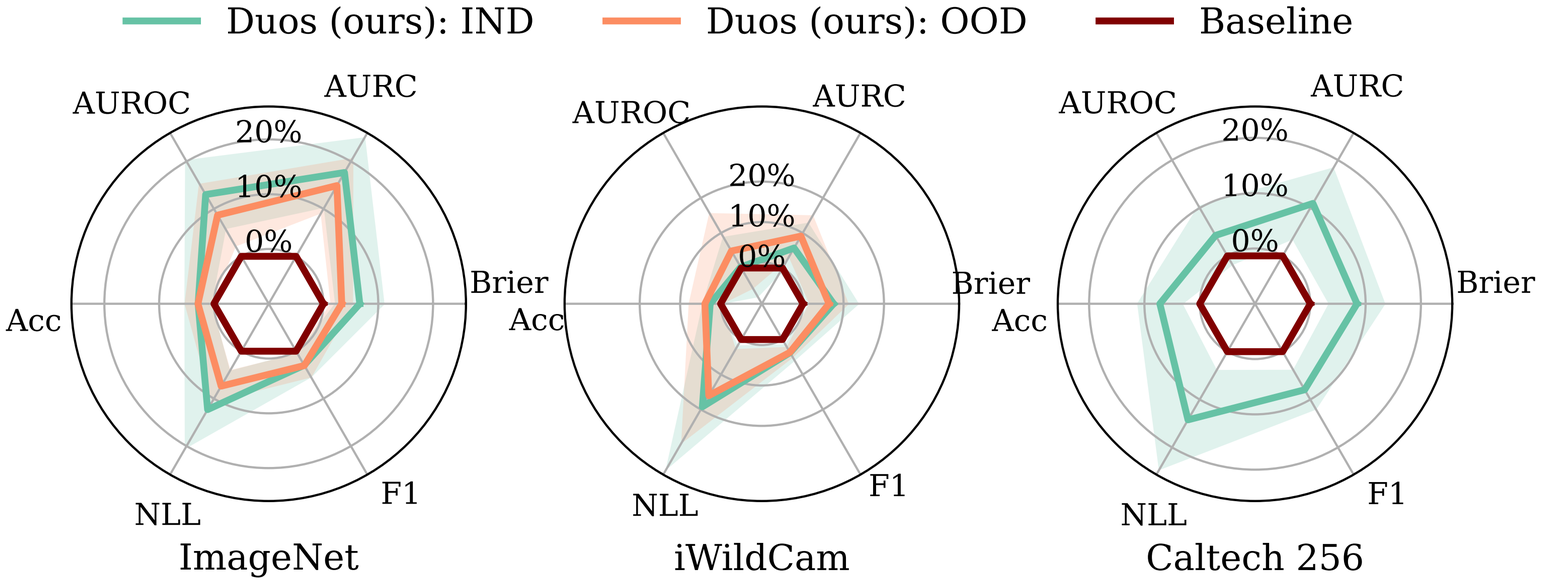
Intuitively, an Asymmetric Duo might not help or might even hurt.

Asymmetric Duos allow us to freely interpolate this correlation and get the most out of our compute budget

Temperature-Weighted Aggregation

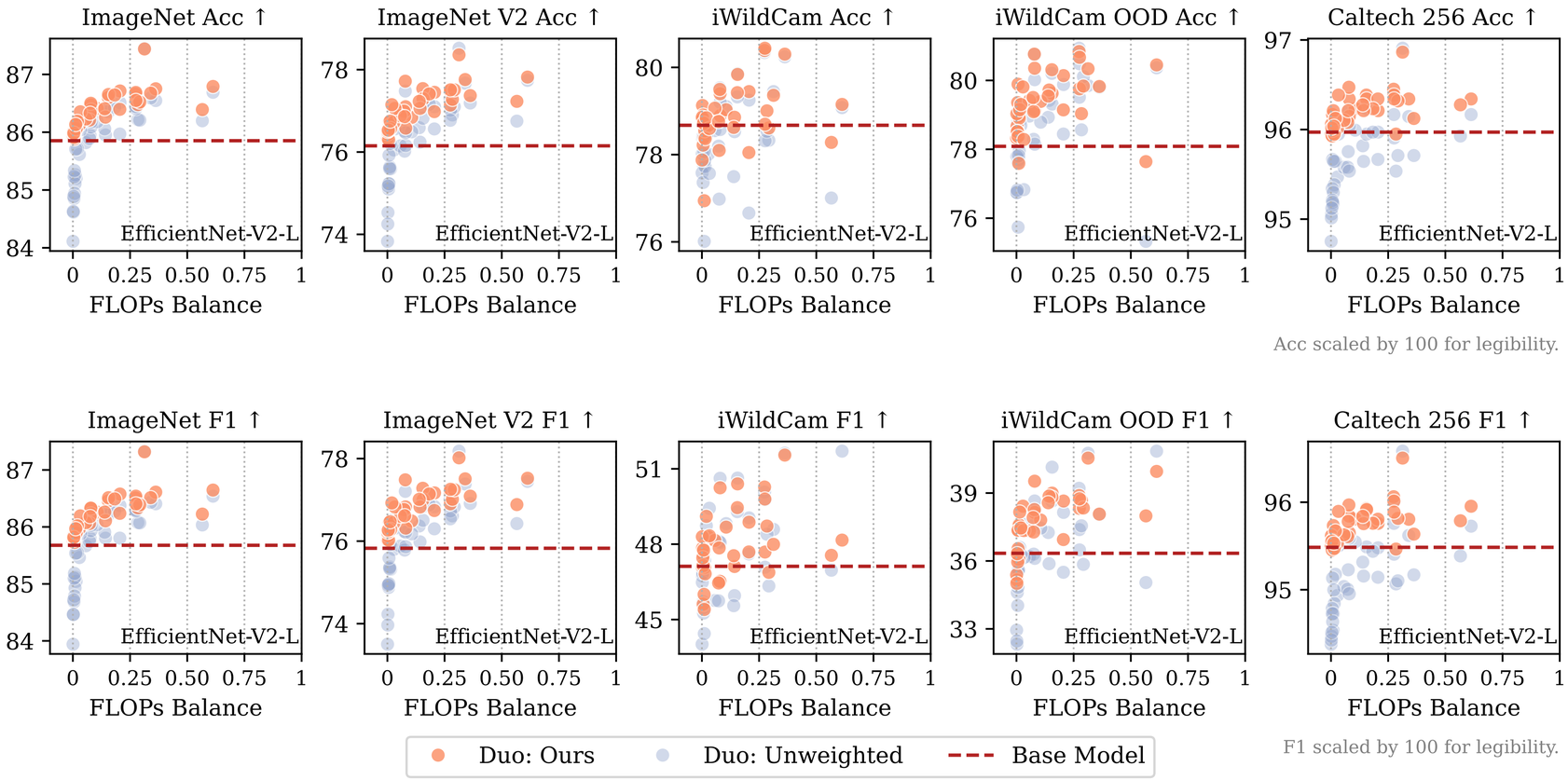
Model size correlates with Accuracy, small model might 'mess up' large model
We show that a simple learned temperature-weighting strategy can effectively combat this.





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

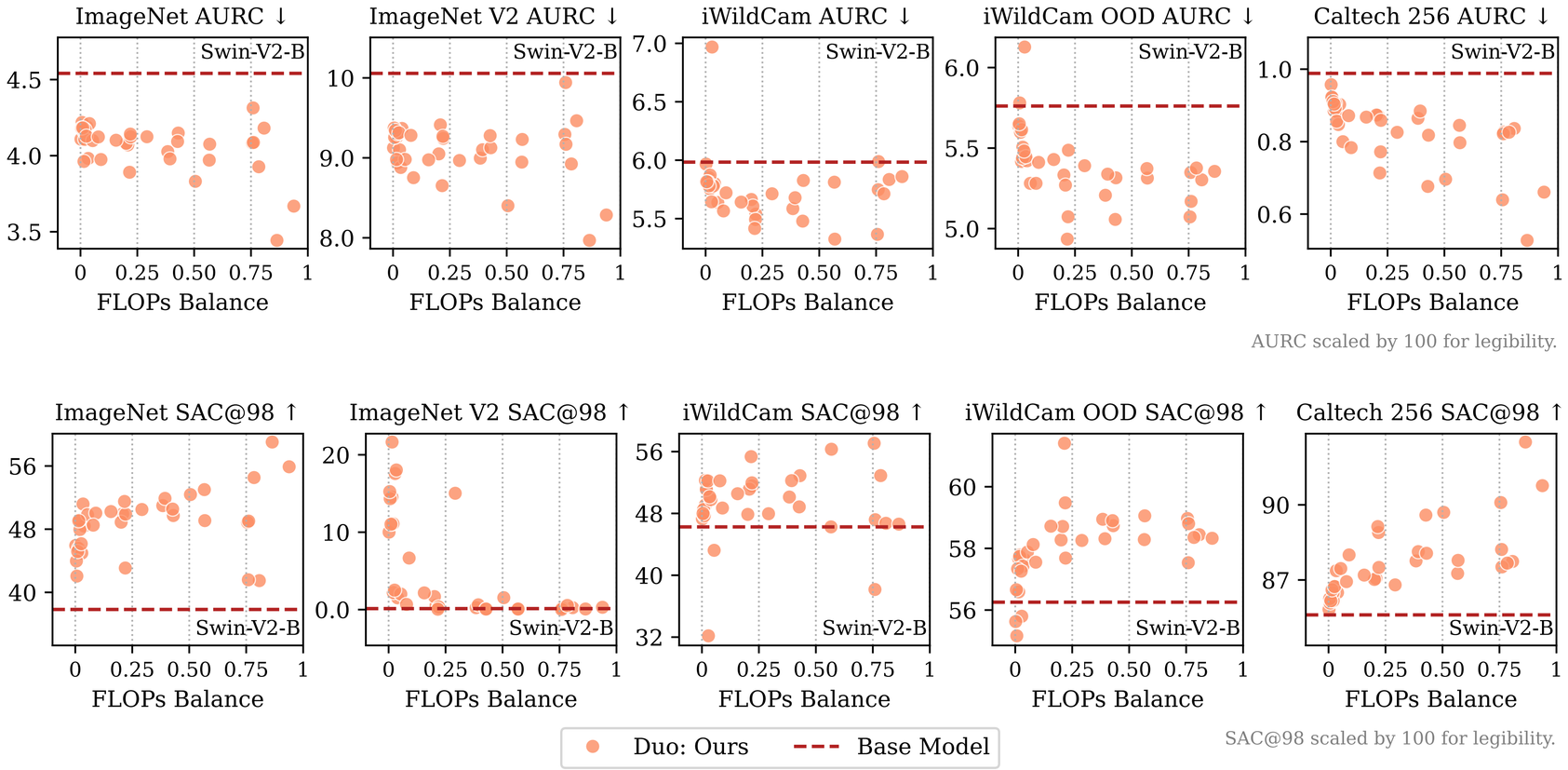
Accuracy / F1



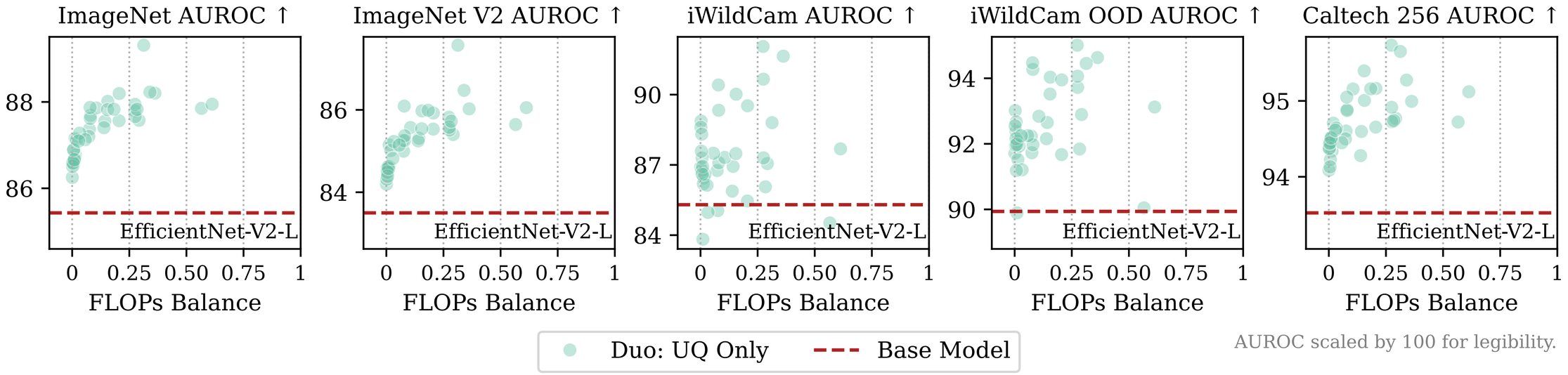
Acc scaled by 100 for legibility.

F1 scaled by 100 for legibility.

Selective Classification—AURC / SAC



Correctness Prediction

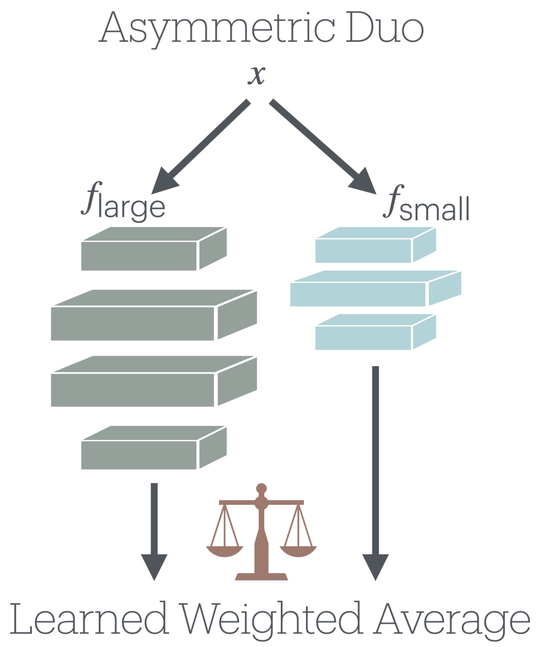


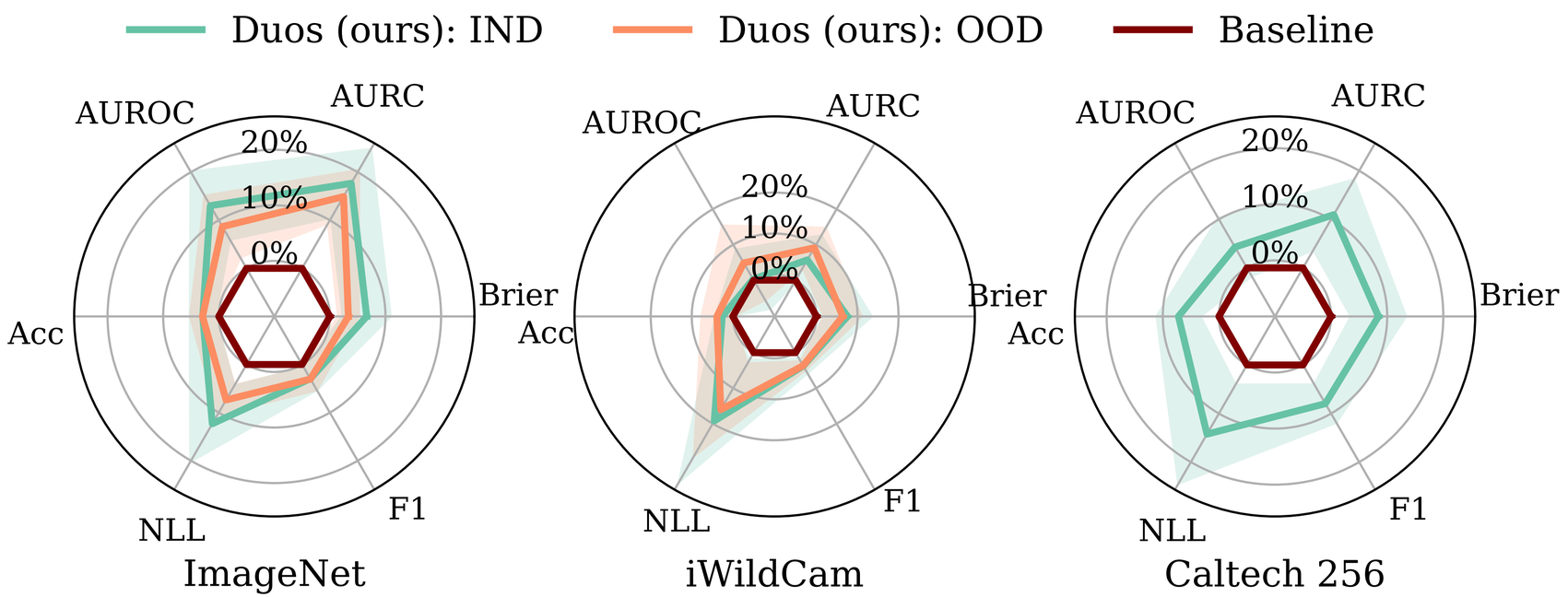
AUROC scaled by 100 for legibility.

UQ-Only Duo variant uses larger model's prediction and temperature-weighted uncertainty.

This shows that the sidekick model f_{small} is effectively re-ranking the large model's prediction.

Universal Improvement Across Metrics





Asymmetric Duos: Sidekicks Improve Uncertainty

Tim G. Zhou, Evan Shelhamer, Geoff Pleiss
University of British Columbia, Vector Institute

NeurIPS 2025