We study the **combination** of **ensembles** and **sparse conditional models**. They are **complementary**, providing **strong predictive performance and uncertainty calibration**. We propose a **new algorithm** with the **best of both worlds**.

# Sparse MoEs meet Efficient Ensembles

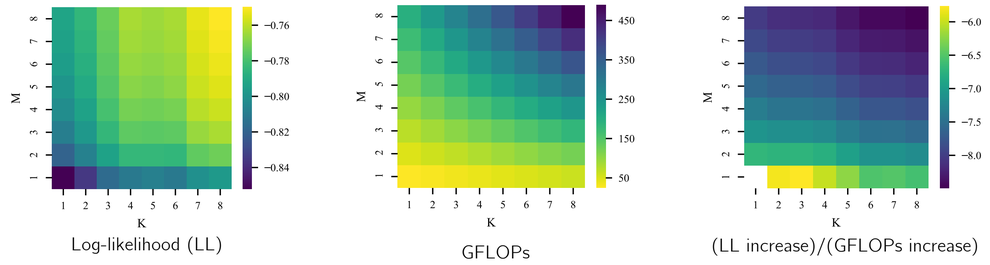

**James U Allingham** (jua23@cam.ac.uk)
Florian Wenzel
Zelda E Mariet
Basil Mustafa
Joan Puigcerver
Neil Houlsby
Ghassen Jerfel
Vincent Fortuin
Balaji Lakshminarayanan
Jasper Snoek
Dustin Tran
Carlos Riquelme
Rodolphe Jenatton (rjenatton@google.com)


Take a picture to see the full paper.

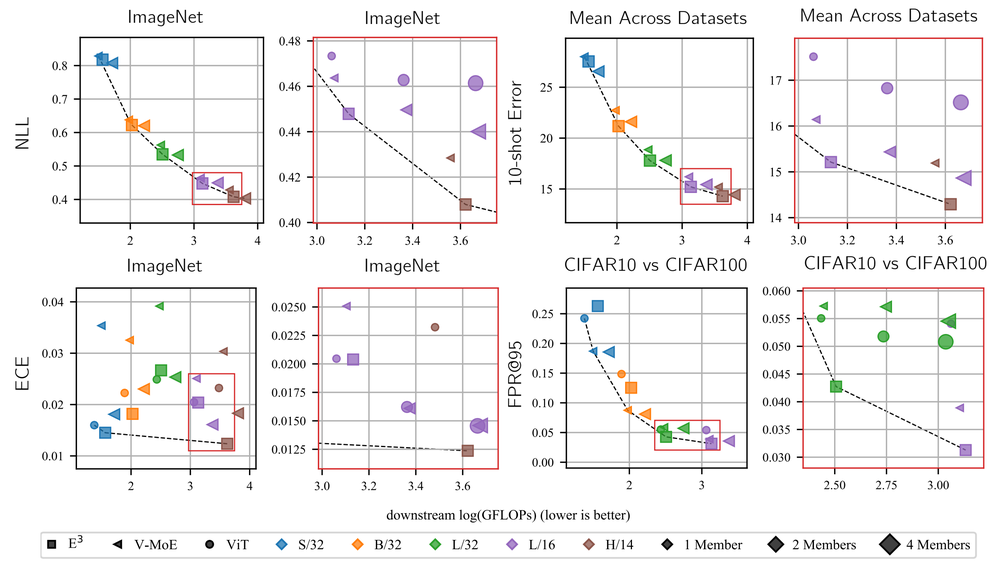
**Sparse MoEs vs Ensembles Summary**

| | Predictions | Combinations | Conditional Computation | Cost |
|---|---|---|---|---|
| Sparse MoEs | Single | Activation level | Yes, adaptively per-input | $\approx$ dense |
| Ensembles | Multiple | Prediction level | No, Static | $>$ dense |
| E$^3$ | Multiple | Activation & prediction level | Yes, adaptively per-input | $\approx$ dense |

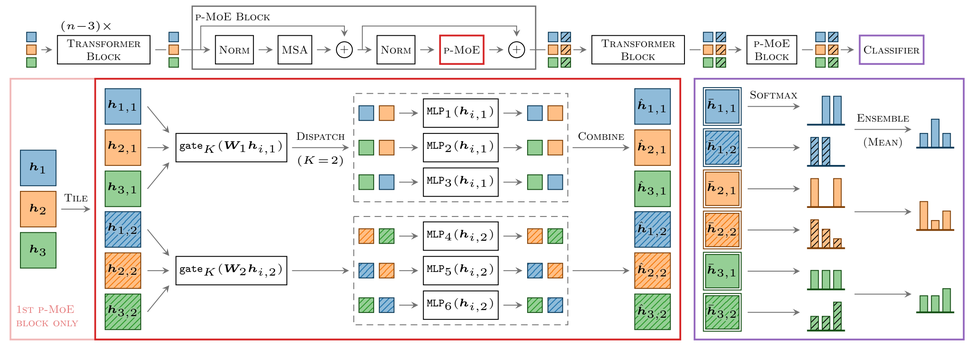
**Static (M) vs Adaptive (K) Ensembling** (yellow is better)




**Highlighted Results** (lower is better)




## E$^3$ Algorithm Overview




UNIVERSITY OF CAMBRIDGE  Google AI