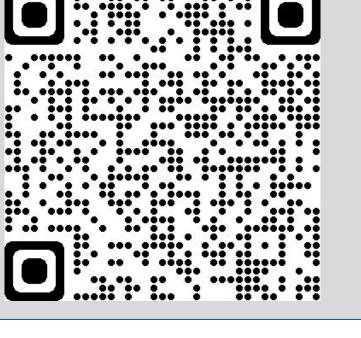


On Adversarial Robustness: A Neural Architecture Search perspective

Chaitanya Devaguptapu¹, Devansh Agarwal^{1*}, Gaurav Mittal^{2*}, Vineeth N Balasubramanian¹ {cs19mtech11025, es16btech11009, vineethnb}@iith.ac.in, gaurav.mittal@microsoft.com

¹Indian Institute of Technology, Hyderabad ²Microsoft





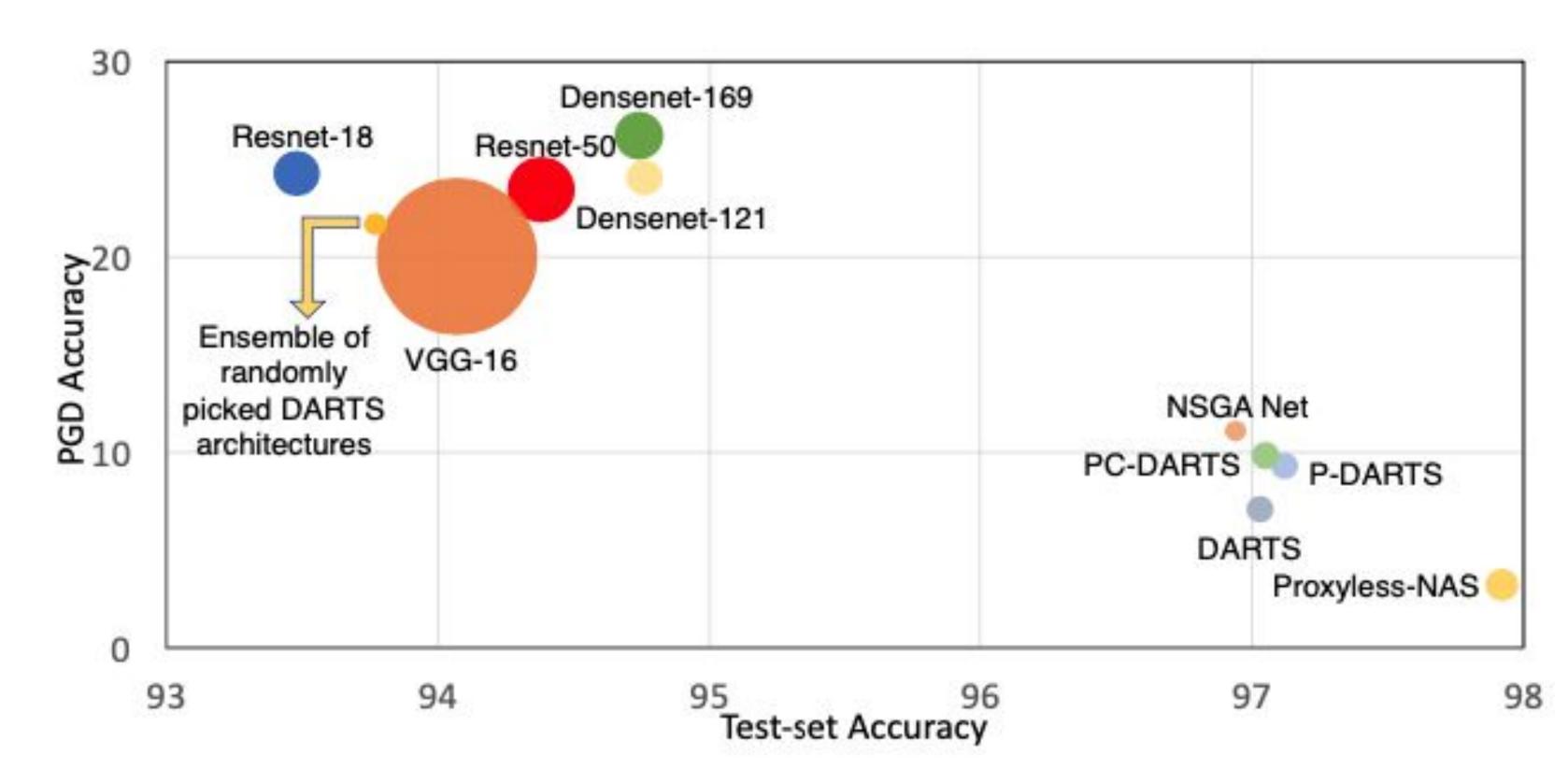
Abstract

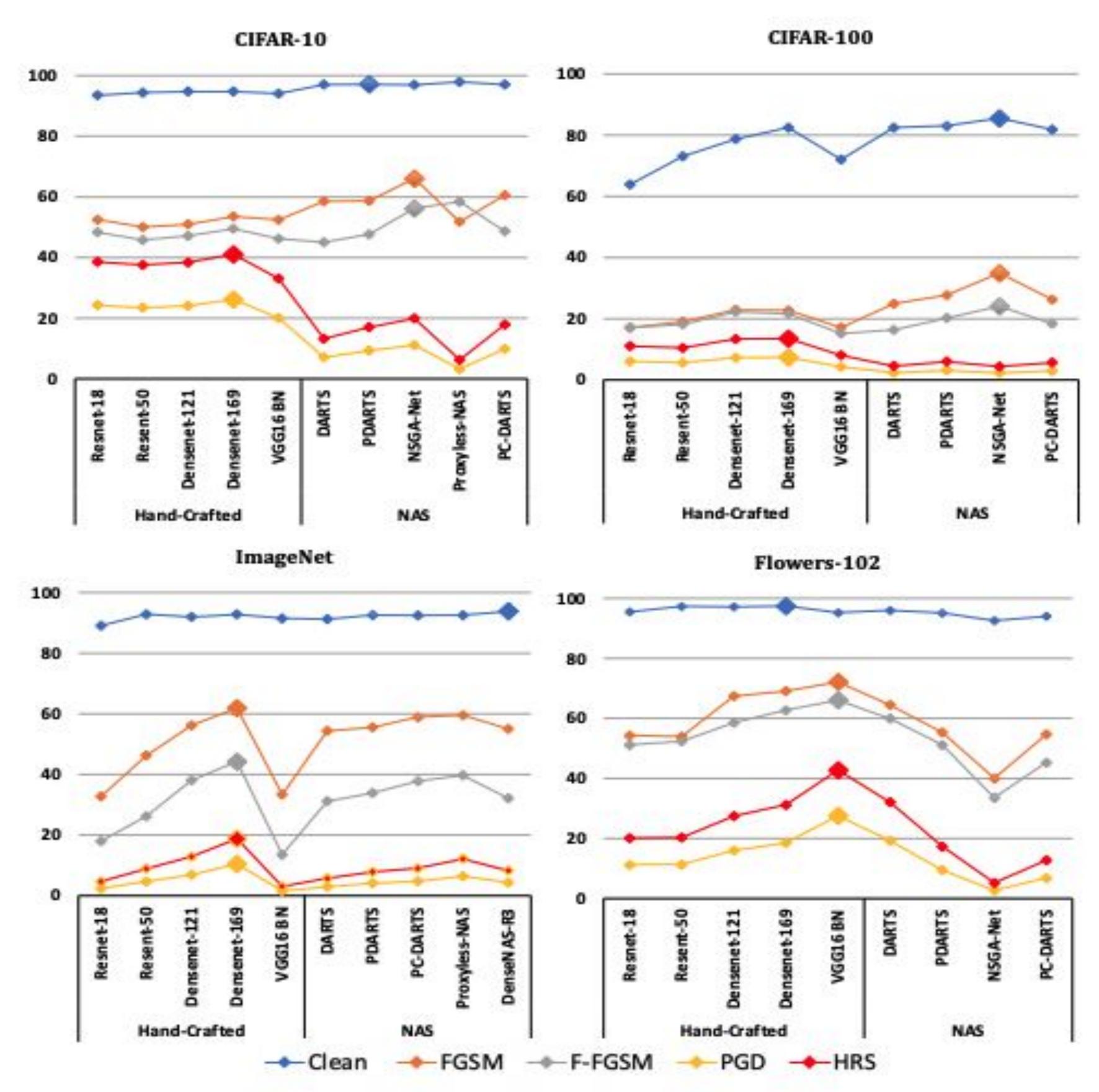
While many approaches have been proposed to improve adversarial robustness, one promising direction for improving adversarial robustness is un-explored, i.e., the complex topology of the neural network architecture. In this work, we empirically understand the effect of architecture on adversarial robustness by experimenting with different hand-crafted and NAS based architectures.

Our Hypothesis

Questions we answer to understand adversarial robustness form an architecture perspective

- How do NAS-based architectures compare with hand-crafted architectures (like ResNets, DenseNets, etc.) in terms of architectural robustness?
- Does an increase in the number of parameters of the architecture help improve robustness?
- **♦** Where does the source of adversarial vulnerability lie for NAS? Is it in the search space or in the way the current methods are performing the search?





Comparison of robustness and clean accuracy of different architectures; As the difficulty of the task or the scale of the dataset increases hand-crafted architectures are more robust

^{*} Equal Contribution