ML in the real world

CS771: Introduction to Machine Learning

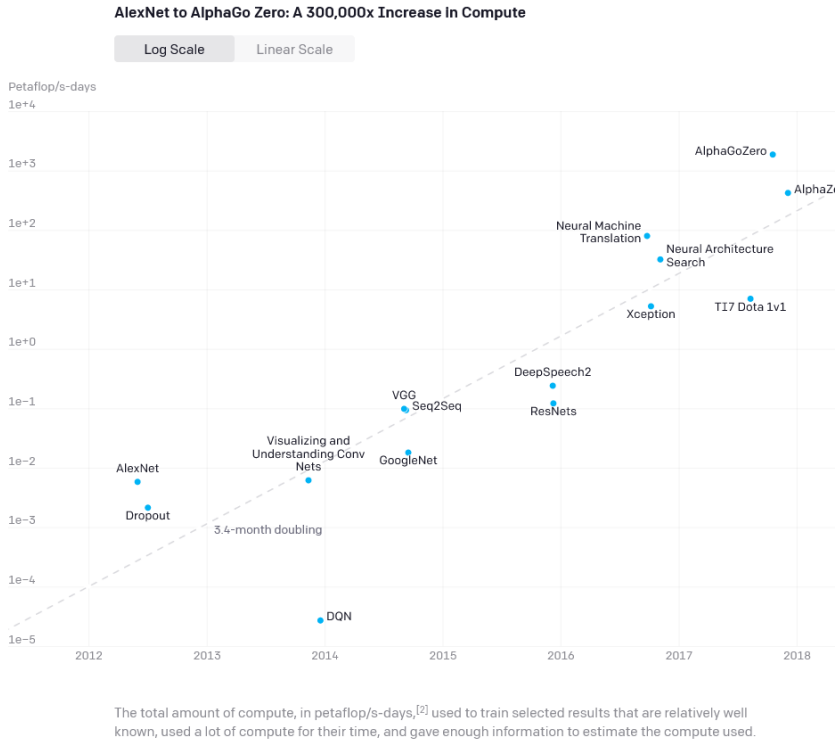
Nisheeth

Challenges

- Reproducibility
- Sustainability
- Validity
- How to fix?

Reproducibility

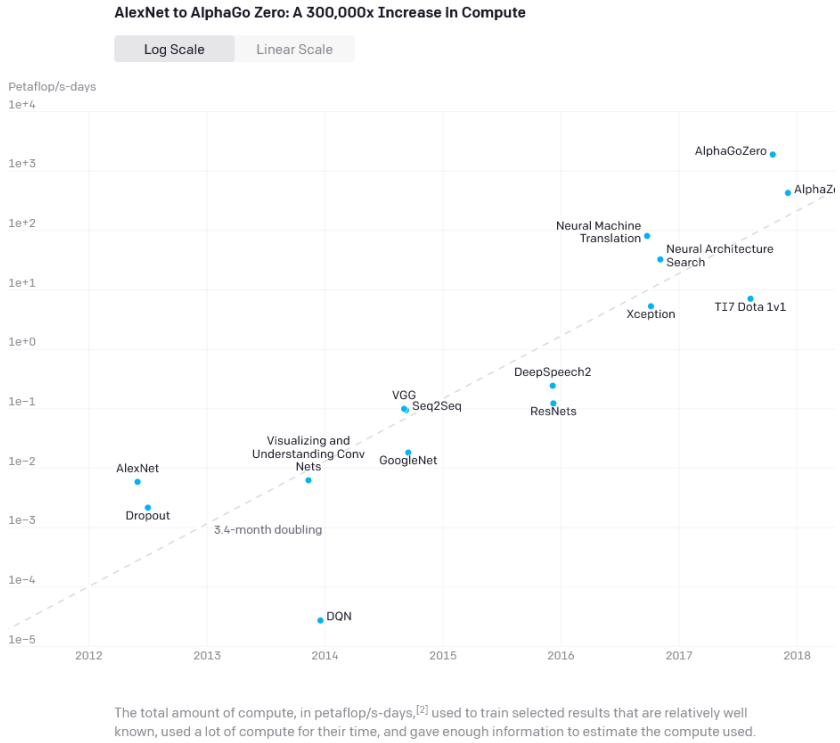
- Compute requirements for ML models are growing exponentially
- Models are too large and too expensive to retrain by other people
- How do we validate presented results?



The total amount of compute, in petaflop/s-days,[2] used to train selected results that are relatively well known, used a lot of compute for their time, and gave enough information to estimate the compute used.

Reproducibility

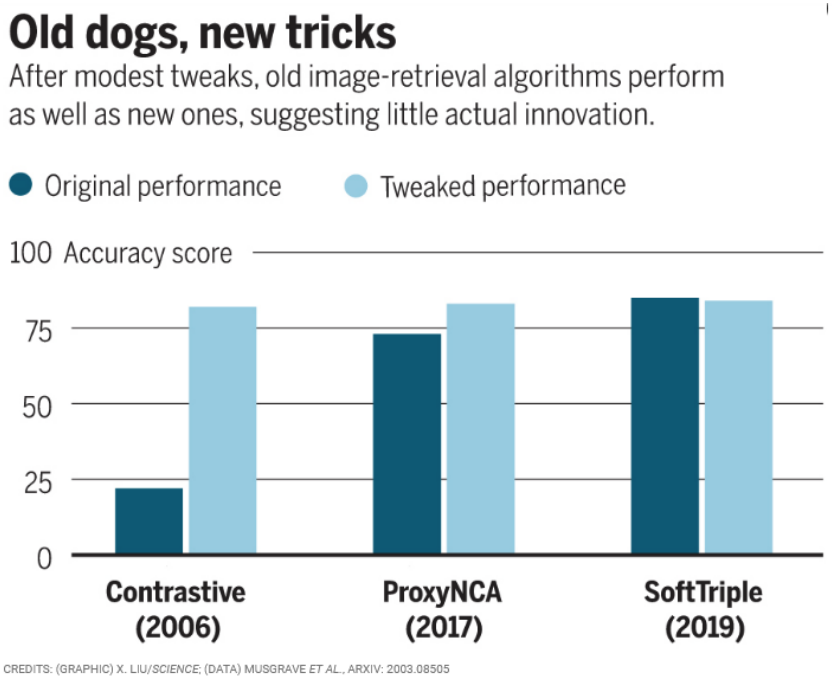
- ML has become alchemy (link)
- We very seldom know why a model works well when it does
- We almost never know why a model is making a mistake on some samples
- Researchers don't share code
  - When code is shared, hyper-parameter settings are frequently missing



The total amount of compute, in petaflop/s-days,[2] used to train selected results that are relatively well known, used a lot of compute for their time, and gave enough information to estimate the compute used.

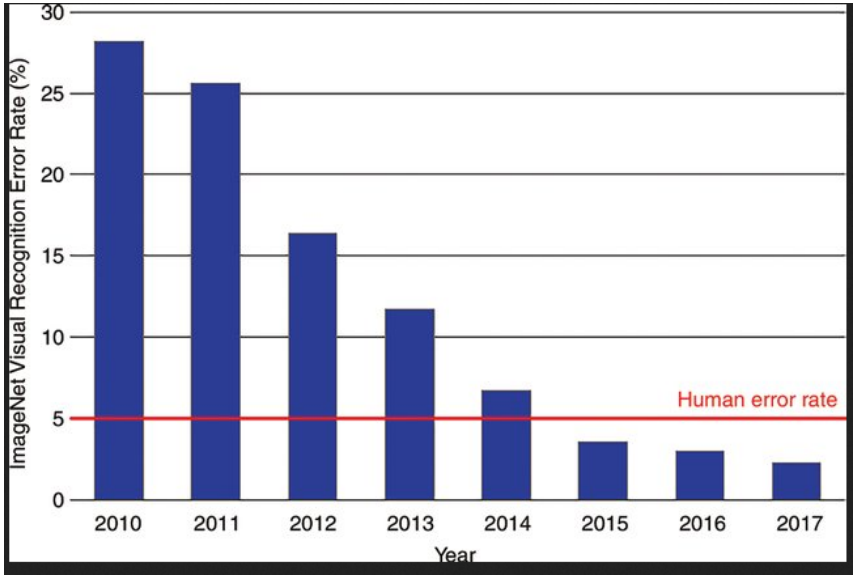
Reproducibility

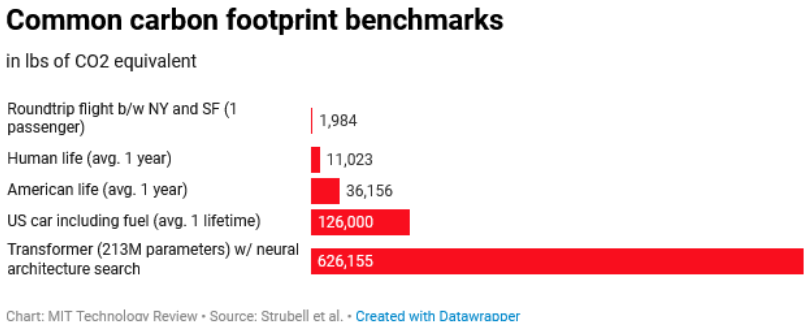
- Publishing standards simply require improvements over ‘state-of-the-art’ (SOTA) models
- Very seldom clear what SOTA is at any point in time
- Or what degree of qualitative improvement your claimed improvement buys
- In some fields of ML, progress reported over the past 10-20 years is pretty illusory (link, link)



Sustainability

- Diminishing results to extra complexity of models
- Energy costs of training models are massive
- High inequality with low reward





Validity

- Are the labels real?
- Do we care about inter-rater reliability
- Distribution shifts are very real