

A Call for Novel Datasets for Deep Learning

Stephen McGough

UK Systems workshop 2021

1/12/2021

Deep Learning is Great ... Right?

- Super-human ability to:

- Classify images
- Speech to text
- Win at Go

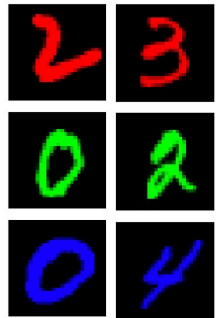
• BUT...

- Deep Learning community is focusing on small number of datasets

ICLR 2019

Datasets	Percentage of Papers
MNIST	27%

MNIST



28x28x1

Deep Learning community is focusing on small number of datasets

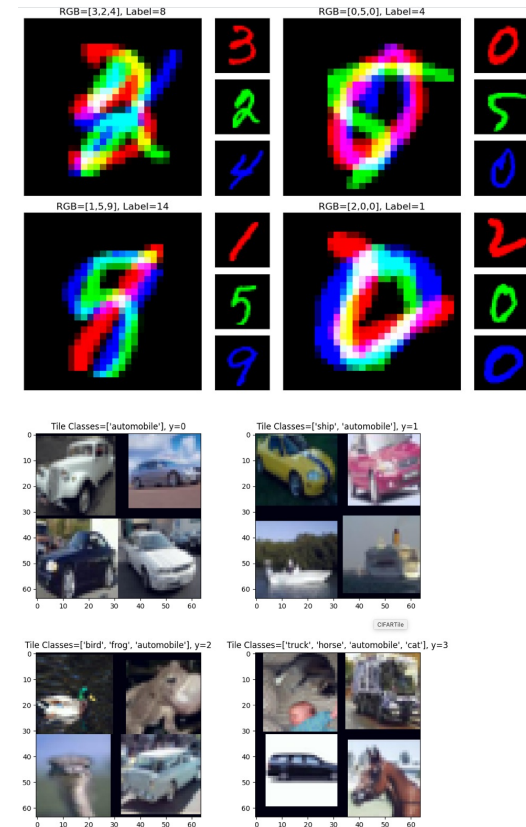
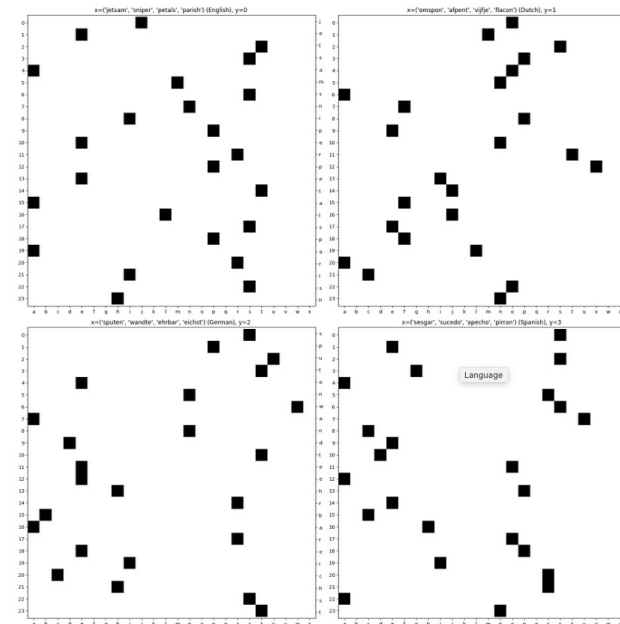
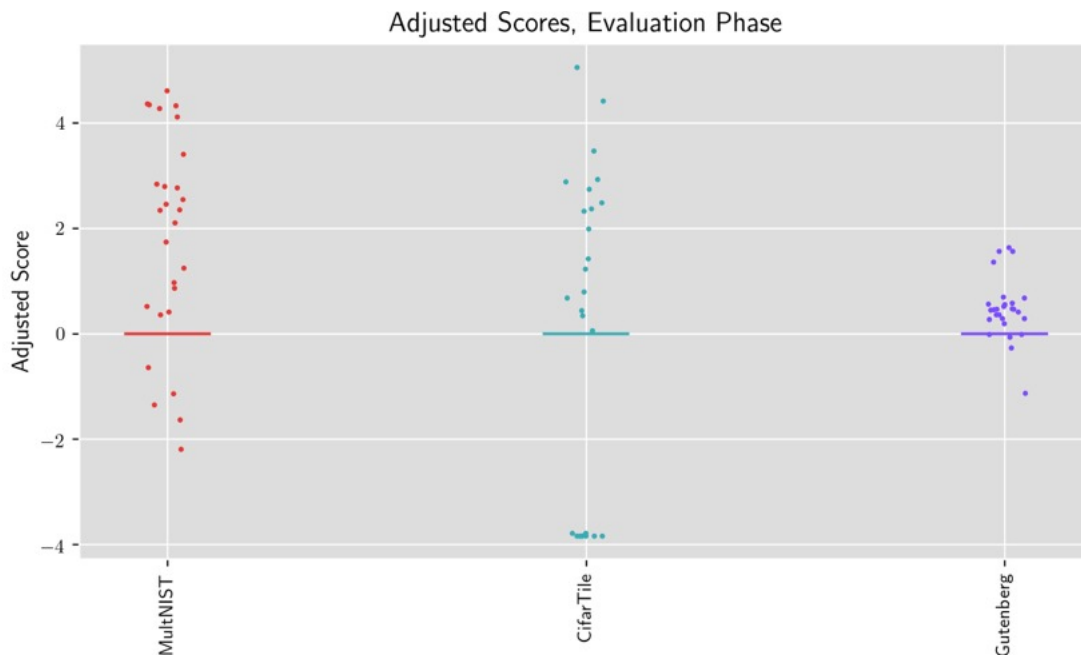
- Datasets are not representative of real—world problems
- Community is practicing ‘navel introspection’
 - Improving the accuracy on the core datasets by 0.001%
 - Lots of cases of dubious practice...
 - Thinking up new ways to use these datasets
 - ... understandable as this is how you get papers published (at the moment)

Neural Architecture Search – prime example

- Neural Architecture Search – finding the ‘best’ neural architecture for a given dataset
- But... that’s not what people are doing
- Pretty much all NAS papers use ImageNet/CIFAR10/CIFAR100
 - Approaches are ‘honed’ to find the ‘best’ architectures for datasets
 - And we already know what these are – so know how to search

So, what are we doing about this?

- Workshop and competition at CVPR 2020 – Unseen NAS
- Competitors had to produce NAS which could work on unknown datasets



So, what are we doing about this?

- David is testing as many NAS approaches as he can
 - On novel datasets
- To see how they work



<u>NAS Algorithm</u>	<u>Their Cells (Cifar-10)</u>	
	%	Error
DARTS	97.239997	2.76
PC-DARTS	97.539998	2.46
DrNAS (Darts Space)	97.279997	2.72
DrNAS (201 Space) Progressive hp = 12		100.00
DrNAS (201 Space) Progressive hp = 200		100.00
DrNAS (201 Space) hp = 12		100.00
DrNAS (201 Space) hp = 200		100.00
ENAS macro-search (no Cutout)		100.00
ENAS micro-search (no Cutout)		100.00
BossNAS		100.00
Bonsai-NET		100.00
NAT		100.00

How can you help?

- Looking for novel datasets
- That we can perform NAS on
 - You get: Optimized DL solution
 - We get: Better understanding of NAS
- Data doesn't need to be public
- Happy to collaborate
 - Joint projects, students, PhDs

stephen.mcgough@newcastle.ac.uk

