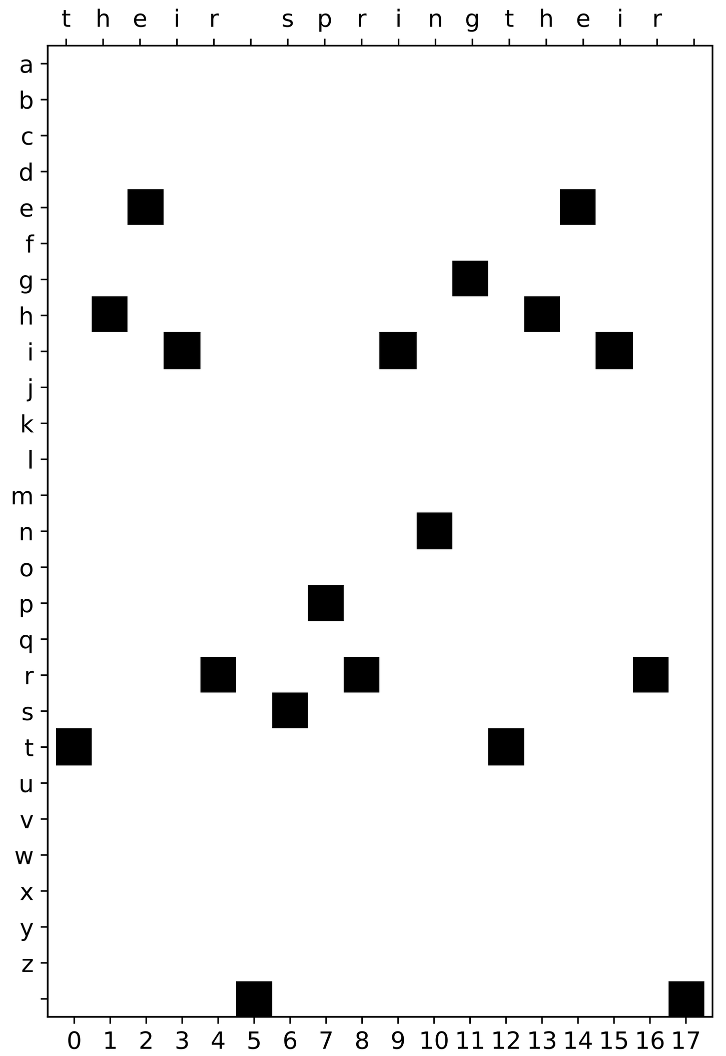
Insights From the Use of Previously Unseen NAS Datasets

Rob Geada*, David Towers* [1], Matthew Forshaw[1] [3], Amir Atapour-Abarghouei[2], A. Stephen McGough[1]

Motivation

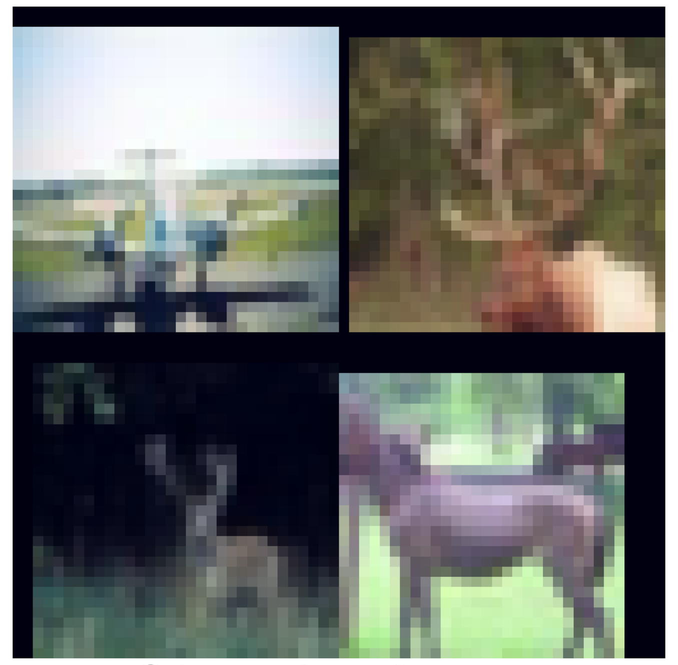
- NAS methods are tested with two datasets in mind — methods "overfit" to these datasets.
- We created the unseen-data competition to address this — and created **8 new** datasets
- These datasets are to help dilute the pool of NAS datasets.
- We want to encourage more datasets to be created for use in NAS evaluation.

Gutenberg



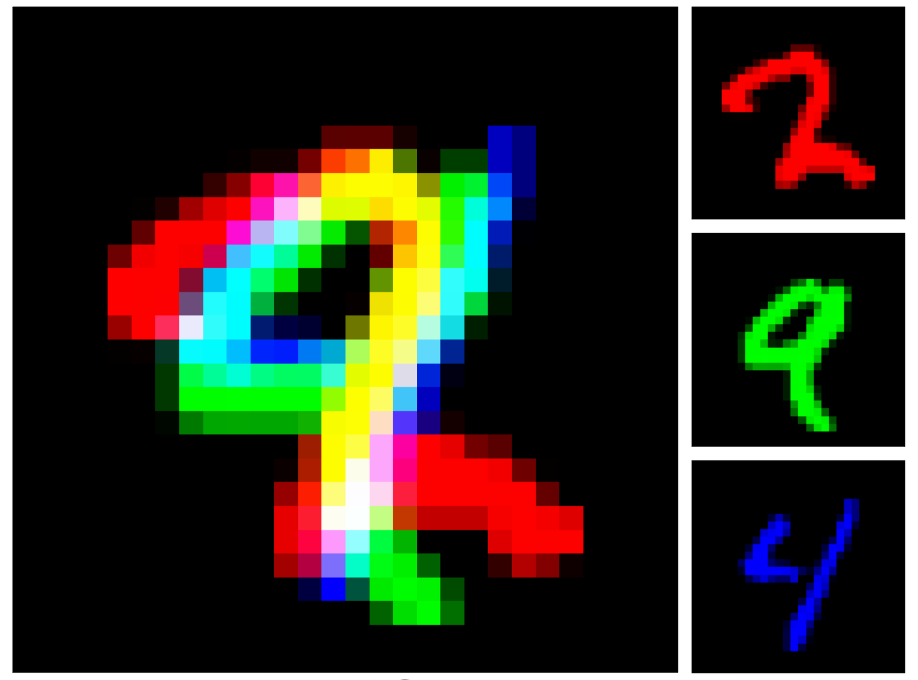
- Encoded phrases from famous books.
- Label: Author

CIFARTile



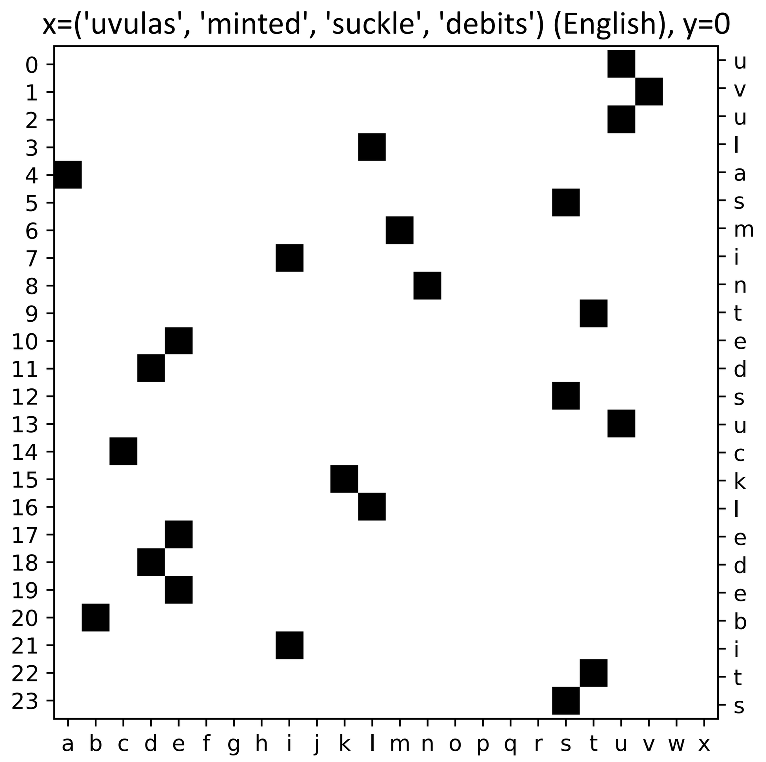
- 4 CIFAR-10 images joined together.
- Label: Number of CIFAR-10 classes.

AddNIST & MultNIST



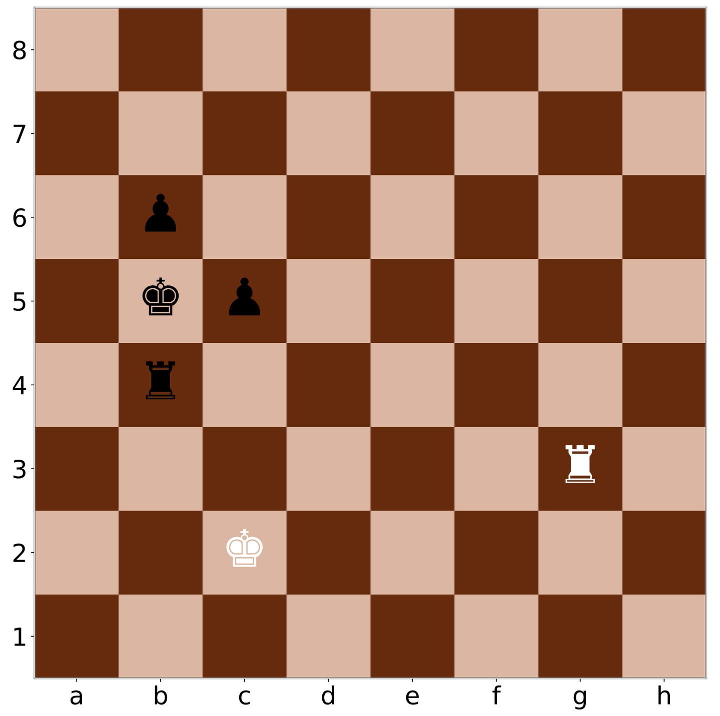
- Three MNIST images in each channel.
- Label: Calculation of the channels.

Language



- Words encoded as an image from a language.
- Label: Language

Chesseract

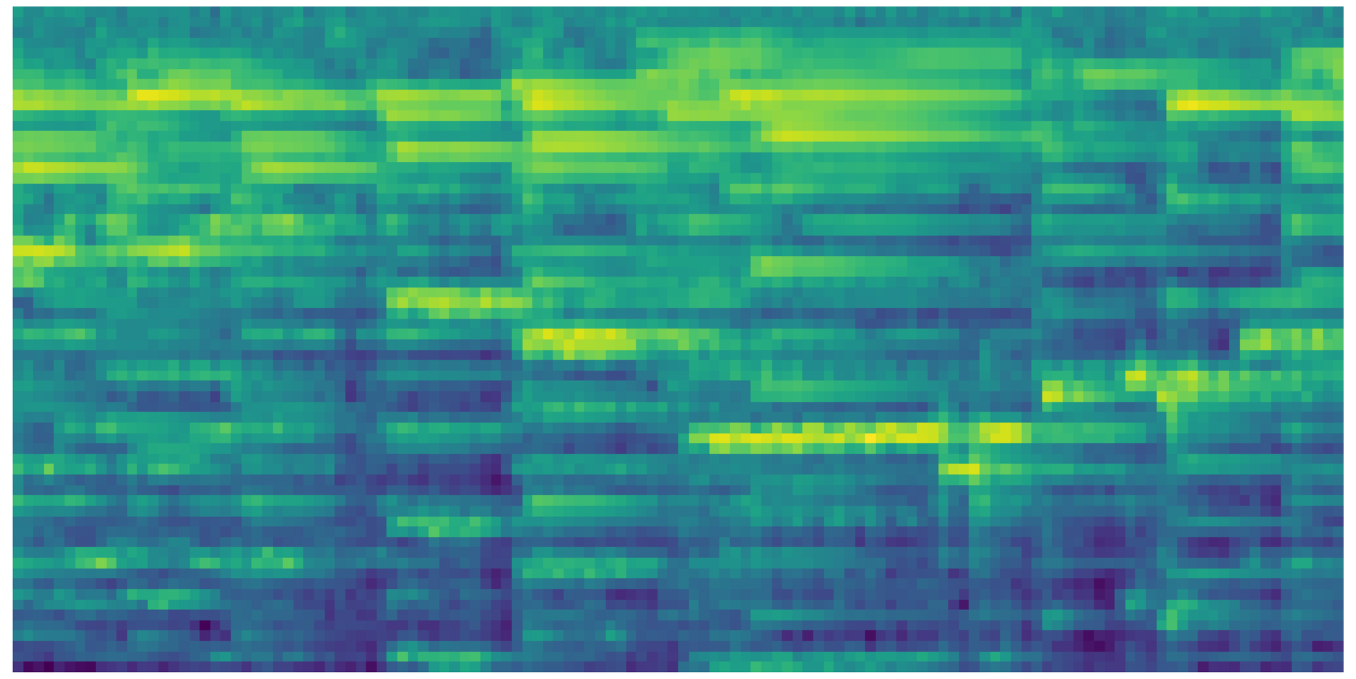


- Encoded games of Grandmaster chess.
- Label: is the outcome of the game.

GeoClassing

- Satellite images of European countries.
- Label: Country.

Isabella



- Spectrograms of music from different musical periods.
- Label: Period of Music.

Results

	AddNIST	Language	MultNIST	CIFARTile	Gutenberg	Isabella	GeoClassing	Chesseract
Conventional DL Methods								
ResNet-18	92.08%	87.00%	91.55%	45.56%	40.98%	62.02%	80.33%	57.83%
AlexNet	**94.87%**	85.00%	**94.01%**	48.88%	**45.53%**	61.37%	92.49%	57.45%
VGG16	92.06%	84.54%	90.43%	24.43%	44.00%	58.13%	83.67%	55.69%
ConvNext	38.06%	83.40%	64.20%	31.06%	31.93%	57.18%	72.76%	52.74%
MNASNet	90.51%	84.63%	87.70%	48.49%	38.00%	60.69%	86.00%	56.26%
DenseNet	93.52%	84.57%	92.81%	**51.28%**	43.28%	**63.27%**	**94.21%**	**59.60%**
ResNeXt	91.42%	**93.97%**	90.57%	46.23%	40.30%	60.46%	89.99%	55.15%
NAS Methods								
PC-DARTS	96.60%	**90.12%**	96.68%	**92.28%**	**49.12%**	**65.77%**	94.61%	57.20%
DrNAS	97.06%	88.55%	**98.10%**	81.08%	46.62%	64.53%	**96.03%**	58.24%
Bonsai-NET	**97.91%**	87.65%	97.17%	91.47%	48.57%	64.08%	95.66%	**60.76%**

Conclusion & Insights

- There is no "one size fits all" approach.
- NAS does generally produce better results than conventional DL.
- However, no NAS method we tested is the best at all datasets.
- More datasets should be made.
- It is reductive to evaluate NAS on few datasets

Contact Us

Feel free to come up and talk to me during the conference.

Or you can email me directly:
d.towers2@ncl.ac.uk

Check out our GitHub (see QR code) and paper for more detail!

*These authors performed equal contribution
[1]Newcastle University
[2]Durham University
[3]The Alan Turing Institute

Scan for our GitHub

Newcastle University

Durham University

The Alan Turing Institute