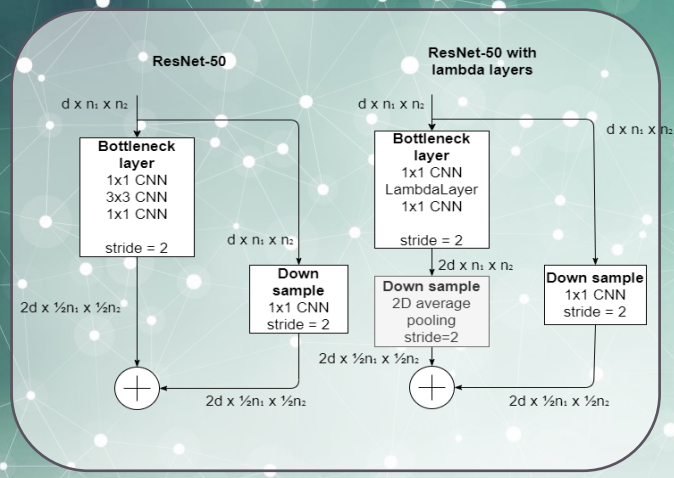
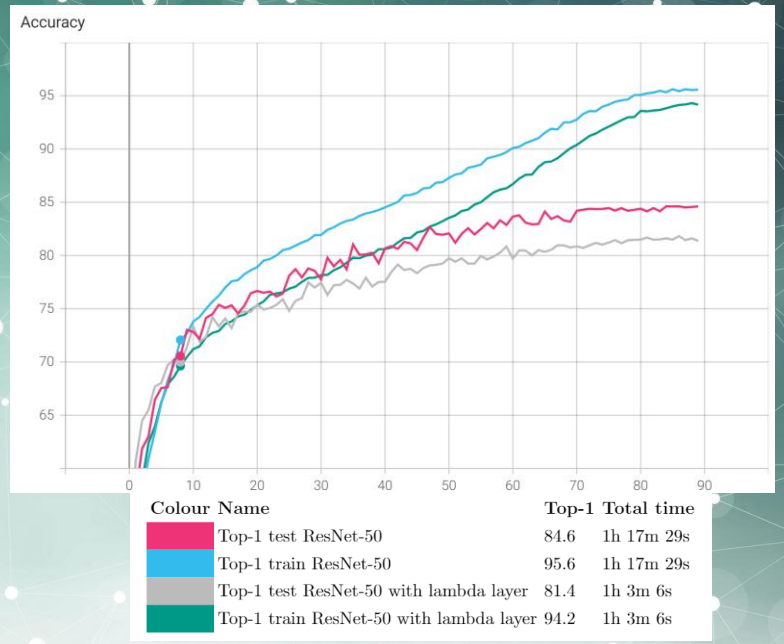
	Test accuracy
ImageNet (Bello)	78.4
CIFAR-10 (Ours)	80.18

	Train	Test	Compute time
ResNet-50	95.58	84.61	77 min
Lambda (Ours)	94.17	81.37	63 min

de Alvear Cárdenas, J.I.
j.i.dealvearcardenas@student.tudelft.nl
de Vries, W.A.J.G.
w.devries-1@student.tudelft.nl
Faculty of Aerospace Engineering
University of Technology Delft

Delft, April 14, 2021





- GPU: 32 TPU v3
- Data: ImageNet
- ResNet-50 conv. layer: Top-1 76.9
- ResNet-50 lambda layer: Top-1 78.4

- ImageNet to CIFAR10
- Implemented of lambda layers
- Learning rate to 0.0005

- GPU: K80/T4/P4s/P100
- Data: CIFAR-10
- ResNet-50 conv. layer: Top-1 84,6
- ResNet-50 lambda layer: Top-1 81.4

LAMBDA-NETWORKS

Efficient & accurate, but also accessible? A reproducibility project with CIFAR-10

Common problems with respect to modeling long-range interactions with self-attention is the generation of attention maps. These require a lot of compute and memory storage; therefore Bello (2021) developed a new method by replacing these maps by linear functions called lambdas. They parallelize nicely, especially on the 32 TPUv3s Bello uses, it slightly increases accuracy (becoming the new state-of-the-art), moreover it is 9 times faster than EfficientNet NoisyStudent (Xie et al., 2020) and Vision Transformer (Dosovitskiy et al., 2020) on the 200 GB large ImageNet dataset. What if you do not have this much compute at your disposal and want to run it from home or on Google Colab?

In that case, as shown by our results, lambda layers do not yield as much of an advantage as the normal ResNet-50 seems superior in performance.