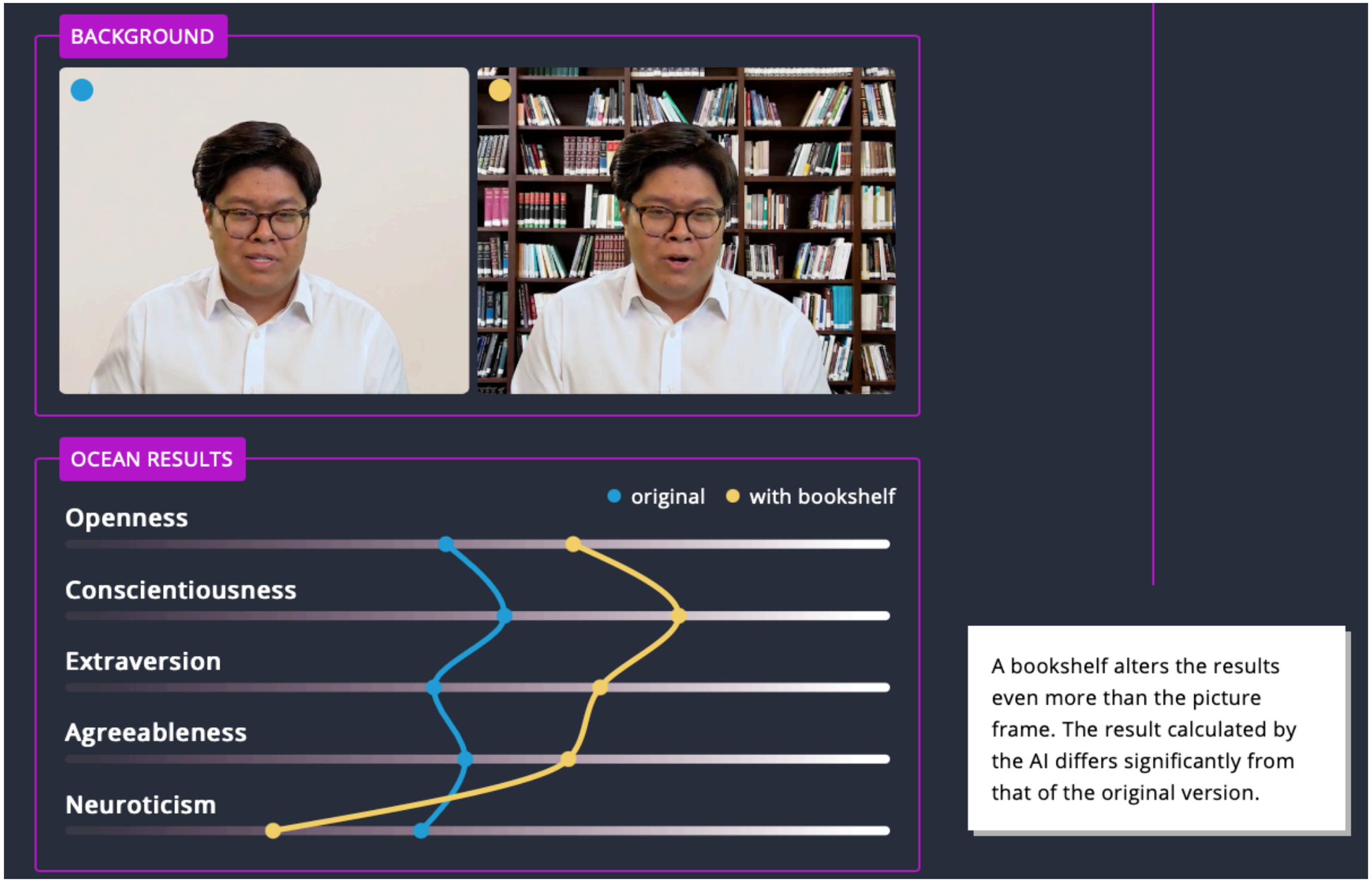
STAT 453: Introduction to Deep Learning and Generative Models

Sebastian Raschka
http://stat.wisc.edu/~sraschka

Deep Learning & AI News #10

Interesting Things Related to Deep Learning

Apr 3, 2021



https://web.br.de/interaktiv/ki-bewerbung/en/

arXiv.org > cs > arXiv:2104.00170

Search...

Help | Advanced

Computer Science > Machine Learning

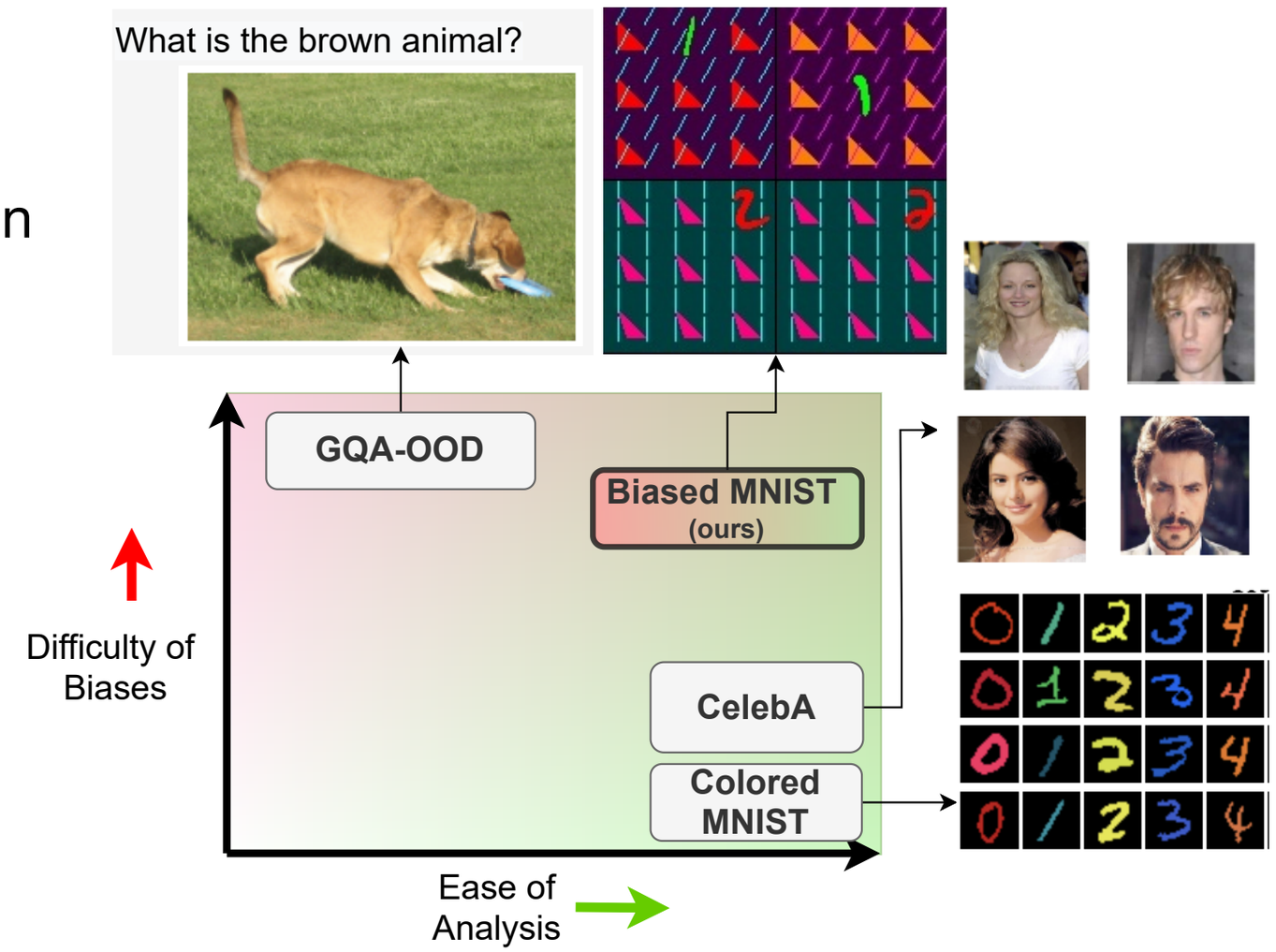
[Submitted on 1 Apr 2021]

An Investigation of Critical Issues in Bias Mitigation Techniques

Robik Shrestha, Kushal Kafle, Christopher Kanan

https://arxiv.org/abs/2104.00170

- Learning inappropriate biases can cause DL models to perform badly on minority groups
- Several methods were developed to address this, but do they work?
- Here:
  - Improved evaluation protocol & dataset
  - Evaluation of 7 methods
  - Biased MNIST dataset
- Code and data: https://github.com/erobic/bias-mitigators



Task: recognize the digit (0 – 9)
Multiple sources of biases:

1) background color,
2) color of the target digit,
3) position of the digit (among 9 grid locations),
4) distractor shapes, which are placed on all cells except the cell with the digit,
5) color of the distractors,
6) type of texture, and
7) texture color.

Each digit co-occurs more frequently with a particular value for each bias type
E.g., "1" is most often green, placed on purple background, co-occurs with right-angled triangles

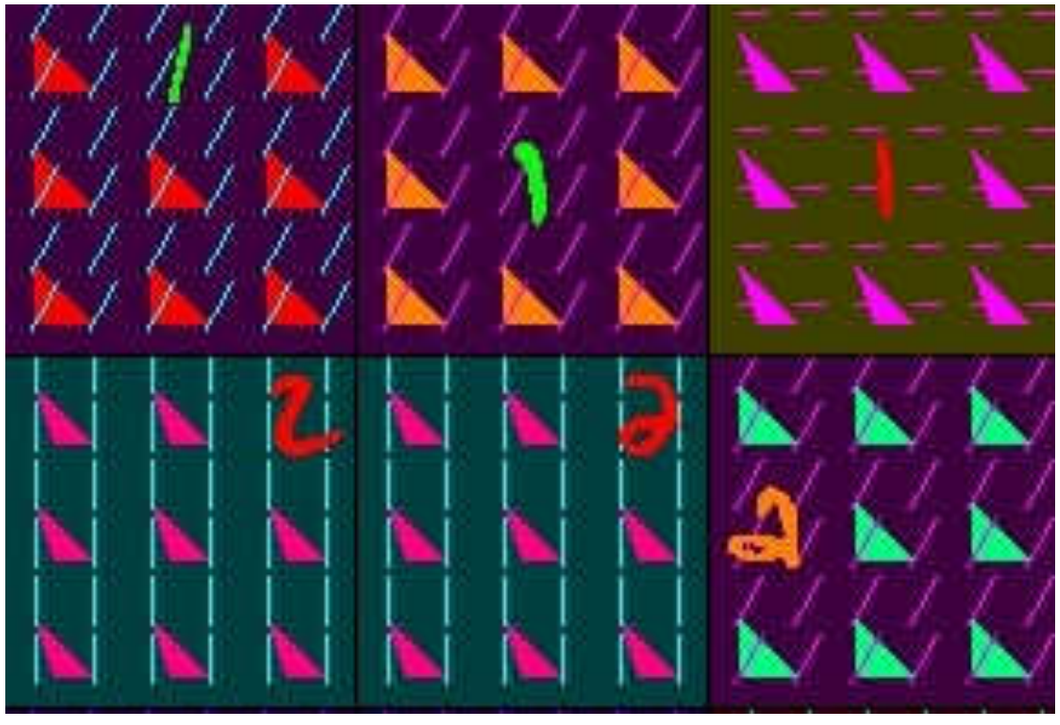
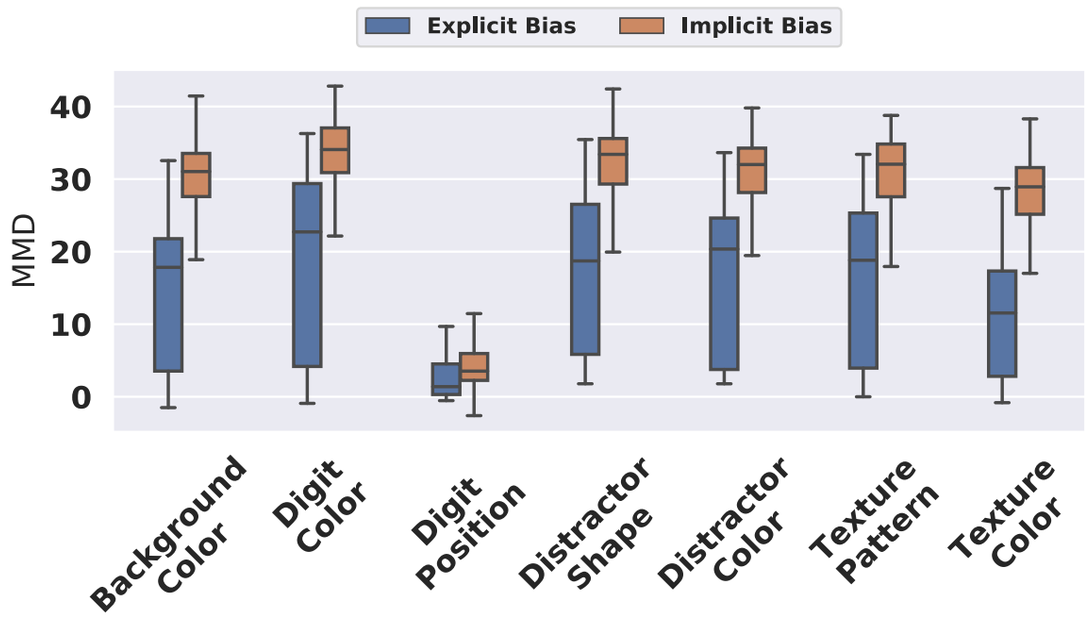
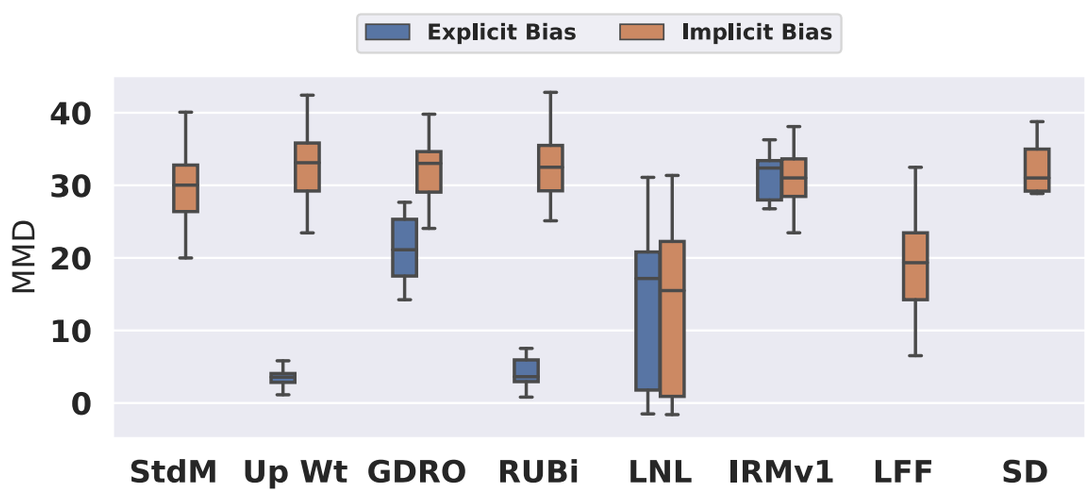


Figure 2: Biased MNIST requires the methods to classify the target digit while remaining invariant to biases.

[treat] "individual variable as the explicit bias in separate experiments, while treating the remaining six as implicit biases"



(a)



(b)

Figure 3: Boxplots of differences between majority and minority groups (MMD) on Biased MNIST over: a) bias variables and b) different methods.

We define two more metrics to help measure bias resistance. **Majority/Minority Difference (MMD)** simply measures the difference between majority and minority groups:

$$MMD = [Acc_{majority} - Acc_{minority}].$$

High MMD indicates that methods rely on factors that work for majority groups, but not for minority groups. The second metric is **Improvement Over the Standard Model (IOSM)**, which measures the difference in group g's accuracy as compared to the standard model (StdM) i.e.,

$$IOSM_g = [Acc_g - Acc_{StdM,g}].$$

Ideal method would obtain high $IOSM_g$ across all groups.

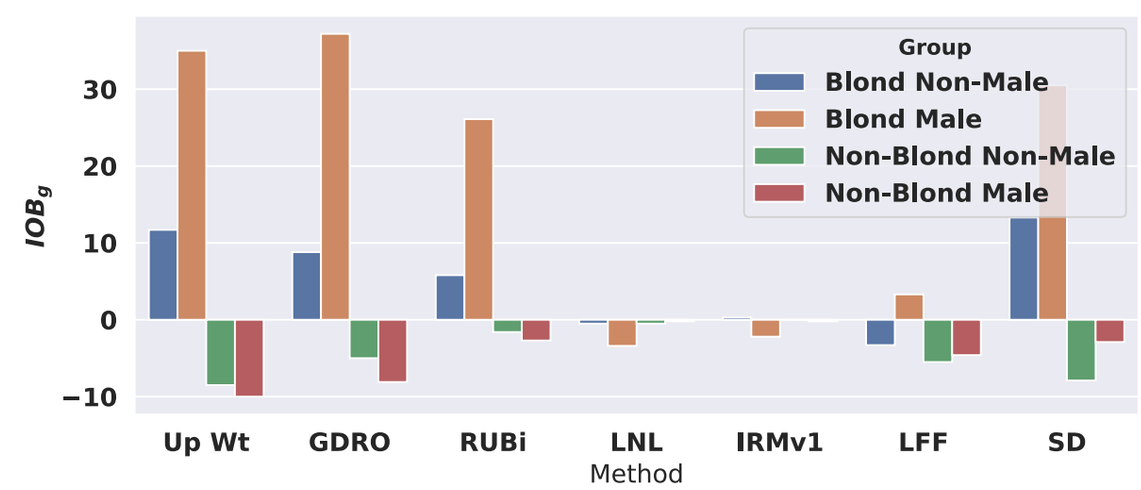


Figure 4: Improvement Over the Standard Model (IOSM) for each group of CelebA.

arXiv.org > cs > arXiv:1905.11946

Search...

Help | Advanced S

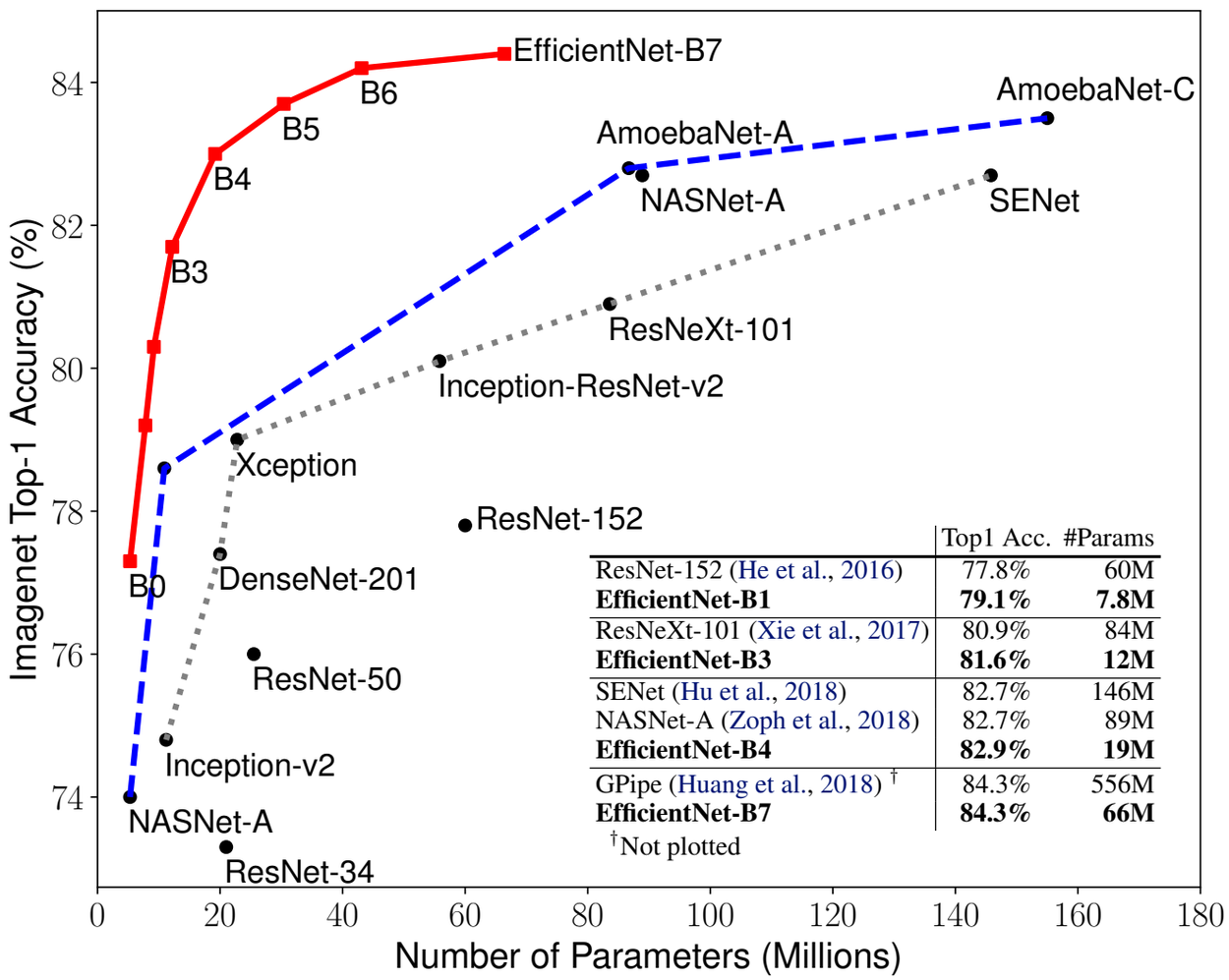
Computer Science > Machine Learning

[Submitted on 28 May 2019 (v1), last revised 11 Sep 2020 (this version, v5)]

EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks

Mingxing Tan, Quoc V. Le

https://arxiv.org/abs/1905.11946



	Top1 Acc.	#Params
ResNet-152 (He et al., 2016)	77.8%	60M
EfficientNet-B1	**79.1%**	**7.8M**
ResNeXt-101 (Xie et al., 2017)	80.9%	84M
EfficientNet-B3	**81.6%**	**12M**
SENet (Hu et al., 2018)	82.7%	146M
NASNet-A (Zoph et al., 2018)	82.7%	89M
EfficientNet-B4	**82.9%**	**19M**
GPipe (Huang et al., 2018) †	84.3%	556M
EfficientNet-B7	**84.3%**	**66M**

†Not plotted

Figure 1. **Model Size vs. ImageNet Accuracy.** All numbers are for single-crop, single-model. Our EfficientNets significantly outperform other ConvNets. In particular, EfficientNet-B7 achieves

- Original **EfficientNets**: Scaling up CNNs with a compound coefficient.
- Common ways to scale:
  (1) More layers,
  (2) Wider layers,
  (3) Higher image resolution
- EfficientNets: fixed compound ratio for scaling all three
- Result: Better accuracy, fewer parameters, faster than reference networks

arXiv.org > cs > arXiv:2104.00298

Search...

Help | Advanced S

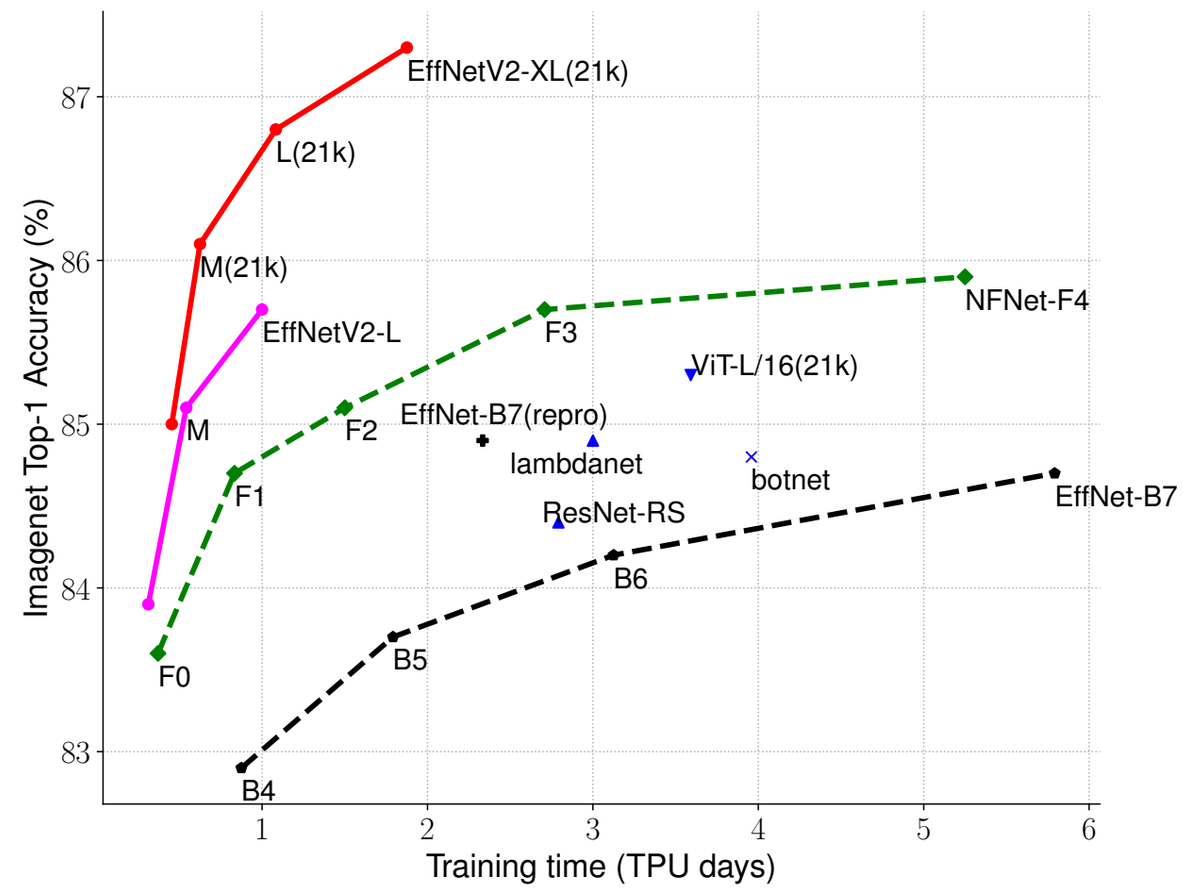
Computer Science > Computer Vision and Pattern Recognition

[Submitted on 1 Apr 2021]

EfficientNetV2: Smaller Models and Faster Training

Mingxing Tan, Quoc V. Le

https://arxiv.org/abs/2104.00298



(a) Training efficiency.

	EfficientNet (2019)	ResNet-RS (2021)	DeiT/ViT (2021)	EfficientNetV2 (ours)
Top-1 Acc.	84.3%	84.0%	83.1%	83.9%
Parameters	43M	164M	86M	24M

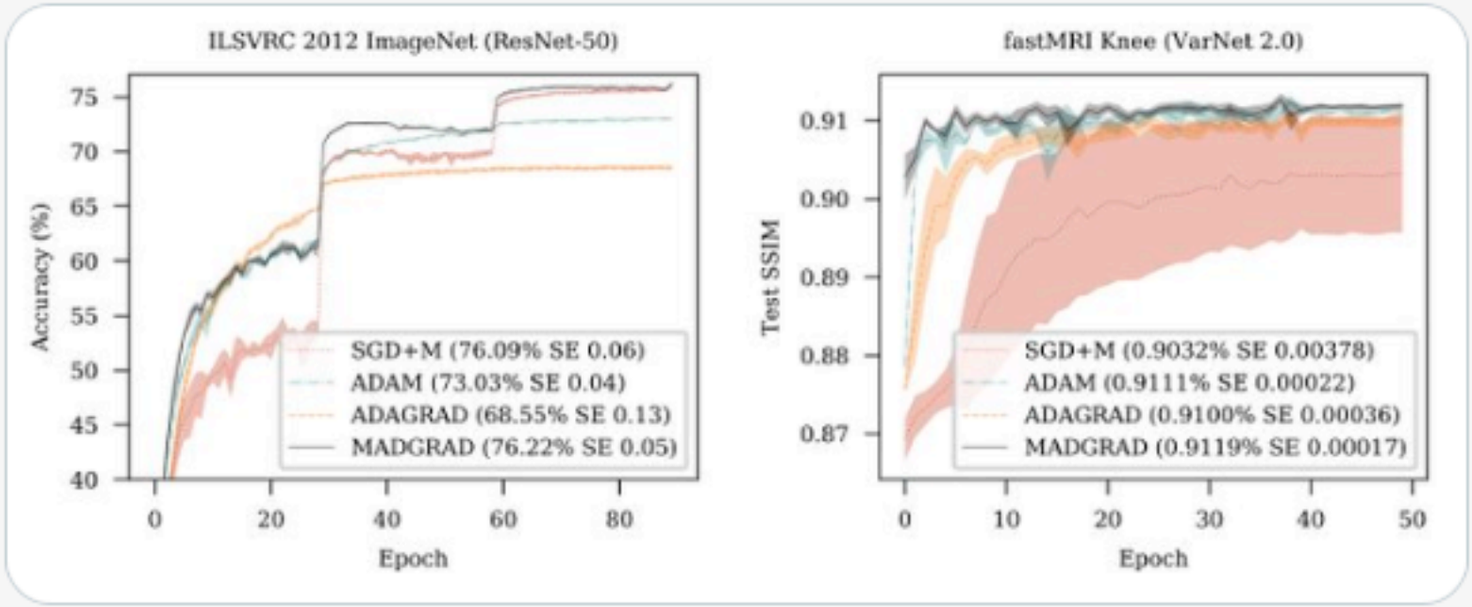
(b) Parameter efficiency.

- **EfficientNetV2:**
  Large improvement over EfficientNets V1
  Also beats Visual Transformers ;)

- Introduces
  new ops such as Fused-MBConv
  progressive increasing of image size during training
  -> adaptively adjusting regularization via dropout and data augmentation

Facebook AI @facebookai · Mar 30

We're introducing an optimizer for deep learning, MADGRAD. This method matches or exceeds the performance of the Adam optimizer across a varied set of realistic large-scale deep learning training problems. github.com/facebookresear...



https://arxiv.org/abs/2101.11075

"Adam doesn't quite reach the goal of being a general-purpose deep learning optimizer. The MADGRAD method is directly designed to address these issues"

Comparison of ML Experiment Tracking Tools

	MLflow	Tensorboard	DVC	ClearML	Guild.ai	Kubeflow	Neptune.ai	Weights & Biases	Comet.ml	SageMaker Experiments	DAGsHub
Open source	✅ Apache	✅ Apache	✅ Apache	✅⚠️ Server has non-standard license	✅ Apache	✅ Apache	❌	❌	❌	❌	– Open source formats
Platform & language agnostic	❌	❌	✅	❌	❌	❌	❌	❌	❌	❌	✅
Experiment data storage Local 📁 / Cloud ☁️	📁+☁️	📁	📁	☁️	📁+☁️	☁️	☁️	☁️	☁️	☁️	📁+☁️
Easy to set up	✅	✅	✅	– Open-source server is hard to set up	✅	❌	✅	✅	✅	✅	✅
Custom visualizations	✅	✅	– Difficult to customize	✅	✅	❌	✅	✅	✅	✅	❌
Scalable for large number of experiments	✅	❌	??	✅	??	✅	✅	✅	✅	✅	✅

To read more, go to https://DAGsHub.com/blog

https://dagshub.com/blog/how-to-compare-ml-experiment-tracking-tools-to-fit-your-data-science-workflow/

- Disclaimer: Graphic made by the DAGsHub developers
- Also see AIM (discussed last week)