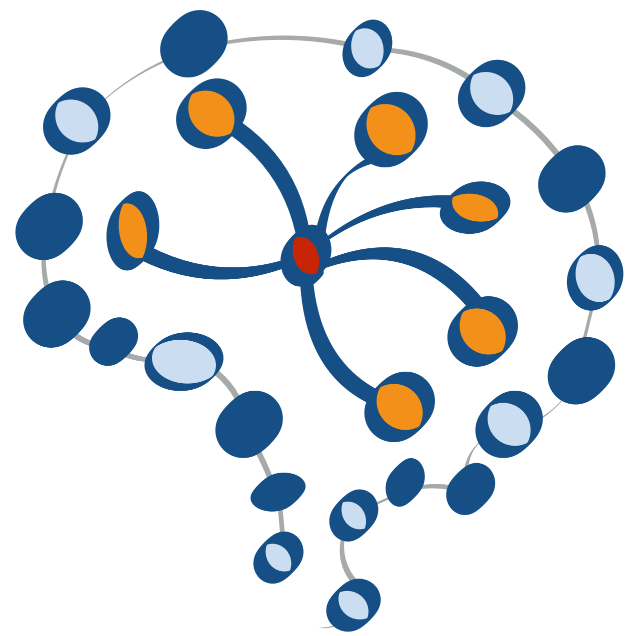
STAT 453: Introduction to Deep Learning and Generative Models

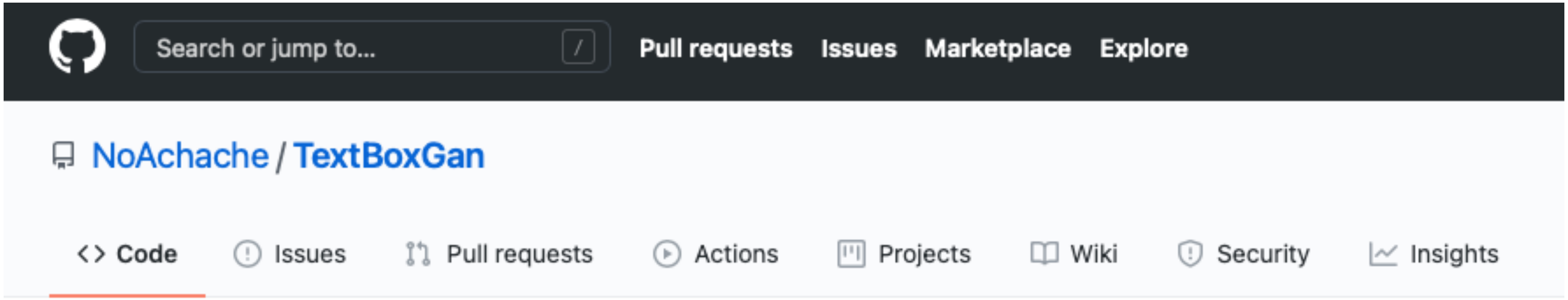
Sebastian Raschka
http://stat.wisc.edu/~sraschka



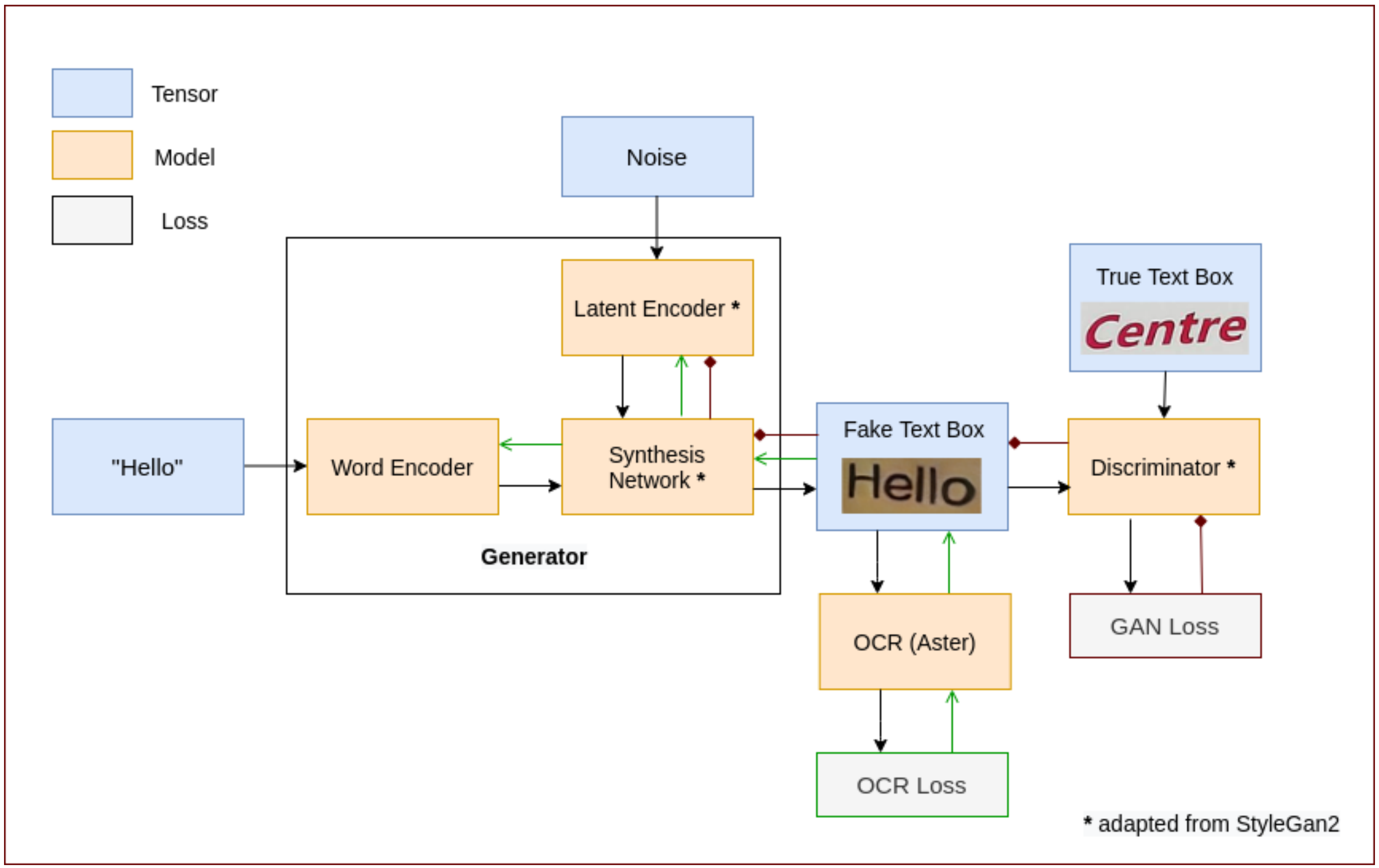
Deep Learning & AI News #7

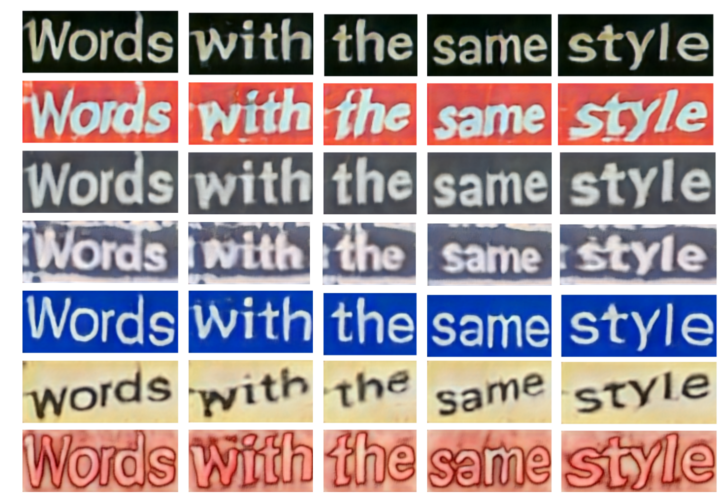
Interesting Things Related to Deep Learning

Mar 13th, 2021



https://github.com/NoAchache/TextBoxGan





arXiv.org > cs > arXiv:2103.00823

Search...

Help | Advanced

Computer Science > Computation and Language

[Submitted on 1 Mar 2021 (v1), last revised 2 Mar 2021 (this version, v2)]

M6: A Chinese Multimodal Pretrainer

Junyang Lin, Rui Men, An Yang, Chang Zhou, Ming Ding, Yichang Zhang, Peng Wang, Ang Wang, Le Jiang, Xianyan Jia, Jie Zhang, Jianwei Zhang, Xu Zou, Zhikang Li, Xiaodong Deng, Jie Liu, Jinbao Xue, Huiling Zhou, Jianxin Ma, Jin Yu, Yong Li, Wei Lin, Jingren Zhou, Jie Tang, Hongxia Yang

https://arxiv.org/abs/2103.00823

- Collaboration between Alibaba and Tsinghua University
- 1.9 Tb of images + 292 Gb of text
- Chinese, not English
- Trained 10 and 100 billion parameter transformers
- Pre-trained model can be used for many tasks: generating descriptions, image search, question answering, poem generation etc.

Image	Source & Text
	*Source:Encyclopedia* 广东草龟是属于曲颈龟亚目龟科的一种草龟。又称黑颈乌龟。 The Guangdong tortoise is a kind of tortoise belonging to Cryptodira. Also known as black-necked turtle.
	*Source:Crawled Webpages* 根据之前信息，马斯克称Cybertruck将配备三种动力版本，其中包括单电机后驱，双电机后驱和三电机全驱版本。 According to previous information, Musk said that Cybertruck will be equipped with three power versions, including a single-motor rear drive, a dual-motor rear drive and a three-motor full-drive version.
	*Source:E-commerce* 柔软的针织面料就能给人一种舒服的感觉，大篇幅的印花以点缀的作用让整体显得更加青春阳光，宽松简约落肩尽显时尚风范，十分适合日常穿搭。 The soft knitted fabric can give people a comfortable feeling. The large-length prints make the whole look more youthful and sunny with the effect of embellishment. The loose and simple shoulders show a fashionable style, which is very suitable for daily wear.

arXiv.org > cs > arXiv:2103.00823

Search...

Help | Advanced

Computer Science > Computation and Language

[Submitted on 1 Mar 2021 (v1), last revised 2 Mar 2021 (this version, v2)]

M6: A Chinese Multimodal Pretrainer

Junyang Lin, Rui Men, An Yang, Chang Zhou, Ming Ding, Yichang Zhang, Peng Wang, Ang Wang, Le Jiang, Xianyan Jia, Jie Zhang, Jianwei Zhang, Xu Zou, Zhikang Li, Xiaodong Deng, Jie Liu, Jinbao Xue, Huiling Zhou, Jianxin Ma, Jin Yu, Yong Li, Wei Lin, Jingren Zhou, Jie Tang, Hongxia Yang

https://arxiv.org/abs/2103.00823

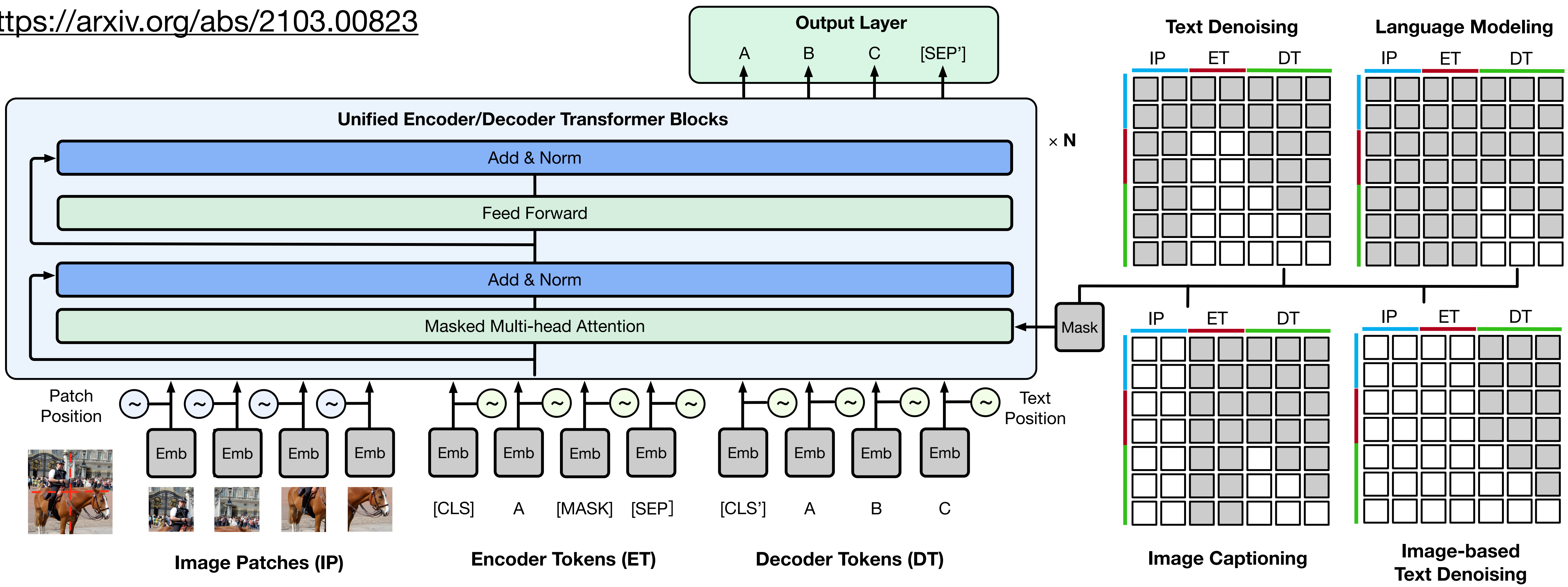


Figure 3: An overview of the pretraining tasks for M6. The design of masking strategies allows the learning of different tasks under the same framework. M6 is pretrained with image-based text denoising, image captioning, text denoising, and language modeling.

Phantom of the **ADAS**:
Phantom Attacks on Driver-Assistance Systems

Ben Nassi[1], Dudi Nassi[1], Raz Ben-Netanel[1], Yisroel Mirsky[1,2], Oleg Drokin[3], Yuval Elovici[1]

Video Demonstration - https://youtu.be/1cSw4fXYqWI
{nassib,nassid,razx,yisroel,elovici}@post.bgu.ac.il, green@linuxhacker.ru
[1] Ben-Gurion University of the Negev, [2] Georgia Tech,[3] Independent Tesla Researcher

ABSTRACT

The absence of deployed vehicular communication systems, which prevents the advanced driving assistance systems (ADASs) and autopilots of semi/fully autonomous cars to validate their virtual perception regarding the physical environment surrounding the car with a third party, has been exploited in various attacks suggested by researchers. Since the application of these attacks comes with a cost (exposure of the attacker's identity), the delicate exposure vs. application balance has held, and attacks of this kind have not yet been encountered in the wild. In this paper, we investigate a new perceptual challenge that causes the ADASs and autopilots of semi/fully autonomous to consider depthless objects (phantoms) as real. We show how attackers can exploit this perceptual challenge to apply phantom attacks and change the abovementioned balance, without the need to physically

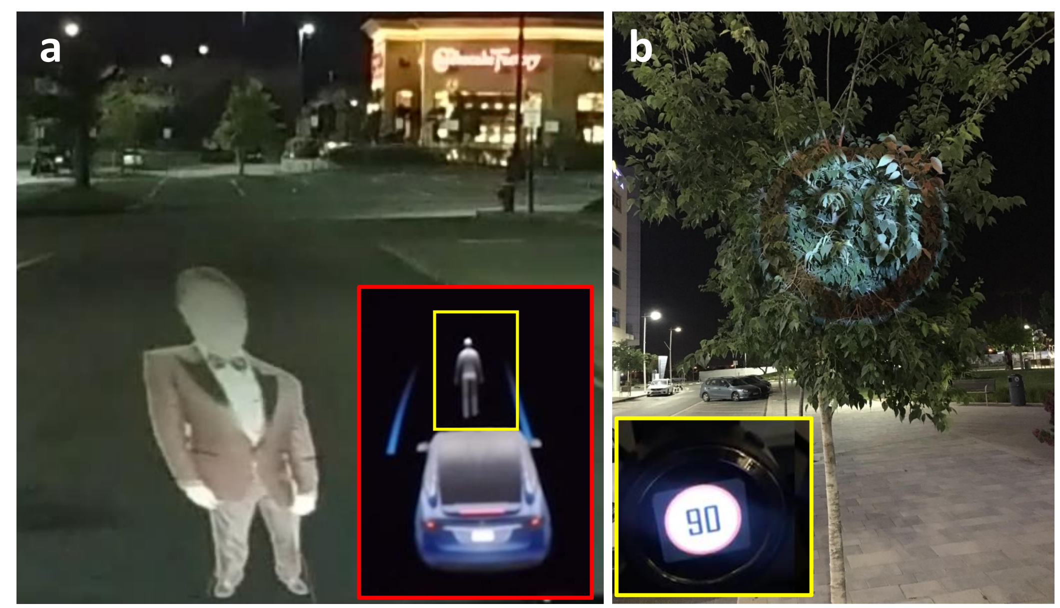


Fig. 1: Perceptual Challenge: Would you consider the projection of the person (a) and road sign (b) real? Telsa considers (a) a real person and Mobileye 630 PRO considers (b) a real road sign.

https://eprint.iacr.org/2020/085.pdf

https://arxiv.org/abs/2103.06504

Adversarial Laser Beam: Effective Physical-World Attack to DNNs in a Blink

Ranjie Duan[1†] Xiaofeng Mao[2] A. K. Qin[1] Yun Yang[1] Yuefeng Chen[2] Shaokai Ye[3] Yuan He[2]
[1]Swinburne University of Technology [2]Alibaba Group
[3]EPFL

Abstract

Though it is well known that the performance of deep neural networks (DNNs) degrades under certain light conditions, there exists no study on the threats of light beams emitted from some physical source as adversarial attacker on DNNs in a real-world scenario. In this work, we show by simply using a laser beam that DNNs are easily fooled. To this end, we propose a novel attack method called Adversarial Laser Beam ($AdvLB$), which enables manipulation of laser beam's physical parameters to perform adversarial attack. Experiments demonstrate the effectiveness of our proposed approach in both digital- and physical-settings. We further empirically analyze the evaluation results and reveal that the proposed laser beam attack may lead to some interesting prediction errors of the state-of-the-art DNNs. We envisage that the proposed $AdvLB$ method enriches the current family of adversarial attacks and builds the founda-

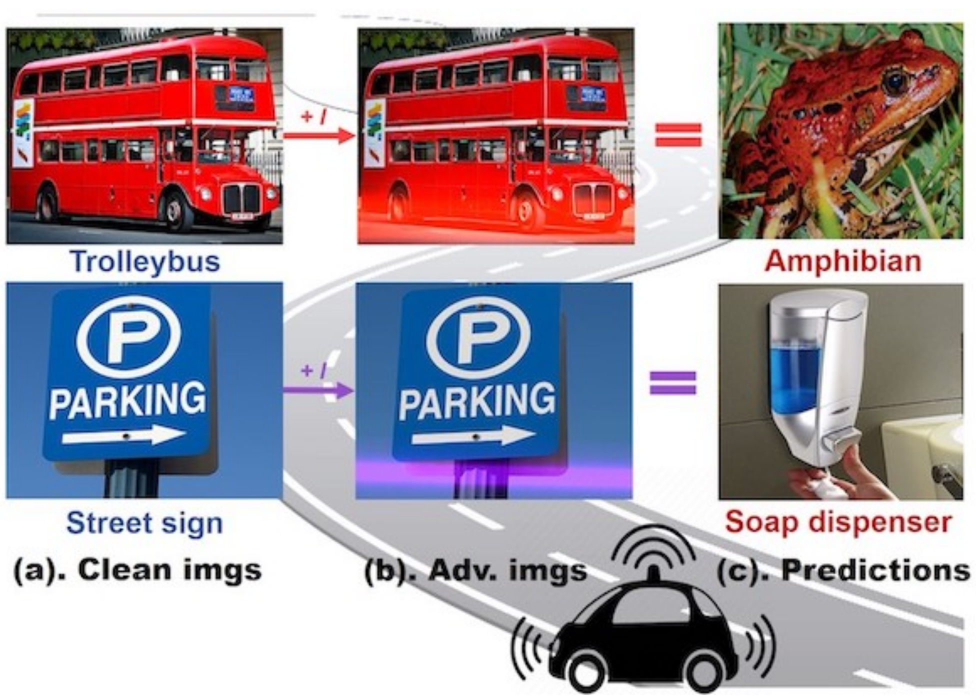


Figure 1: **An example.** When the camera of self-driving car captures object shot by the laser beam, it recognizes "trolleybus" as "amphibian" and "street sign" as "soap dispenser".

RESEARCH

Fairness On The Ground: Applying Algorithmic Fairness Approaches To Complex Production Systems

March 11, 2021

https://ai.facebook.com/research/publications/applying-algorithmic-fairness-approaches-to-production-systems

5.2 Label fairness

As with most approaches to fair supervised learning, the approach described in the previous section assumes the outcome being predicted (Y) is measured accurately in the data used to assess the system. There are some cases where this assumption is reasonable—for example, websites can perfectly measure whether users click on a given button. However, in many cases, such as identifying bullying, the labels themselves are generated through human judgement, and may thus embed human biases. This is of concern for at least three reasons. First, accurate labels are needed to compute most model fairness metrics, including the metric in Section 5.1. Second, supervised learning systems trained on biased labels will learn those biases. Finally, labelers' decisions might be used to directly intervene in the system. In this section, we describe how our high level fairness approach can be applied to assess human decision making, in the case where decisions can be compared to a ground truth.

In our bullying and harassment example, the decision being made by human labelers is whether a given post violates a bullying policy as written. These decisions won't always be correct—labelers may misunderstand the policy or the post, make a mistake, or be misled by implicit or explicit biases. To track these errors, we also collect (for a subset of posts) the judgement of an expert in applying the written policy, whose decisions provide the ground truth for each

A New Lens on Understanding Generalization in Deep Learning

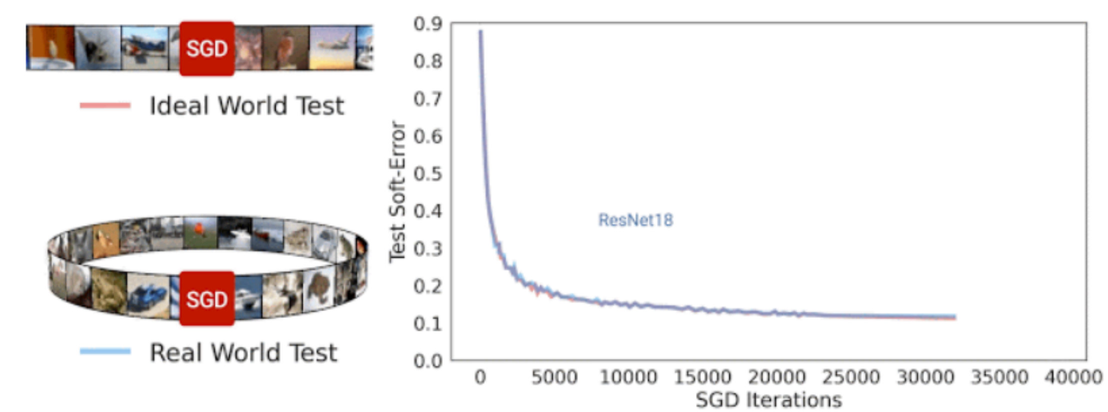
Wednesday, March 10, 2021

Hanie Sedghi, Google Research and Preetum Nakkiran, Harvard University

https://ai.googleblog.com/2021/03/a-new-lens-on-understanding.html

The Deep Bootstrap Framework: Good Online Learners are Good Offline Generalizers: https://arxiv.org/abs/2010.08127

- *Real World (N, T):* Train a model on *N* train samples from a distribution, for *T* minibatch stochastic gradient descent (SGD) steps, re-using the same *N* samples in multiple epochs, as usual. This corresponds to running SGD on the *empirical loss* (loss on training data), and is the standard training procedure in supervised learning.
- *Ideal World (T):* Train the same model for *T* steps, but use fresh samples from the distribution in each SGD step. That is, we run the exact same training code (same optimizer, learning-rates, batch-size, etc.), but sample a fresh train set in each epoch instead of reusing samples. In this ideal world setting, with an effectively infinite "train set", there is no difference between train error and test error.

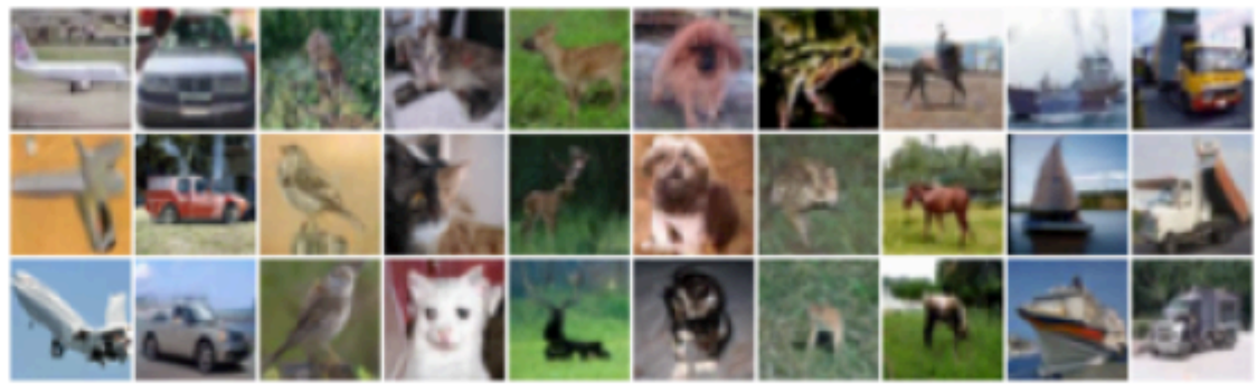


Test soft-error for ideal world and real world during SGD iterations for ResNet-18 architecture. We see that the two errors are similar.

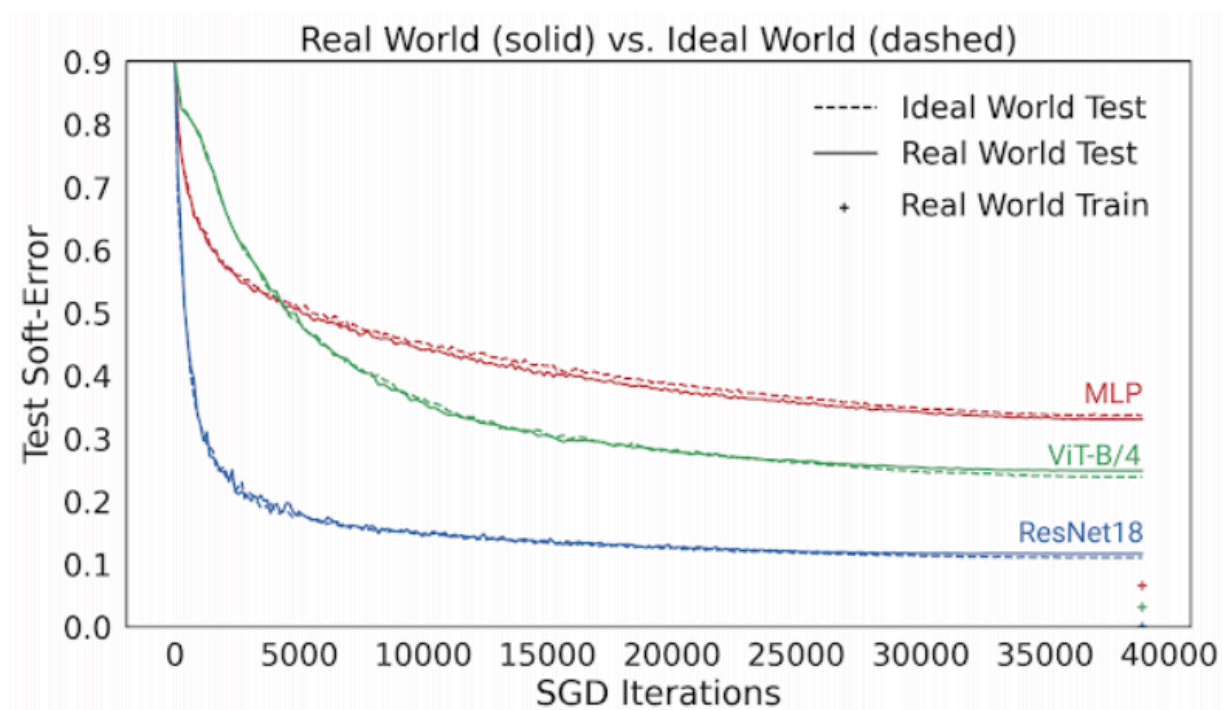
CIFAR-5m. We construct a dataset of 6 million synthetic CIFAR-10-like images, by sampling from the CIFAR-10 Denoising Diffusion generative model of Ho et al. (2020), and labeling the unconditional samples by a 98.5% accurate Big-Transfer model (Kolesnikov et al., 2019). These are

We now claim that for all t until the Real World converges, these two models f_t, f_t^{iid} have similar test performance. In our main claims, we differ slightly from the presentation in the Introduction in that we consider the "soft-error" of classifiers instead of their hard-errors. The soft-accuracy of classifiers is defined as the softmax probability on the correct label, and (soft-error) $:= 1 -$ (soft-accuracy).

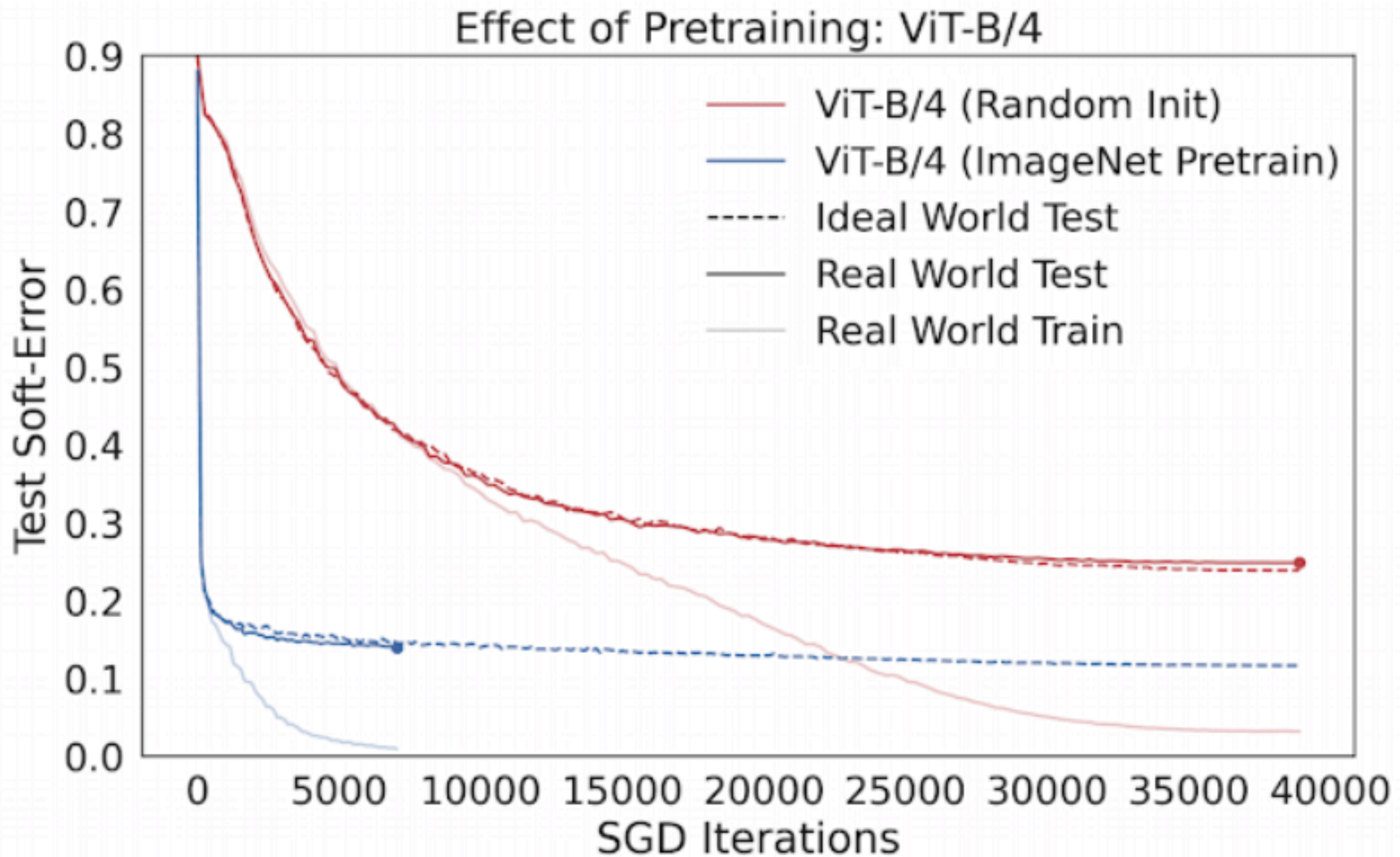
In order to quantify this observation, we simulated an ideal world setting by creating a new dataset, which we call CIFAR-5m. We trained a generative model on CIFAR-10, which we then used to generate ~6 million images. The scale of the dataset was chosen to ensure that it is "virtually infinite" from the model's perspective, so that the model never resamples the same data. That is, in the ideal world, the model sees an entirely fresh set of samples.



Samples from CIFAR-5m



The real world model is trained on 50K samples for 100 epochs, and the ideal world model is trained on 5M samples for a single epoch. The lines show the test error vs. the number of SGD steps.



Effect of pre-training — pre-trained ViTs optimize faster in the ideal world.

RESEARCH | COMPUTER VISION

SEER: The start of a more powerful, flexible, and accessible era for computer vision

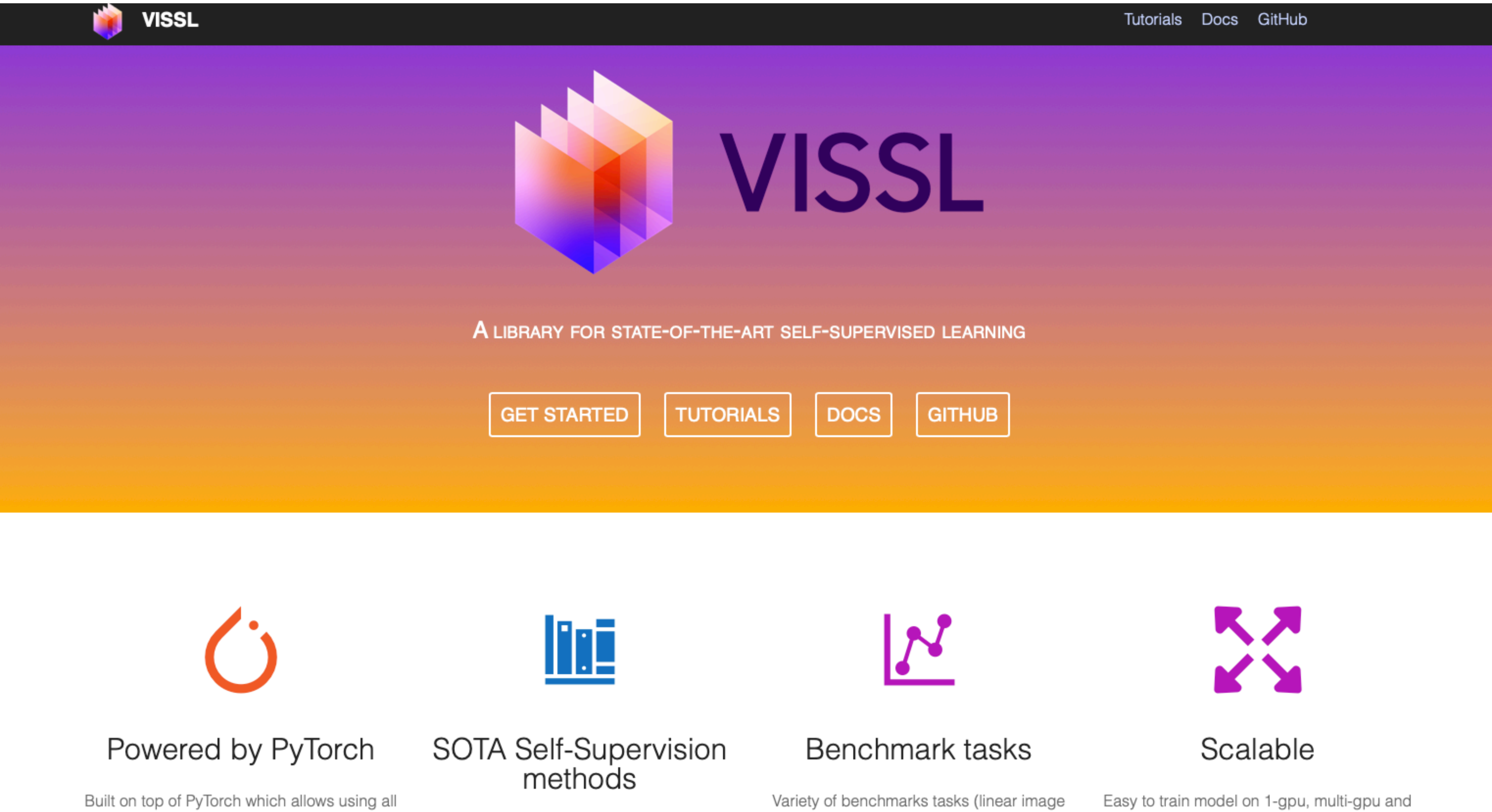
March 4, 2021

https://ai.facebook.com/blog/seer-the-start-of-a-more-powerful-flexible-and-accessible-era-for-computer-vision

Self-supervised Pretraining of Visual Features in the Wild: https://arxiv.org/abs/2103.01988

- SEER = SElf-supERvised
- new billion-parameter self-supervised computer vision model
- pretraining on a billion random, unlabeled and uncurated public Instagram images
- self-supervised SOTA: reaching 84.2 percent top-1 accuracy on ImageNet
- SwAV (https://arxiv.org/abs/2006.09882) uses online clustering to rapidly group images with similar visual concepts and leverage their similarities (doesn't need pair-wise comparisons; fast)

https://arxiv.org/abs/2006.09882



https://vissl.ai