STAT 453: Introduction to Deep Learning and Generative Models

Sebastian Raschka

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

Deep Learning & AI News #8

Interesting Things Related to Deep Learning

Mar 20th, 2021

Do you like math?

Metamath Theorem Proving

To prove: $n \in \mathbb{N} \wedge \frac{n+1}{2} \in \mathbb{N} \implies \exists m \in \mathbb{N} : n = 2m + 1$.
GPT-f's generated proof:

```
|- ((N e. NN0 /\ ((N + 1)/2) e. NN0) ->
   ((N - 1) / 2) e. NN0)
|- (N e. NN0 -> N e. CC)
|- 1 e. CC
|- ((N e. CC /\ 1 e. CC) ->
   (N - 1) e. CC )
```

$\vdots$

DeepMind Mathematics Dataset

```
Problem: Divide 1136975704 by -142121963
Answer: -8
Problem: Calculate ((-2)/3)/(-1-(-24)/9)
Answer: -2/5
Problem: Let k(u) = u**2+u-4. Find k(0)
Answer: -4
Problem: Sort 2, 4, 0, 6
Answer: 0, 2, 4, 6
Problem: Solve 4 - 4 - 4 = 188*m for m
Answer: -1/47
```

MATH Dataset (Ours)

Problem: Tom has a red marble, a green marble, a blue marble, and three identical yellow marbles. How many different groups of two marbles can Tom choose?
Solution: There are two cases here: either Tom chooses two yellow marbles (1 result), or he chooses two marbles of different colors ($\binom{4}{2} = 6$ results). The total number of distinct pairs of marbles Tom can choose is $1 + 6 = \boxed{7}$.

Problem: If $\sum_{n=0}^{\infty} \cos^{2n} \theta = 5$, what is $\cos 2\theta$?
Solution: This geometric series is $1 + \cos^2 \theta + \cos^4 \theta + \cdots = \frac{1}{1-\cos^2 \theta} = 5$. Hence, $\cos^2 \theta = \frac{4}{5}$. Then $\cos 2\theta = 2\cos^2 \theta - 1 = \boxed{\frac{3}{5}}$.

Problem: The equation $x^2 + 2x = i$ has two complex solutions. Determine the product of their real parts.
Solution: Complete the square by adding 1 to each side. Then $(x+1)^2 = 1 + i = e^{\frac{i\pi}{4}}\sqrt{2}$, so $x + 1 = \pm e^{\frac{i\pi}{8}}\sqrt[4]{2}$. The desired product is then $\left(-1 + \cos\left(\frac{\pi}{8}\right)\sqrt[4]{2}\right)\left(-1 - \cos\left(\frac{\pi}{8}\right)\sqrt[4]{2}\right) = 1 - \cos^2\left(\frac{\pi}{8}\right)\sqrt{2} = 1 - \frac{\left(1+\cos\left(\frac{\pi}{4}\right)\right)}{2}\sqrt{2} = \boxed{\frac{1-\sqrt{2}}{2}}$.

Hendrycks D, Burns C, Kadavath S, Arora A, Basart S, Tang E, Song D, Steinhardt J. Measuring Mathematical Problem Solving With the MATH Dataset. arXiv preprint arXiv:2103.03874. 2021

Computer Science > Machine Learning

[Submitted on 5 Mar 2021]

Measuring Mathematical Problem Solving With the MATH Dataset

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, Jacob Steinhardt

Paper: https://arxiv.org/abs/2103.03874
Dataset & PyTorch DataLoaders: https://github.com/hendrycks/math/

"MATH, a dataset of math problems that contemporary Transformer-based models can't solve (yet)."

- Consists of 12,500 problems from high school math competitions
- Plus a ~100,000 Khan Academy solutions with step-by-step solutions
- Plus 5 millions problems generated via Mathematica

MATH Dataset (Ours)

Problem: Tom has a red marble, a green marble, a blue marble, and three identical yellow marbles. How many different groups of two marbles can Tom choose?
Solution: There are two cases here: either Tom chooses two yellow marbles (1 result), or he chooses two marbles of different colors ($\binom{4}{2} = 6$ results). The total number of distinct pairs of marbles Tom can choose is $1 + 6 = \boxed{7}$.
Problem: If $\sum_{n=0}^{\infty} \cos^{2n} \theta = 5$, what is $\cos 2\theta$?
Solution: This geometric series is $1 + \cos^2 \theta + \cos^4 \theta + \cdots = \frac{1}{1-\cos^2 \theta} = 5$. Hence, $\cos^2 \theta = \frac{4}{5}$. Then $\cos 2\theta = 2\cos^2 \theta - 1 = \boxed{\frac{3}{5}}$.
Problem: The equation $x^2 + 2x = i$ has two complex solutions. Determine the product of their real parts.
Solution: Complete the square by adding 1 to each side. Then $(x+1)^2 = 1 + i = e^{\frac{i\pi}{4}}\sqrt{2}$, so $x + 1 = \pm e^{\frac{i\pi}{8}}\sqrt[4]{2}$. The desired product is then $\left(-1 + \cos\left(\frac{\pi}{8}\right)\sqrt[4]{2}\right)\left(-1 - \cos\left(\frac{\pi}{8}\right)\sqrt[4]{2}\right) = 1 - \cos^2\left(\frac{\pi}{8}\right)\sqrt{2} = 1 - \frac{\left(1+\cos\left(\frac{\pi}{4}\right)\right)}{2}\sqrt{2} = \boxed{\frac{1-\sqrt{2}}{2}}$.

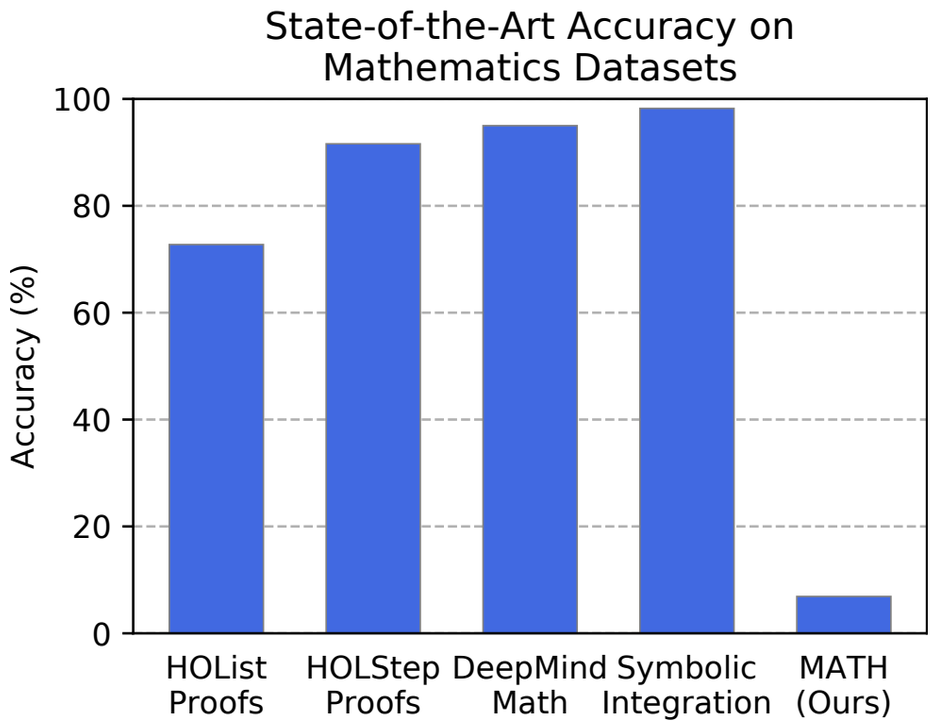


Figure 2: Compared to existing proof and plug-and-chug tasks, our mathematical problem solving task is considerably more challenging. HOList results are from Wu et al. (2021). HOLStep results are from Crouse et al. (2019). DeepMind Math accuracy is the median IID accuracy from Henighan et al. (2020). Symbolic Integration accuracy is from Lample and Charton (2020).

Hendrycks D, Burns C, Kadavath S, Arora A, Basart S, Tang E, Song D, Steinhardt J. Measuring Mathematical Problem Solving With the MATH Dataset. arXiv preprint arXiv:2103.03874. 2021

https://github.com/hendrycks/math/blob/main/modeling/equivalent.py

Model	Prealgebra	Algebra	Number Theory	Counting & Probability	Geometry	Intermediate Algebra	Precalculus	Average
GPT-2 (0.1B)	5.2	5.1	5.0	2.8	5.7	6.5	7.3	5.4 (+0%)
GPT-2 (0.3B)	6.7	6.6	5.5	3.8	6.9	6.0	7.1	6.2 (+15%)
GPT-2 (0.7B)	6.9	6.1	5.5	5.1	8.2	5.8	7.7	6.4 (+19%)
GPT-2 (1.5B)	8.3	6.2	4.8	5.4	8.7	6.1	8.8	6.9 (+28%)
GPT-3 (2.7B)	2.8	2.9	3.9	3.6	2.1	2.5	2.6	2.9 (−46%)
GPT-3 (175B)	7.7	6.0	4.4	4.7	3.1	4.4	4.0	5.2 (−4%)

Table 2: MATH accuracies across subjects for GPT-2 and *few-shot* GPT-3 models. The character 'B' denotes the number of parameters in billions. The gray text indicates the *relative* improvement over the 0.1B baseline. All GPT-2 models pretrain on AMPS, and all values are percentages. A 15× increase in model parameters increased accuracy by 1.5%, a 28% relative improvement. Model accuracy is increasing very slowly, so much future research is needed.

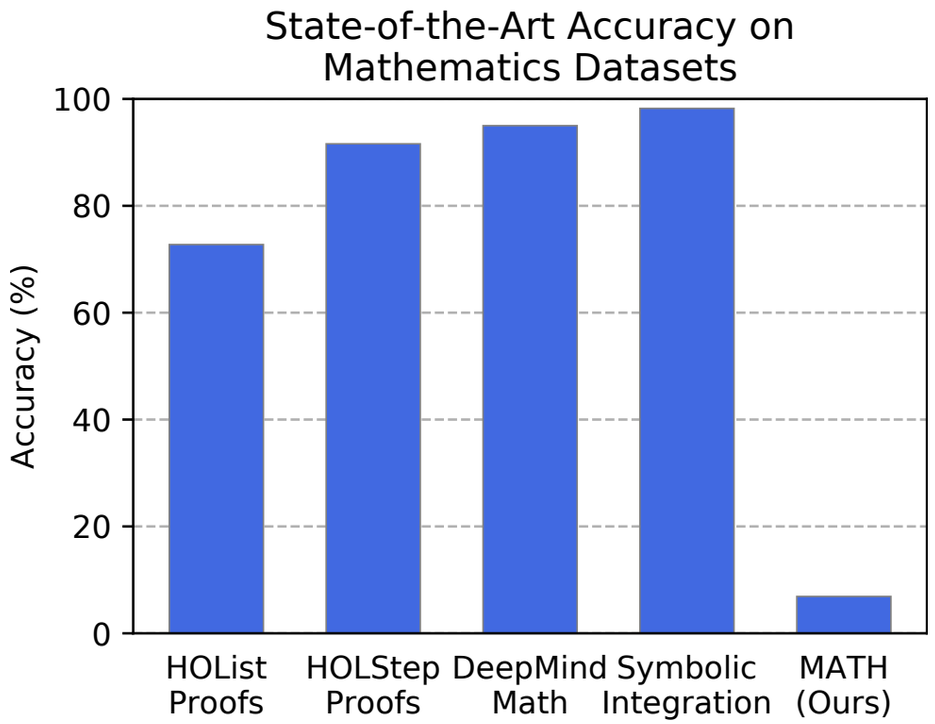


Figure 2: Compared to existing proof and plug-and-chug tasks, our mathematical problem solving task is considerably more challenging. HOList results are from Wu et al. (2021). HOLStep results are from Crouse et al. (2019). DeepMind Math accuracy is the median IID accuracy from Henighan et al. (2020). Symbolic Integration accuracy is from Lample and Charton (2020).

Hendrycks D, Burns C, Kadavath S, Arora A, Basart S, Tang E, Song D, Steinhardt J. Measuring Mathematical Problem Solving With the MATH Dataset. arXiv preprint arXiv:2103.03874. 2021

We also evaluated humans on MATH, and found that a computer science PhD student who does not especially like mathematics attained approximately 40% on MATH, while a three-time IMO gold medalist attained 90%, indicating that MATH can be challenging for humans as well.

Model	Prealgebra	Algebra	Number Theory	Counting & Probability	Geometry	Intermediate Algebra	Precalculus	Average
GPT-2 (0.1B)	5.2	5.1	5.0	2.8	5.7	6.5	7.3	5.4 (+0%)
GPT-2 (0.3B)	6.7	6.6	5.5	3.8	6.9	6.0	7.1	6.2 (+15%)
GPT-2 (0.7B)	6.9	6.1	5.5	5.1	8.2	5.8	7.7	6.4 (+19%)
GPT-2 (1.5B)	8.3	6.2	4.8	5.4	8.7	6.1	8.8	6.9 (+28%)
GPT-3 (2.7B)	2.8	2.9	3.9	3.6	2.1	2.5	2.6	2.9 (−46%)
GPT-3 (175B)	7.7	6.0	4.4	4.7	3.1	4.4	4.0	5.2 (−4%)

Table 2: MATH accuracies across subjects for GPT-2 and *few-shot* GPT-3 models. The character 'B' denotes the number of parameters in billions. The gray text indicates the *relative* improvement over the 0.1B baseline. All GPT-2 models pretrain on AMPS, and all values are percentages. A 15× increase in model parameters increased accuracy by 1.5%, a 28% relative improvement. Model accuracy is increasing very slowly, so much future research is needed.

Revisiting Data Augmentation

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 9 Mar 2021]

Thumbnail: A Novel Data Augmentation for Convolutional Neural Network

Tianshu Xie, Xuan Cheng, Minghui Liu, Jiali Deng, Xiaomin Wang, Ming Liu

Pasting a smaller version of the image inside itself

https://arxiv.org/abs/2103.05342

(a) Original Sample

(b) Cutout

(c) Mixup

(d) CutMix

(e) Thumbnail

(a) Original Sample

(b) Self Thumbnail

(c) Mixed Single Thumbnail

(d) Mixed Multiple Thumbnails

(e) Other Strategy

Model	Method	Accuracy(%)
ResNet56	baseline	73.71±0.12
	+Cutout	74.64±0.15
	+Mixup	75.97±0.26
	+CutMix	76.57±0.13
	+ST (ours)	75.58±0.11
	+MST (ours)	**76.78±0.08**
WideResNet-28-10	baseline	80.90±0.06
	+Cutout	81.86±0.08
	+Mixup	82.57±0.12
	+CutMix	83.13±0.06
	+ST (ours)	80.81±0.04
	+MST (ours)	**83.35±0.05**

TABLE III: Comparison of accuracy of ResNet50 and WideResNet-28-10 on the CIFAR100 validation set. We report average over 3 runs.

Do GANs learn meaningful structural parts of objects during their attempt to reproduce those?

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 7 Mar 2021 (v1), last revised 9 Mar 2021 (this version, v2)]

Repurposing GANs for One-shot Semantic Part Segmentation

Nontawat Tritrong, Pitchaporn Rewatbowornwong, Supasorn Suwajanakorn

Paper: https://arxiv.org/abs/2103.04379

Demos: https://repurposegans.github.io

Figure 1: One-shot segmentation results. In each task, our segmentation network is given only one example of part labels.

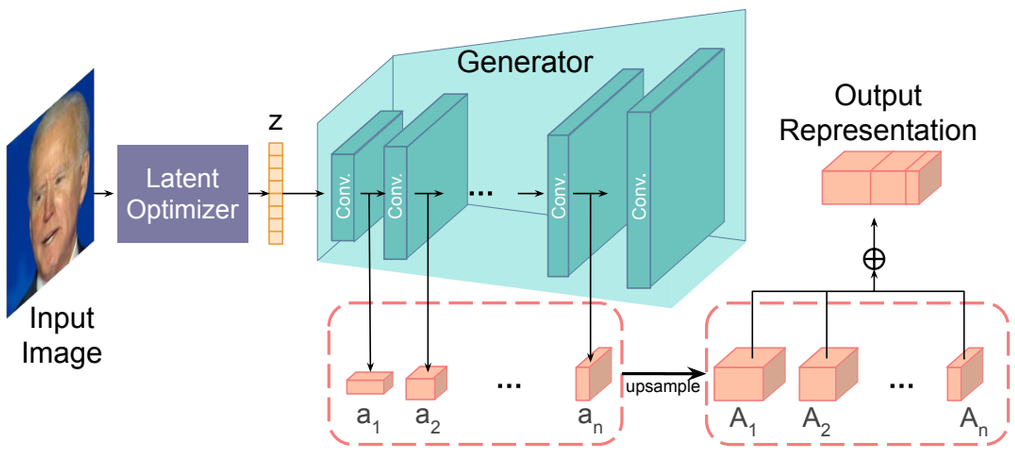


Figure 2: **Representation extraction** To extract a representation from an image, we embed the image into the latent space of GAN by optimizing for the latent z that reproduces the input image. z is then fed to the generator and we collect multiple activation maps $a_1, a_2, ..., a_n$ of dimensions $(h_1, w_1, c_1), ..., (h_n, w_n, c_n)$. Each of these maps is upsampled to A_i with dimension (h_n, w_n, c_i). The representation is a concatenation of all A_i along the channel dimension.

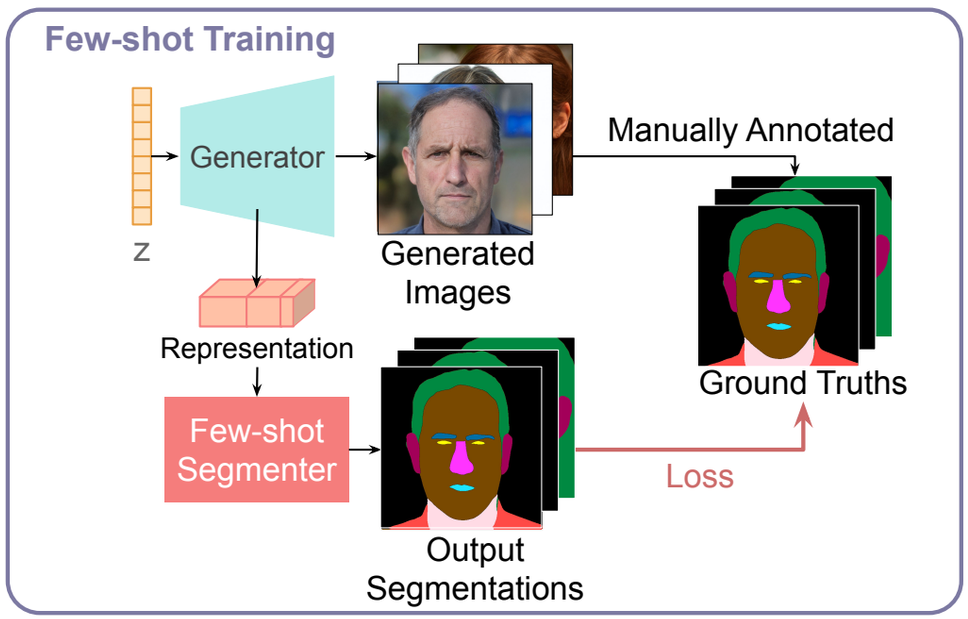


Figure 3: **Few-shot segmentation pipeline** For training, we use a trained GAN to generate a few images along with their representations by feeding random latent codes. Then, we manually annotate these images and train our few-shot segmenter to output segmentation maps that match our annotated masks. For inference, we extract a representation from a test image (Figure 2) then input it to the few-shot segmenter to obtain a segmentation map.

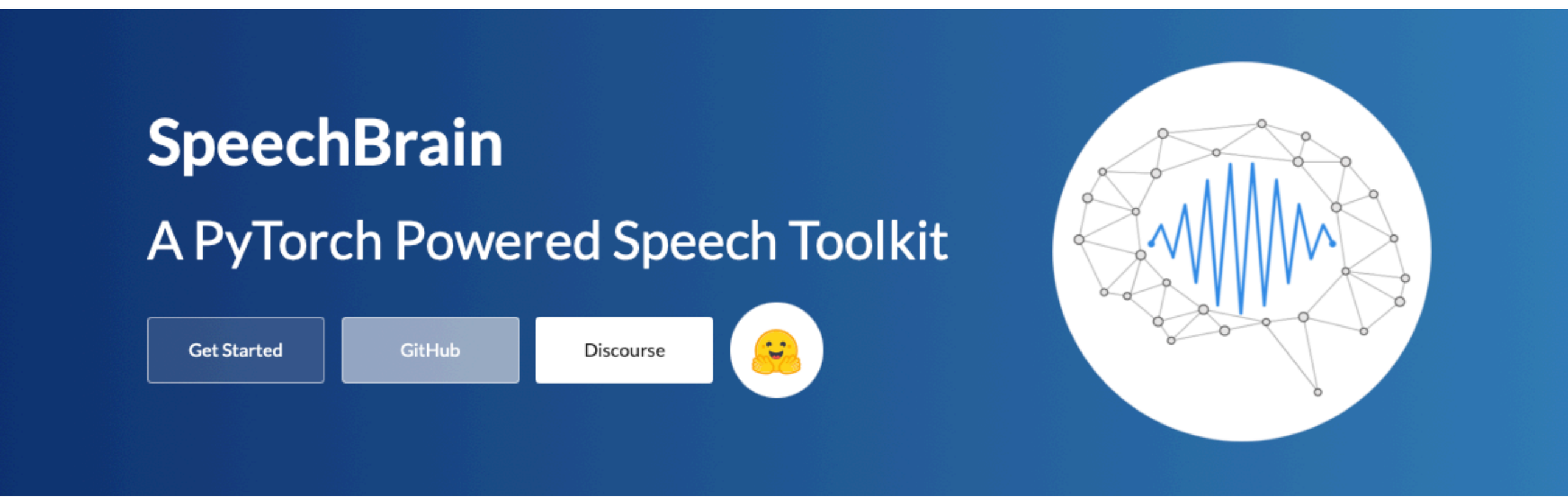
CADE METZ BACKCHANNEL 03.16.2021 07:00 AM

The Secret Auction That Set Off the Race for AI Supremacy

How the shape of deep learning—and the fate of the tech industry—went up for sale in Harrah's Room 731, on the shores of Lake Tahoe.

https://www.wired.com/story/secret-auction-race-ai-supremacy-google-microsoft-baidu/

HINTON STOPPED THE auction because finding the right home for his research was ultimately more important to him than commanding the maximum price. When he told the bidders at Google he was stopping the auction at $44 million, they thought he was joking—that he couldn't possibly give up the dollars that were still coming. He wasn't joking, and his students saw the situation much as he did. They were academics, not entrepreneurs, more loyal to their idea than to anything else.



https://speechbrain.github.io/

SpeechBrain is an open-source toolkit designed to speedup research and development of speech technologies. It is flexible, modular, easy-to-use and well documented