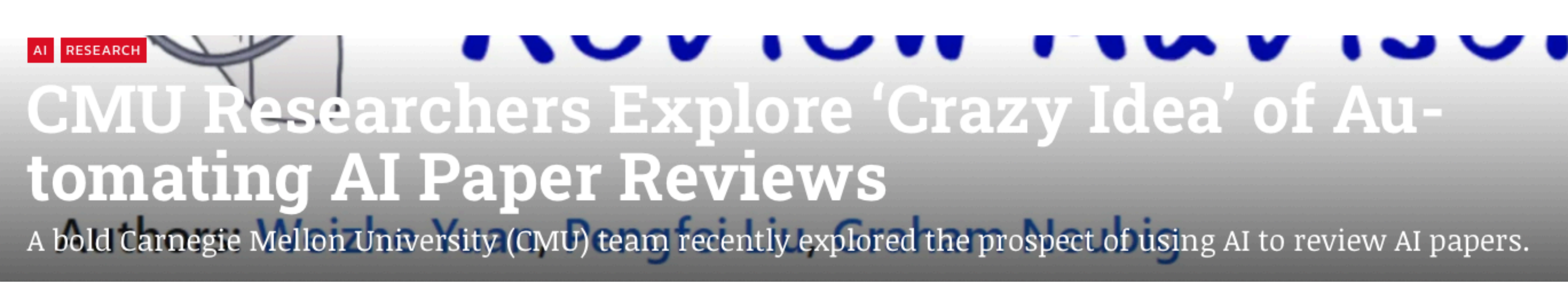
Sebastian Raschka

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

Deep Learning & AI News

Interesting Things Related to Deep Learning

Week 2: Feb 6th, 2021



https://syncedreview.com/2021/02/04/cmu-researchers-explore-crazy-idea-of-automating-ai-paper-reviews/

the most frequently mentioned qualities of a good review:

- *Decisiveness: A good review should take a clear stance, selecting high-quality submissions for publication and suggesting others not be accepted.*
- *Comprehensiveness: A good review should be well-organized, typically starting with a brief summary of the paper's contributions, then following with opinions gauging the quality of a paper from different aspects.*
- *Justification: A good review should provide specific reasons for its assessment, particularly whenever it states that the paper is lacking in some aspect.*
- *Accuracy: A review should be factually correct, with the statements contained therein not being demonstrably false.*
- *Kindness: A good review should be kind and polite in language use*

the system also tends to generate non-factual statements in its paper assessments

Uses BART, a denoising autoencoder for pretraining sequence-to-sequence models

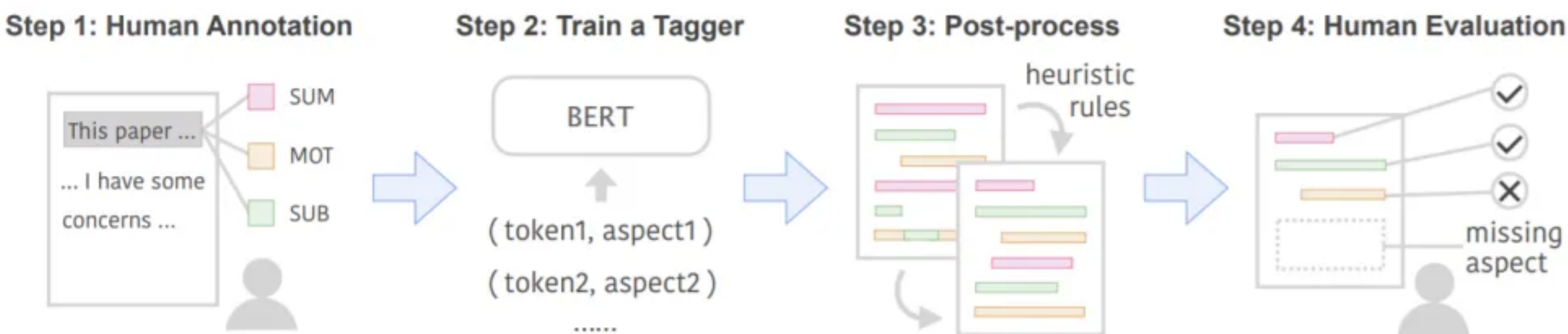


Figure 1: Data annotation pipeline.

modifications, which are *Summary* (SUM), *Motivation/Impact* (MOT) , *Originality* (ORI), *Soundness/Correctness* (SOU), *Substance* (SUB), *Replicability* (REP), *Meaningful Comparison* (CMP) and *Clarity* (CLA). The detailed elaborations of

Code: https://github.com/neulab/ReviewAdvisor

Paper: https://arxiv.org/pdf/2102.00176.pdf

You've probably never wondered what a knight made of spaghetti would look like ...

https://www.wired.com/story/ai-go-art-steering-self-driving-car/

DALL-E: a version of a 12-billion param GPT-3

inputs: images and descriptions

outputs: mashups from text input

DALL-E paper & code are not available

Original blog article (Jan 5): https://openai.com/blog/dall-e/

CLIP: Connecting Text and Images

We're introducing a neural network called CLIP which efficiently learns visual concepts from natural language supervision. CLIP can be applied to any visual classification benchmark by simply providing the names of the visual categories to be recognized, similar to the "zero-shot" capabilities of GPT-2 and GPT-3.

https://openai.com/blog/clip/

(not fine-tuned to the lower 2 datasets)

DALL-E is based on CLIP (*Contrastive Language–Image Pre-training*)

which is based on

- zero-shot transfer,
- natural language supervision,
- multimodal learning.

What: Predict novel image class (zero-shot) based on word embedding

Why: Useful for searching databases of image and video content (e.g., when developing self driving cars)

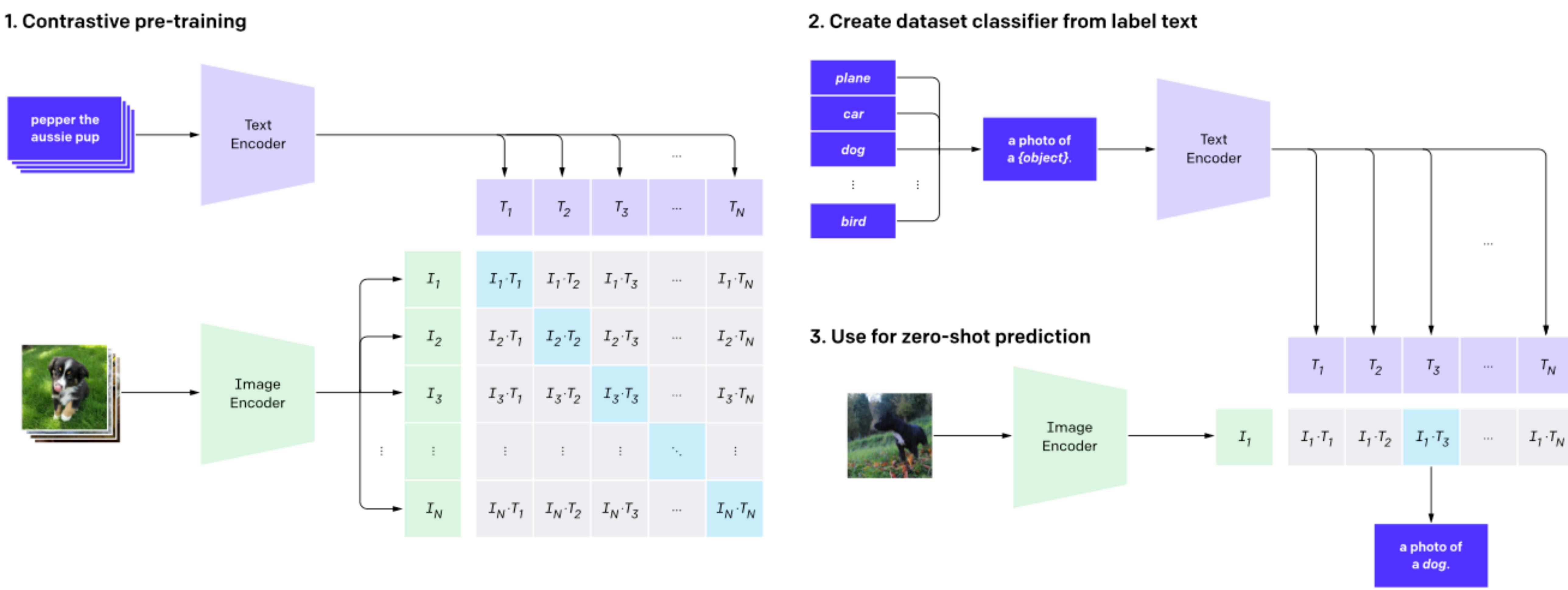
Paper: https://cdn.openai.com/papers/Learning_Transferable_Visual_Models_From_Natural_Language_Supervision.pdf

Code: https://github.com/openai/CLIP

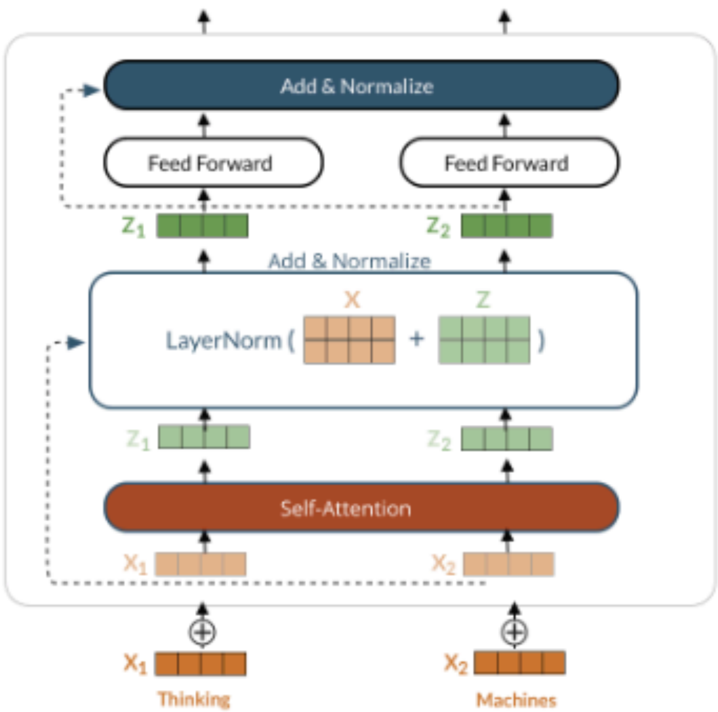
CLIP: Connecting Text and Images

We're introducing a neural network called CLIP which efficiently learns visual concepts from natural language supervision. CLIP can be applied to any visual classification benchmark by simply providing the names of the visual categories to be recognized, similar to the "zero-shot" capabilities of GPT-2 and GPT-3.

https://openai.com/blog/clip/



EleutherAI Home About Blog Publications Get Involved



GPT-Neo

GPT-Neo is the code name for a series of transformer-based language models loosely styled around the GPT architecture that we plan to train and open source. Our primary goal is to replicate a GPT-3 sized model and open source it to the public, for free.

Along the way we will be running experiments with alternative architectures and attention types, releasing any intermediate models, and writing up any findings on our blog.

Our models are built in Tensorflow-mesh, which will allow us to scale up to GPT-3 sizes and beyond using simultaneous model and data parallelism.

Progress:

- We have the bulk of the model built, GPT-2 size models trained, and several experimental architectures implemented.
- Our current codebase should be able to scale up to GPT-3 sized models

Next Steps:

- We are currently working on wrapping up GPT-2-sized model replication, looking mostly at evaluations there.
- The largest model we've gotten to train for a single step so far has been 200B parameters.

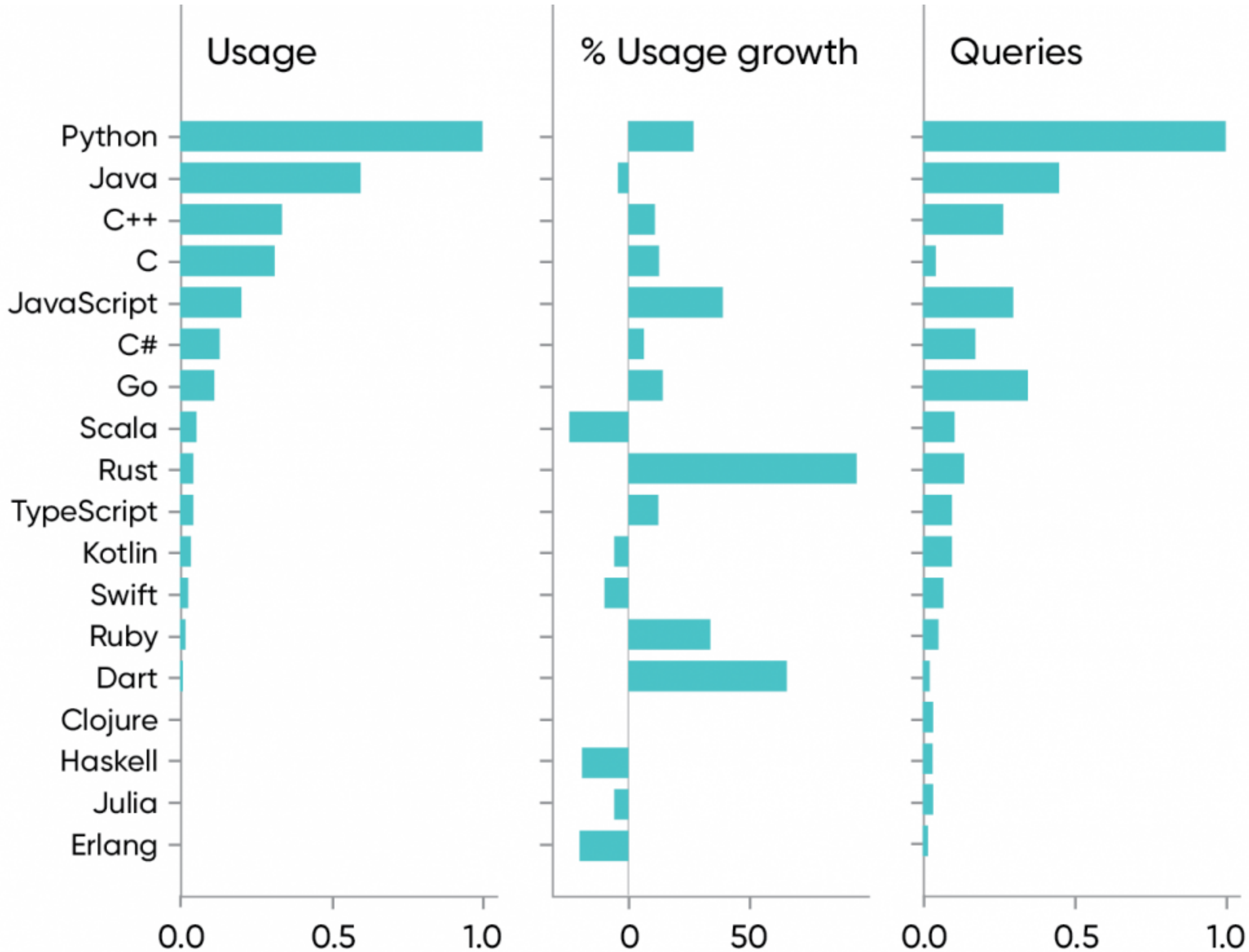
https://www.eleuther.ai/projects/gpt-neo/

Training on "The Pile," an **825 GB** language modeling dataset from various sources (YouTube, PubMed, etc.)

Radar / AI & ML

Where Programming, Ops, AI, and the Cloud are Headed in 2021

Following O'Reilly online learning trends to see what's coming next.



https://www.oreilly.com/radar/where-programming-ops-ai-and-the-cloud-are-headed-in-2021/

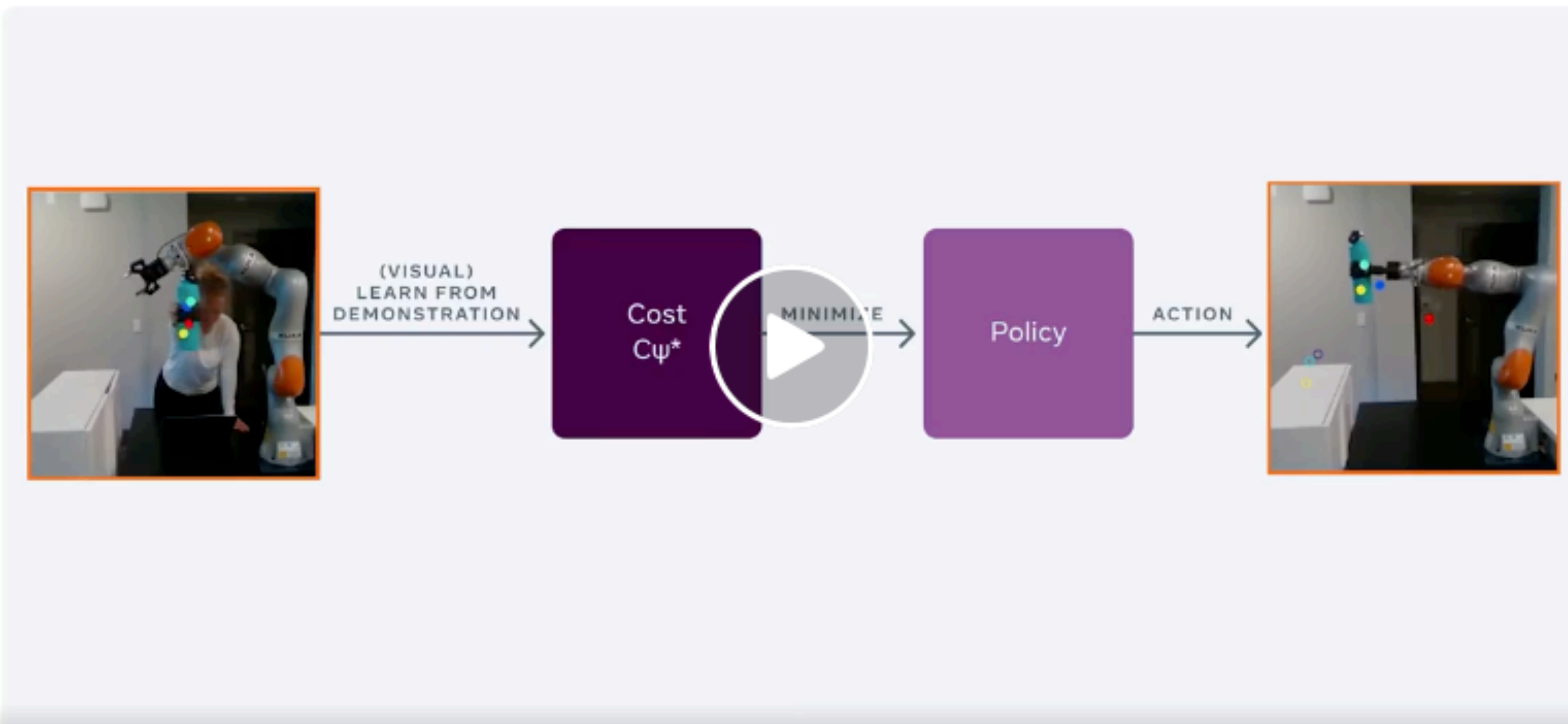
RESEARCH | OPEN SOURCE

Teaching AI to manipulate objects using visual demos

model-based inverse reinforcement learning (IRL) — using visual demonstrations on a physical robot

January 25, 2021

Training AI to manipulate objects from visual demonstrations



This work on training a visual dynamics model using self-supervision techniques provides an important test bed for us to push self-supervision forward

https://ai.facebook.com/blog/teaching-ai-to-manipulate-objects-using-visual-demos/

Code: https://github.com/facebookresearch/LearningToLearn

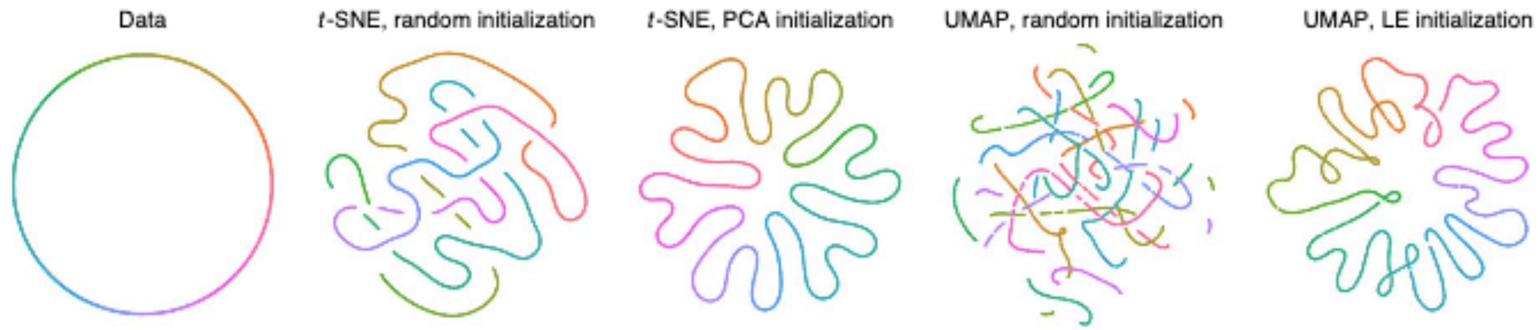


Fig. 1 | *t*-SNE and UMAP with random and non-random initialization. Embeddings of $n = 7{,}000$ points sampled from a circle with a small amount of Gaussian noise ($\sigma = r/1{,}000$, where r is the circle's radius). We used random and PCA initialization for *t*-SNE (openTSNE[11] v.0.4.4) and random and LE initialization for UMAP (v.0.4.6). All other parameters were kept as default. For this dataset, PCA and LE give the same initialization. Note that openTSNE scales PCA initialization to have s.d. = 0.0001, which is the default s.d. for random initialization in *t*-SNE[2]; similarly, UMAP scales the LE result to have a span of 20, which is the value it uses for random initialization.

Table 1 | Performance of *t*-SNE and UMAP with random and informative initialization using datasets and evaluation metrics from Becht et al.

	Preservation of pairwise distances			Reproducibility of large-scale structures[a]		
Dataset	**Samusik et al.[8]**	**Wong et al.[9]**	**Han et al.[10]**	**Samusik et al.[8]**	**Wong et al.[9]**	**Han et al.[10]**
UMAP, LE initialization	**0.70**	0.57	**0.30**	0.94	**0.98**	0.49
UMAP, random initialization	0.41	0.38	0.14	0.24	0.21	0.22
t-SNE, PCA initialization	0.59	**0.66**	0.28	**0.95**	**0.98**	**0.92**
t-SNE, random initialization	0.32	0.36	0.18	0.29	0.33	0.06

[a]For the reproducibility metric, the average over three random subsamples of size $n = 200{,}000$ is reported. Bold numbers denote the maximum value in each column.

nature biotechnology

Explore Content Journal Information Publish With Us

nature > nature biotechnology > matters arising > article

Matters Arising | Published: 01 February 2021

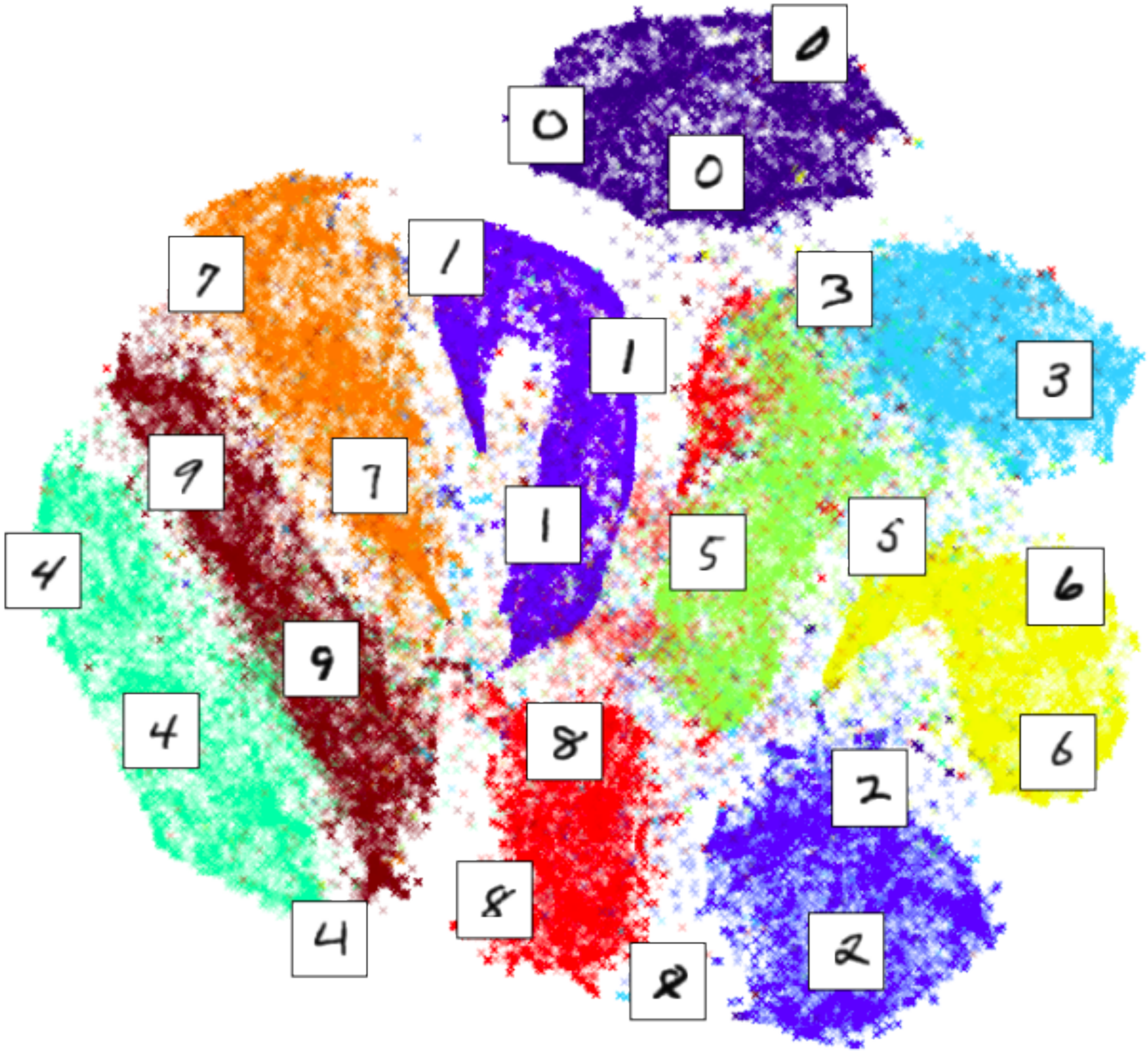
Initialization is critical for preserving global data structure in both *t*-SNE and UMAP

Dmitry Kobak & George C. Linderman

Nature Biotechnology (2021) | Cite this article

The Original Article was published on 03 December 2018

Article: https://www.nature.com/articles/s41587-020-00809-z



Source: https://nlml.github.io/in-raw-numpy/in-raw-numpy-t-sne/

Install User Guide API Examples More

Prev Up Next

scikit-learn 0.24.1
Other versions

Please **cite us** if you use the software.

`sklearn.manifold.TSNE`
Examples using `sklearn.manifold.TSNE`

`sklearn.manifold`.TSNE

class `sklearn.manifold.`**`TSNE`**(*n_components=2, *, perplexity=30.0, early_exaggeration=12.0, learning_rate=200.0, n_iter=1000, n_iter_without_progress=300, min_grad_norm=1e-07, metric='euclidean', init='random', verbose=0, random_state=None, method='barnes_hut', angle=0.5, n_jobs=None, square_distances='legacy'*) ¶ [source]

t-distributed Stochastic Neighbor Embedding.

t-SNE [1] is a tool to visualize high-dimensional data. It converts similarities between data points to joint probabilities and tries to minimize the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-

init : *{'random', 'pca'} or ndarray of shape (n_samples, n_components), default='random'*

Initialization of embedding. Possible options are 'random', 'pca', and a numpy array of shape (n_samples, n_components). PCA initialization cannot be used with precomputed distances and is usually more globally stable than random initialization.

Discussion and tips for choosing hyperparams and default values:

https://github.com/scikit-learn/scikit-learn/issues/18018