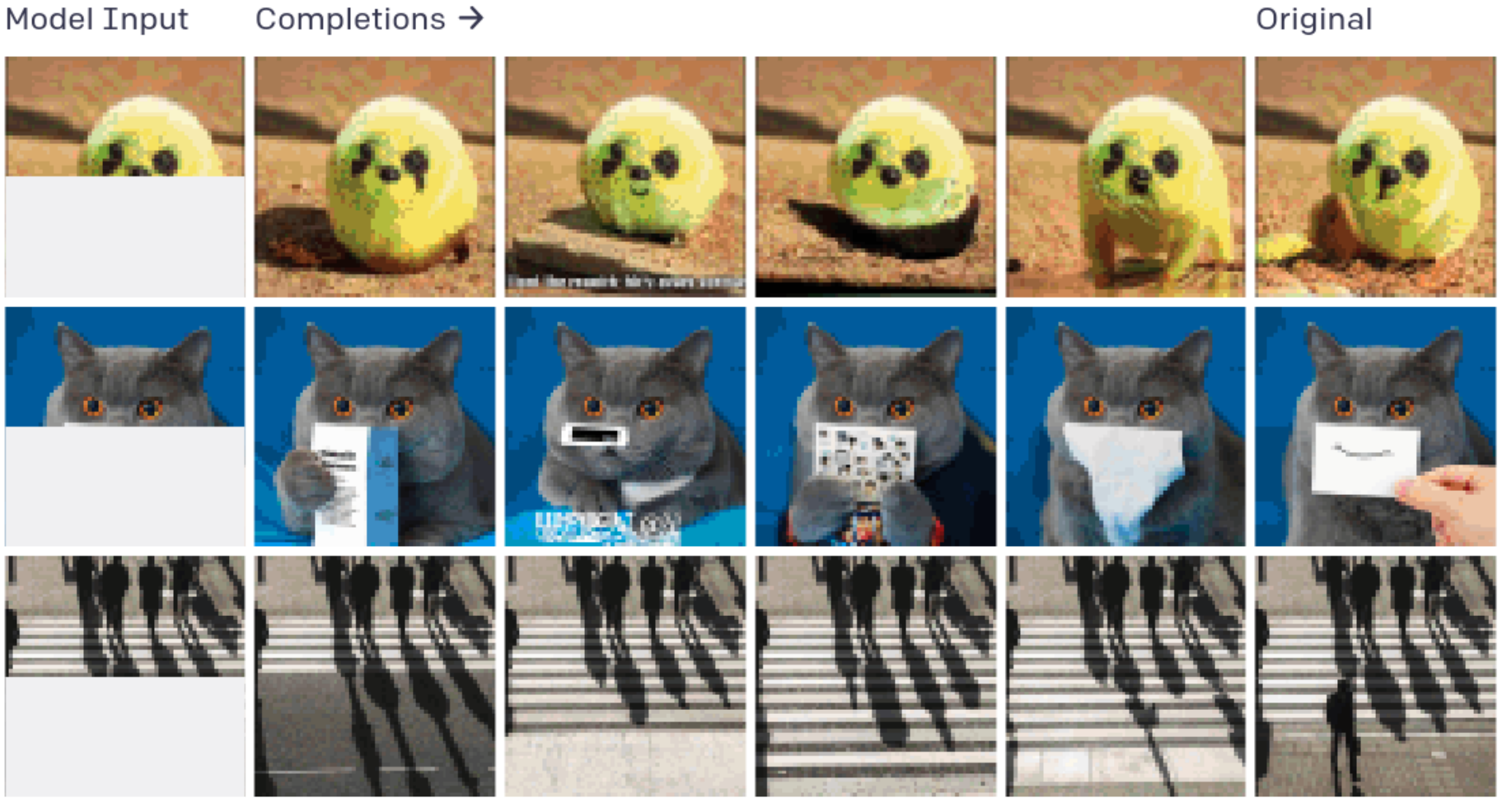
imageGPT

Generative Pretraining From Pixels

Author: Kemal Erdem

What iGPT is not?



iGPT autoregressive examples, Source: https://openai.com/blog/image-gpt/

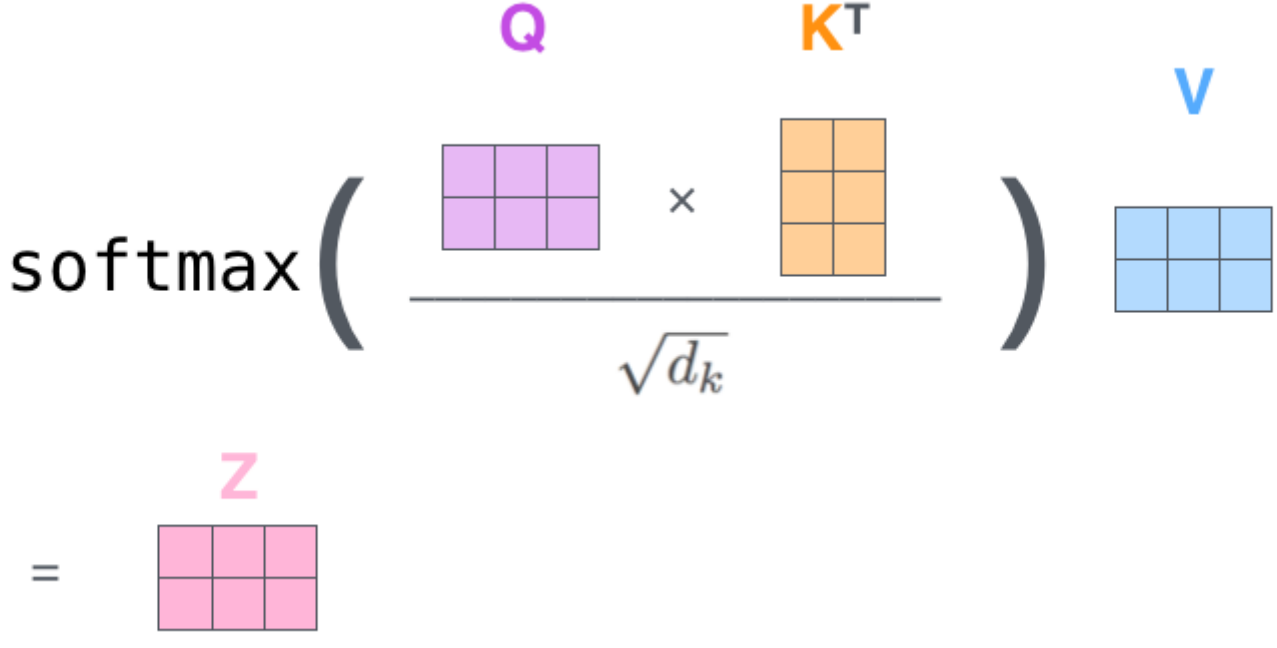
If it's not about generating images then what?

			Pre-trained on ImageNet	
Evaluation	**Model**	**Accuracy**	**w/o labels**	**w/ labels**
CIFAR-10 Linear Probe	ResNet-152[paper]	94.0		✓
	SimCLR[paper]	95.3	✓	
	iGPT-L 32x32	**96.3**	✓	
CIFAR-100 Linear Probe	ResNet-152	78.0		✓
	SimCLR	80.2	✓	
	iGPT-L 32x32	**82.8**	✓	
STL-10 Linear Probe	AMDIM-L[paper]	94.2	✓	
	iGPT-L 32x32	**95.5**	✓	
CIFAR-10 Fine-tune	AutoAugment[paper]	98.5		✓
	SimCLR	98.6	✓	
	GPipe[paper]	**99.0**		✓
	iGPT-L	**99.0**	✓	
CIFAR-100 Fine-tune	iGPT-L	88.5	✓	
	SimCLR	89.0	✓	
	AutoAugment	89.3		✓
	EfficientNet[paper]	**91.7**		✓

A comparison of linear probe and fine-tune accuracies, Source: OpenAI iGPT

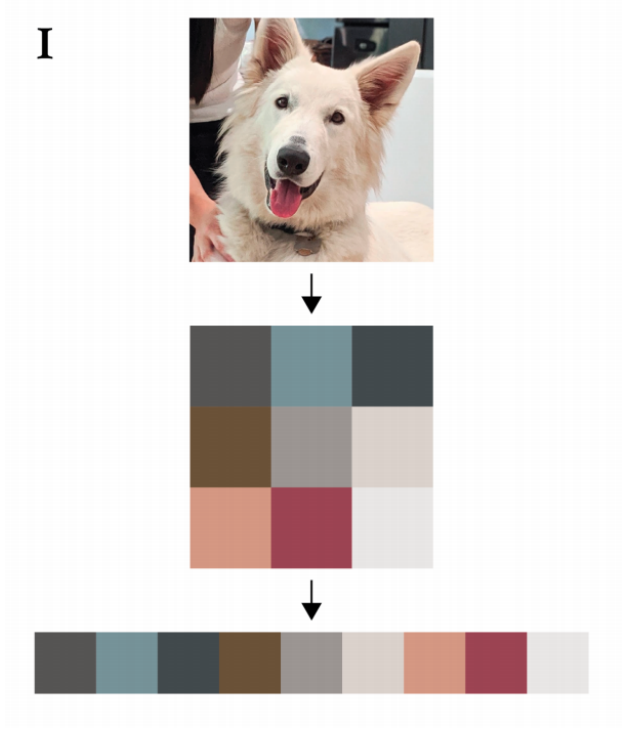
Training process (2048 TPU cores)

Problem with attending over an entire image



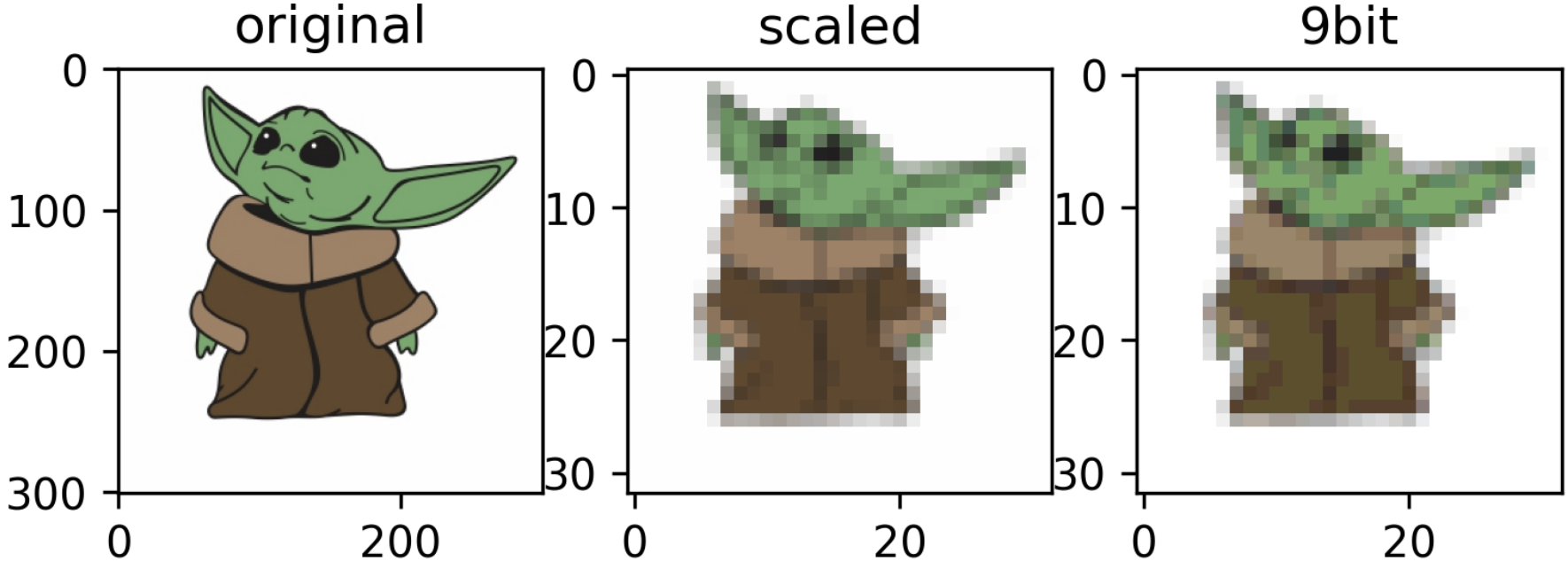
The self-attention calculation in matrix form, Source: The Illustrated Transformer by Jay Alammar

Input scaling with 9bit color palette



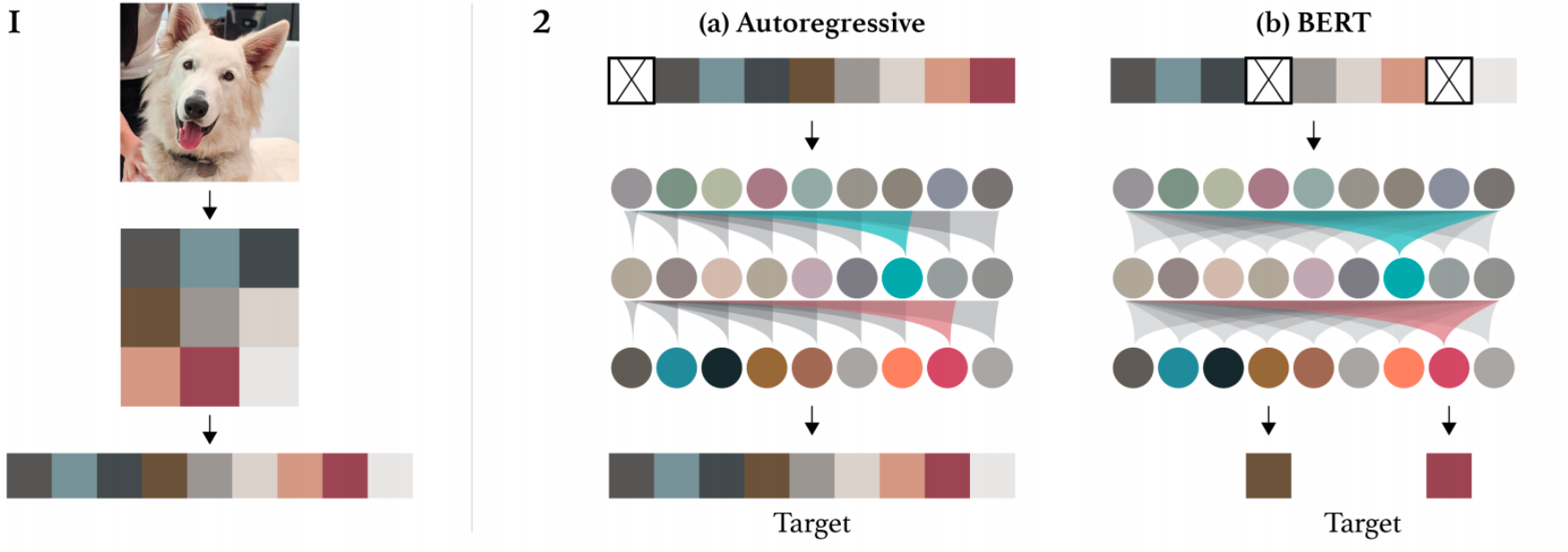
Input preprocessing, Source: Generative Pretraining From Pixels

How color encoding works in practice?



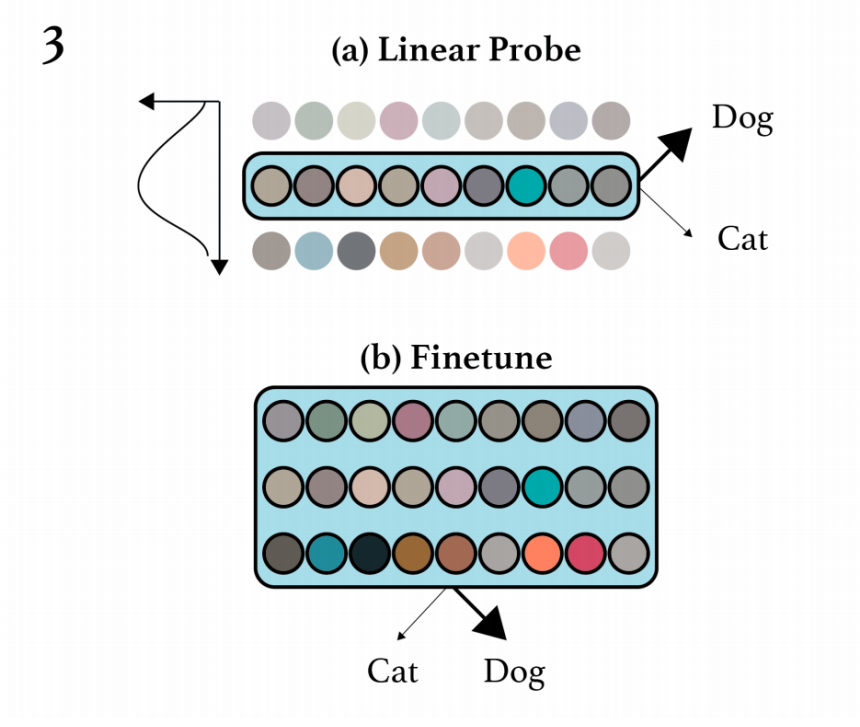
Original image | scaled image | color encoded image

How to train it?



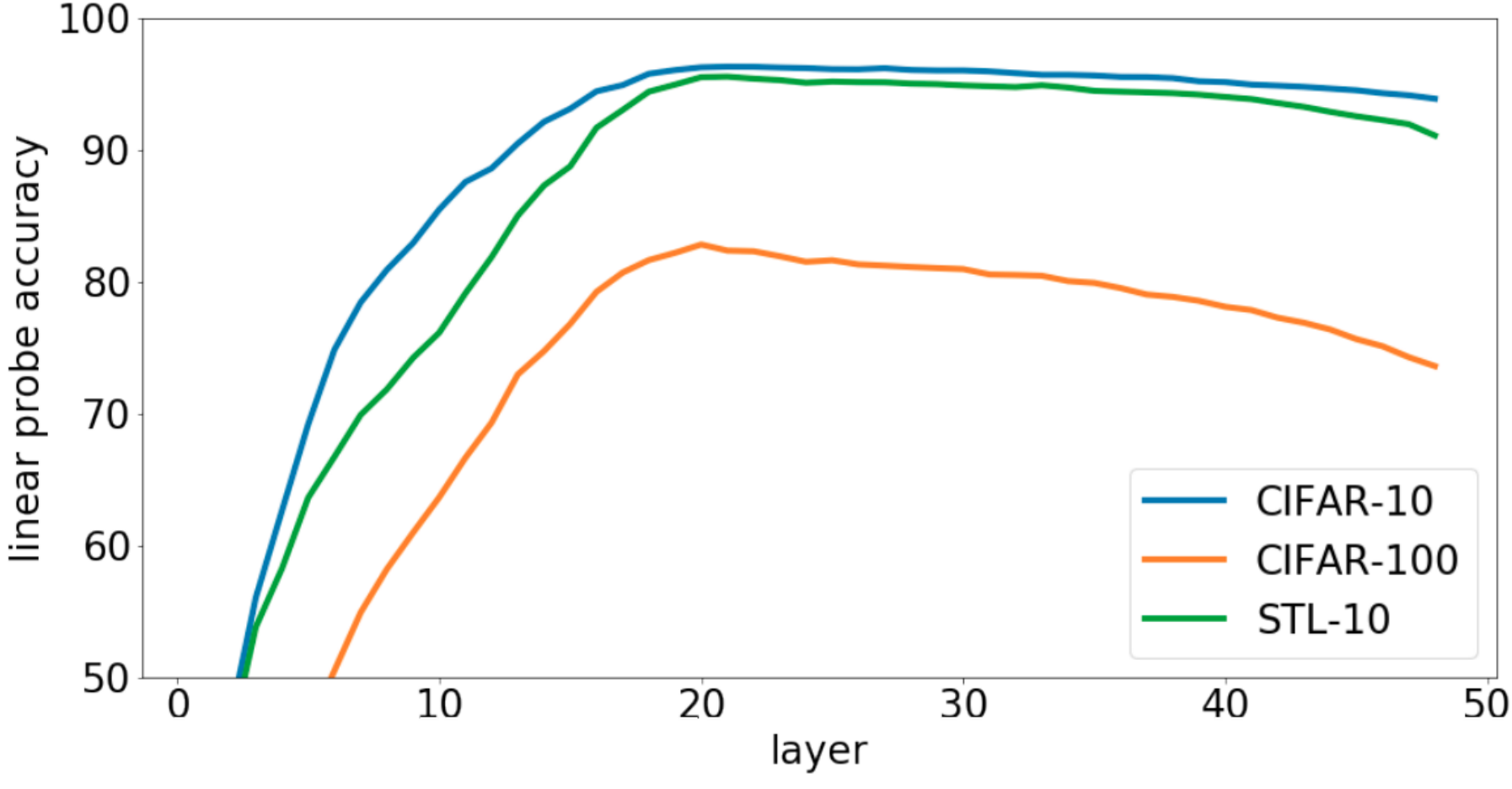
Training process, Source: Generative Pretraining From Pixels

Classification methods



Source: Generative Pretraining From Pixels

Results using linear probe



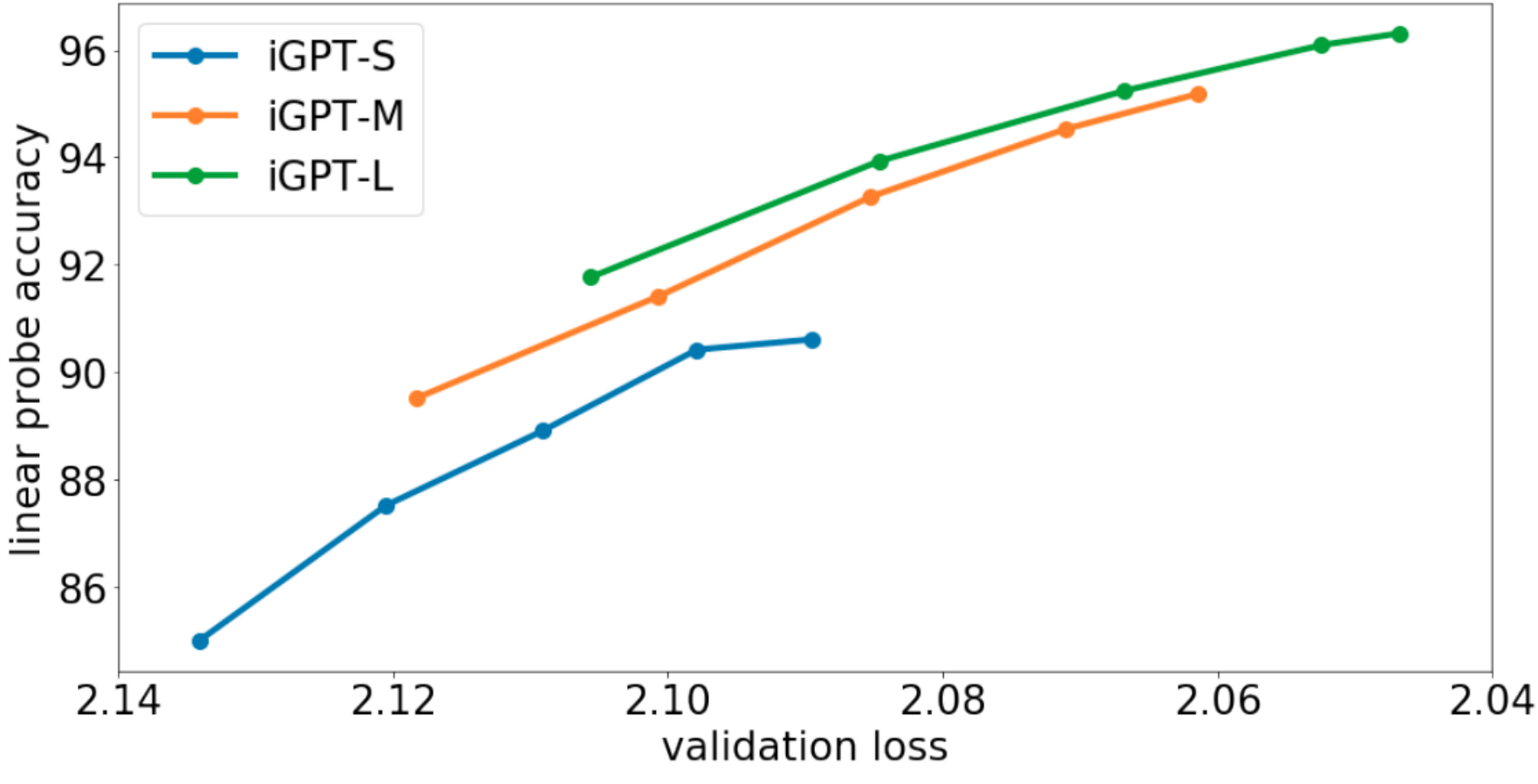
Linear probe results from iGPT-L model, Source: Generative Pretraining From Pixels

Results using linear probe

Method	Input Resolution	Features	Parameters	Accuracy
Rotation[paper]	original	8192	86M	55.4
iGPT-L	32x32x3	1536	1362M	60.3
BigBiGAN[paper]	original	16384	86M	61.3
iGPT-L	48x48x3	1536	1362M	65.2
AMDIM[paper]	original	8192	626M	68.1
MoCo[paper]	original	8192	375M	68.6
iGPT-XL	64x64x3	3072	6801M	68.7
SimCLR[paper]	original	2048	24M	69.3
CPC v2[paper]	original	4096	303M	71.5
iGPT-XL	64x64x3	3072 x 5	6801M	72.0
SimCLR	original	8192	375M	**76.5**

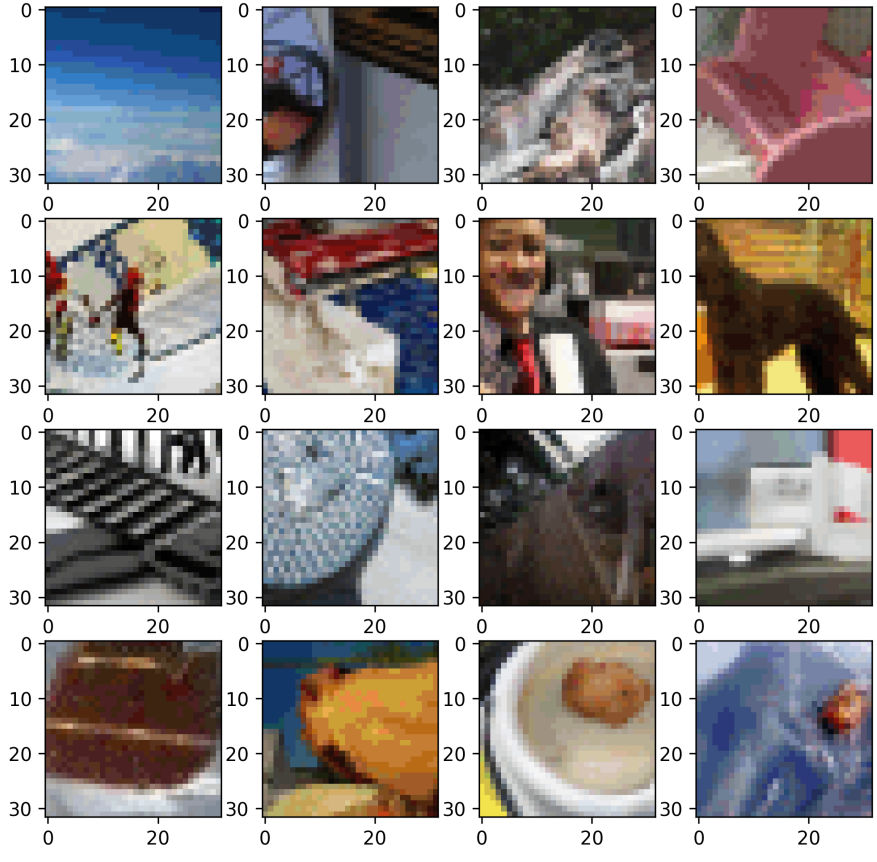
Linear probe results from different models against SOTA self-supervised models, Source: Generative Pretraining From Pixels

Accuracy and model size dependency

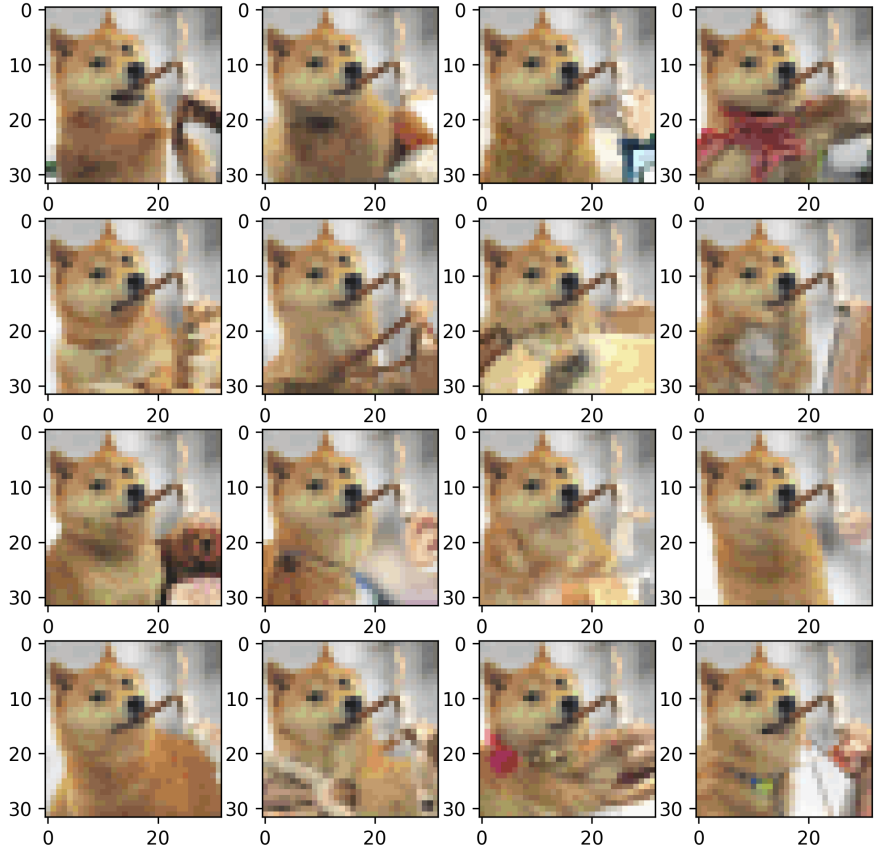


Validation generative loss, each model has a checkpoint at **65K, 131K, 262K, 524K, and 1000K steps**, Source: Generative Pretraining From Pixels

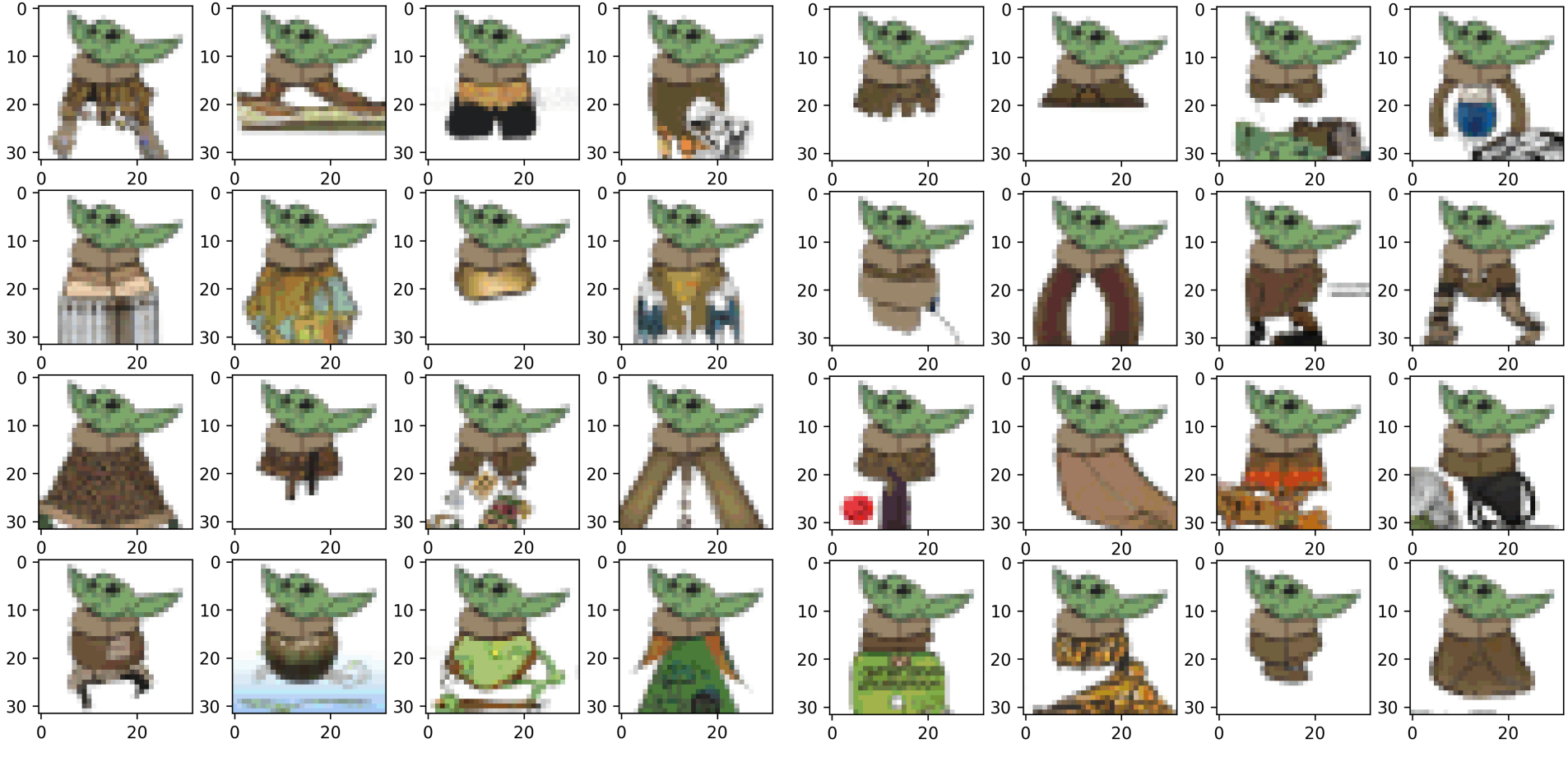
Autoregressive image generation



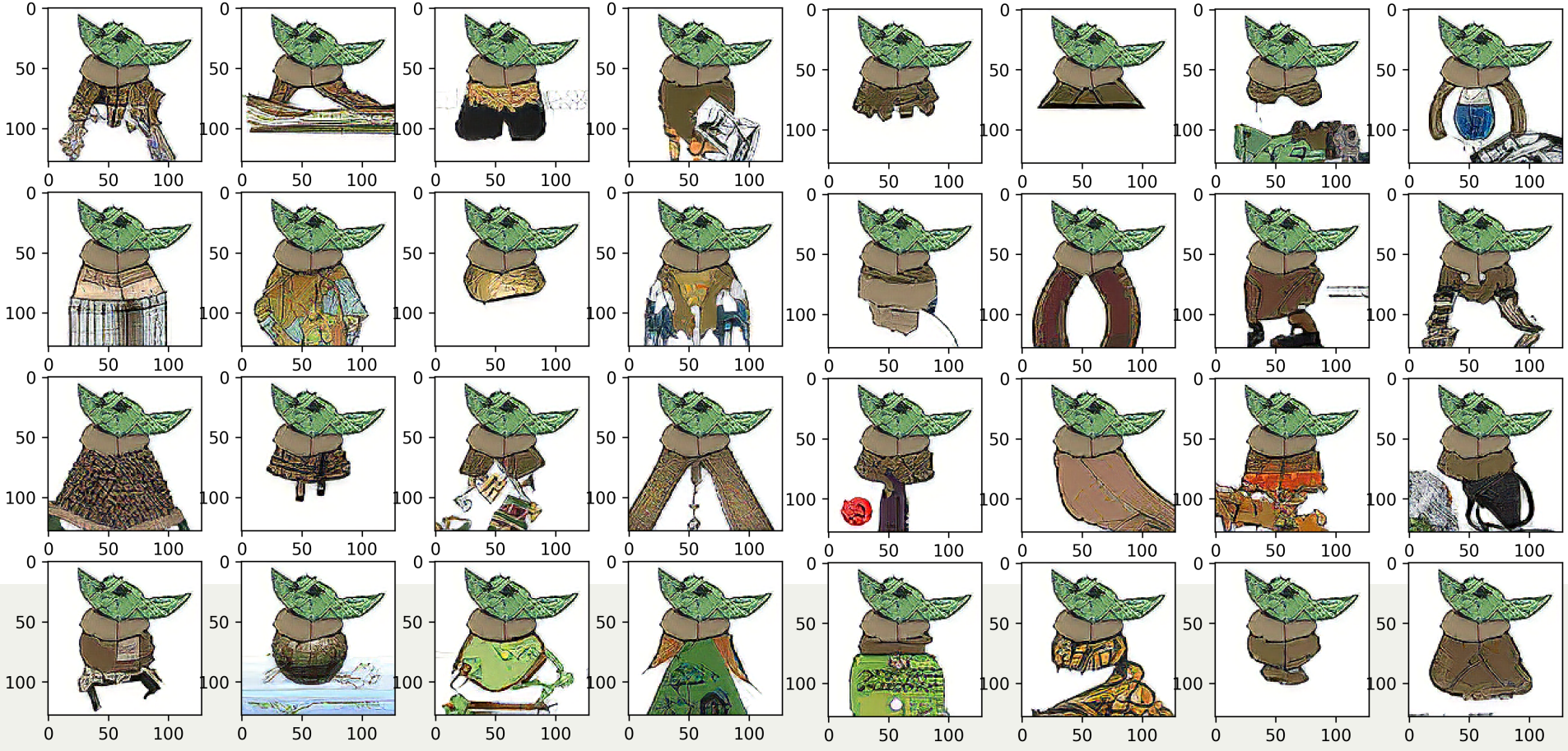
Autoregressive image generation with primer



Autoregressive image generation with primer



Resolution enhancement using ESRGAN



Thanks

"There's no such thing as a stupid question!"

Author: Kemal Erdem

Check out the code and generate your own images:
https://github.com/burnpiro/image-gpt