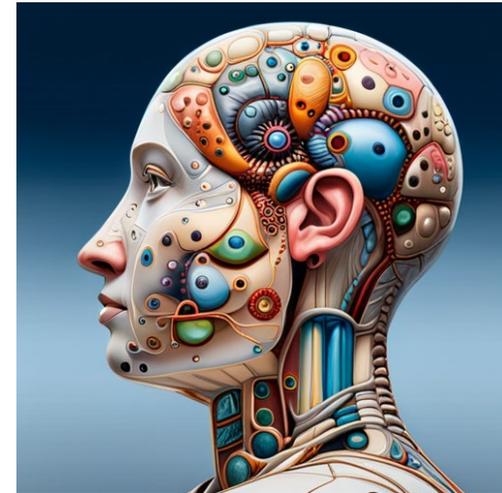


# テキストからの画像生成の ブレイクスルー

相澤清晴

東京大学 電子情報工学専攻, VRセンター



テキストから画像を  
生成するAI  
(DreamStudio)



東京大学 バーチャルリアリティ 教育研究センター  
The University of Tokyo Virtual Reality Educational Research Center

# 生成AI

- **言語系生成AI**

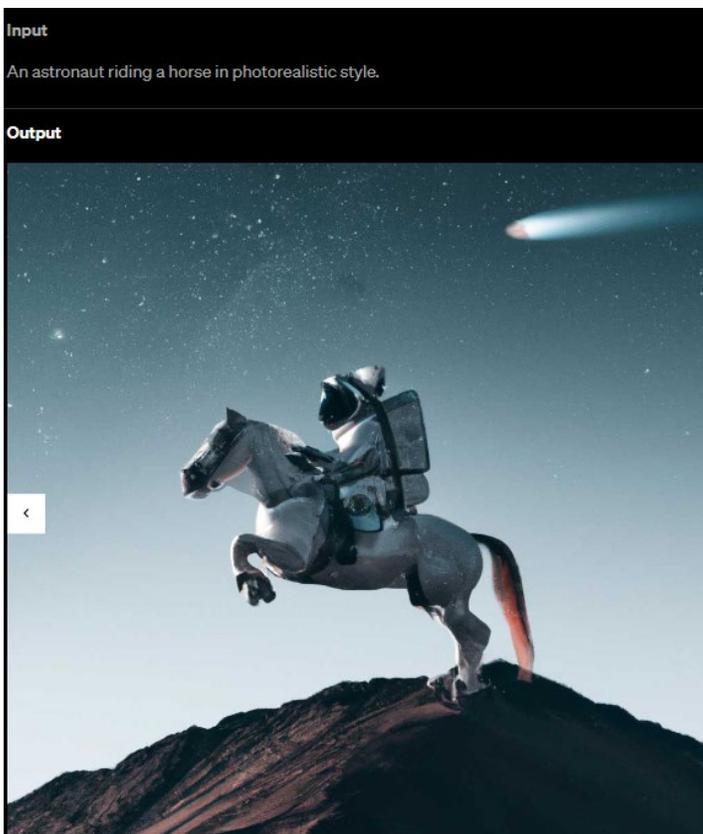
- ChatGPT, GPT-4, Bard
- 2022から爆発
- 大規模言語モデル (LLMs)

- **画像系生成AI**

- Dall-E2, Imagen や Stable Diffusion (ex. DreamStudio) 等
- 2022から爆発
- 大規模vision-languageモデル
- 拡散モデル

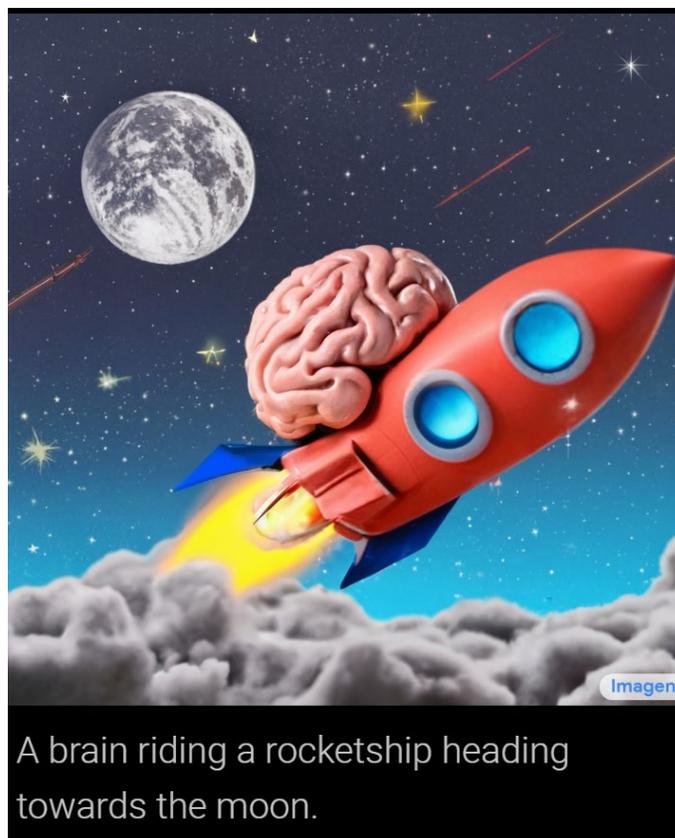
# テキストからの画像生成

- DALL•E2 (OpenAI)



<https://openai.com/dall-e-2>

- Imagen (Google)



A brain riding a rocketship heading towards the moon.

<https://imagen.research.google/>

- Stable Diffusion (Stability AI)



<https://github.com/Stability-AI/stablediffusion>

# カンブリア的爆発

## 拡散モデル (DDPM) 2020

- 2022.4 Dall-E2 (Open AI)
- 2022.5 Imagen (Google)
- 2022.8 Stable Diffusion (Stability AI)  
オープンソース  
モデル公開

数多の派生が現れ続けている。

## 敵対的生成 (GAN) 2014 ネットワーク

- Conditional GAN 2014
- Cycle GAN 2017
- Progressive GAN 2017
- Style GAN 2018
- Self Attention GAN 2018

# 2つの技術的ブレークスルー

- 大規模なVision-Language Model (VL model)

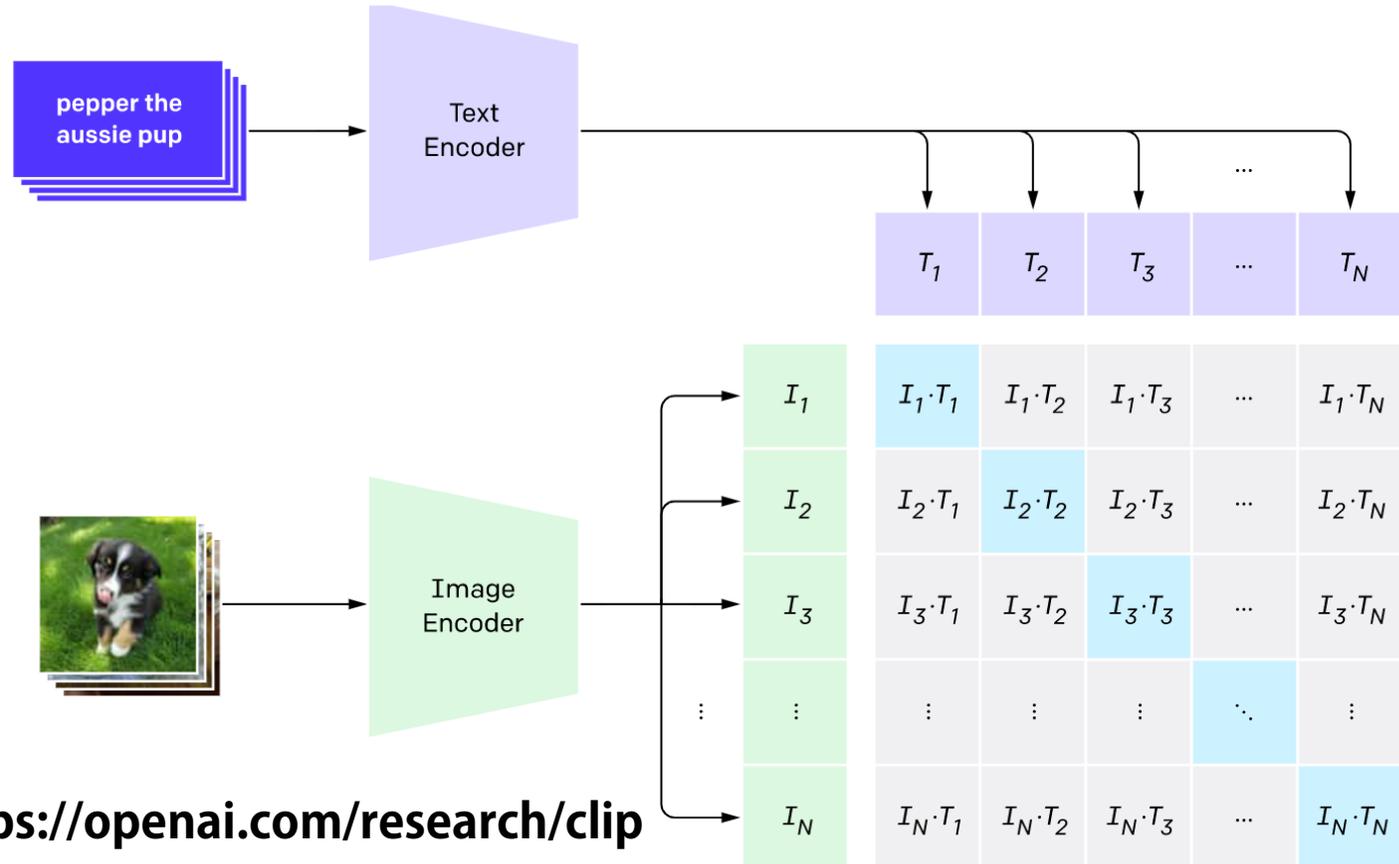
CLIPといった言語と画像のセマンティクスをつなげるモデルが実用的なレベルに到達したこと。モデルは公開されて、広く利用可能となっている。

- 拡散モデルによる画像生成

デノイジング拡散モデルによる画像生成手法の誕生。安定に学習でき、品質がよく、条件付けしやすい画像生成が可能となったこと。コード、モデルが公開され、開発が加速的に進んでいる。

# CLIP: 大規模VLモデル (OpenAI 2021)

- インターネットから収集した膨大な画像テキスト対(4億)を用いる。
- 膨大な量のデータを用いたシンプルな学習アルゴリズム。N個のペアに対して、本来のペアが近づくように、テキストと画像のエンコーダーを学習する。
- 画像特徴と言語特徴を近づけることができる。



Interacting with CLIP.ipynb

<https://openai.com/research/clip>

# CLIP : 汎用性. 特定の学習なしでの認識 (ゼロショット学習)

Food101

**guacamole** (90.1%) Ranked 1 out of 101 labels



✓ a photo of **guacamole**, a type of food.

✗ a photo of **ceviche**, a type of food.

✗ a photo of **edamame**, a type of food.

✗ a photo of **tuna tartare**, a type of food.

✗ a photo of **hummus**, a type of food.

Caltech-101

**kangaroo** (99.8%) Ranked 1 out of 102 labels



✓ a photo of a **kangaroo**.

✗ a photo of a **gerenuk**.

✗ a photo of a **emu**.

✗ a photo of a **wild cat**.

✗ a photo of a **scorpion**.

SUN397

**television studio** (90.2%) Ranked 1 out of 397 labels



✓ a photo of a **television studio**.

✗ a photo of a **podium indoor**.

✗ a photo of a **conference room**.

✗ a photo of a **lecture room**.

✗ a photo of a **control room**.

ImageNet-R (Rendition)

**Siberian Husky** (76.0%) Ranked 1 out of 200 labels



✓ a photo of a **siberian husky**.

✗ a photo of a **german shepherd dog**.

✗ a photo of a **collie**.

✗ a photo of a **border collie**.

✗ a photo of a **rottweiler**.

<https://openai.com/research/clip>

# Open CLIPによるゼロショットでの ImageNet (1000クラス) 認識



<https://laion.ai/blog/giant-openclip/>

# 画像-テキスト対のデータセット

- データ量が、ネットワークアーキテクチャ以上に重要。  
OpenAIは、データを公開していない。

**LAION (Large-scale Artificial Intelligence Open Network)**  
<https://laion.ai/> (since Aug. 2021)

画像-テキスト対のデータセット, モデル (OpenCLIP) の公開

LAION-400M (4億) OpenAIの4億対のデータセットに相当

LAION-5B (58.5億) 多言語

23億英語、23億の多言語(100+)、13億 その他

うち、日本語は、1.3億.

# LAION 5B cat vs 猫

cat CAT



Gray Cat Paintings - Grey Lady by Darren Robinson



Cute Cat Photoshop Brushes



卡哇伊呆萌猫咪PS卡通猫图案笔刷



黑色猫咪图像图案PS笔刷



Tom Cat Ink 4x5 \$50



Yawn like hell



Inagaki Tomoo (Japanese, 1902-1980), Cat Gathering...



CONTROVERSIAL CAT



Cat Brush



BooksChocolate\_cats



黑色猫咪图像图案PS笔刷



Ink Cat 1



The Art of Cat-a-login



ACT - Cat Breeds1



MALE CAT



Cat Page-A-Day Gallery Calendar 2018



LIFEAPP喵星人-猫宠物用品



اختيار نفهم الموضوع لاطفال cat pdf



Funny Quotes About Cat Lovers : Funny Cat Quotes: ...



Vintage Chinese Painting of Cat



gatos bajo la luna



個性豊かな猫が集合



Cat - 8x10 - watercolor



Белый кот и черная кошка



Caturday Art: Sketches and Glows



Paul Gosselin - Cats



Marie Christine Legeay - CAT THE CAT



wanderingcat



卡哇伊呆萌猫咪PS卡通猫图案笔刷



Cat Brushes by Helenartathome



Cooper the cat (booklover713) Tags: lighting cat k...



Siamese Gaze, Watercolor Painting Animal Cat Art

猫 猫



二一株 日記



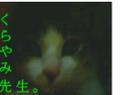
The Spirit of Your Eyes-Tsai Chin



猫街ポスター



にほんブログ村 コレクションブログ 御朱印へ



c0052756\_18251327.jp



にほんブログ村 猫ブログ MIX白黒猫へ



にほんブログ村 猫ブログ 老猫・高齢猫へ



にほんブログ村 猫ブログ MIX白黒猫へ



IMG\_3960\_a.jpg



にほんブログ村 猫ブログ 猫絵・猫漫画へ



banner090127.jpg



2011秋ばんさんバーナー.jpg



にほんブログ村 猫ブログ 老猫・高齢猫へ



にほんブログ村 猫ブログ 猫へ



にほんブログ村 猫ブログ 猫写真へ



にほんブログ村 猫ブログ 猫写真へ



にほんブログ村 猫ブログ 猫カフェへ



家猫怎么喂养



C0132438\_22412349\_イベント限定カラフル黄金麺(麺大盛り)



imga7d6306ezikezj-



【楽天ブックスならいつでも送料無料】巨猫団【えみち】



にほんブログ村 猫ブログ 猫写真へ



にほんブログ村 猫ブログ ソマリへ



banner090214.jpg



にほんブログ村 地域生活(街) 中国地方ブログ 尾道情報へ



image/sakuragumi-2006-01-28T13:58:13-1.jpg



にほんブログ村 猫ブログ 猫カフェへ



にほんブログ村 猫ブログ 猫カフェへ



にほんブログ村 猫ブログ 猫写真へ



山のトムさん DVD-BOX



ネコちゃん達が 怖い寝るかも!? ★人気猫のチョコビちゃん「お部屋で遊んだり怖い寝が楽しみだニヤニヨ」と申しますっ(^^)



にほんブログ村 猫ブログ

# LAION 5B Jizo vs お地蔵さん

## Jizo



ohenrosunafukin(冬の道路 | Social Profile



【風景写真】六角堂(頂法寺)



Little Buddha Statue



Buddhist guardian of travelers, Jizo Bosatsu



Jizo Statue Buddha Concrete Statues Buddhist Prote...



Statue guardian wearing red hat in Tokyo daytime c...



jizo statue buddha concrete statues buddhist prote...



PSL Khakkhara Ksitigarbha Handmade statue buddha j...



復讐 セックスレス 東京都 折柄師 神宮 司龍峰



id:mizu-kuki



statuette jizo au temple Ruriko-ji, Yamaguchi



jizo statue concrete statues of jizo s protectors ...



Jizo Statue Locations File:jizo Statue at Zojo-ji



美しい仏像の画像を貼るトピ♪



momokurayama (3 von 22)



大光山 真徳寺-3 お地蔵さまに心通わせ・・穏やかに☆平和の火守り・鐘撞きも随意に



mumblealice17's avatar



Rokkaku-dō



sculptures



20110815\_名古屋・犬山\_一目



jizo statue buddha concrete statues buddhist prote...



Buddhist guardian of travelers, Jizo Bosatsu



新潟県村上散歩 2014年3月28日



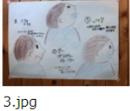
Buddhist Statue Royalty Free Stock Photo



## お地蔵さん



第62回「キモノショップあいこ(セルバ3F)」



3.jpg



熊本・大分の皆様への画像(プリ画像)



<特別企画!>はまぐり堂代表・亀山とオンラインで話せる券



Day15〜滋賀遠征マンションだよ〜の記事に添付されている画像



f0043559\_19412454.jpg



korokoroのブログ



Yoshiyuki U



七七五小江 お友だちと待ち合わせ(小学3年生)



第49回「エムバイセンソユニコ(セルバ3F)」



なつたま(夏珠)



靴下、こよろっ!!の記事に添付されている画像



代表 田辺 行雄さん (39)



オモシロTシャツ18



Yuki Ochiai



皆様から愛されるようなアットホームなお店を目指しております★電車でお越しの方、常盤線、土浦駅までお迎え...



f:fid:genron:20171106011514j:image



f:fid:kimurakatsunori:20



age 高橋毅さん



♡お饅頭同盟♡ | (おもちゃ)



avatar



[無料版]しょーちゅーでちゅ(1)



人気の「Q-CHAN牧師」動画 2,654本 - Q-CHAN牧師の伝道部屋



鈴木尚広 Twitter



東京・山手線の中の動画に出演させてもらいました



中学生徒総会1.jpg



f0031037\_18523530.jp



喜良先生のTシャツ ☆



伊藤 祐次朗



2016-0911-488.jpg



坂上忍オフィシャルブログ「綺麗好きでなにが悪い!」 Powered by Ameba

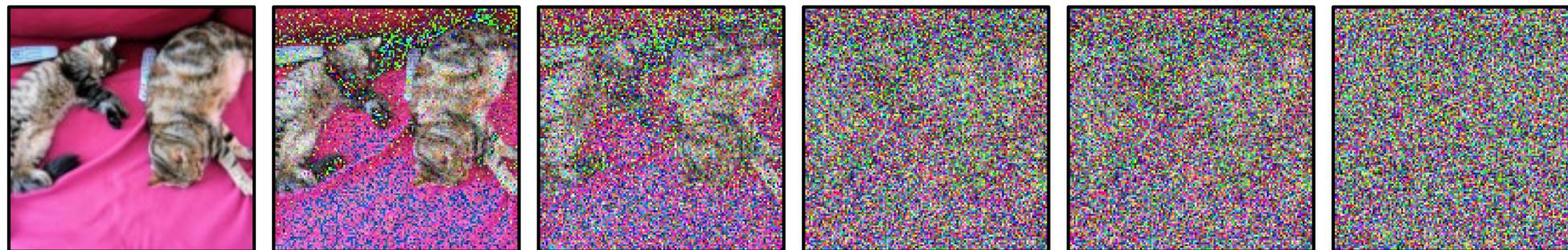


鈴木尚広 Twitter

# 拡散モデルによる画像生成

 Hugging Face  
demo

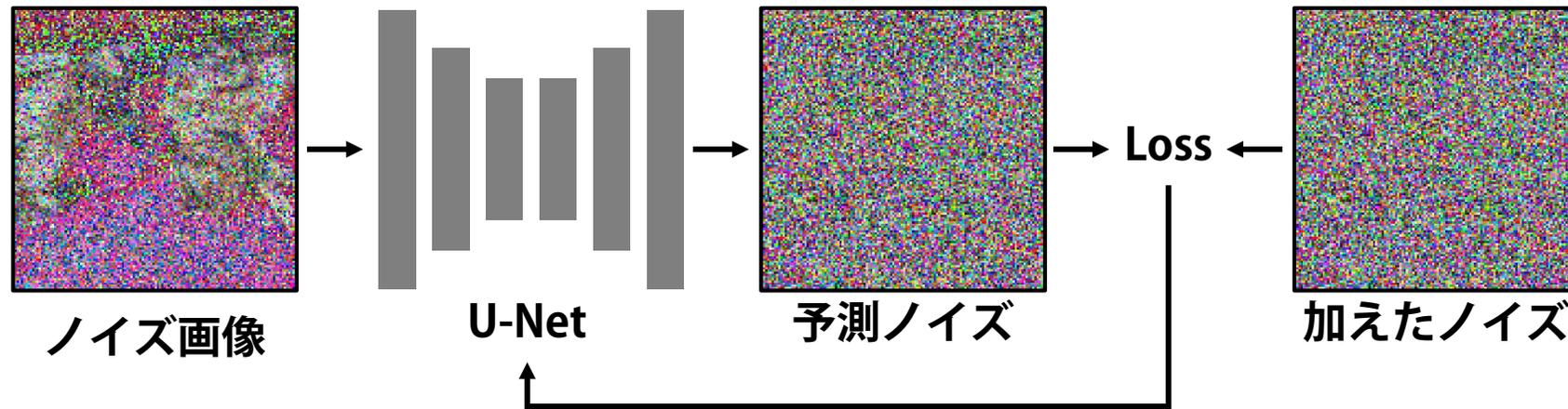
拡散過程：トレーニングのためにノイズを付与



逆拡散過程：ノイズ除去により画像生成

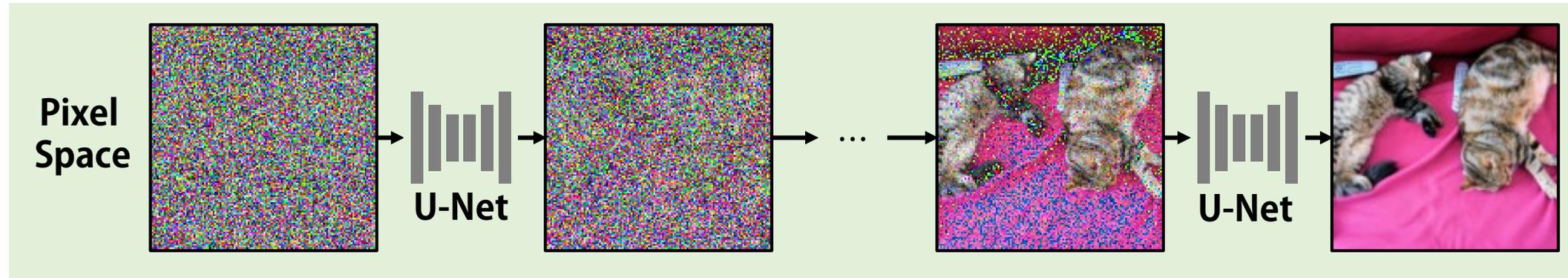
- ノイズと画像の関係を学習
- 拡散過程では、画像に徐々にノイズを付与し、ノイズに変換
- 生成時には、ノイズとして与えられる初期画像から画像を復元

# 学習：ノイズの予測

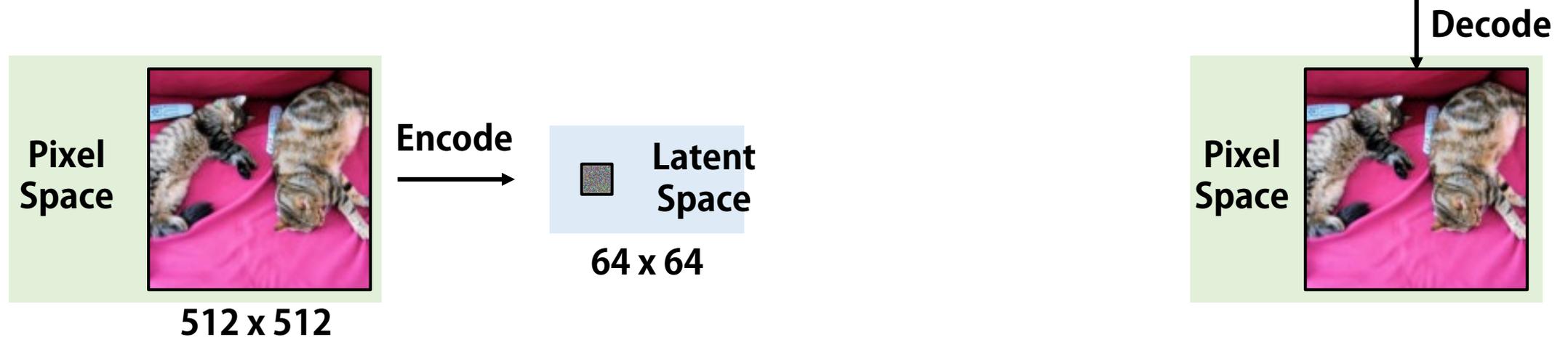
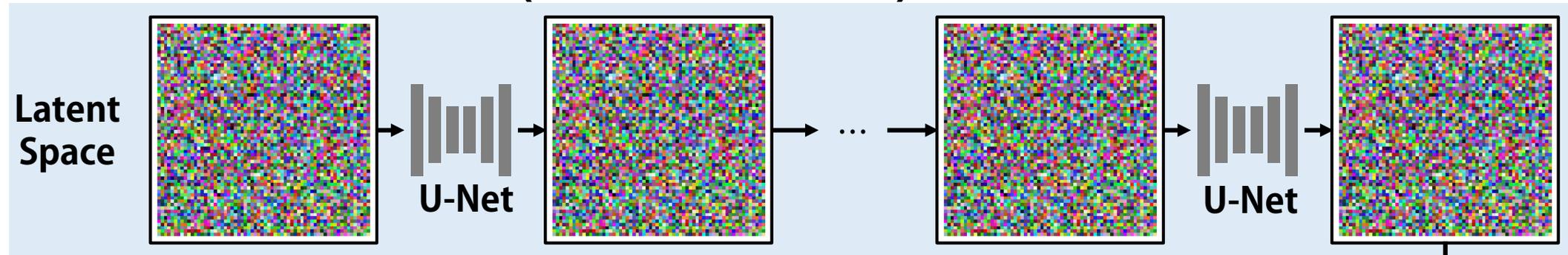


- ノイズ画像から、ノイズを予測。
- ニューラルネットワーク (U-Net) により、ノイズの予測を学習。
- 最小平均2乗推定(MMSE)では、通常、元信号になんらかの事前分布を仮定して、ノイズ信号から元信号 (あるいはノイズ) を予測する。一方、事前分布を使わずに、ノイズ信号とその尤度からMMSEできる (Miyasawa 61) 。

## Diffusion Model



## Stable Diffusion Model ( Latent Diffusion )



# 多くの派生モデルが誕生，公開

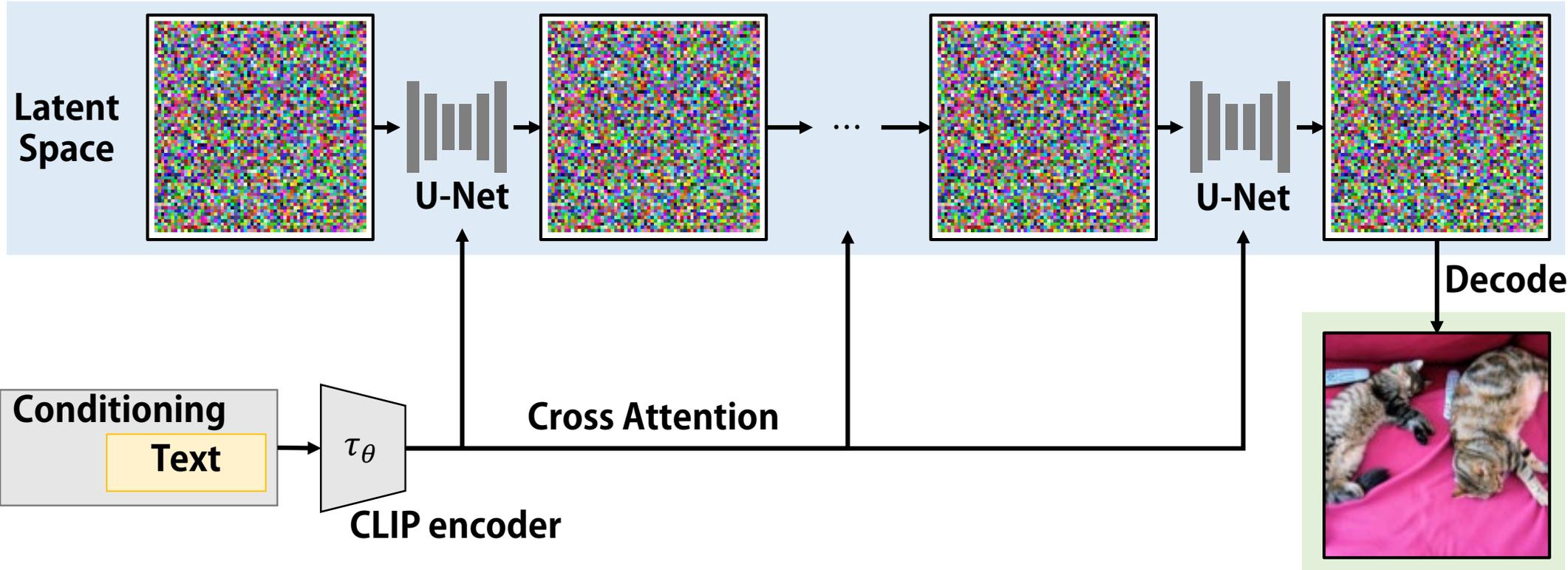
2022年8月にコード，学習済みモデルが公開された (Stability AI).

Hugging Faceにて，3422 モデルが公開(2023.6.5).

The screenshot shows the Hugging Face website interface. At the top, there is a search bar with the text "Search models, datasets, users...". Below the search bar, there are navigation links for "Models", "Datasets", "Spaces", "Docs", "Solutions", "Pricing", "Log In", and "Sign Up". The main content area displays a list of models under the "Models" tab, with a total of 3,422 models. The list is filtered by name and sorted by "Most Downloads". The models listed include:

- runwayml/stable-diffusion-v1-5 (Updated about 1 month ago, 2.66M downloads, 8.01k likes)
- CompVis/stable-diffusion-v1-4 (Updated 20 days ago, 924k downloads, 5.58k likes)
- SG161222/Realistic\_Vision\_V1.4 (Updated May 2, 902k downloads, 187 likes)
- stabilityai/sd-vae-ft-mse (Updated Jan 5, 888k downloads, 169 likes)
- prompthero/openjourney (Updated 20 days ago, 700k downloads, 2.61k likes)
- stabilityai/stable-diffusion-2-1-base (Updated 20 days ago, 687k downloads, 374 likes)
- stabilityai/stable-diffusion-2-1 (Updated 20 days ago, 351k downloads, 2.63k likes)
- runwayml/stable-diffusion-inpainting (Updated 20 days ago, 343k downloads, 1.17k likes)
- llyasviel/sd-controlnet-canny (Updated May 2, 332k downloads, 61 likes)
- hakurei/waifu-diffusion (Updated Mar 15, 323k downloads, 2.12k likes)
- stabilityai/stable-diffusion-2-inpainting (Updated 20 days ago, 199k downloads, 298 likes)
- stabilityai/stable-diffusion-x4-upscaler (Updated Apr 24, 193k downloads, 390 likes)
- dreamlike-art/dreamlike-photoreal-2.0
- llyasviel/control\_v11p\_sd15\_canny

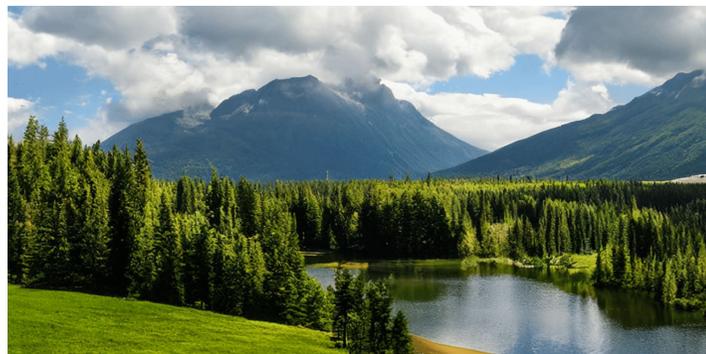
# テキストによる条件付け



# テキスト以外の条件付けも可能

hot air balloon, ユーザの手書きスケッチ

Semantic map

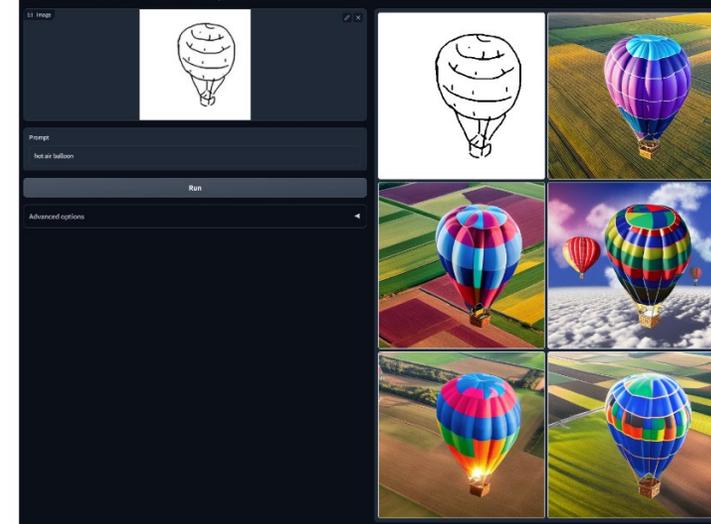


Inpaint



Prompt: "hot air balloon"

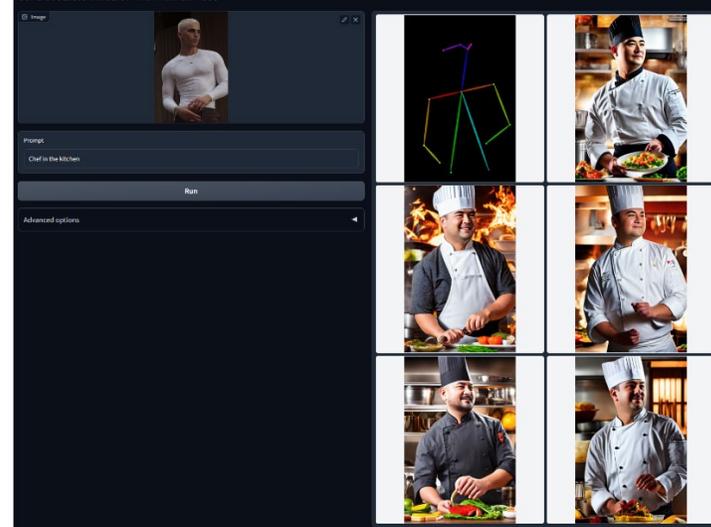
Control Stable Diffusion with Scribble Maps



Chief in the kitchen, 人の姿勢

Prompt: "Chief in the kitchen"

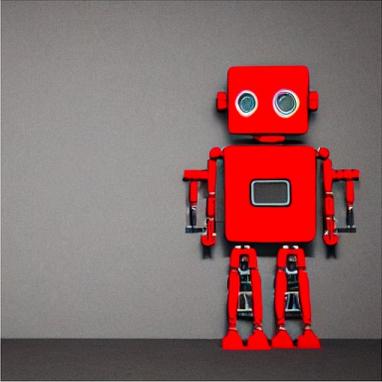
Control Stable Diffusion with Human Pose



<https://github.com/CompVis/latent-diffusion>

<https://github.com/lllyasviel/ControlNet>

# 拡散モデルの追加学習 (fine tune)



A robot, humanoid,  
red and gray



a photograph of an  
astronaut riding a  
horse

- [pokemon dataset](#)  
(416)
- iteration: 15,000
- denoising steps: 1000



A robot, humanoid,  
red and gray



a photograph of an  
astronaut riding a  
horse



a drawing of a yellow  
and gray pokemon  
character

# まとめ

テキストからの画像生成. “4つ”のブレークスルー.

- 大規模な Vision Language モデルによる画像と言語のセマンティクスの対応.
- デノイジング拡散モデルによる高品質で、条件付けの容易な画像生成.
- 画像テキスト対の大規模データの公開.
- オープンソース（コード、モデル）による民主化で、加速的な開発.

コンテンツ（データ）の多様性は、今後さらに、重要.