

at-4.ディープラーニングの利用 例

(ディープラーニングのシステムとプログラミング)
(全12回)

<https://www.kkaneko.jp/ai/at/index.html>

金子邦彦



1. AIによる生成：Stable Diffusionなどを用いた画像生成、高解像度化、GANの仕組み
2. セグメンテーションの章：パノプティック、インスタンス、セマンティックの3種類の手法。学習用の訓練データ
3. 顔情報処理：顔検出、アラインメント、ランドマーク検出、表情推定など。
4. オンラインデモを活用した演習。実践的な応用例。

A blurred background image showing several people in an office setting, likely a meeting or collaborative work environment. The image is out of focus, emphasizing the text in the foreground.

アウトライン

1. 生成
2. セグメンテーション
3. 顔情報処理

12-1 生成

人工知能による生成の例

実在しない人間の顔画像を生成



実在



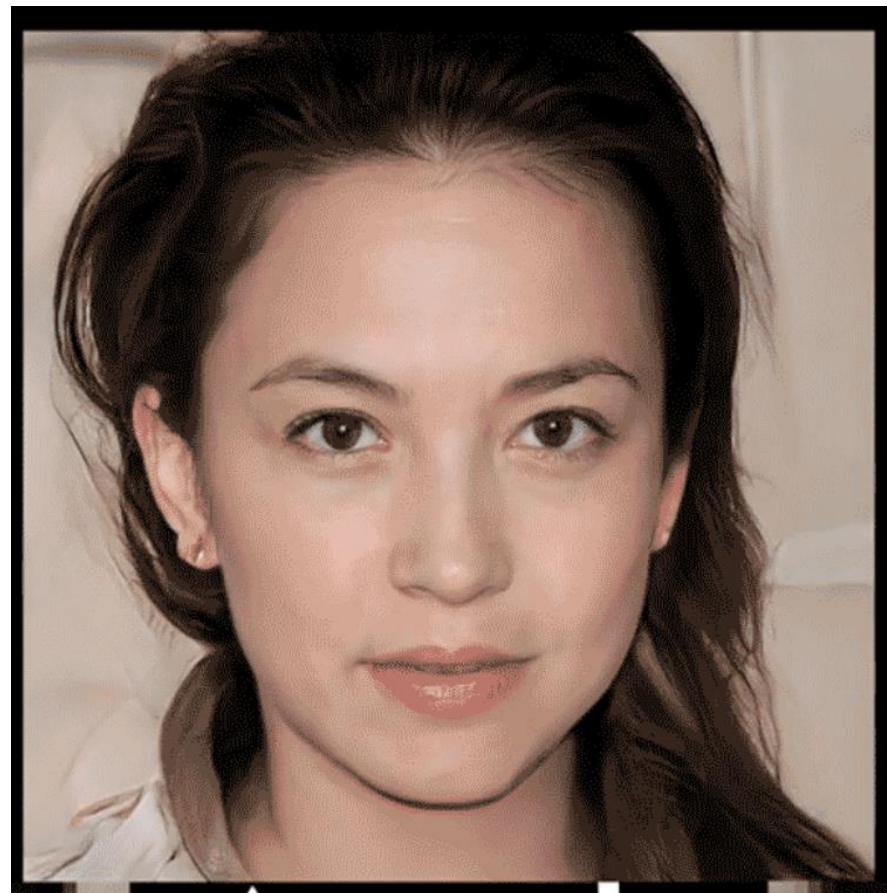
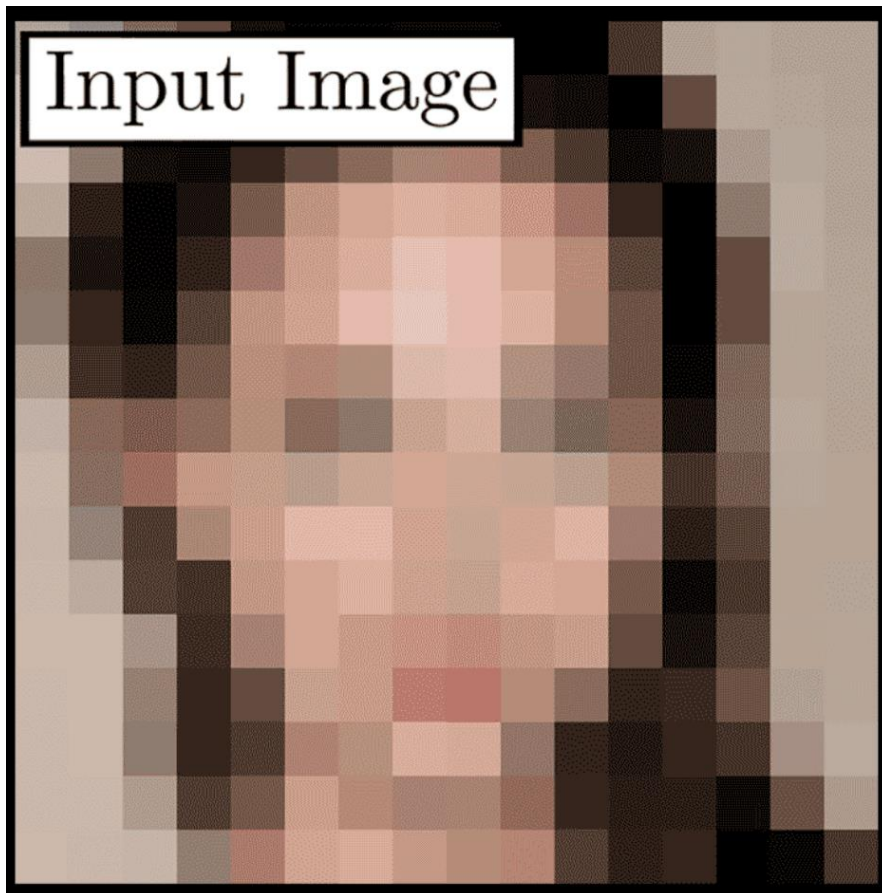
フェイク

Web ブラウザで動く

<https://www.whichfaceisreal.com/>

人工知能による生成

低解像度の画像を，高解像度の画像に変換



研究成果，プログラムのソースコードはオンラインで公開されている

<http://pulse.cs.duke.edu/>

文献 Menon, Sachit and Damian, Alex and Hu, McCourt and Ravi, Nikhil and Rudin, Cynthia,

PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models, 6

The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

Stable Diffusion

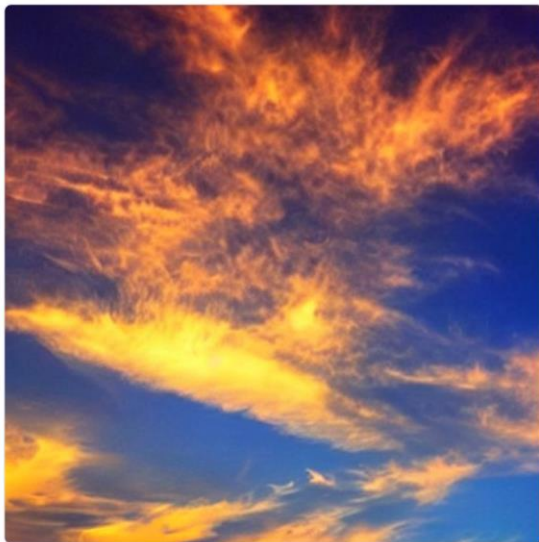
- 2022年発表
- Text-to-Image (**英語の文章などを画像に変換**) の人工知能. 他の画像生成等のタスクの基礎
- **オンラインデモ**有り

<https://stablediffusionweb.com/>

Due to the large number of users, the server may experience problems. If you encounter an error, please try again.

beautiful sky

Generate image



英語の文章などを画像に変換

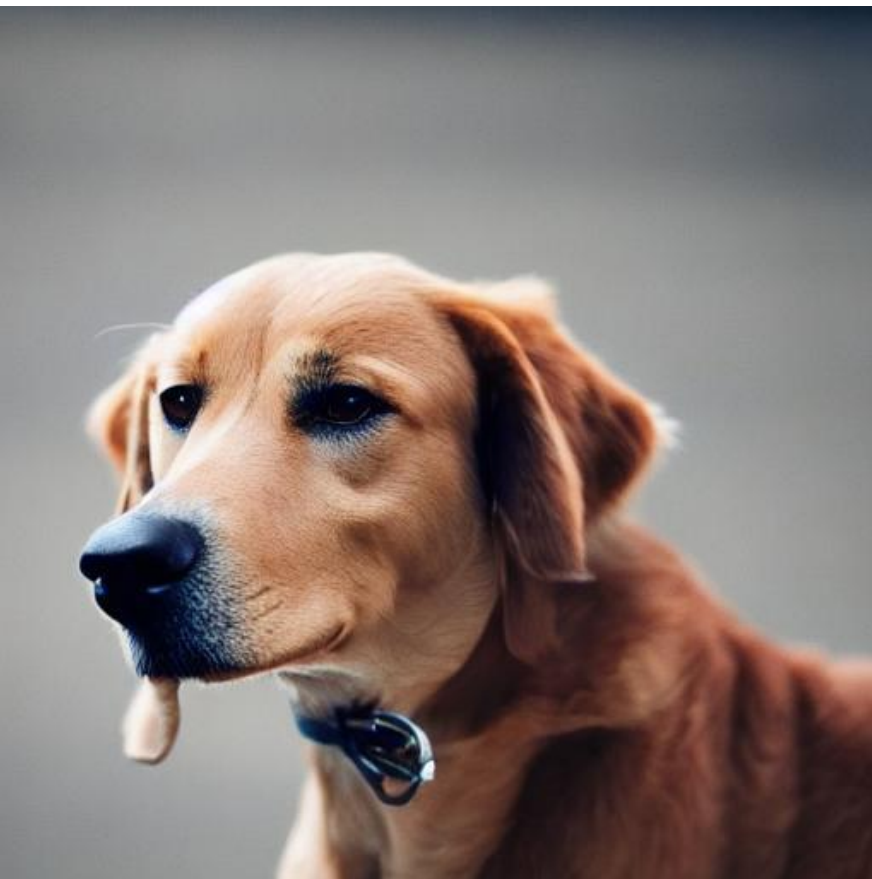
人間が文章「**a scenic landscape**」を与える。人工
知能が画像を**生成**



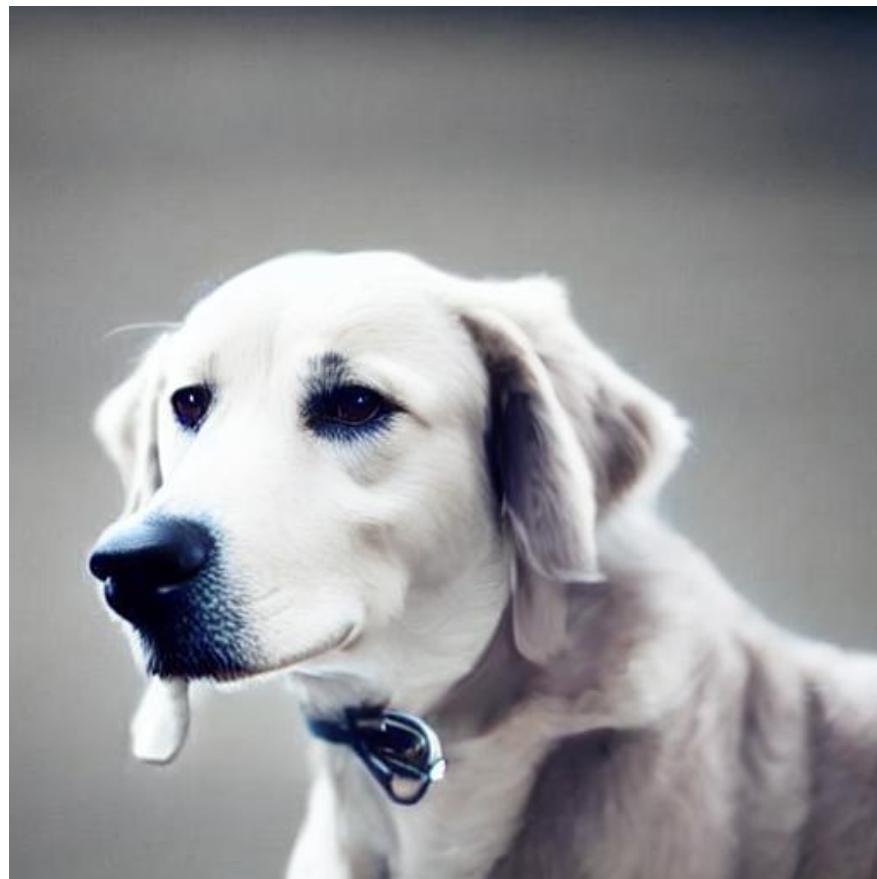
画像編集

人間が、文章「**make a dog white**」を与える。

人工知能が画像編集を行う



元画像



人工知能による画像編集結果

画像編集

人間が、文章「**make a man long hair**」を与える。
人工知能が画像編集を行う



元画像



人工知能による画像編集結果

ImaginAIry (<https://github.com/brycedrennan/imaginAIry>) を使用

画像編集

人間が、文章「**add glasses**」を与える。

人工知能が画像編集を行う



元画像

人工知能による画像編集結果

ImaginAIry (<https://github.com/brycedrennan/imaginAIry>) を使用

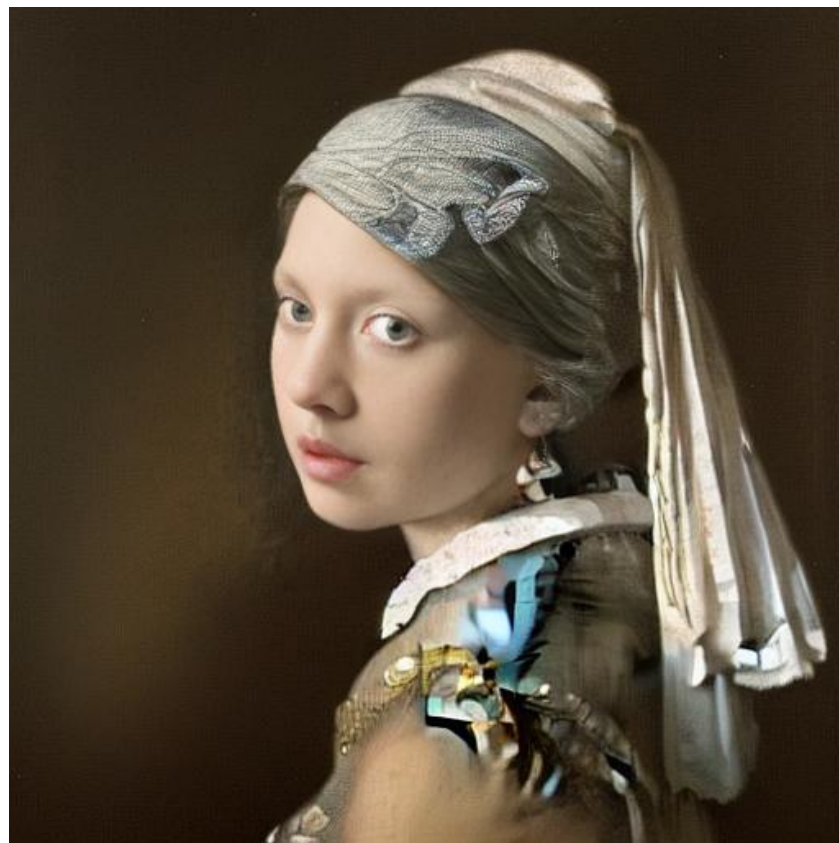
画像編集

人間が、文章「**white dressed**」を与える。

人工知能が画像編集を行う



元画像



人工知能による画像編集結果

画像編集

人間が、文章「**animal doll**」を与える。

人工知能が**画像編集**を行う



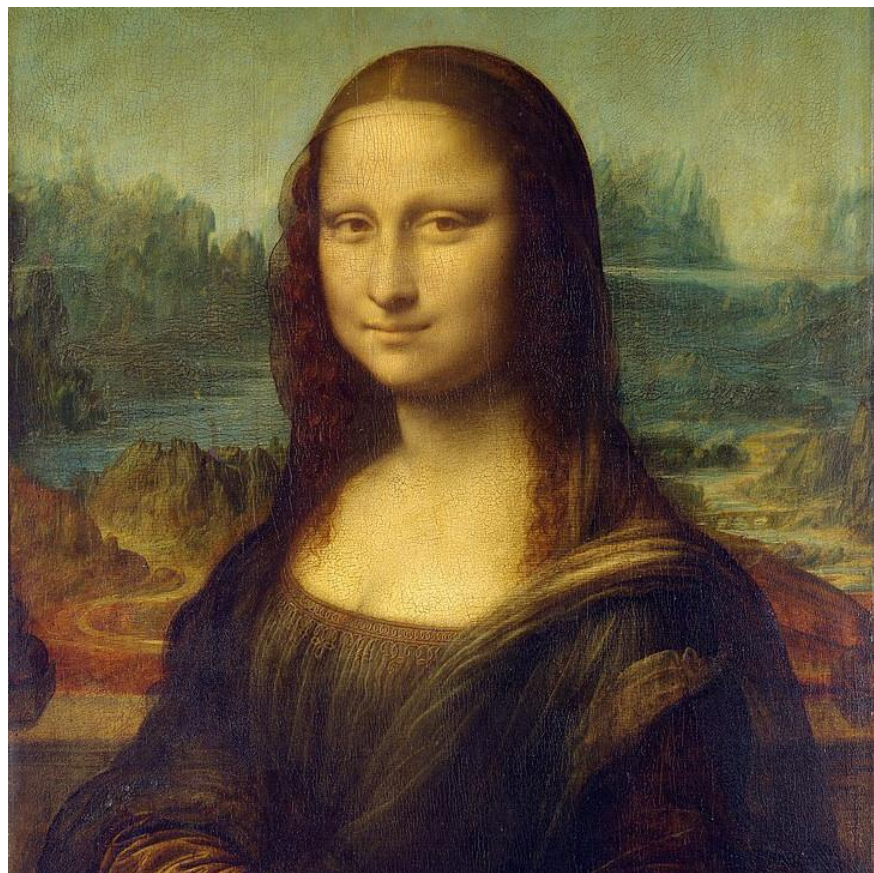
元画像



人工知能による画像編集結果

画像編集

人工知能が実写風になるように画像編集を行う



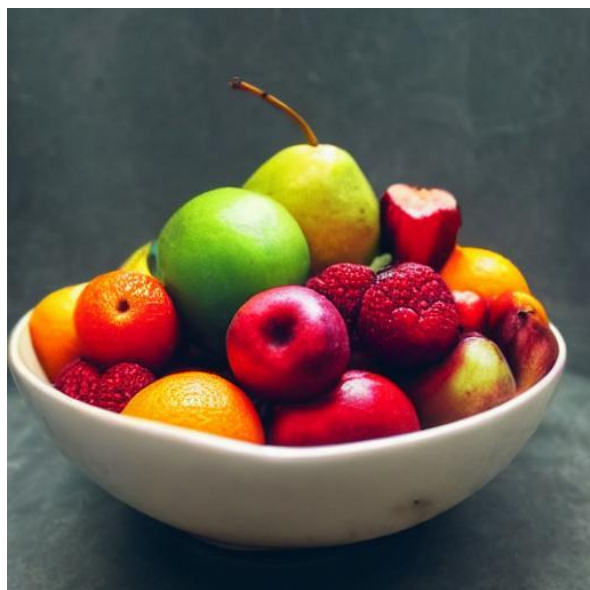
元画像

人工知能による画像編集結果

ImaginAIry (<https://github.com/brycedrennan/imaginAIry>) を使用

まとめ

- Text-to-Image (英語の文章などを画像に変換) の人工知能



元画像



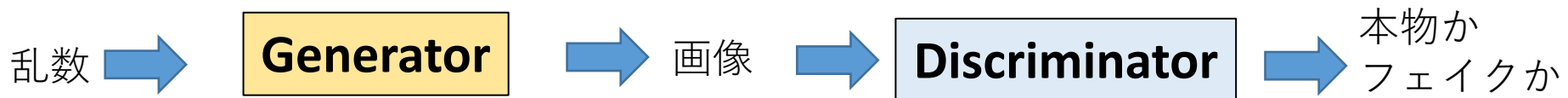
人工知能による画像編集結果

ページ16から22の結果は、パソコンでの動作結果である。手順は次のページで説明

<https://www.kkaneko.jp/ai/win/imaginairy.html>

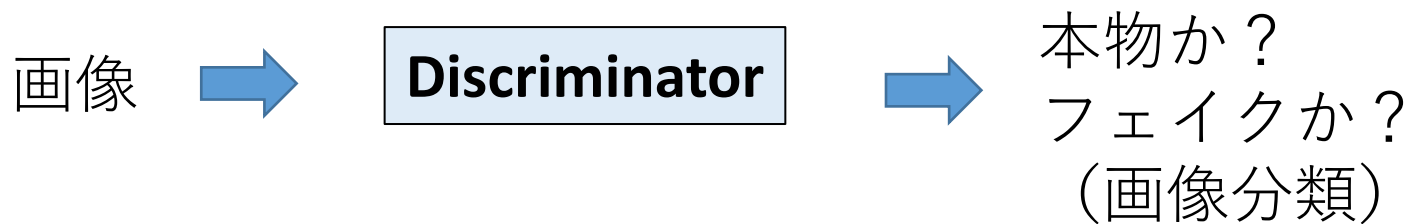
GAN

- GAN (Generative Adversarial Network) は、画像生成に利用されるディープラーニングのアーキテクチャ。
- GANは、Discriminator (判別器) とGenerator (生成器) と呼ばれる2つのディープニューラルネットワークで構成。



Discriminator

- Discriminator（判別器）は、画像を分類するために使用。
- 与えられた画像が**本物の画像**であるか、それとも**生成されたフェイクの画像**であるかを**判別**することです。
- ディープラーニングにより、画像の特徴を学習し、判別能力を高める。



Generator

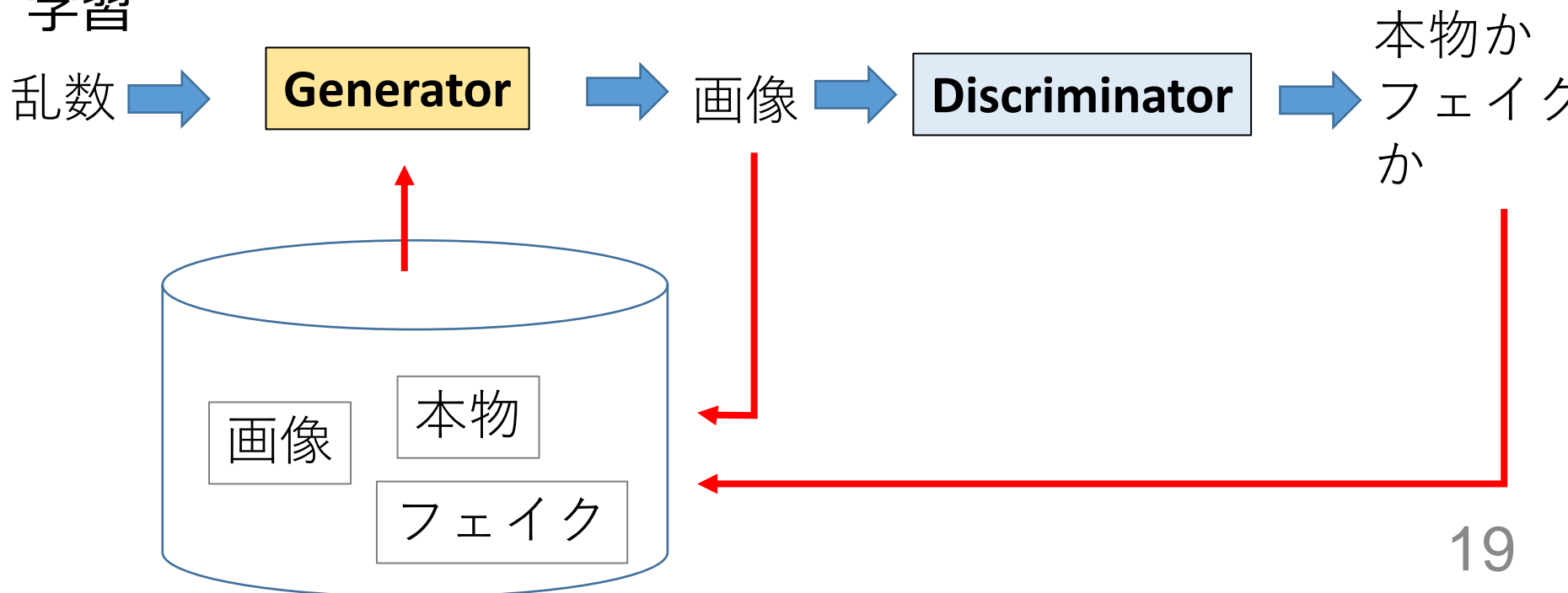
- Generator（生成器）は、乱数を種として画像を生成
- 通常、オートエンコーダと呼ばれるディープラーニングのアーキテクチャを使用
- Generatorの目標は、本物の画像と見分けがつかないような新しい画像を生成すること。



Generator の学習

- Generator の学習は、Discriminator を使用。
- Generator の訓練データは、生成された画像とそれに対する本物かフェイクかの判別結果。
- **Generatorは生成した画像をDiscriminatorに与え、その結果を使って自身を改善。**

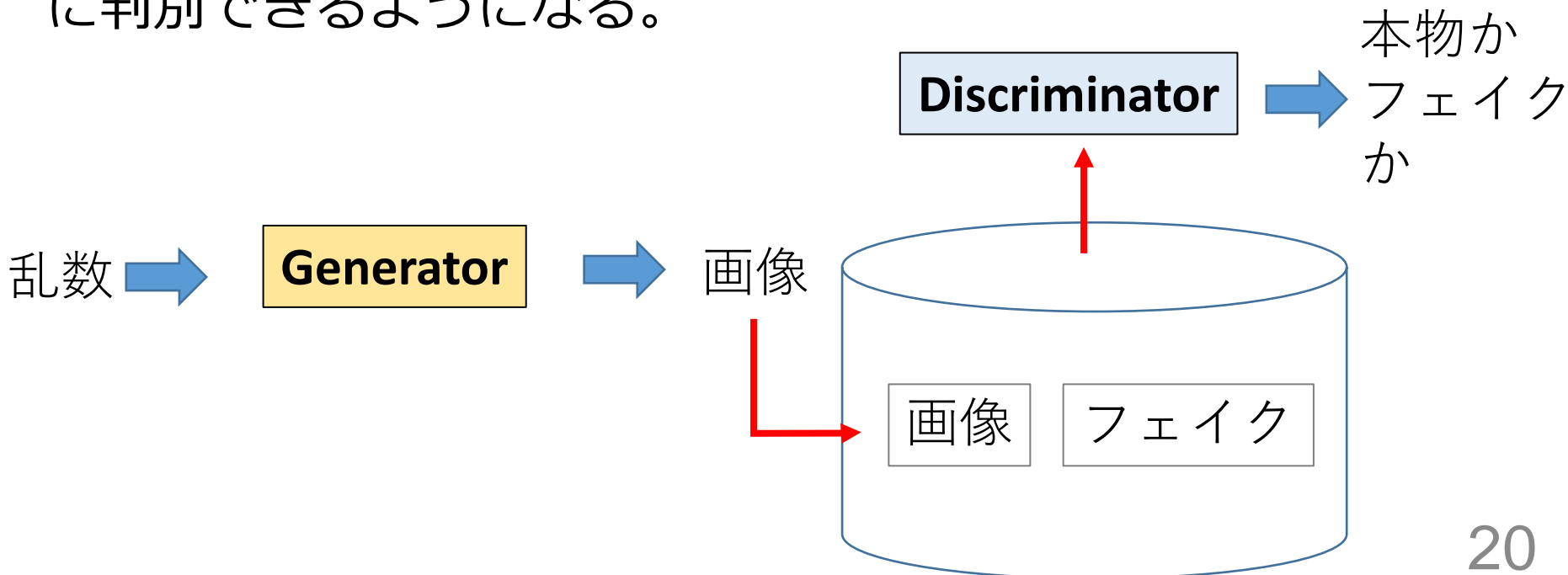
→ この過程により、Generatorは本物の画像に近づくように学習



Discriminator の学習

- Discriminatorの学習は、Generatorを使用。
- Generator は「**フェイク**」の画像を生成し、それを**Discriminator**に与える。
- Discriminator はそれらの画像が**本物かフェイクか**を判別することを学習

→ Discriminatorは本物の画像とフェイクの画像をより正確に判別できるようになる。





演習

**Stable Diffusion 2.1 デモによる
画像生成**

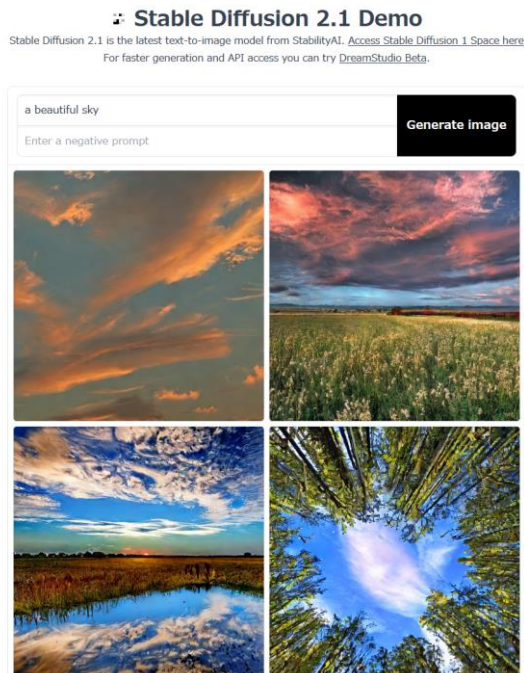
【トピックス】

- **Stable Diffusion**
- **プロンプト**

① 次のページを開く

<https://huggingface.co/spaces/stabilityai/stable-diffusion>

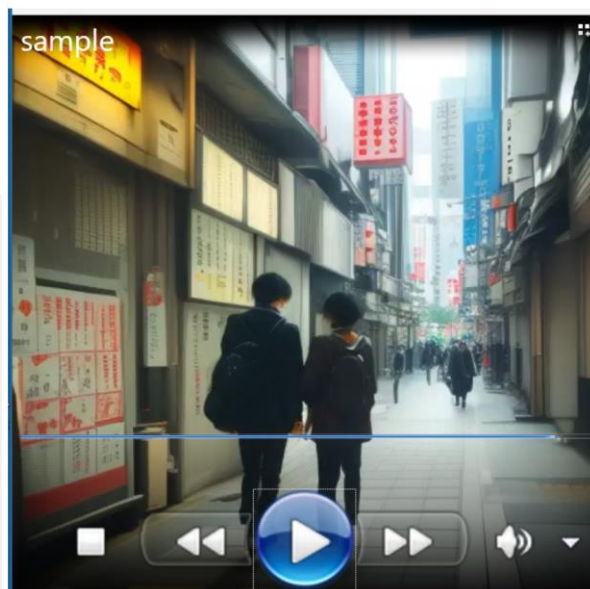
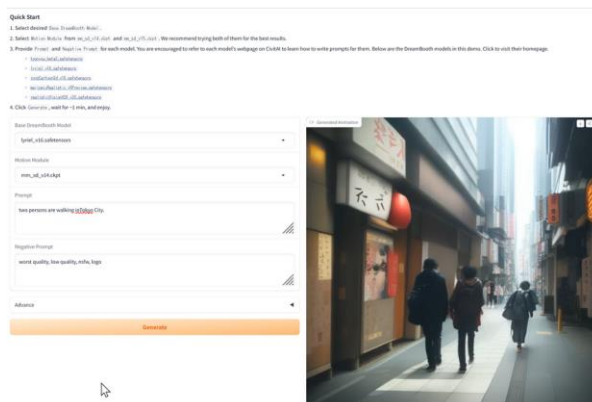
② 英文や英単語の並びを入れ「Generate Image」



別のサイト

③次のサイトは、プロンプトからアニメーションを生成する

<https://huggingface.co/spaces/guoyww/AnimateDiff>



12-2 セグメンテーション

セグメンテーション



物体の形を画素単位で抜き出し



ラベルを得ることもできる

画像理解の主な種類

① 画像分類

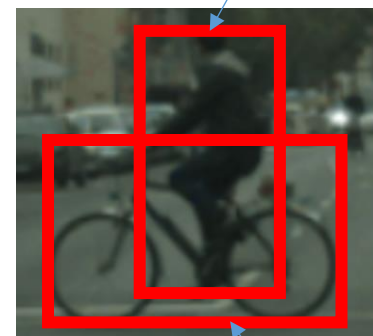
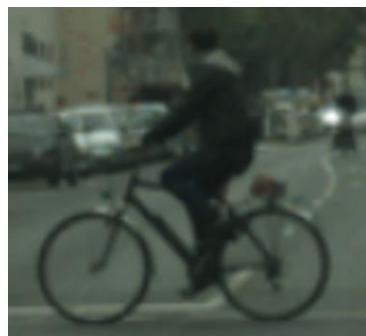
「何があるか」を理解



person
bicycle

② 物体検出

場所と大きさも理解

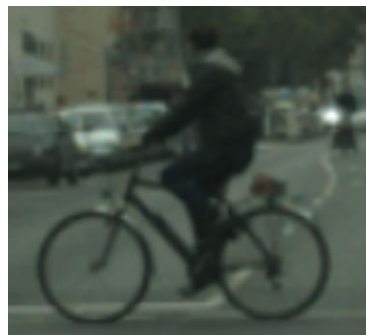


person

bicycle

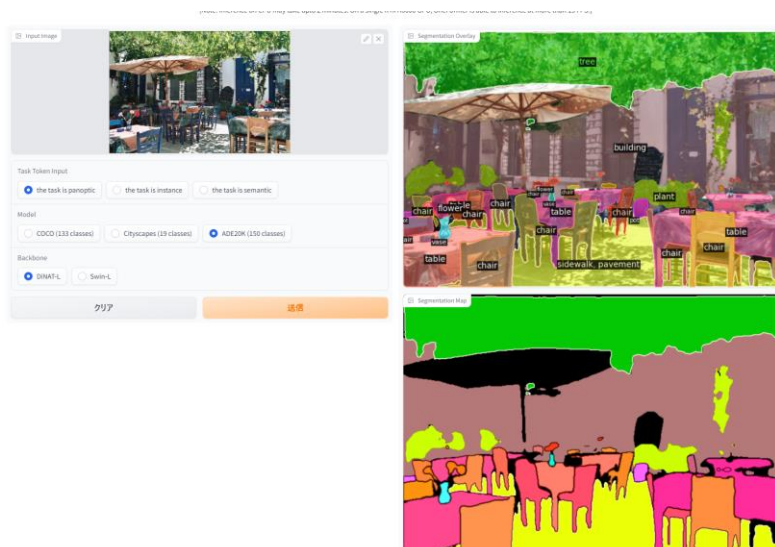
③ セグメンテーション

画素単位で領域を理解



セグメンテーションを試すことができるオンラインのサイト

- OneFormer のデモサイト
- URL: <https://huggingface.co/spaces/shi-labs/OneFormer>
- セグメンテーションの種類：パノプティック, インスタンス, セマンティック
- データセット：COCO (133 クラス) , Cityscapes (19 クラス) , ADE20K (150クラス)



文献: Jitesh Jain, Jiachen Li, MangTik Chiu, Ali Hassani, Nikita Orlov, Humphrey Shi, OneFormer: One Transformer to Rule Universal Image Segmentation, arXiv:2211.06220, 2022.

訓練データにより結果が変わってくる



元画像



訓練データは Cityscapes



訓練データは COCO



訓練データは ADE20K

- OneFormer のデモサイトを使用
- URL: <https://huggingface.co/spaces/shi-labs/OneFormer>
- パノプティック・セグメンテーションを実行
- バックボーンは DiNAT-L を使用

ADE20K

- アノテーション済みの画像データ
- オブジェクト(car や person など) も, 背景領域も (grass, sky など) , 画素単位でアノテーションされている
- 画像数 : 30,574
- クラス数: 3,688

ADE20K データセットの URL:

<http://groups.csail.mit.edu/vision/datasets/ADE20K/>

文献: Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, Antonio Torralba,
Scene Parsing Through ADE20K Dataset,
CVPR 2017, also CoRR, abs/1608.05442, 2017.

CityScapes

- アノテーション済みの画像データ
- 50都市の数ヶ月間（春，夏，秋）の日中，良好な/中程度の天候のもとで撮影，計測
- 画像数：24,998
- クラス数: 30

road, sidewalk, parking, rail track, person, rider, car, truck, bus, on rails, motorcycle, bicycle, caravan, trailer, building, wall, fence, guard rail, bridge, tunnel, pole, pole group, traffic sign, traffic light, vegetation, terrain, sky, ground, dynamic, static

CityScapes データセットの URL:

<https://www.cityscapes-dataset.com/>

文献: Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Frank $\text{\textcircled{e}}$, Stefan Roth, Bernt Schiele,

The Cityscapes Dataset for Semantic Urban Scene Understanding, CVPR 2016, also CoRR, abs/1604.01685, 2016.

COCO

- 画像データ, 人体のランドマーク, 人体姿勢のデータ
- ラベル付け済みの画像数: 200,000以上
- オブジェクトのクラス数: 80
- ランドマーク: 左目、鼻、右腰、右足首などの 17 のキーポイント

COCO データセットのURL: <https://cocodataset.org/>

文献: Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollr,
Microsoft COCO: Common Objects in Context,
CoRR, abs/1405.0312, 2014.

① 次のページで公開されているページを利用

- OneFormer のデモサイト

- URL: <https://huggingface.co/spaces/shilabs/OneFormer>

② 画像は各自で準備

③ このサイトで次を設定できる

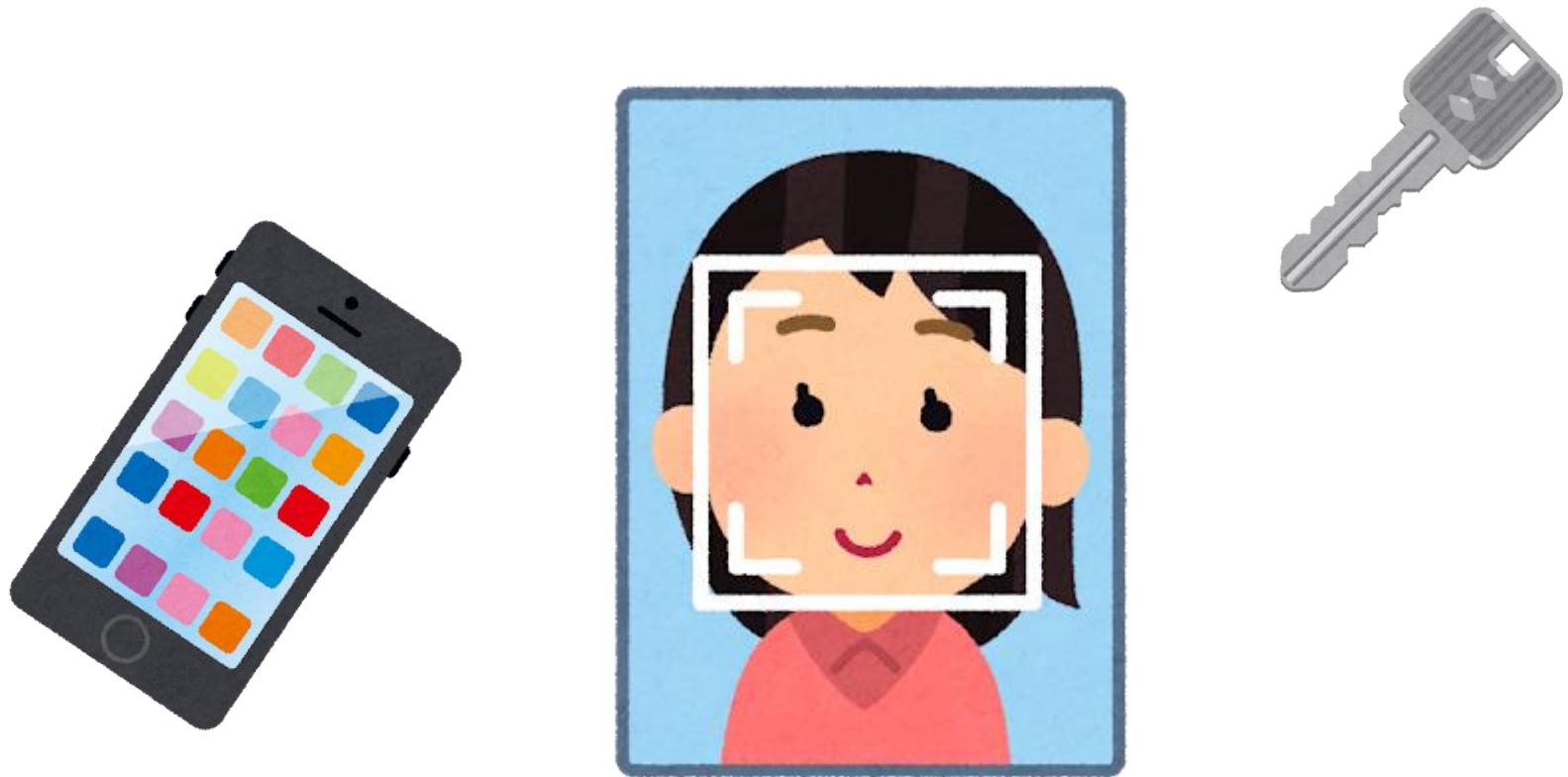
panoptic, instance, semantic (3通り)

COCO, Cityscapes, ADK20K (3通り) の組み合わせ

計**9通り**を試す

④ Backbone のところは「DiNAT-L」で実施

12-3. 顔情報処理



- **顔識別（本人の特定）** を行い，鍵代わりに使用

群衆のカウント

- 群衆のカウント（画像内の人数を数える）
- 監視等に役立つ。



元画像



FIDTM 法による群衆のカウント

FIDTM 法（2021年発表）は、それ以前の手法よりも、さまざまな大きさの顔を精度よく検出できるとされている

顔検出 (face detection)

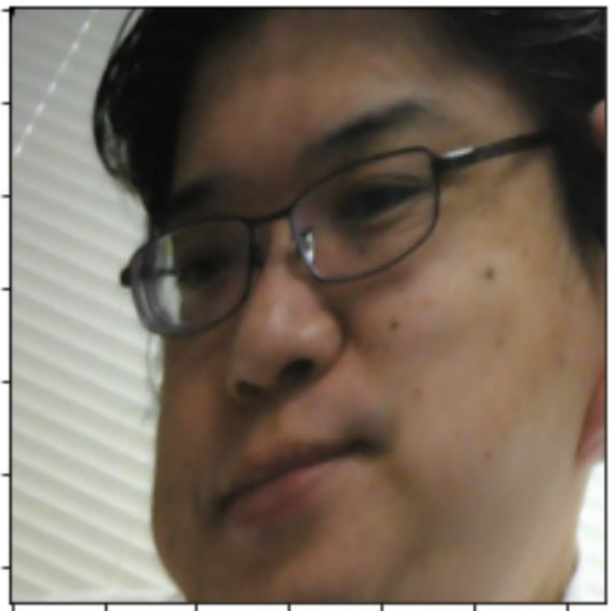
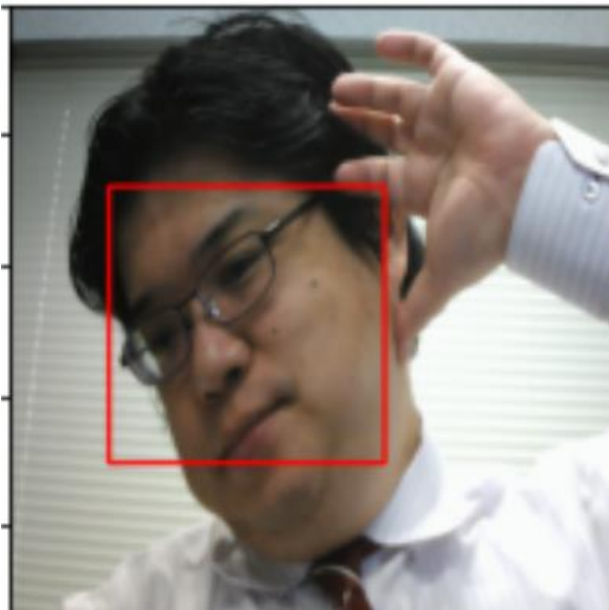


顔検出

写真やビデオの中から、
顔を検出すること

- ・ **顔の場所と領域**を検出
- ・ 複数ある場合にはすべて識別

顔のアライメント



顔のアライメント

顔の向き, 大きさをそろえる

顔ランドマーク (facial landmark)

顔ランドマーク

まゆげ, 目, 鼻, くち,
顔の輪郭など, **顔の目
印となるポイント**

**何をランドマークとす
るかで, さまざま種類
がある**



顔の 68 ランドマーク

顔ランドマークと顔のコード化



顔ランドマーク
(顔の目印となるポイント)

【用途】

顔のコード化,
顔識別 (本人の特定) や顔認識,
年齢の推定, 性別の推定,
表情の推定, 顔の 3次元再構成
など, **さまざまな用途**に

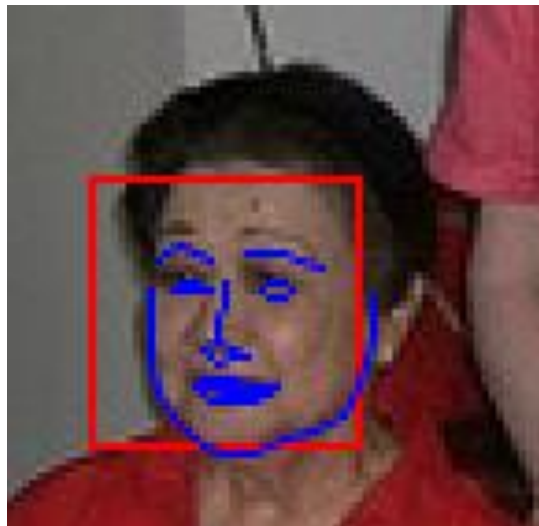


```
[ 0.00455161 0.18412845 0.03993299 -0.00525442 -0.07139583 0.13386713  
-0.00805785 -0.05532961 0.23680274 -0.08285979 0.23483047 0.00531268  
-0.20885959 -0.08143201 -0.07200904 0.08402673 -0.11594398 -0.08790354  
-0.14109443 -0.06426463 -0.01909781 0.04940728 -0.09086579 0.04928832  
-0.21027854 -0.29970455 -0.0172264 -0.11317077 0.01727176 -0.16984728  
-0.09449591 0.13467805 -0.11701771 -0.05477776 -0.00075595 0.02513538  
-0.14347562 -0.03627753 0.16524641 -0.01949969 -0.12205475 0.08029588  
0.0263175 0.27922362 0.26064846 0.00236352 -0.00611193 -0.09485988  
0.21062998 -0.27925995 0.02801728 0.21664803 0.10859946 0.12886725  
0.13740893 -0.2259727 -0.04096841 0.08036234 -0.15912706 0.04860799  
-0.05707375 -0.0537034 -0.0382829 -0.04455713 0.08876641 0.08084331  
-0.08434555 -0.12838244 0.14743124 -0.16956419 0.05809632 0.13801341  
-0.08743402 -0.27705559 -0.17667291 0.12853563 0.40629697 0.25780377  
-0.18405011 0.0308622 -0.04818188 -0.08573647 0.05904602 -0.00438196  
-0.12593931 -0.09314068 -0.07387269 0.06559635 0.18202811 0.02111054  
-0.0110196 0.23951159 -0.00382505 -0.03864232 0.03268222 0.00397856  
-0.16739431 0.05639653 -0.11059792 -0.01466753 0.09382112 -0.13222337  
0.00125152 0.08060078 -0.16358295 0.22464372 -0.00883166 0.03960528  
0.02616193 -0.04999878 -0.14178708 0.06188779 0.25730038 -0.27515596  
0.19043125 0.04089454 0.02481355 0.17113039 0.10915946 0.05673687  
-0.00430411 -0.0710548 -0.14189517 -0.08640159 0.0943445 -0.08253401  
0.04771694 -0.04451045]
```

顔のコード化
顔による本人確認や,
個人の特定の基礎

顔のコード化

- **顔のコード**は、複数の数値（ふつう100以上）の組み合わせ
- **顔画像**から、**顔ランドマーク**を求め、顔のコードを得る
- **さまざまな用途**：**顔識別（本人の特定）**，**顔認識**，**年齢の推定**，**性別の推定**，**表情の推定**，**顔の3次元再構成**など



顔検出，
顔ランドマーク



```
0.0512202  
0.0150111  
0.0642803  
0.0438789  
0.0485441  
0.0107222  
0.0653341  
0.144676  
0.156388  
0.12738  
0.137432  
0.059849  
0.1569  
0.0690487  
0.0250859  
0.215287  
0.134682  
0.212719  
0.0921698  
0.019872  
0.0154232  
0.0199377  
0.0035686  
0.0199529
```

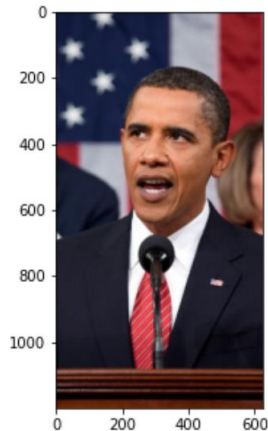
数値化 40

顔検証 (face verification)

requirement already satisfied: error = 0
[True]



requirement already satisfied: error = 19.7
[False]



同一人物である

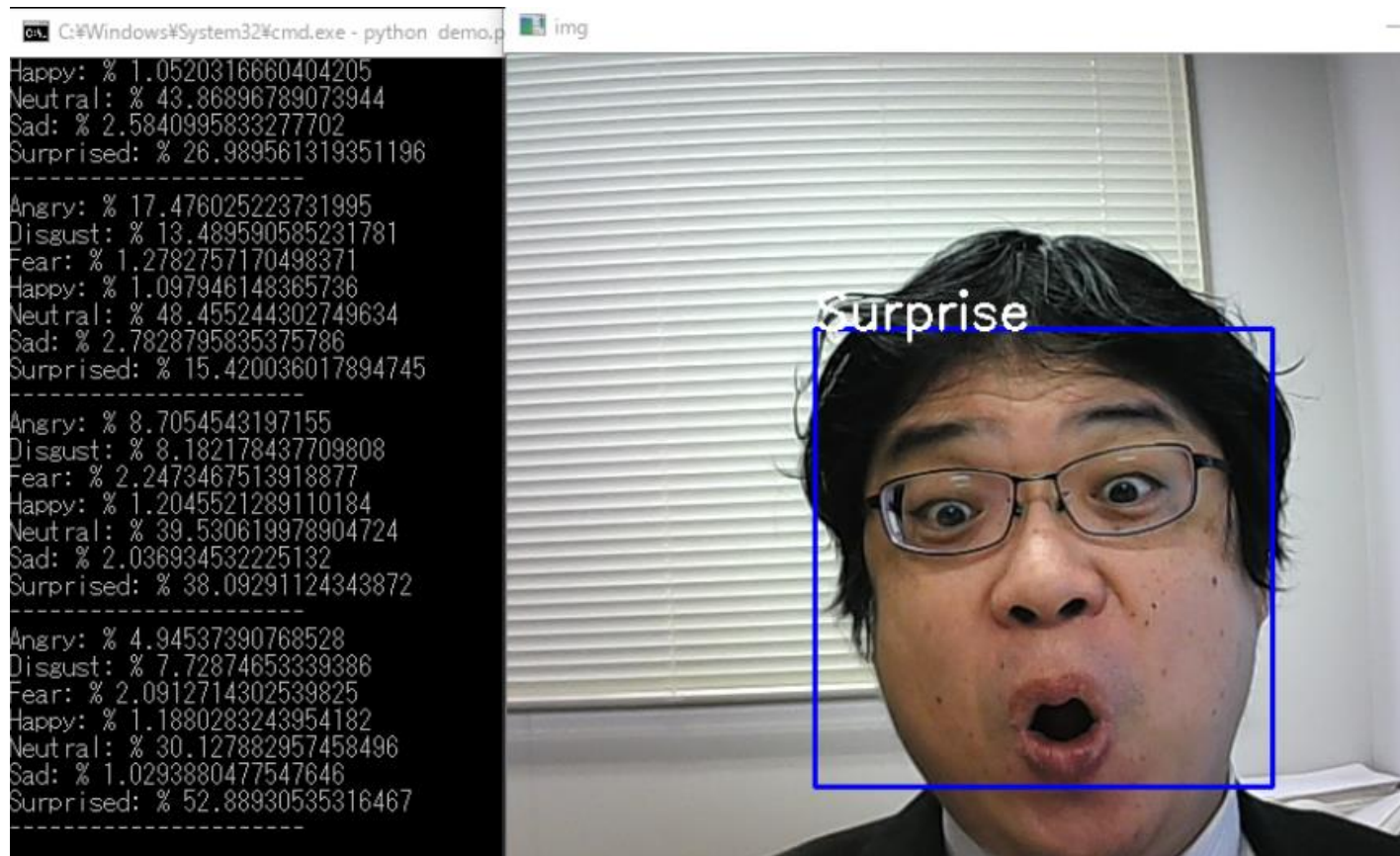
同一人物でない

顔検証

2つの別の写真あるいはビデオを照合し、同一人物であるかを判定すること

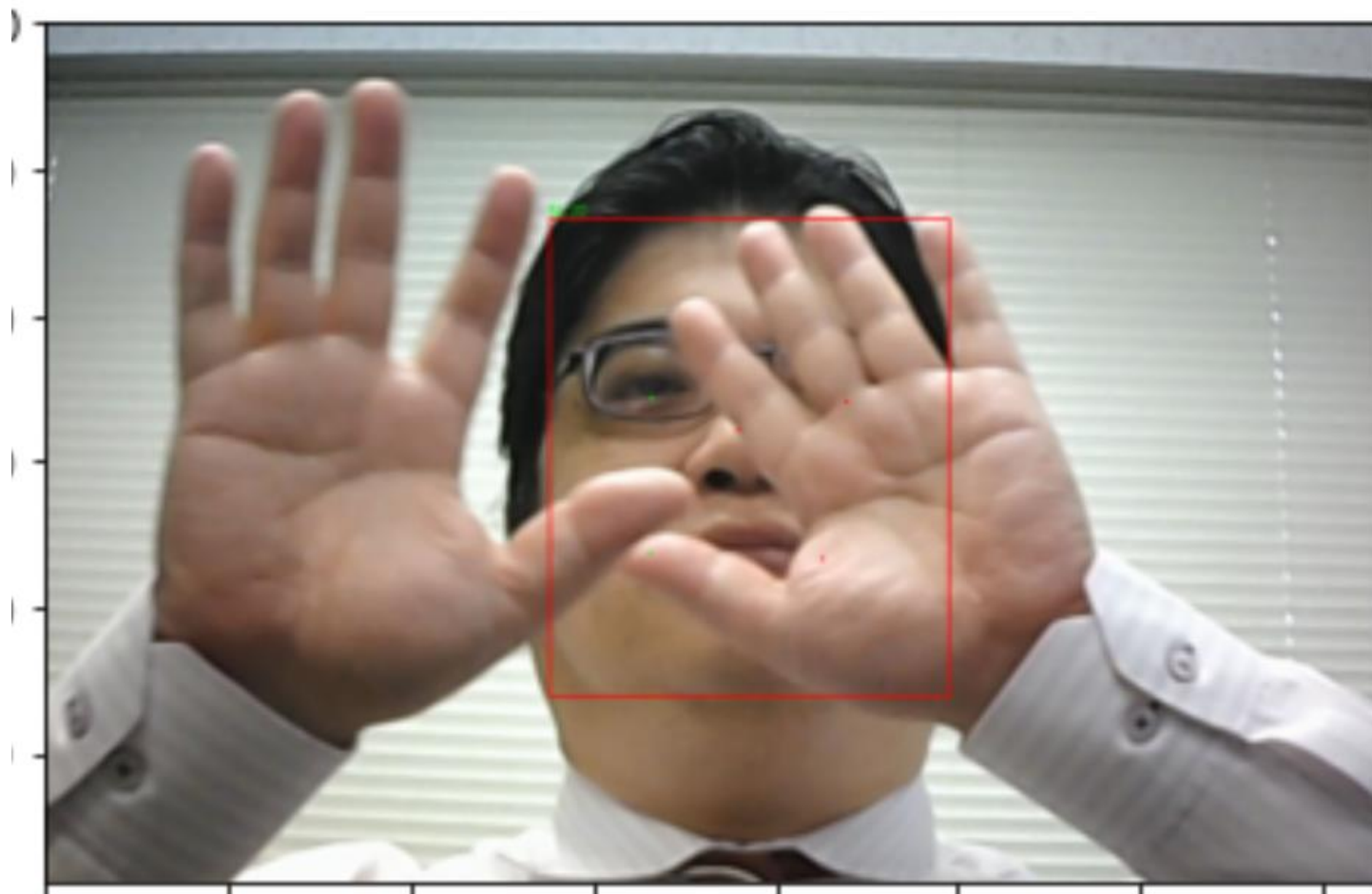
※ **顔識別 (本人の特定)** や **顔認識** は類似の処理

表情の推定



- ここでは、**7種の表情** Angry, Disgust, Fear, Happy, Neutral, Sad, Surprised の**それぞれの確率**を推定
- <https://github.com/ezgiakcora/Facial-Expression-Keras>で公開されている成果物

年齢の推定, 性別の推定



'gender' : 1, 'age' : 30, 'e

顔情報処理のバリエーション

- 顔検出
- 顔のアライメント
- 顔ランドマーク検出
- 顔のコード化
- 顔検証
- 顔識別（本人の特定）や顔認識
- 年齢の推定，性別の推定，表情の推定
- さまざまな応用
 - 群衆のカウント，マスク有り顔に関する処理，
 - 顔の3次元再構成など



演習

顔情報処理のオンラインデモ

【トピックス】

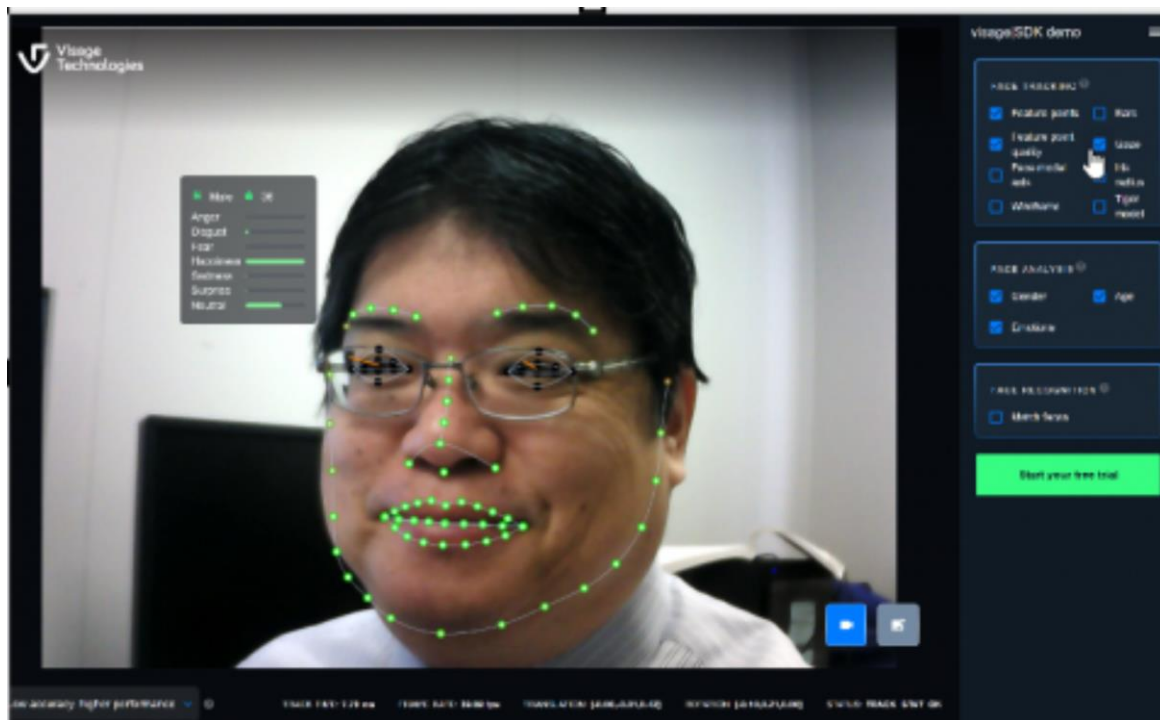
- 顔検出
- 顔ランドマーク
- 表情推定
- 顔の向き推定

オンラインデモ

URL: <https://visagetechologies.com/demo/>

【使い方】

- ① カメラ付きのノートパソコンを準備
- ② 上の URL を Web ブラウザで開き, 「Try out demo」 をクリック



顔検出,
顔ランドマーク,
表情推定,
顔の向き推定

全体まとめ

①新しいデータや画像の生成

実在しない人間の顔画像を生成したり、低解像度の画像を高解像度に変換したり、プロンプトをもとに画像に生成したりなど

②画像のセグメンテーション（領域分割）

パノプティックセグメンテーションやインスタンスセグメンテーション、セマンティックセグメンテーションなど

③顔情報処理

顔検出、顔のアライメント（顔の向きや大きさの補正）、顔ランドマーク（顔の目印となるポイント）の検出、顔のコード化（顔の特徴を数値化）、顔検証（同一人物かの判定）、年齢や性別、表情の推定など

オンラインデモや公開されているモデルを利用することで、実際にこれらの技術を試すことも可能

授業の学ぶ意義と満足感

- 最新のAI技術の実例。ディープラーニングの可能性と有用性を実感。学ぶ楽しさと能力の成長。
- AIアルゴリズムの仕組みや応用範囲を広く知る
- オンラインデモで最先端技術を体験的に理解することで、面白さと奥深さを実感でき、学習意欲が高まる。
- ディープラーニングの応用について、視野を広げ、実力を身につけ、AI時代に求められる人材となる。