

# 7. ディープラーニングでの画像理解② (セグメンテーション、NoShot 学習)

(ディープラーニング, Python を使用)  
(全 15 回)

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

金子邦彦





- ①テクノロジーへの洞察
- ②問題解決能力の向上
- ③実用的なスキルの獲得



# アウトライン

1. イントロダクション
2. 画像セグメンテーションの基礎
3. 画像セグメンテーションの技術
4. 画像セグメンテーションのバリエーション
5. 画像の前処理とデータ拡張
6. NoShot 学習

# Google Colaboratory



Colaboratory へようこそ

Colaboratory は、完全にクラウドで実行される Jupyter ノートブック環境です。設定不要で、無料でご利用になれます。

```
x = [5, 4, 1, 3, 2]
for i in x:
    print(i * 120)
```

600  
480  
120  
360  
240

PRO ファイル 編集 表示 挿入 ランタイム

+ コード + テキスト | ドライブに...

```
[1] x = 100
```

```
if (x > 20):
    print("big")
else:
    print("small")
```

big

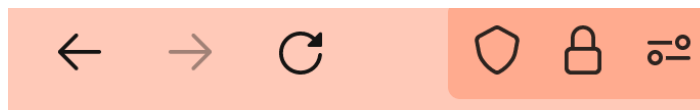
```
s = 0
for i in [1, 2, 3, 4, 5]:
    s = s + i
print(s)
```

15

URL: <https://colab.research.google.com/>

- オンラインで動く
- Python のノートブックの機能を持つ
- Python や種々の機能がインストール済み
- 本格的な利用には, Google アカウントが必要

# Google Colaboratory の全体画面



Colab の定期購入を最大限に活用する  
ファイル 編集 表示 挿入 ランタイム ツール ヘルプ

メニュー

+ コード + テキスト

コードセル, テキストセル  
の追加



メニュー

(目次, 検索と置換,  
変数, ファイル)

1. 変数

```
[2] x = 100  
y = 200
```

2. 式

```
▶ print(x + y)  
print(3 * x + y)  
  
300  
500
```

3. 条件分岐

```
[4] if (x > 50):  
    print('big')  
else:  
    print('small')  
  
big
```

コードセル,  
テキストセルの  
並び

Web ブラウザの画面

# Google Colaboratory のノートブック



## コードセル, テキストセルの2種類

- **コードセル** : Python プログラム, コマンド, 実行結果
- **テキストセル** : 説明文, 図

2. 式

← テキストセル

```
[5] print(x + y)
     print(3 * x + y)
```

← コードセル

```
300
500
```

3. 条件分岐

← テキストセル

```
▶ if (x > 50):
    print('big')
else:
    print('small')
```

← コードセル

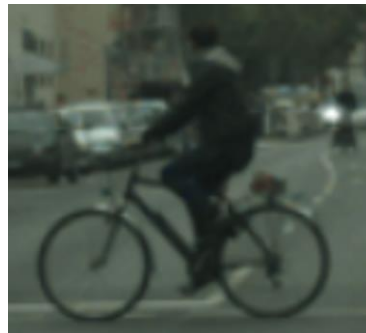
```
big
```

# 7-1. イントロダクション

# 画像理解の主な種類

## ① 画像分類

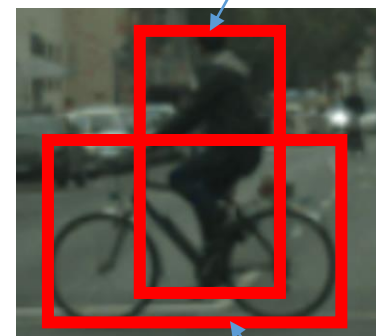
「何があるか」を理解



person  
bicycle

## ② 物体検出

場所と大きさも理解

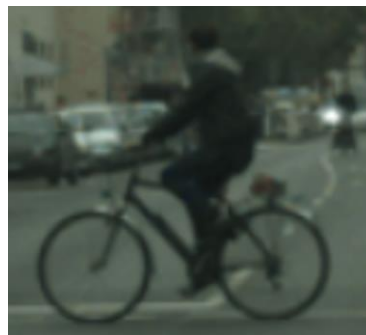


person

bicycle

## ③ セグメンテーション

画素単位で理解





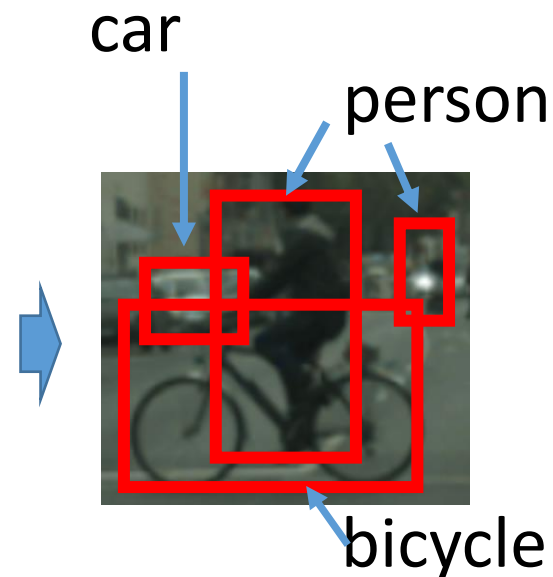
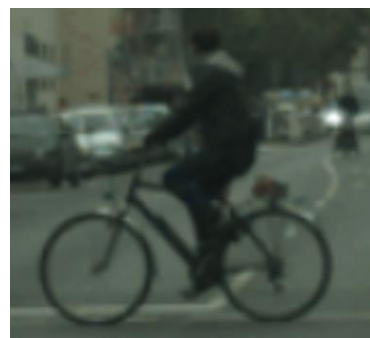
# ① 画像分類



```
Score 0.9827020168304443, Label lab_coat  
Score 0.0030872616916894913, Label syringe  
Score 0.0024311079178005457, Label beaker  
Score 0.0016609227750450373, Label stethoscope  
Score 0.00037950885598547757, Label plate
```

**画像分類の結果は、ラベルと確率**  
※ 5つの候補 (top 5) が表示されている

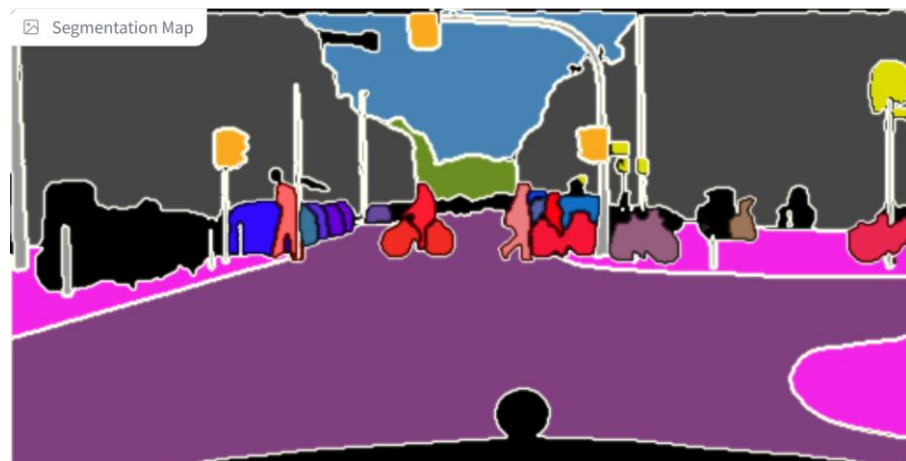
## ② 物体検出



バウンディングボックス,  
ラベルを得る

**バウンディングボックス**は、  
物体を囲む最小のボックス（四角形）

### ③ 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

# ここまでのまとめ

## 画像理解の主な種類

### • 画像分類

「何があるか」を理解。結果は「ラベル」として識別  
各ラベルに対する「確率」も提供

### • 物体検出

物体の種類、場所、大きさを理解。場所と大きさについての結果は、物体を囲むバウンディングボックス。

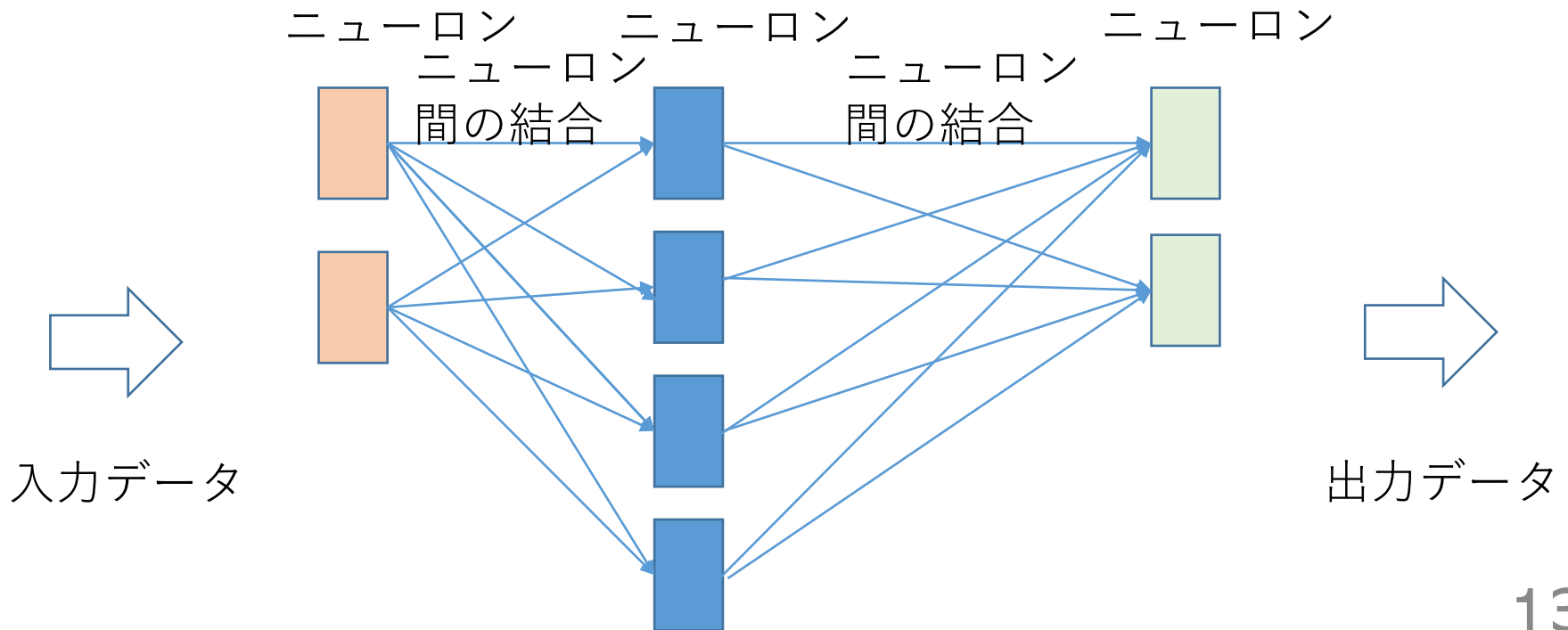
### セグメンテーション

画素単位で理解。

物体の「形」を詳細に抽出。

# ニューロンとニューラルネットワーク

- **ニューロン**は、**ニューラルネットワーク**の基本的な構成要素
- 一つ一つの**ニューロン**は、データの受け取り、処理、伝達を行う
- **ニューラルネットワーク**は、これらの**ニューロン**が多数組み合わさったもの

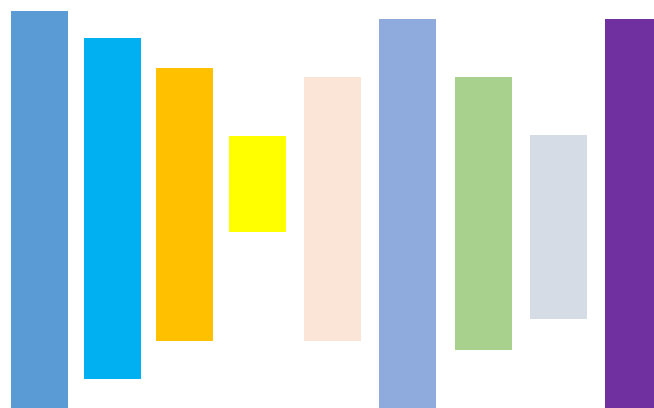


# ディープラーニング

ディープラーニングに「ディープ」とついているのは、多層のニューラルネットワークを使用するため



層の数が少ない



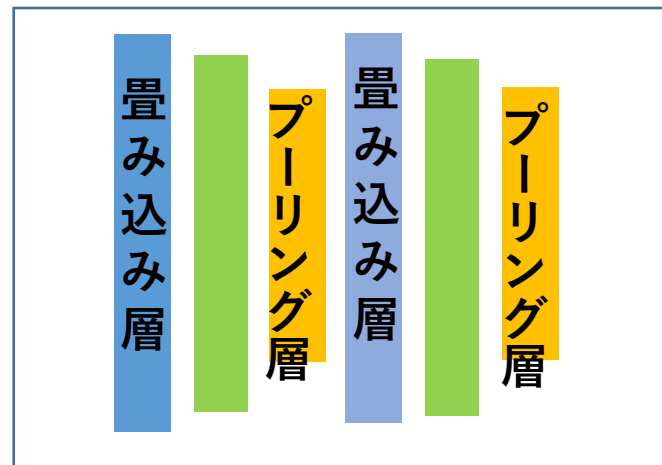
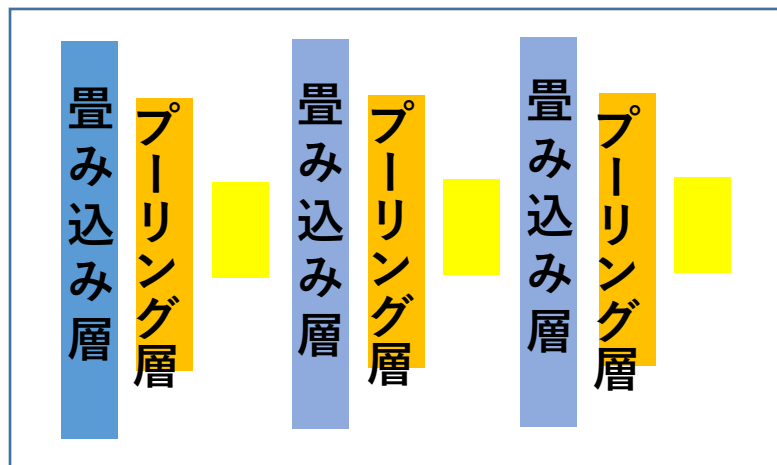
層の数が多 (ディープ)

# 畳み込みニューラルネットワーク (CNN)



- **畳み込みニューラルネットワーク (CNN)** は**画像理解**や**画像の分析**に特化した**ディープラーニング**の一種。
- CNNは主に**畳み込み層**、**プーリング層**、**全結合層**の3種類  
(注) これら3種類以外のもさまざまある
- **畳み込み層**：画像の**局所的な特徴をとらえる**ための層。**特徴**は、画像内の**顕著なパターン**や**属性** (例：エッジ、テクスチャ)
- **プーリング層**：**特徴マップのサイズを縮小**するための層。**過学習を防止**。**計算効率を向上**
- **全結合層**：全ニューロンが前の層のすべてのニューロンと接続された層。畳み込み層とプーリング層を通過した後の特徴を基に、**画像の分類**や**回帰**を行う

# 畳み込みニューラルネットワーク (CNN) の例



さまざまなバリエーション



# 畳み込みニューラルネットワーク (CNN) のメリット



- **画像理解、画像の分析の高精度化**
- **画像の特徴を自動的に抽出**
- **幅広い分野での応用**

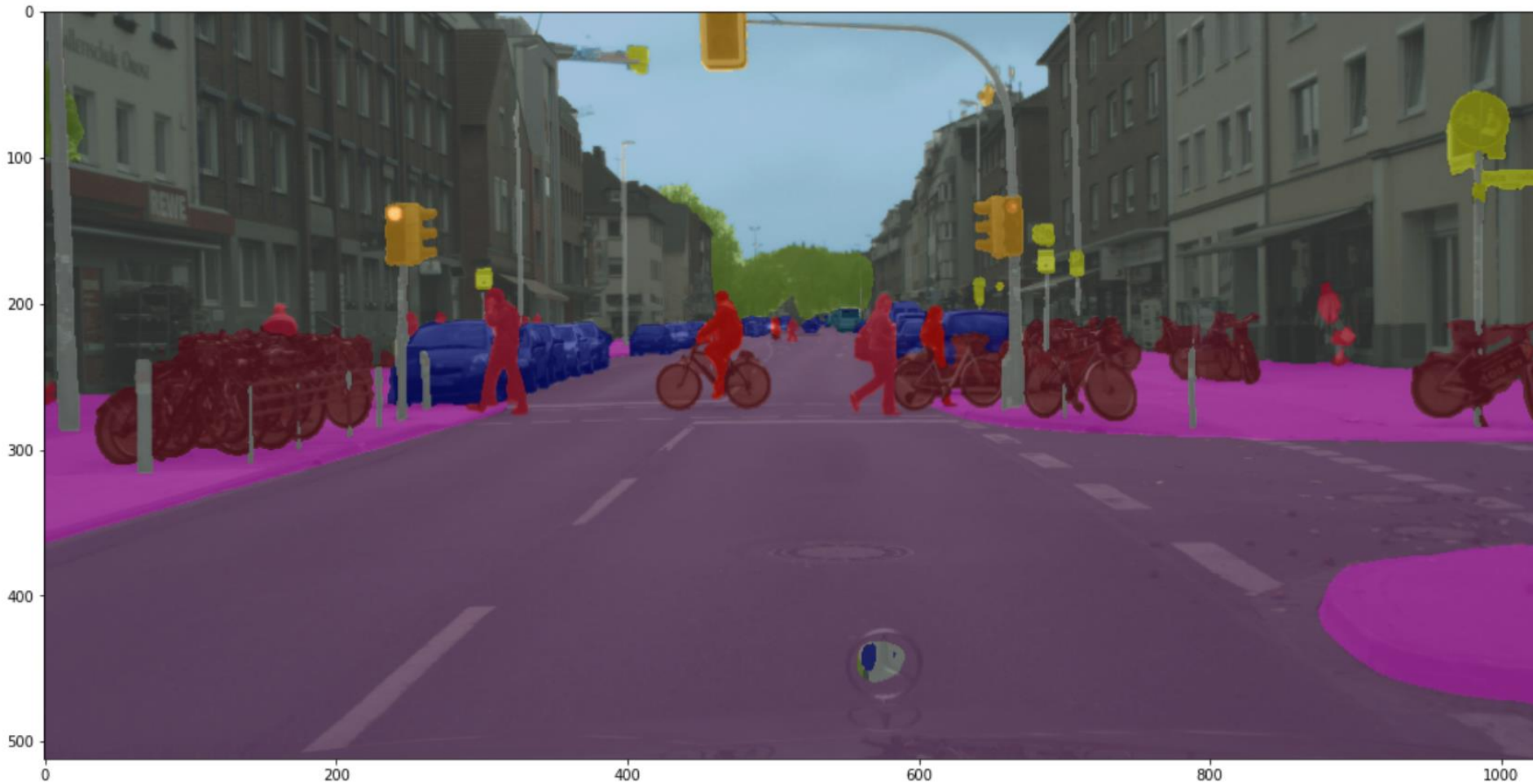
## 7-2. 画像セグメンテーションの基礎

# 画像セグメンテーションの例



画像セグメンテーションは、**画像の画素単位で、種類を識別**

# 画像セグメンテーションの例



# 画像セグメンテーション



## セグメンテーションの目的

- 画像を、**意味のある領域ごと**に分割
- 物体の「形」を詳細に解析

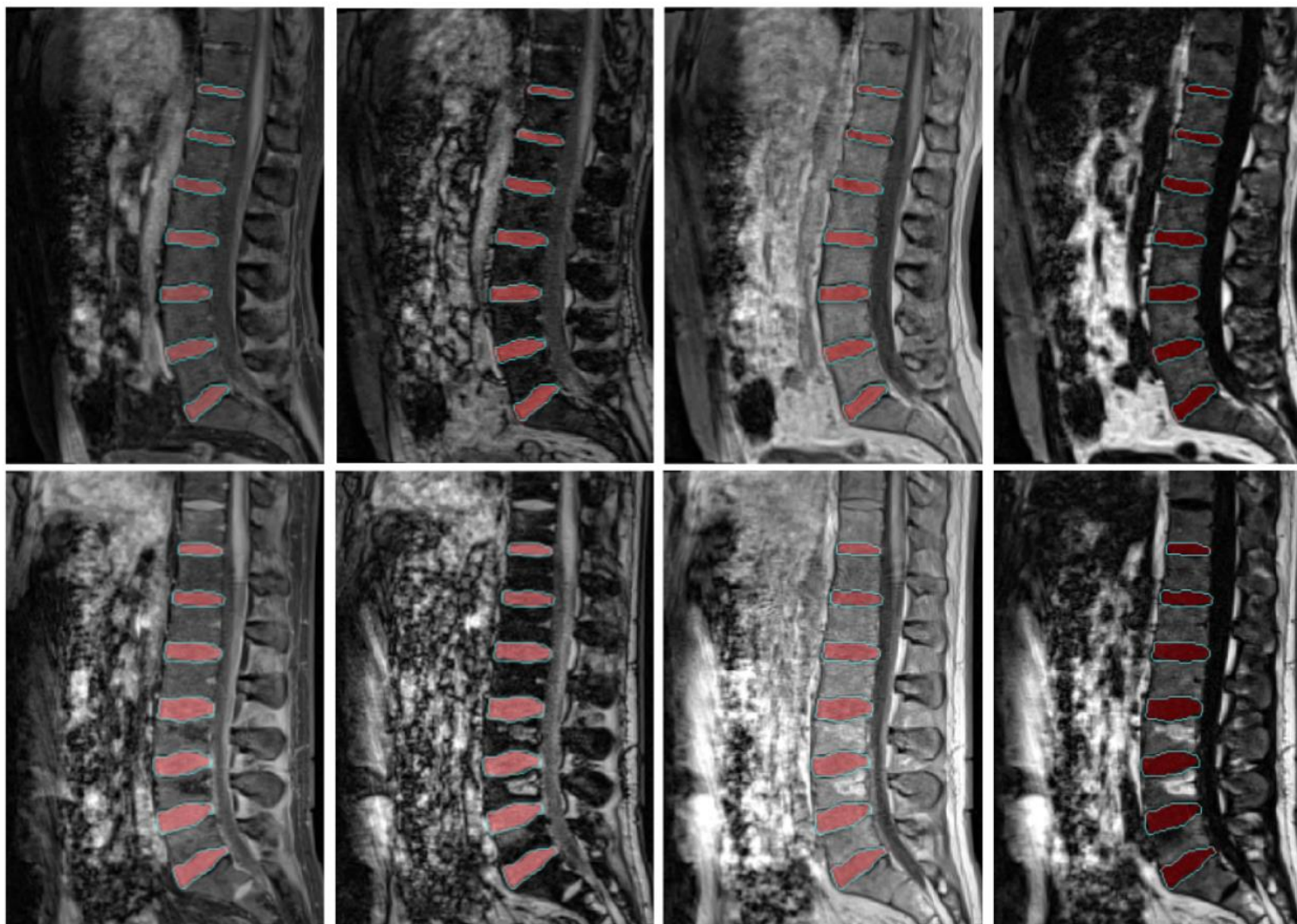
## 機能

- **同じ種類の画素**が同じラベルに分類される

## セグメンテーションの利点

- 物体が**画像のどこにあるか**を正確に知ることができる
- 物体の**形や大きさ**を**数値化**するための基礎

# 医療画像のセグメンテーション



Water

Out-Phase

In-Phase

Fat

# 自動運転での画像セグメンテーション

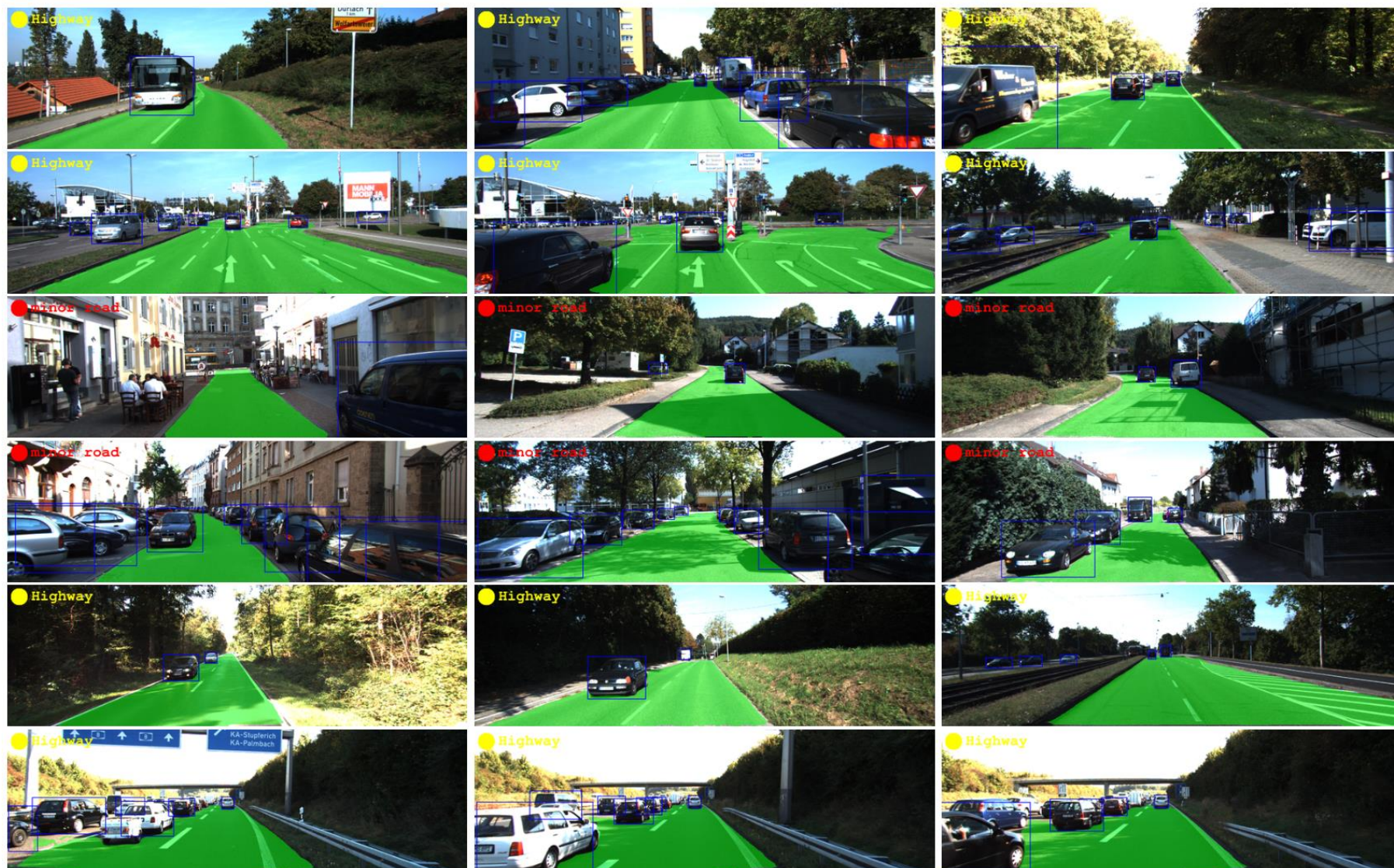
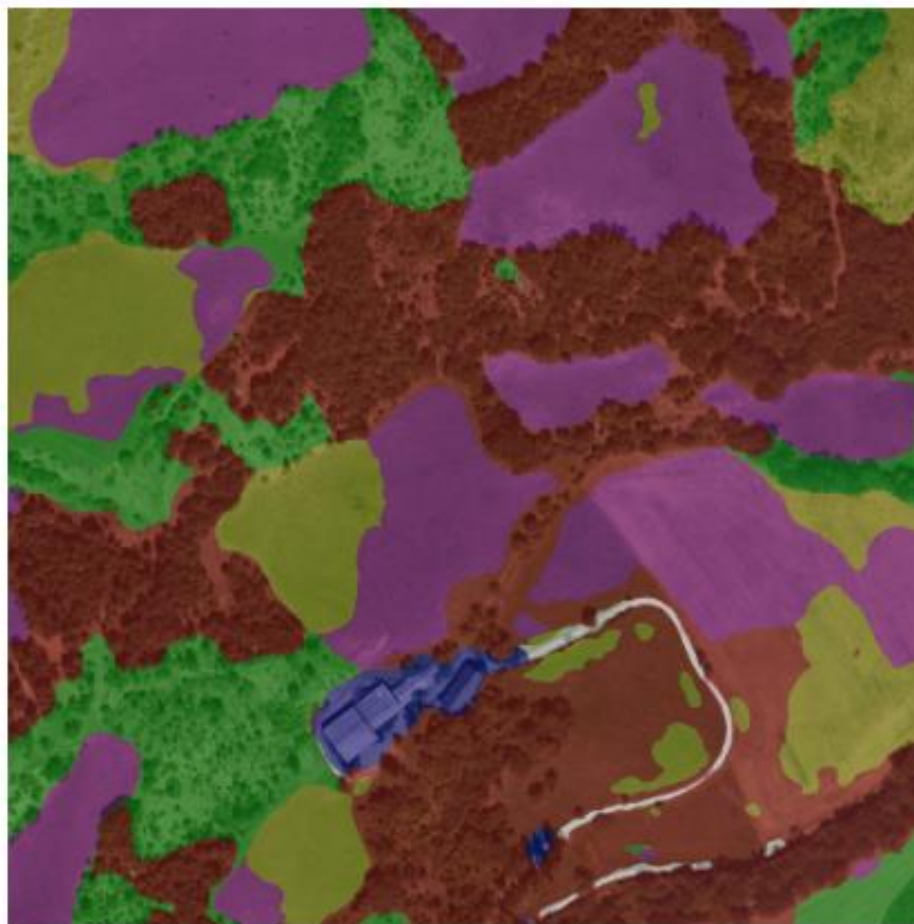


Figure 6: Visualization of the MultiNet output.

<https://arxiv.org/pdf/1612.07695v2.pdf> より転載

# 衛星画像のセグメンテーション



■ Dense forest | ■ Sparse forest | ■ Moor | ■ Herbaceous formation | ■ Building | □ Road

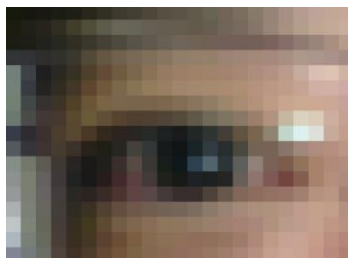
(■ No information)

<https://arxiv.org/pdf/2110.05812.pdf> より転載



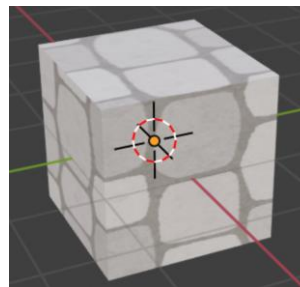
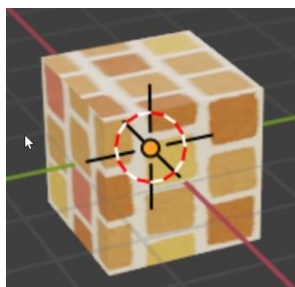
## 画素

- 画素は、画像を構成する最小の単位。各画素には、色、明るさの情報がある



## テクスチャ

- 画像内のパターン。セグメンテーションの手がかりとして利用される



## ラベル付け

- セグメンテーションでは、各画素に、**識別のための名前**として**ラベル**を割り当てる

## 演習 1

オンラインデモによる  
画像セグメンテーションの体験

【トピックス】

- 画像セグメンテーション
- OneFormer のデモ

# ① OneFormer のデモページを開く

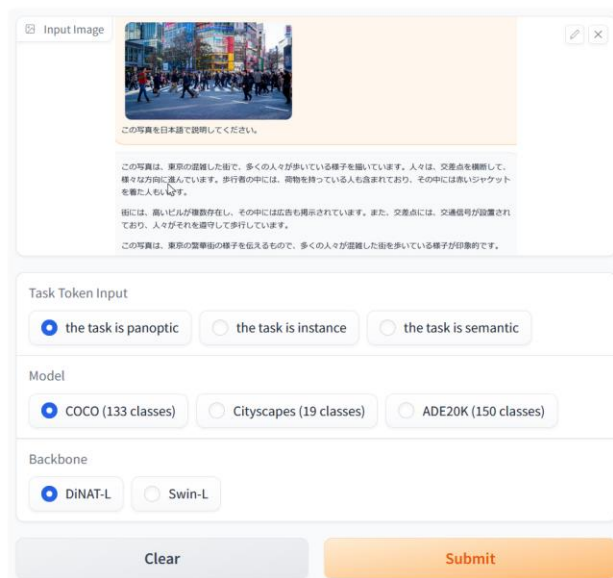


<https://huggingface.co/spaces/shi-labs/OneFormer>

オンラインサービスであり、混雑時などは動かない場合がある。授業中で動かなかった場合には、後日試してほしい

② 画面の「Input Image」で画像ファイルを設定するか、下の「Examples」で画像を選ぶ

③ 「Submit」 をクリック。結果を確認。



## 画像セグメンテーション

- 画像セグメンテーションは、**画像の画素単位で、種類を識別**

## セグメンテーションの目的

- 画像を、**意味のある領域ごと**に分割
- 物体の「形」を詳細に解析

## セグメンテーションの利点

- **物体が画像のどこにあるか**を正確に知ることができる
- **物体の形や大きさを数値化**するための基礎

## 7-3. 画像セグメンテーションの技術

# 画像セグメンテーションの手法



## ディープラーニングによる画像セグメンテーション

- FCN (Fully Convolutional Network), 2015年発表
- U-Net, 2015年発表

## ディープラーニング利用の特徴

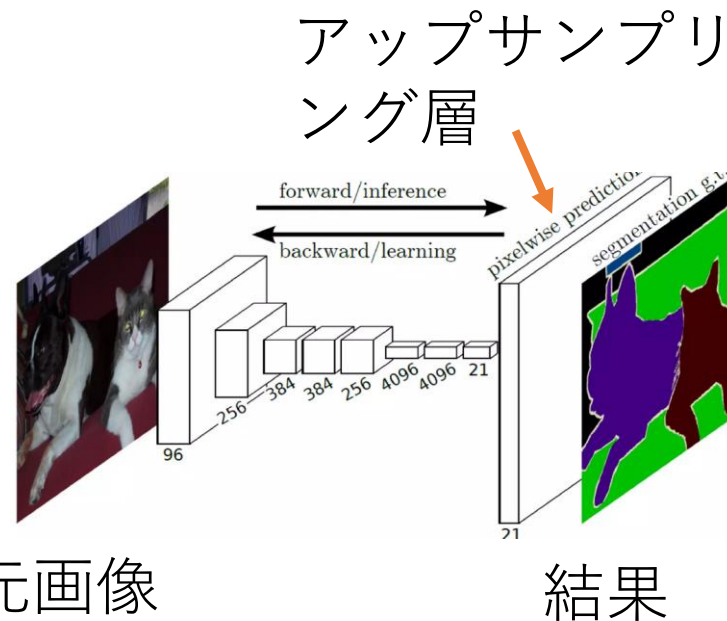
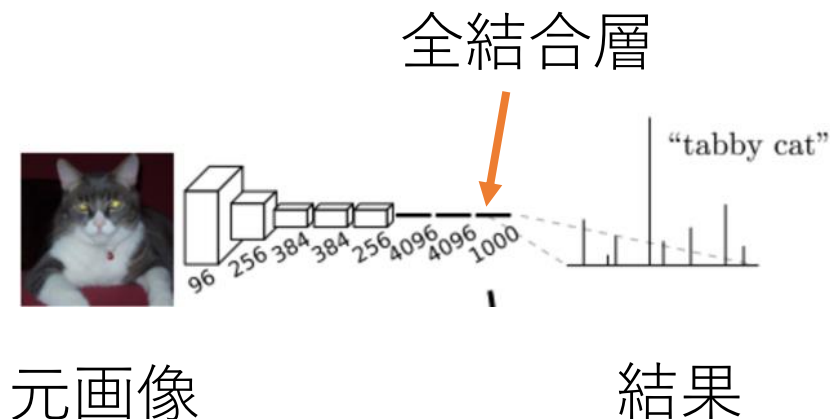
- 精度の向上
- 学習を行うことで、さまざまな種類に対応可能
- 教師あり学習であり、訓練データが必要

- 畳み込みニューラルネットワークの改良
- 画素ごとに結果を出力する

通常のニューラルネットワーク	<b>画像全体</b> に対する結果を出力 (画像分類など)
FCN	<b>画素ごと</b> に結果を出力 (セグメンテーションなど)

# FCN の仕組み

- 畳み込み層、プーリング層、アップサンプリング層で構成
- アップサンプリング層を持つことが特徴である。アップサンプリング層では、特徴マップを、元の入力画像の大きさまで拡大



畳み込みニューラルネットワークによる画像分類

FCN による画像セグメンテーション



# FCN まとめ

- **畳み込み層**: 局所的特徴抽出
- **プーリング層**: 画像縮小と過学習防止
- **アップサンプリング層**: 畳み込み層とプーリング層で得られた特徴マップを、**元の入力画像の大きさまで拡大**

## FCNのアイデア

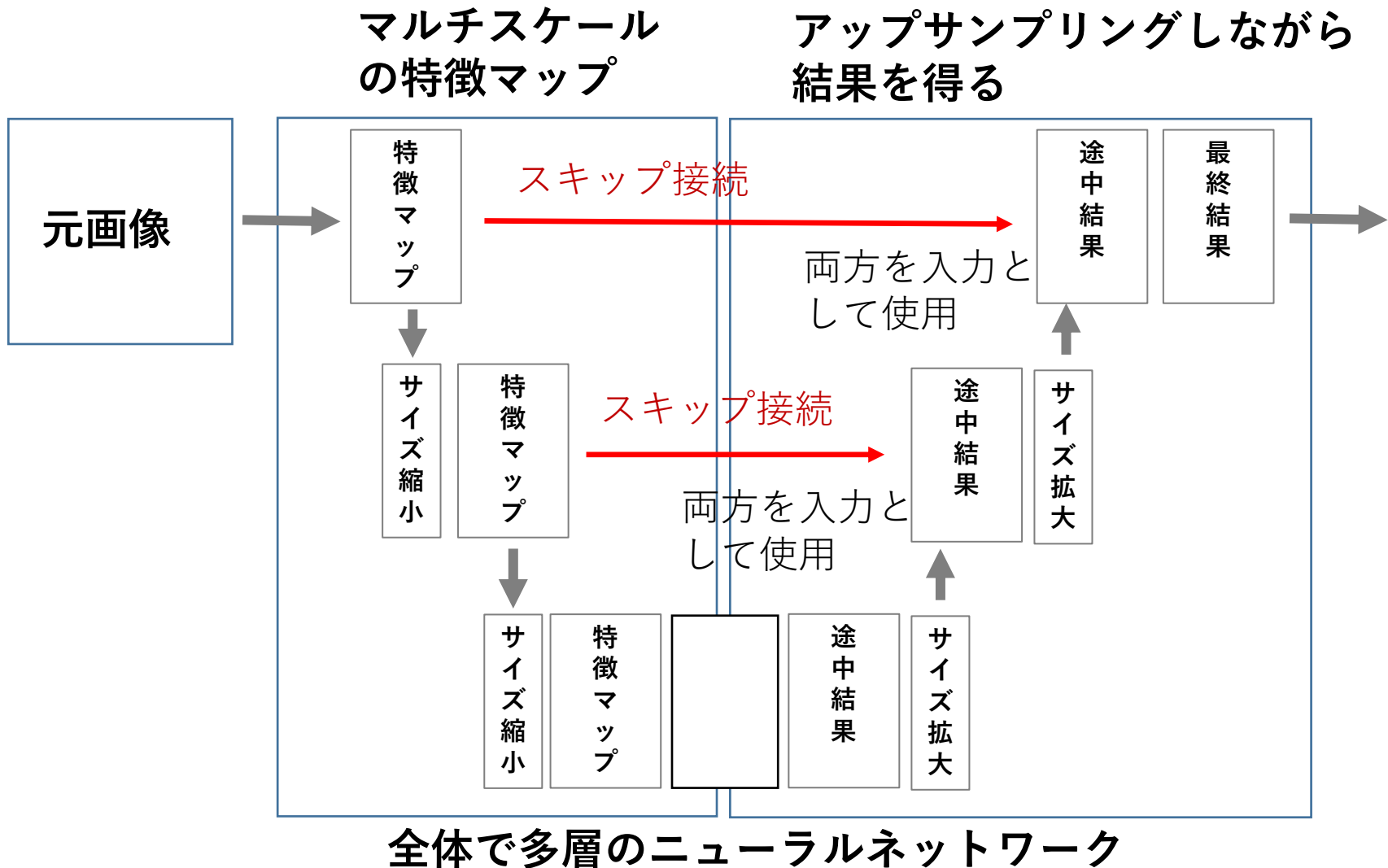
全結合層を使用せずに、アップサンプリング層によって、画素ごとの結果出力を可能に

# U-Net

- FCN の改良
- **ダウンサンプリングとアップサンプリングの組み合わせ**
  - **ダウンサンプリング**（プーリング層）：  
画像全体のサイズ縮小により、大域的な特徴を捉えることを可能に。
  - **アップサンプリング**（アップサンプリング層）：  
特徴マップのサイズ拡大で、低解像度の特徴マップと、高解像度との特徴マップを統合
- **スキップ接続**  
ダウンサンプリングにおいて、**輪郭情報が失われるという問題を解決.**

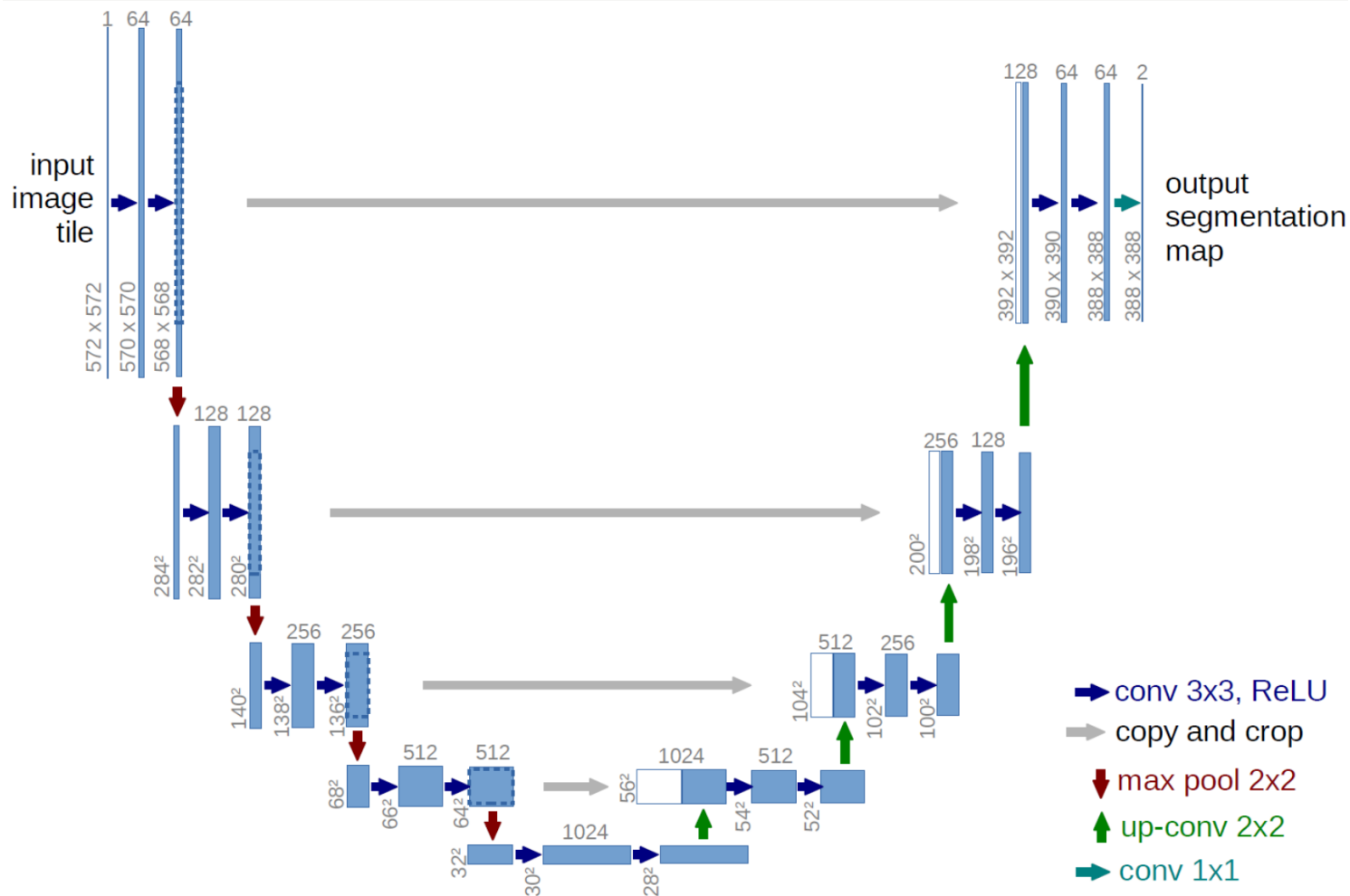
# U-Net の仕組み

- ダウンサンプリング：サイズ縮小。
- アップサンプリング：サイズ拡大。

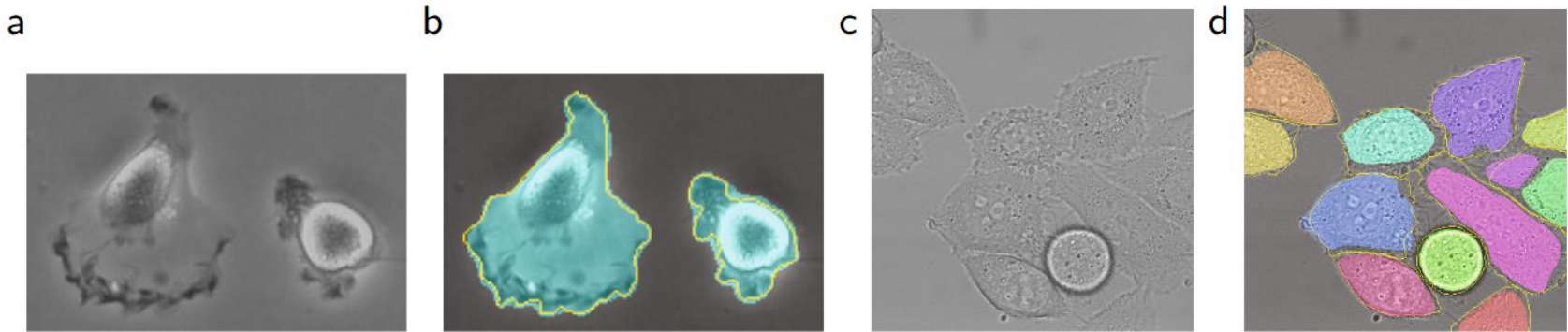


# U-Net の仕組み

青矢印：畳み込み層、赤矢印：プーリング層、  
緑矢印：アップサンプリング層



# U-Net によるセマンティック・セグメンテーション



**Fig. 4.** Result on the ISBI cell tracking challenge. (a) part of an input image of the “PhC-U373” data set. (b) Segmentation result (cyan mask) with manual ground truth (yellow border) (c) input image of the “DIC-HeLa” data set. (d) Segmentation result (random colored masks) with manual ground truth (yellow border).

細胞のモノクロ画像のセグメンテーション

# ここまでのまとめ

## 畳み込みニューラルネットワークを利用した画像セグメンテーション

- FCN (Fully Convolutional Network) - 2015年発表
- U-Net - 2015年発表

## ディープラーニングによる画像セグメンテーションの特徴

- 精度の向上
- 学習を行うことで、**さまざまな種類に対応可能**
- 教師あり学習であり、訓練データが必要

## FCNの概要

- 畳み込みニューラルネットワークの改良
- **画素レベルでの結果出力**
- 畳み込み層、プーリング層、アップサンプリング層から構成

## U-Netの概要

- FCN の改良
- スキップ接続を使用し、輪郭情報の損失を防ぐ

## 7-4. 画像セグメンテーションのバリエーション

# セグメンテーションの種類



## セマンティック・セグメンテーション

画像内のすべての画素にラベルを付ける。  
「個々の物体を検出する」という概念はない



## インスタンス・セグメンテーション

物体を検出する。同じ種類の複数の物体がある場合は、別々のものとして認識。各物体に対してセグメンテーションを実施。

「物体として識別できない部分は結果がない」ということもある



## パノプティック・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と  
インスタンス・セグメンテーションの同時実行



# インスタンス・セグメンテーションの例

画像を、インスタンス（物体）に分けセグメンテーションを実行



# パノプティック・セグメンテーションの例

物体ではないエリアも対象である

(インスタンス・セグメンテーションでは、このようにはならない)



物体と物体の切れ目が判別できている

(セマンティック・セグメンテーションでは、このようにはならない)



## 演習 2

### インスタンス・セグメンテーションのプログラム例

#### 【トピックス】

- インスタンス・セグメンテーション
- **Google Colaboratory**

# ① Google Colaboratory のページを開く



[https://colab.research.google.com/drive/1Ri1v\\_O4A5vXRSSEBgt4umKC1I3SoWp9U?usp=sharing](https://colab.research.google.com/drive/1Ri1v_O4A5vXRSSEBgt4umKC1I3SoWp9U?usp=sharing)

# ② インスタンス・セグメンテーションのプログラムや説明や実行結果が掲載されている

2人の人間は「違うもの」として識別される



## 演習 3

### 画像セグメンテーションの バリエーション

【トピックス】

- 画像セグメンテーション
- OneFormer のデモ

① OneFormer のデモページを開く

<https://huggingface.co/spaces/shi-labs/OneFormer>

オンラインサービスであり、混雑時などは動かない場合がある。授業中で動かなかった場合には、後日試してほしい

② 画面の「**Input Image**」で画像ファイルを設定するか、下の「**Examples**」で画像を選ぶ

③ 今後は、「Task Token Input」のところ、種類を選ぶ

the task is panoptic: **パノプティック・セグメンテーション**

the task is instance: **インスタンス・セグメンテーション**

the task is semantic: **セマンティック・セグメンテーション**

Task Token Input

the task is panoptic     the task is instance     the task is semantic

④ 「**Submit**」 をクリック。結果を確認。

# 次のような結果が得られる



元画像



インスタンス・セグメンテーション



パノプティック・セグメンテーション

- OneFormer のデモサイトを使用
- URL: <https://huggingface.co/spaces/shi-labs/OneFormer>
- 訓練データは COCO
- バックボーンは DiNAT-L を使用



セマンティック・セグメンテーション

# 自習

次のページにパノプティック・セグメンテーションのプログラムを掲載している。余裕のある人は各自で確認

目的：AIについて理解を深める

使用するページ:

[https://colab.research.google.com/drive/1xWaQuJt50LqYwyw9ohsYERZ\\_lx1gy1rN?usp=sharing#scrollTo=az0NragleQUI](https://colab.research.google.com/drive/1xWaQuJt50LqYwyw9ohsYERZ_lx1gy1rN?usp=sharing#scrollTo=az0NragleQUI)



# セグメンテーションのバリエーション

- セマンティック・セグメンテーション

「空」、「道路」などの広い領域の解析

- インスタンス・セグメンテーション

物体の形状の分析、数の分析、変化の分析

- パノプティック・セグメンテーション

複雑なシーンにおいて、セマンティックとインスタンスの両方の情報を同時に取得

## 7-5. 画像の前処理とデータ拡張

# ディープラーニングでの画像データの前処理



ディープラーニングモデルの学習効率と精度を最適化するため、画像データの前処理は重要

## 画像の前処理

- ノイズ除去
- コントラスト調整
- 学習に使う画像のサイズ（縦、横）をそろえる



左：調整前  
右：調整後

- 実データをもとに、多様なデータを合成し、訓練データの量を増やす
- ディープラーニングの過学習の抑制を目的とする

## 【画像データのデータ拡張】

- **回転**：さまざまな角度で回転
- **反転**：水平または垂直に画像を反転
- **切り取り（クロッピング）**：部分を切り取る
- **色調**：明るさ、コントラストなどを変更

# 画像データの拡張の例



多様なデータを合成



## 7-6. NoShot 學習

# イントロダクション



## 教師あり学習

- 教師あり学習は、**訓練データ**に**正解**が含まれる

例：人間、自動車、信号機

## 訓練データの準備

画像や画像内の物体や画素に、人間、自動車、信号機といった具体的なラベルを付ける。

## 従来の学習方法の特徴

- **学習時に使用されたラベル**（例：人間、自動車、信号機）  
**の範囲内でのみ識別・分類が可能**

## NoShot 学習の特徴

- 広範な出力能力
- **学習時のラベルにはない新しいラベルに対しても、識別・分類が可能。**



# NoShot のセグメンテーションの例



Text Prompt

hard object

画像と英語のプロンプトを  
AI に与える。

プロンプトは自由

## 汎用性

- 再学習をしなくても、新しいラベルでの物体検出やセグメンテーションが可能

## 効率化

- 訓練データの準備のコストを削減

## 多様な応用

- 自由なラベルに対応できる

## 埋め込み

- **埋め込み**は、**オブジェクト**を**数値のベクトル**で表現すること
- 数値を用いて、オブジェクトの**類似性**や**関連性**を計算
- 未知のオブジェクトに関連する情報を推測可能に

## 属性学習

- オブジェクトの属性を学習
- **共通の属性を持つオブジェクト同士**を関連付けることが可能になる
- 属性を用いて未知のオブジェクトを識別可能に  
例：「**羽がある**」などの属性から未知の鳥を識別

## 演習 4

### NoShot の画像理解

#### 【トピックス】

- NoShot
- 英語のプロンプト

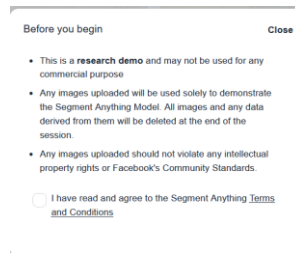
# ① Segment Anything のデモページを開く



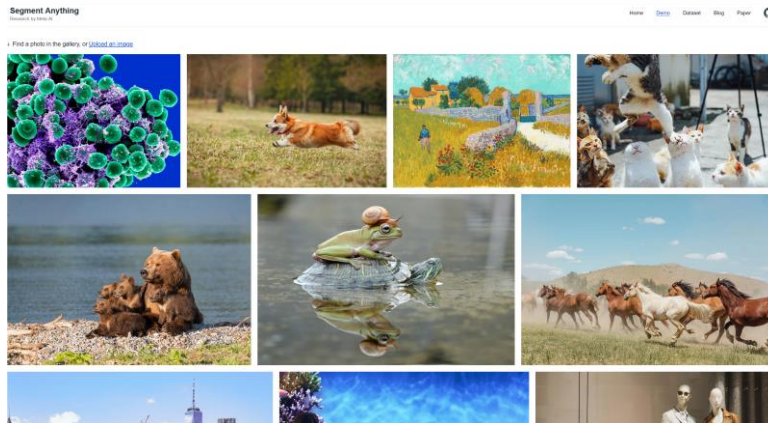
<https://segment-anything.com/demo#>

オンラインサービスであり、混雑時などは動かない場合がある。授業中で動かなかった場合には、後日試してほしい

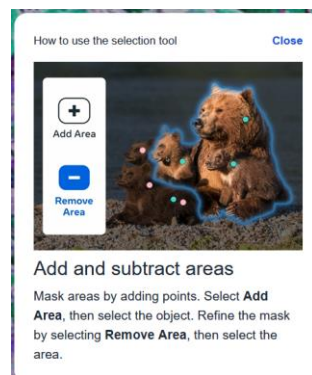
## ②最初「I have ...」のところをチェック



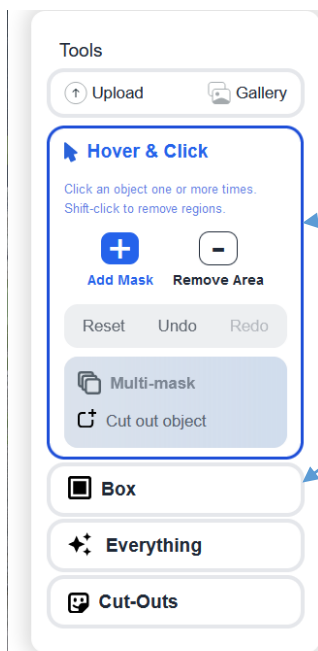
## ③画像を選ぶ。「Upload an image」により画像のアップロードも可能



## ④ 使用法の説明が出る。「close」をクリック



## ⑤ 右側のメニューとマウスで操作



マウスでクリック

範囲選択

画像全体を一括で処理

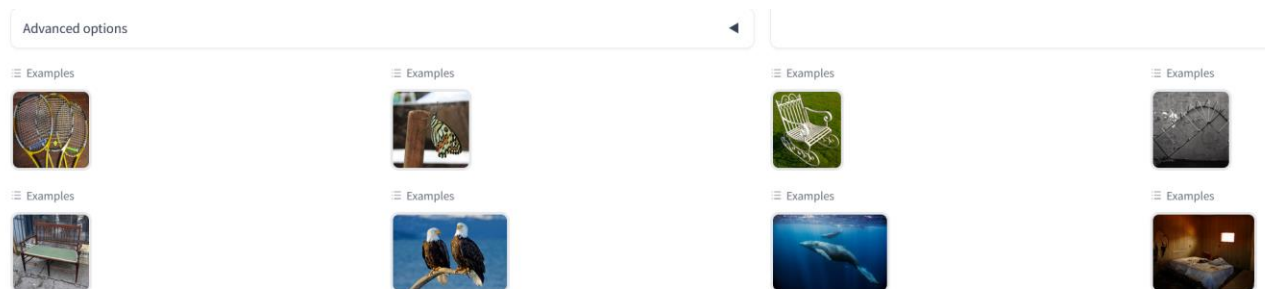


- 別サイトで、NoShot の能力を確認

## ① 次のページを使用

<https://huggingface.co/spaces/sam-hq-team/sam-hq>

## ② 下のところで画像を選ぶ



## ③ 「Text Prompt」 に英語の単語や文書を入れ「Run」

Text Prompt



# 全体まとめ①

## セグメンテーションの基本

- 画像を、意味のある領域ごとに分割

## 機能

- 同じ種類の画素が同じラベルに分類される

## セグメンテーションの利点

- 物体が画像のどこにあるかを正確に知ることができる
- 物体の形や大きさを数値化するための基礎

## メンテーションのメリット

## セグメンテーション技術で、畳み込みニューラルネットワークを基礎とするもの

- FCN: 画素ごとに結果出力を行う
- U-Net: FCNを改良し、輪郭情報の損失を防ぐ。

# 全体まとめ②

## セグメンテーションの種類

- セマンティック: 画素にラベル付け
- インスタンス: 物体を検出し、そのうえでセグメンテーションを実施
- パノプティック: セマンティックとインスタンスの同時実行

## データ拡張

- 実データをもとに、多様なデータを合成し、訓練データの量を増やす
- ディープラーニングの過学習の抑制を目的とする

## データ拡張の方法

- 回転
- 反転
- 切り取り (クロッピング)
- 色調

## NoShot学習

- **学習時のラベルにはない新しいラベル**に対しても、**識別・分類が可能**