5. ディープラーニングでの画像理解、画像分析の基本

（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ディープラーニングでの画像理解

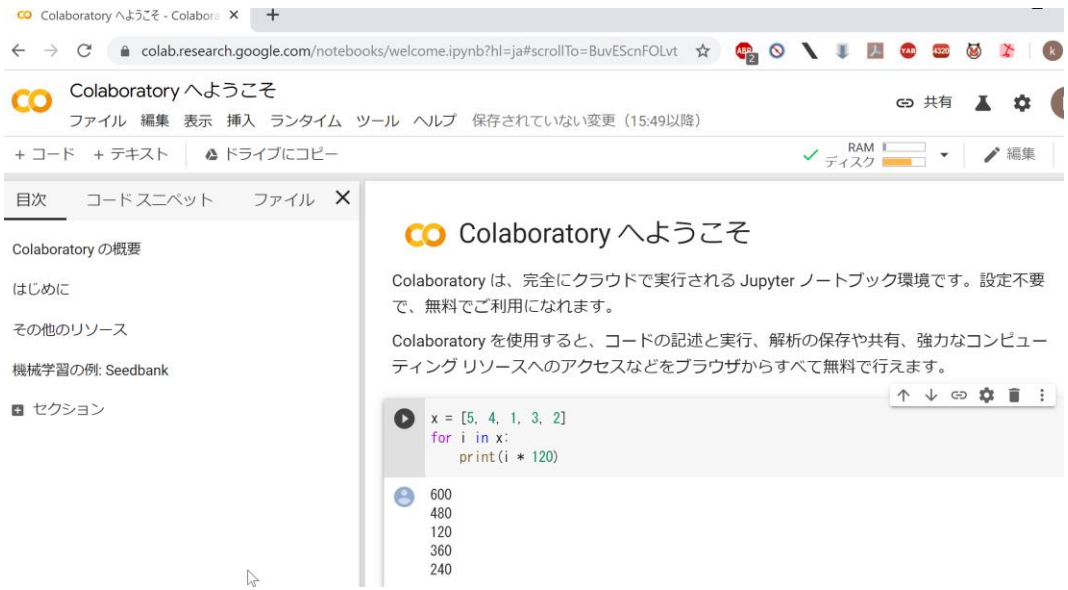
②ディープラーニングの高度な技術の理解

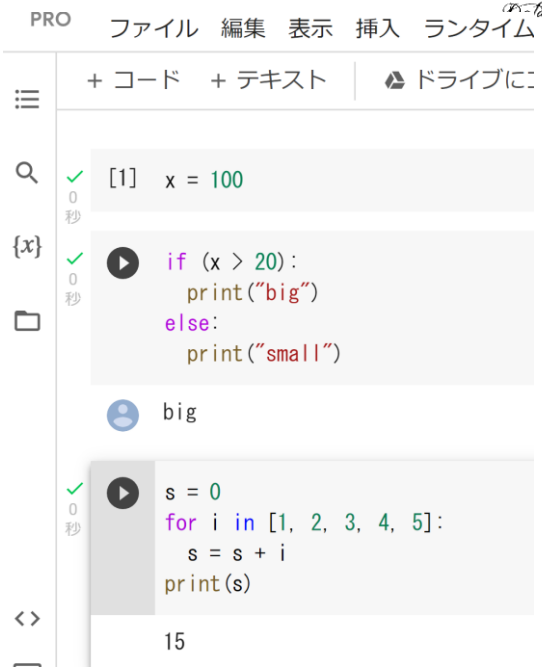
③興味、関心の向上

アウトライン

1. イントロダクション
2. コンピュータによる画像理解
3. 画像データの扱い
4. 畳み込みニューラルネットワーク(CNN)の基礎
5. 畳み込み
6. 全結合層
7. 畳み込み層
8. プーリング層
9. CNN Explainer のデモ

Google Colaboratory

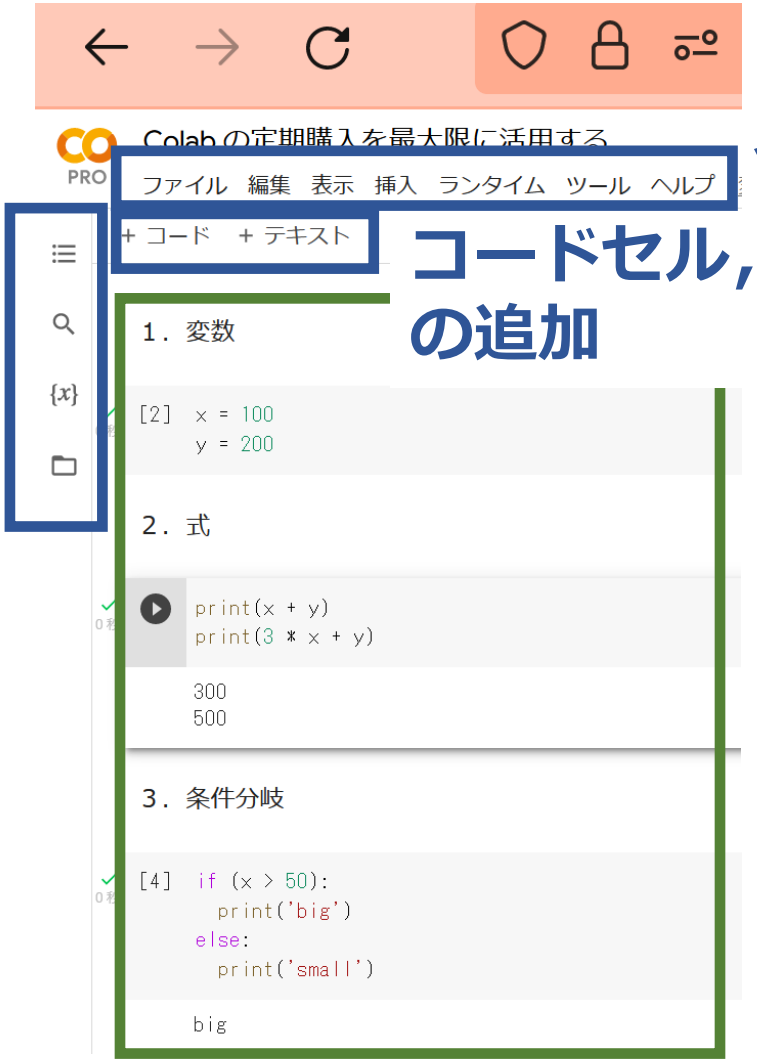




URL: https://colab.research.google.com/

- **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



メニュー

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）

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
```

5-1. イントロダクション

人工知能

知的なITシステム

機械学習

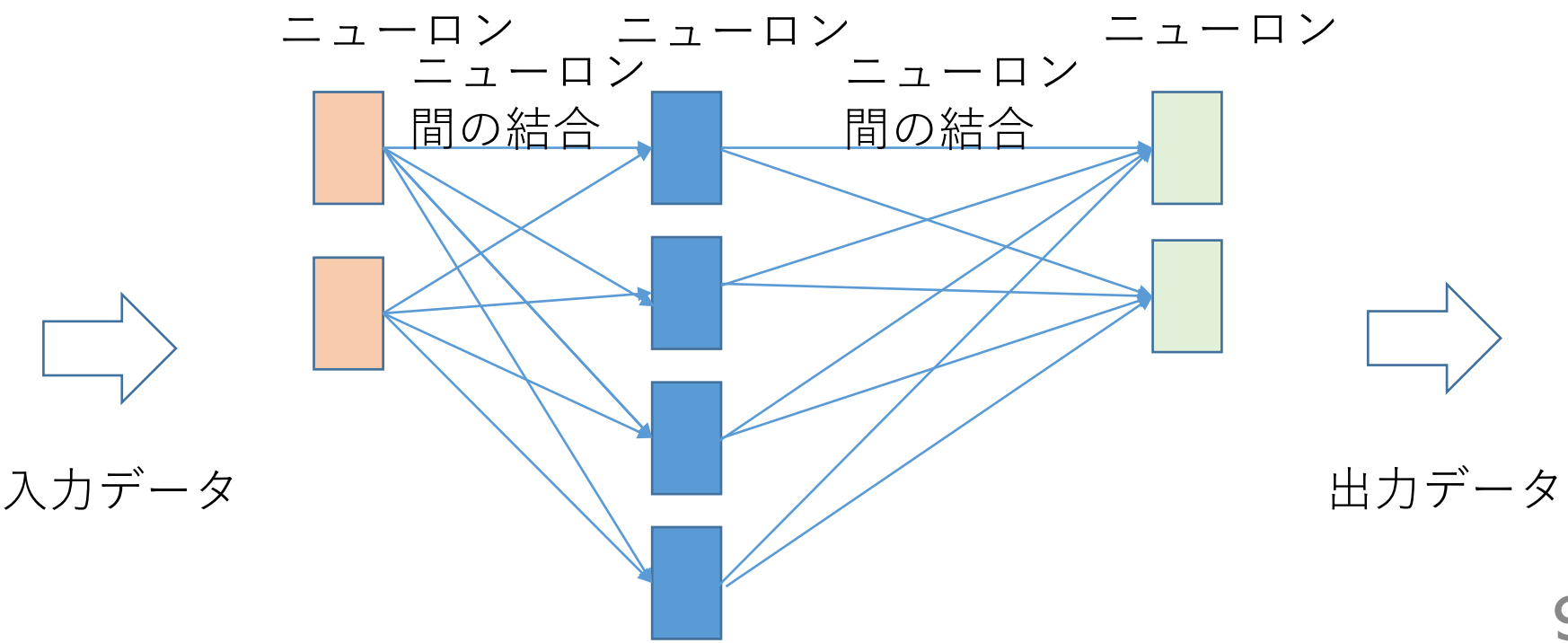
データから**学習**し、知的能力を向上

ディープラーニング

データから**学習**し、複雑なタスクを実行。**多層のニューラルネットワーク**を使用

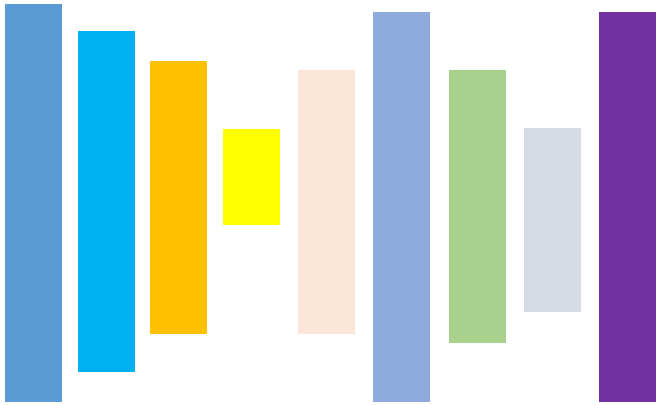
ニューロンとニューラルネットワーク

- **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



層の数が多い（ディープ）

ディープラーニングまとめ

- **ディープラーニング**は**機械学習**の一種であり、人工**ニューラルネットワーク**を使用して**データから学習**し、複雑な**タスクを実行**する技術

- 「ディープ」の名前は、**多層のニューラルネットワーク**を使用することに由来

- ディープラーニングが広く利用される理由は、**多様なデータに適用**でき、**さまざまなタスク**で高性能を発揮するため。例：**画像認識**、**自然言語処理**、**音声認識**など。

5-2. コンピュータによる画像理解

コンピュータによる画像理解

コンピュータが画像を理解する

コンピュータによる画像理解

- **一般的な画像が対象となる**
  （実験室で撮影などの制約が無い）
- **さまざまな応用**：スマホ，デジカメ，自動車，ロボット
- **さまざまな種類**：画像分類，物体検出，セグメンテーション，顔画像処理，姿勢推定など

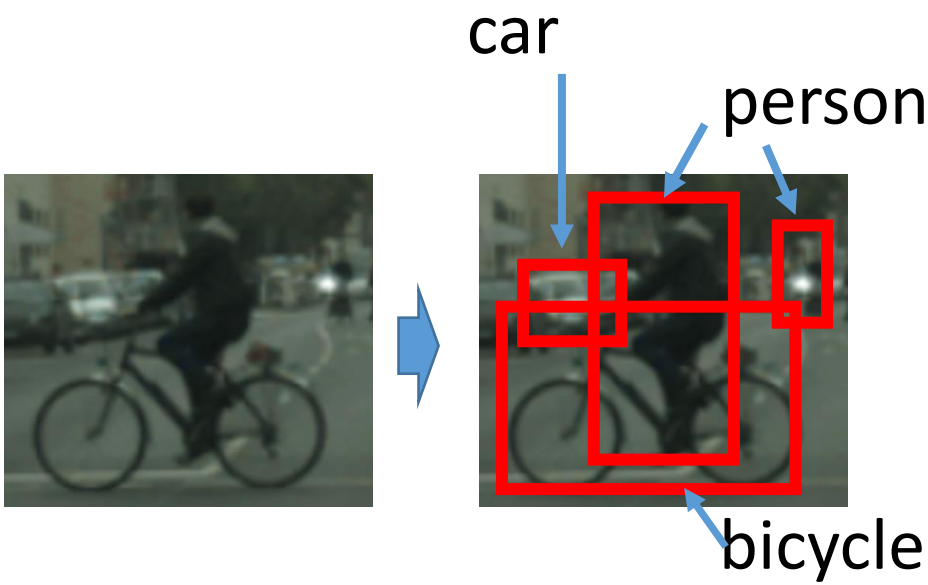
① 画像分類

```
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

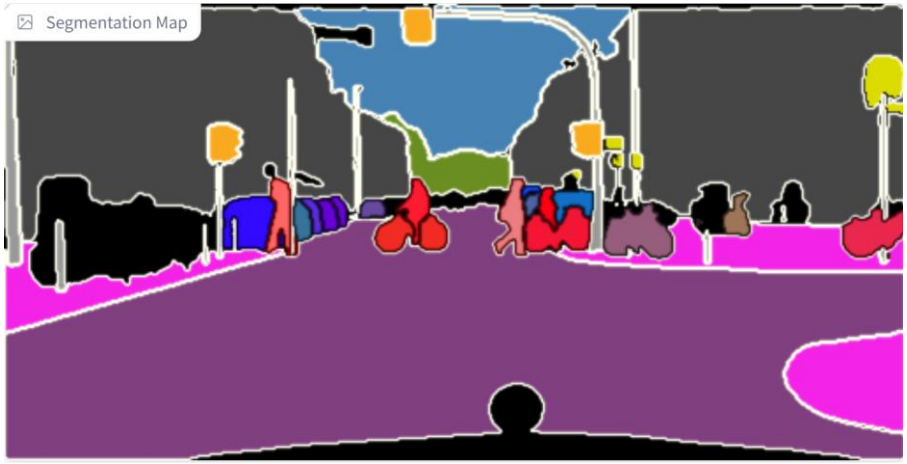
② 物体検出



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

画像理解の主な種類

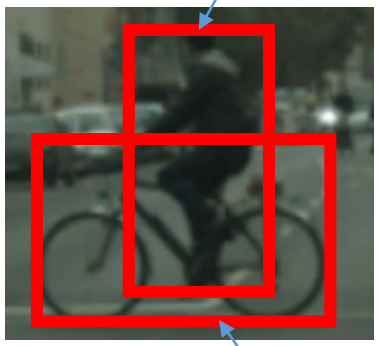
① **画像分類**

「何があるか」を理解

person
bicycle

② **物体検出**

場所と大きさも理解



③ **セグメンテーション**

画素単位で理解

画像分類の精度の向上

ImageNet データセットの画像分類の結果

- **ディープラーニング**の進展
- **画像分類は，場合によっては，AI が人間と同等の精度**　とも考えらるように

画像分類の誤り率 (top 5 error)

人間: 5.1 %

PReLU による画像分類: 4.9 %

（2015年発表）

文献: Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun,
Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification
arXiv:1502.01852, 2015.

ここまでのまとめ

<u>画像理解の主な種類</u>

- **画像分類**

「何があるか」を理解。結果は「ラベル」として識別

- **物体検出**

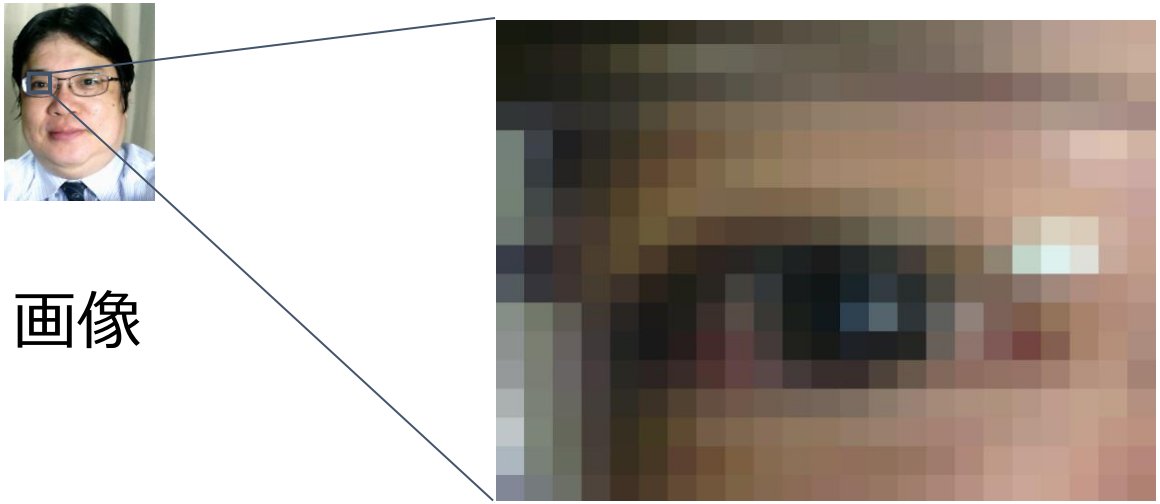
物体の種類、場所、大きさを理解。

- **セグメンテーション**

画素単位で理解。

5-3. 画像データの扱い

画像と画素



画像

それぞれの格子が画素

画像の種類

カラー画像

輝度と**色**の情報

濃淡画像

輝度のみの情報

濃淡画像でのコード化

画像の輝度の情報

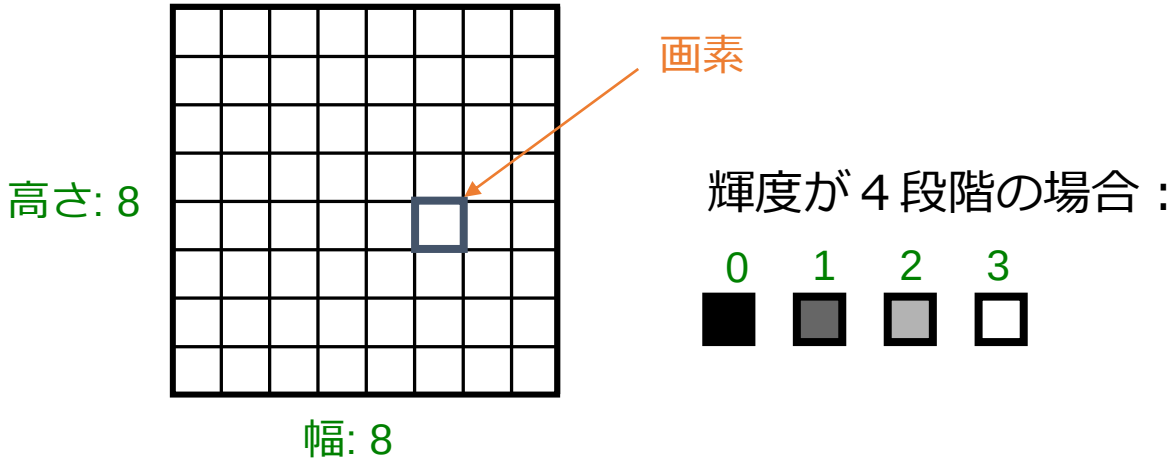
例えば：　黒　＝　０，

暗い灰色　＝　１，

明るい灰色　＝　２，

白　＝　３

のように**コード化**



R（赤）成分，G（緑），B（青）成分で考える場合

R（赤）成分　　G（緑）成分　　B（青）成分

画素ごとに
１つの数値

画素ごとに
１つの数値

画素ごとに
１つの数値

すべてあわせて，画素ごとに３つの数値

演習 1

カラー画像の画素

【トピックス】

カラー画像

画素

配列

① Google Colaboratory のページを開く

https://colab.research.google.com/drive/1MMhlrh08-0Byq-U1JlTdDVwMe6dyUcaq?usp=drive_link

（このページは、演習１，２で使用する）

② **次**について、プログラムや説明や実行結果が掲載されていることを確認。各自でよく読む。

１．カラー画像の画素

１．カラー画像の画素

【プログラムの説明】

image_data は3x3ピクセルのカラー画像を模倣したもので、RGB形式で色が保存されています。

特定の位置 (x, y) の画素の色を取得するためのコードも示しています。

```
# カラー画像の画素例（RGB形式）
# 例: 3x3ピクセルのカラー画像
image_data = [
    [(255, 0, 0), (0, 255, 0), (0, 0, 255)],
    [(0, 255, 255), (255, 0, 255), (255, 255, 0)],
    [(128, 128, 128), (64, 64, 64), (32, 32, 32)]
]

# 特定の画素の色を取得
x, y = 1, 1
pixel_color = image_data[y][x]
print(f"Pixel color at ({x}, {y}): {pixel_color}")
```

```
Pixel color at (1, 1): (255, 0, 255)
```

自習

目的：配列になっている画像データについて、特定の位置の画素の色を取得する方法を学び、理解を深めること。

指示：異なる (x, y) の位置の画素を取得して、期待通りであるか確認してみよう

解答例：

x, y = 2, 2

のときは次のように表示される

Pixel color at (2, 2): (32, 32, 32)

```
# 特定の画素の色を取得
x, y = 2, 2
pixel_color = image_data[y][x]
print(f"Pixel color at ({x}, {y}): {pixel_color}")
```

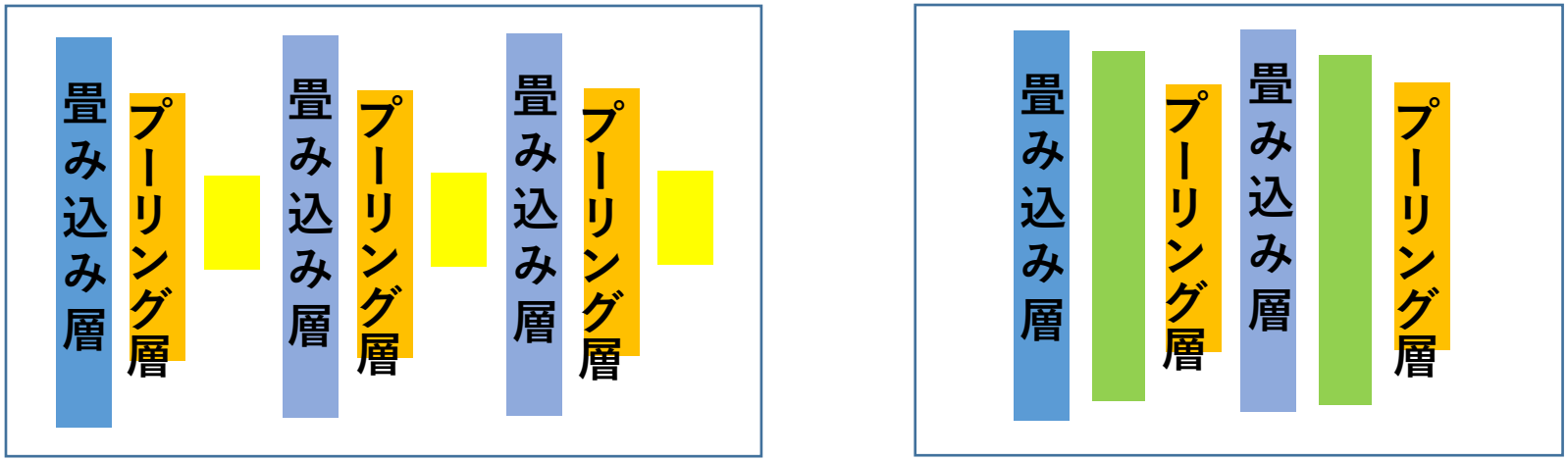
```
Pixel color at (2, 2): (32, 32, 32)
```

5-4. 畳み込みニューラルネットワーク(CNN)の基礎

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）の用途

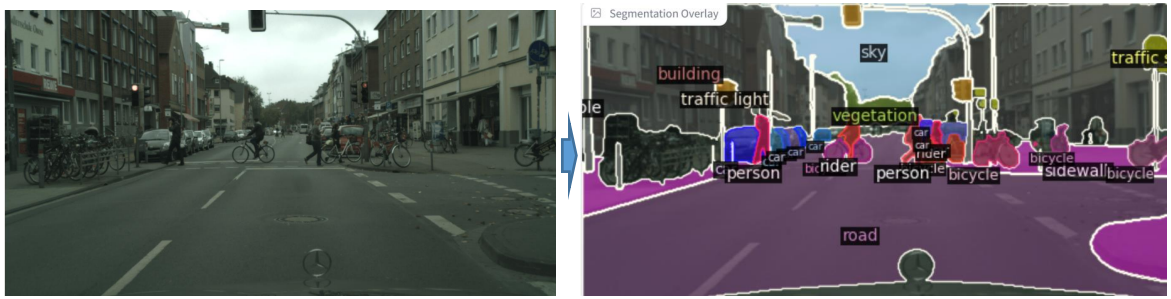
画像分類，物体検出，セグメンテーションなどで高い性能・機能を発揮する場合がある

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 0.0003795088559854775 7, Label plate
```

物体検出

セグメンテーション



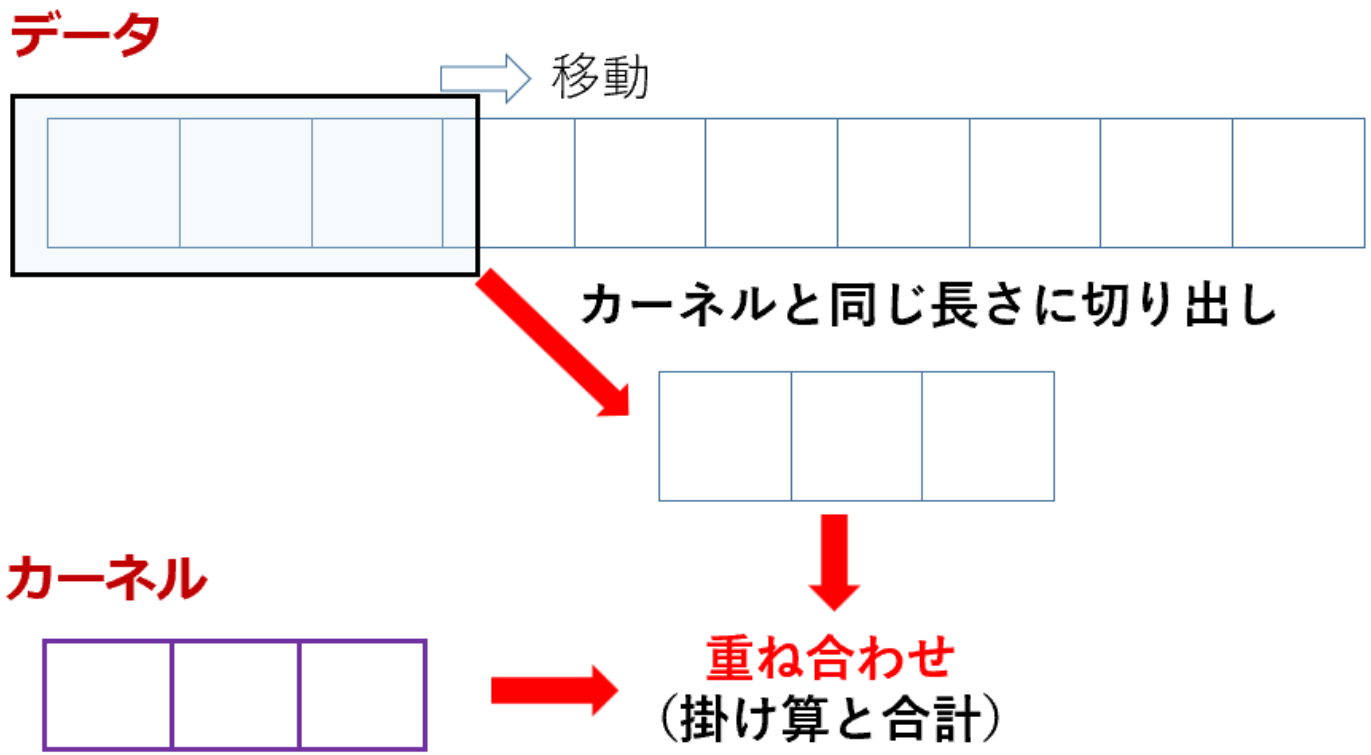
畳み込みニューラルネットワーク（CNN）のメリット

- **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**

5-5. 畳み込み

畳み込みの役割

畳み込みは、**データ上を移動**しながら、**カーネル**との**重ね合わせ**を行い、**局所的な特徴**を**抽出**



畳み込みの応用例

- ぼかし，エッジ抽出，テクスチャ分析などさまざまな処理

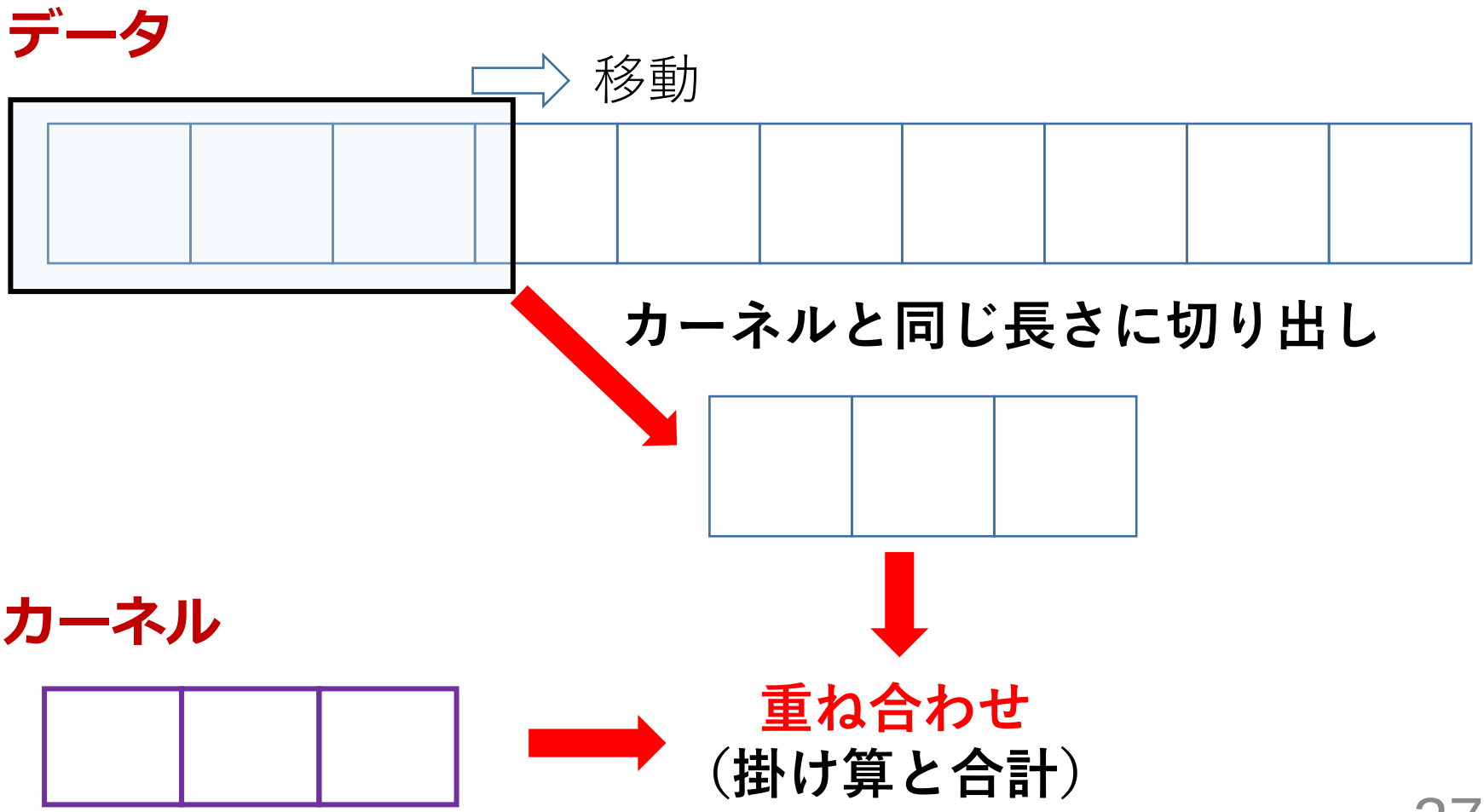
元画像

畳み込みによる
ぼかし

畳み込みによる
エッジ抽出

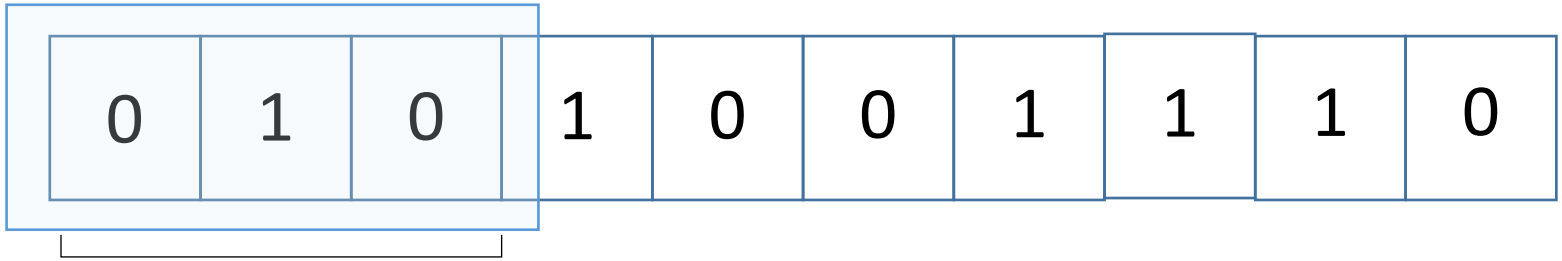
畳み込みの仕組み

畳み込みは，**データ上を移動**しながら，**カーネル**との**重ね合わせ**を行う．重ね合わせの結果は１つの値になる．



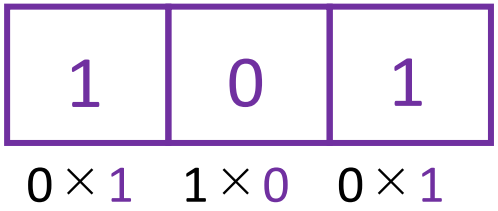
畳み込みの例

データ



この部分を切り出す

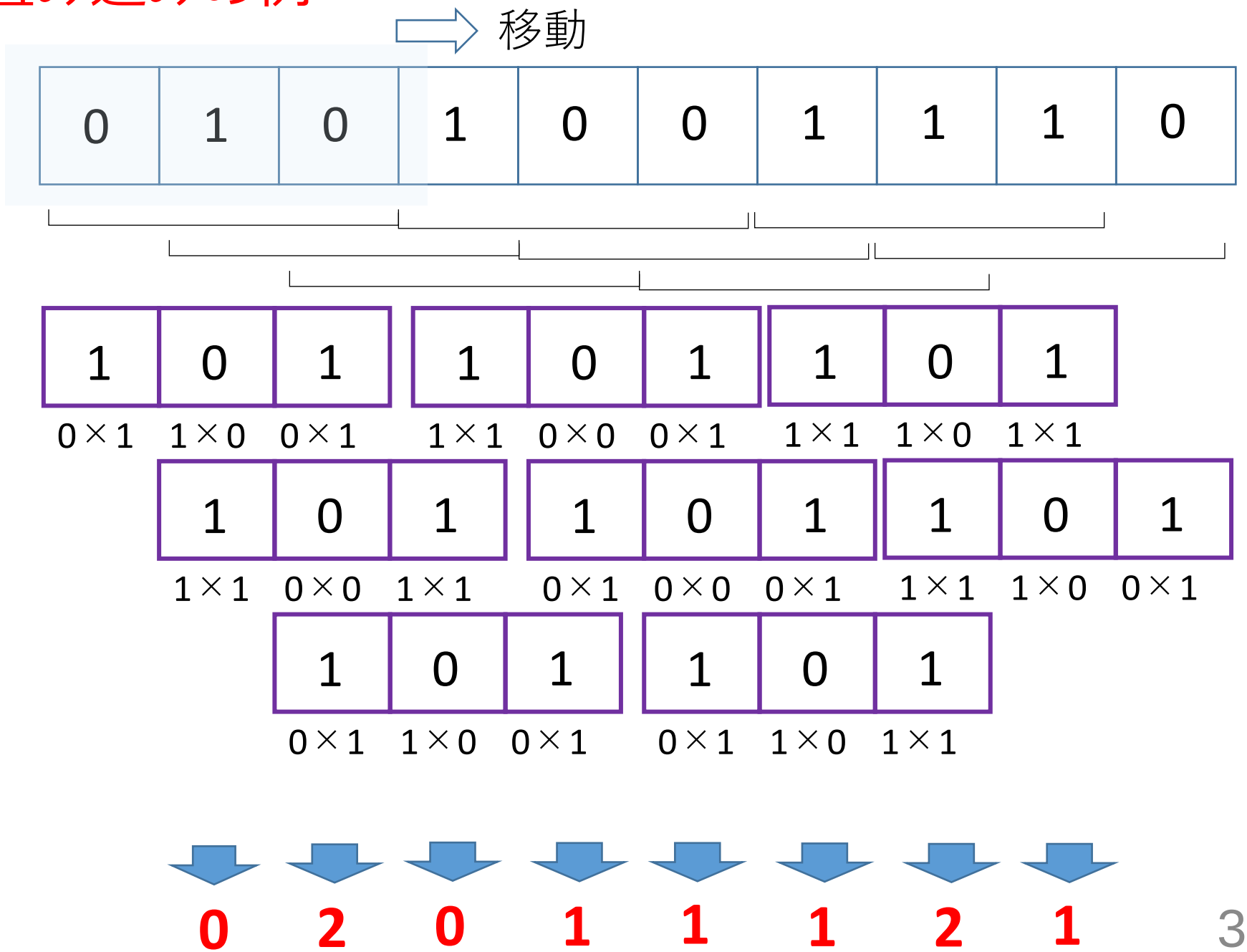
カーネル



0

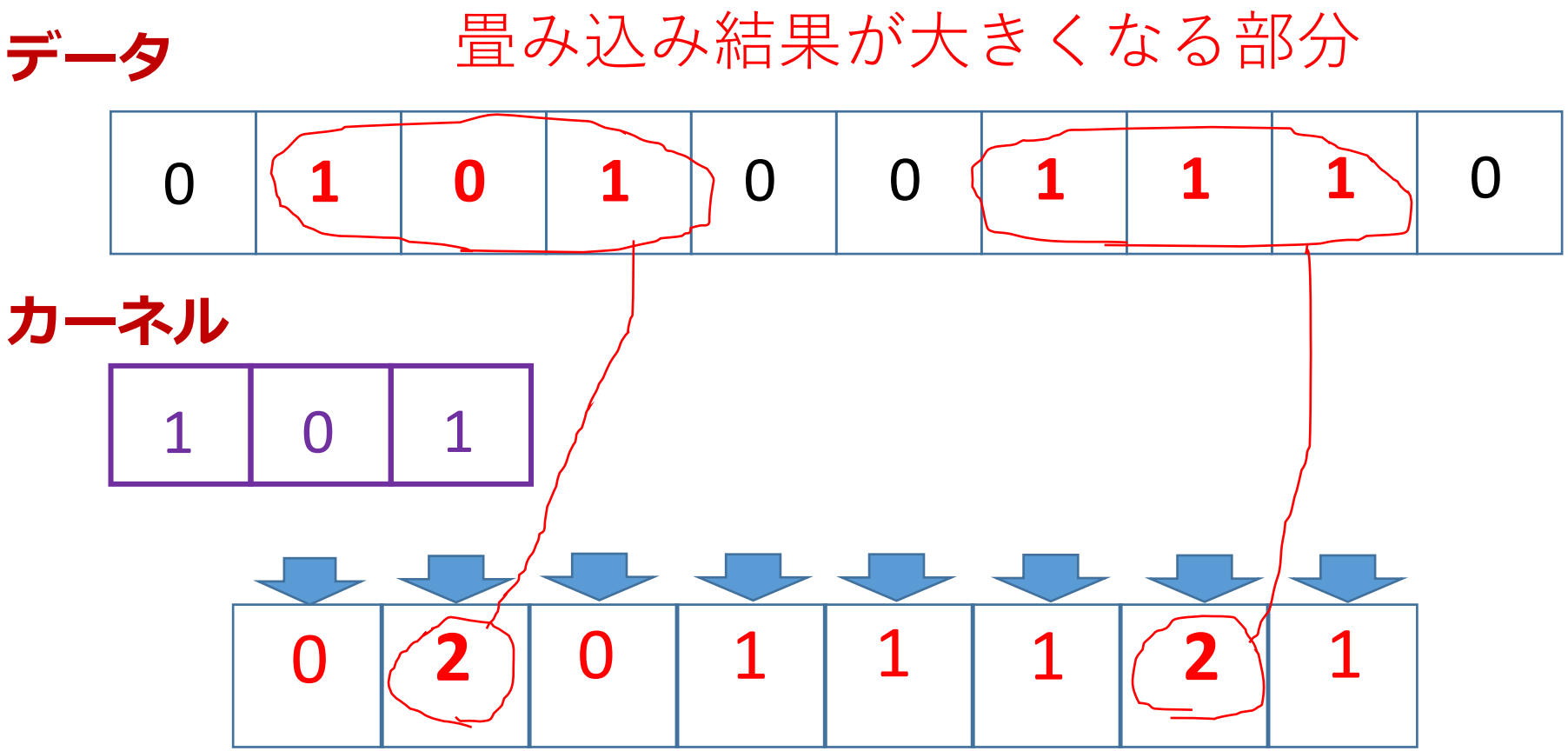
重ね合わせの結果：　0×1 ＋ 1×0 ＋ 0×1 = 0

畳み込みの例


移動
0 1 0 1 0 0 1 1 1 0
1 0 1
0×1 1×0 0×1
1 0 1
1×1 0×0 0×1
1 0 1
1×1 0×0 1×1
1 0 1
0×1 0×0 0×1
1 0 1
1×1 1×0 1×1
1 0 1
0×1 1×0 0×1
1 0 1
0×1 1×0 1×1
0 2 0 1 1 1 2 1

畳み込みの例と、局所的な特徴の抽出

畳み込みは，「**特定の局所的な特徴に強く反応する**」と考えることもできる



画像の畳み込み

Input

0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1

元画像（５×５マス）

Filter / Kernel

1	0	1
1	1	1
0	0	1

カーネル（３×３マス）

出典: https://serokell.io/blog/introduction-to-convolutional-neural-networks

画像での畳み込み

0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1

元画像（５×５マス）

1	0	1
1	1	1
0	0	1

カーネル
（３×３マス）

0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1

切り出し（３×３マス）
カーネルと同じサイズで切り出す

切り出した部分とカーネルの掛け算の合計

0×1	1×0	1×1
0×1	1×1	1×1
0×0	1×0	1×1

合計: 4（これが畳み込み結果）

畳み込み

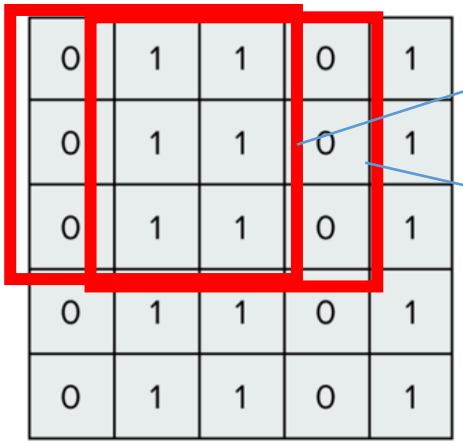
画像での畳み込み

0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1

元画像（５×５マス）

1	0	1
1	1	1
0	0	1

カーネル
（３×３マス）



切り出し（３×３マス）

切り出し領域を横にずらす

0×1	1×0	1×1
0×1	1×1	1×1
0×0	1×0	1×1

合計: 4

1×1	1×0	0×1
1×1	1×1	0×1
1×0	1×0	0×1

合計: 3

4	3	

畳み込み結果

畳み込み結果

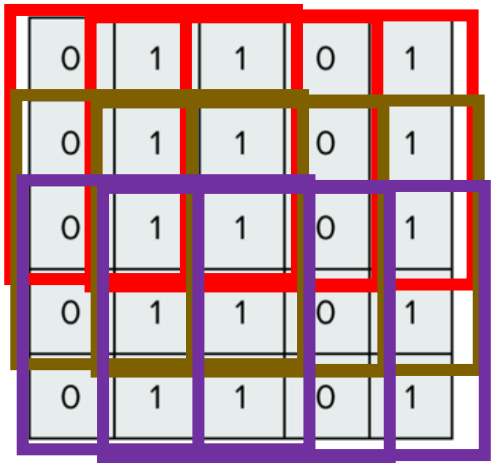
画像での畳み込み

0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1
0	1	1	0	1

元画像（５×５マス）

1	0	1
1	1	1
0	0	1

カーネル
（３×３マス）



切り出し（３×３マス）

画像全体について
畳み込み

4	**3**	**5**
4	**3**	**5**
4	**3**	**5**

畳み込み結果

切り出し領域を縦横にずらす

演習２

畳み込み

【トピックス】

畳み込み

二値画像（０，１の画像）の畳み込み

① Google Colaboratory のページを開く

https://colab.research.google.com/drive/1MMhlrh08-0Byq-U1JlTdDVwMe6dyUcaq?usp=drive_link

（このページは、演習１，２で使用する）

② **次**について、プログラムや説明や実行結果が掲載されていることを確認。各自でよく読む。

２．畳み込み

３．二値画像（画素値が０，１のみ）の畳み込み

```
x = [0, 1, 0, 1, 0, 0, 1, 1, 1, 0]
k = [1, 0, 1]
y = [0, 0, 0, 0, 0, 0, 0, 0]

y[0] = x[0] * k[0] + x[1] * k[1] + x[2] * k[2]
y[1] = x[1] * k[0] + x[2] * k[1] + x[3] * k[2]
y[2] = x[2] * k[0] + x[3] * k[1] + x[4] * k[2]
y[3] = x[3] * k[0] + x[4] * k[1] + x[5] * k[2]
y[4] = x[4] * k[0] + x[5] * k[1] + x[6] * k[2]
y[5] = x[5] * k[0] + x[6] * k[1] + x[7] * k[2]
y[6] = x[6] * k[0] + x[7] * k[1] + x[8] * k[2]
y[7] = x[7] * k[0] + x[8] * k[1] + x[9] * k[2]

print(y)
```

```
[0, 2, 0, 1, 1, 1, 2, 1]
```

```
k = [[0, 1, 1, 0, 1],
     [0, 1, 1, 0, 1],
     [0, 1, 1, 0, 1],
     [0, 1, 1, 0, 1],
     [0, 1, 1, 0, 1]]
k = [[1, 0, 1],
     [1, 1, 1],
     [0, 0, 1]]
y = [[0, 0, 0],
     [0, 0, 0],
     [0, 0, 0]]

def conv(i, j):
    return x[i + 0][j + 0] * k[0][0] + x[i + 0][j + 1] * k[0][1] + x[i + 0][j + 2] * k[0][2] ¥
    + x[i + 1][j + 0] * k[1][0] + x[i + 1][j + 1] * k[1][1] + x[i + 1][j + 2] * k[1][2] ¥
    + x[i + 2][j + 0] * k[2][0] + x[i + 2][j + 1] * k[2][1] + x[i + 2][j + 2] * k[2][2]

y[0][0] = conv(0, 0)
y[0][1] = conv(0, 1)
y[0][2] = conv(0, 2)
y[1][0] = conv(1, 0)
y[1][1] = conv(1, 1)
y[1][2] = conv(1, 2)
y[2][0] = conv(2, 0)
y[2][1] = conv(2, 1)
y[2][2] = conv(2, 2)

print(y)
```

```
[[4, 3, 5], [4, 3, 5], [4, 3, 5]]
```

自習 2

目的: 元データの変化が、畳み込みの結果がどのように影響するか確認

指示: 演習 1 のプログラムについて、x を [0, 1, 2, 1, 0, 0, 1, 1, 1, 0] に変更して結果を確認してください

解答例: [2, 2, 2, 1, 1, 1, 2, 1]

自習 3

目的: 元データとカーネルの変化が、畳み込みの結果がどのように影響するか確認

指示: 演習 1 のプログラムについて、x を [0, 1, 2, 1, 0, 0, 1, 1, 1, 0], kを [1,2,1]にして結果を確認してください

解答例: [4, 6, 4, 1, 1, 3, 4, 3]

畳み込みのまとめ

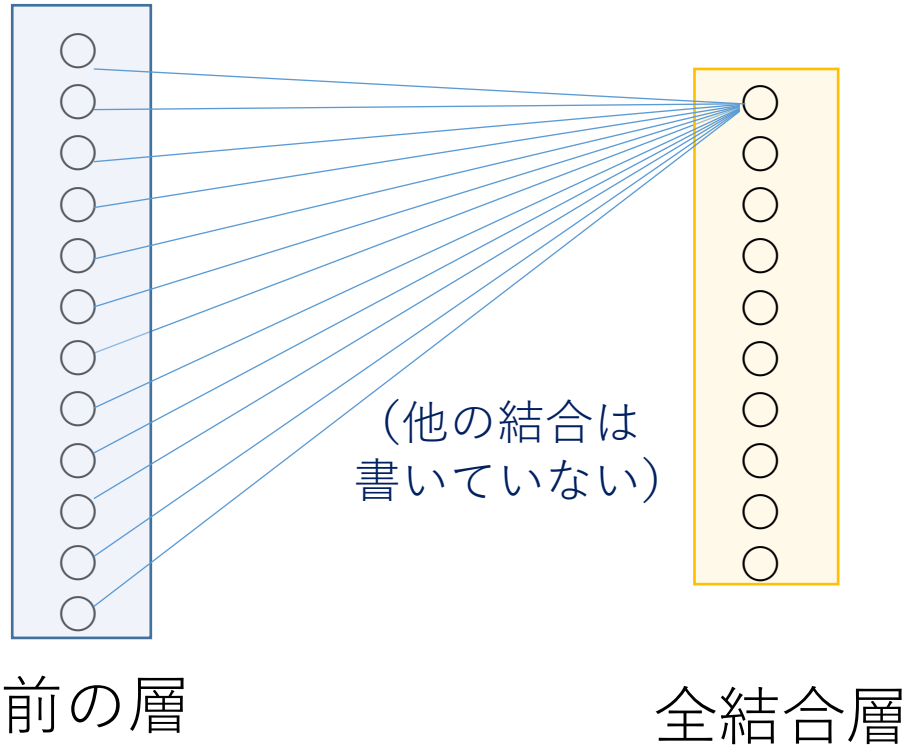
- **畳み込みニューラルネットワーク（CNN）**は画像理解や画像の分析に特化。
- CNNの主層：**畳み込み層**、**プーリング層**、**全結合層**。
- **畳み込み層**は**画像の局所的な特徴**を捉える。

  例：エッジ、テクスチャなど
- **畳み込み**は、**データ上を移動**しながら、**カーネル**との**重ね合わせ**を行い、**局所的な特徴**を**抽出**

5-6. 全結合層

全結合層

- **全結合層**の**ニューロン**は，**前の層**の**すべてのニューロンと結合**している
- 畳み込みニューラルネットワークだけでなく、通常のニューラルネットワークにも用いられる



5-7. 畳み込み層

演習 3

畳み込み層

【トピックス】

畳み込み

畳み込み層

①「Animated AI」のサイトにアクセスする

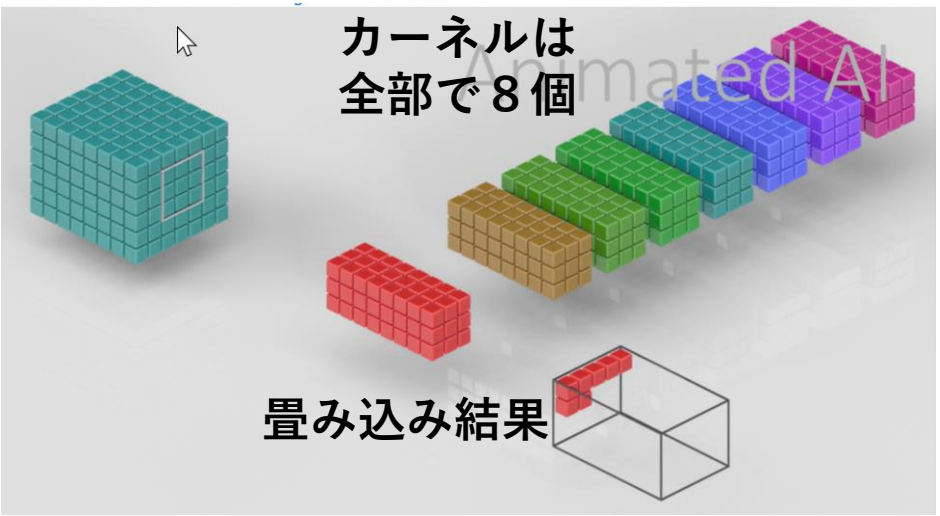
https://animatedai.github.io/

②このサイトでは、さまざまな**ニューラルネットワーク**の**アルゴリズム**が解説されている

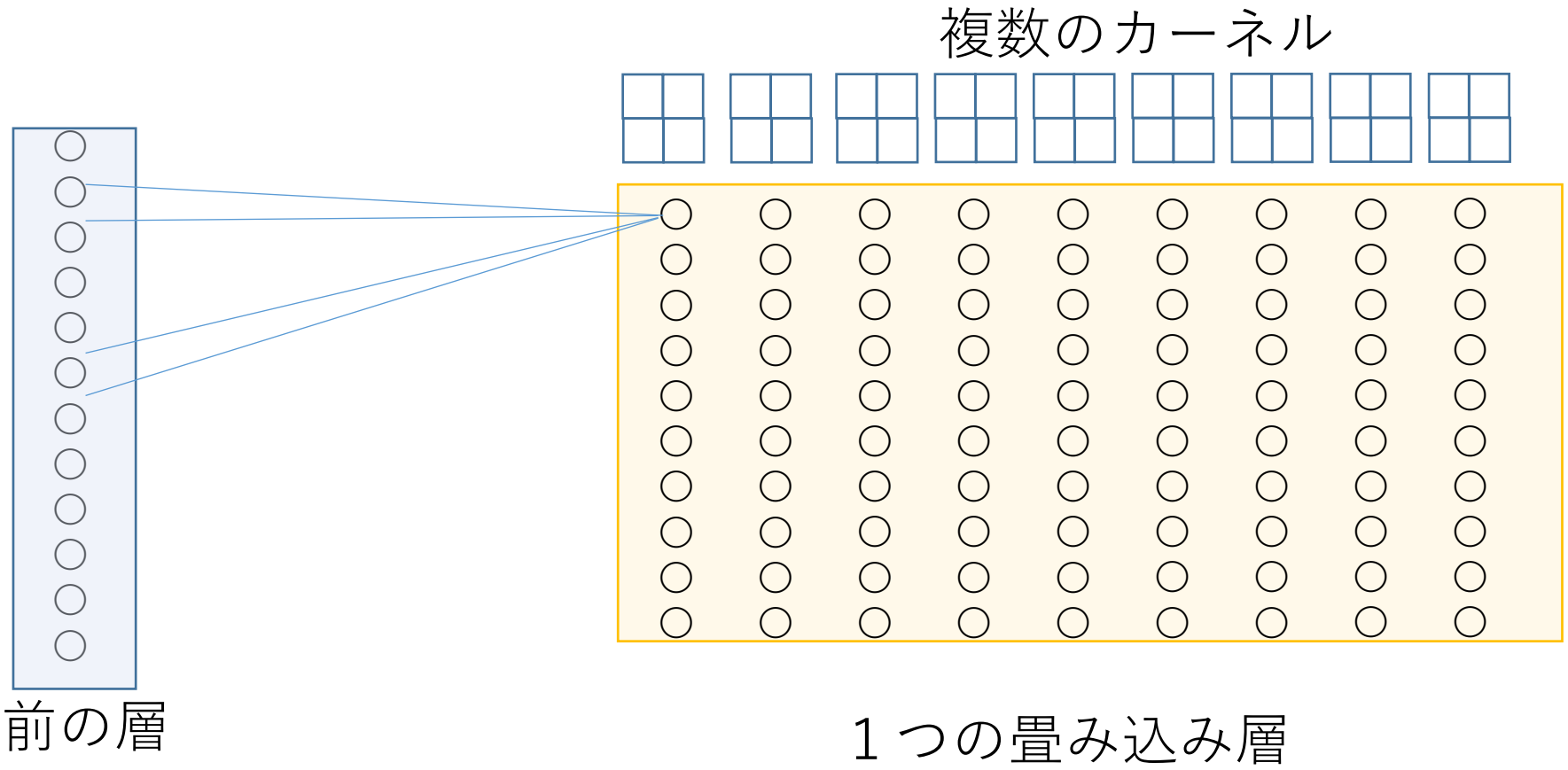
③トップの「The Basic Algorithm」は、**畳み込み**をアニメーションで示している。

- 元データとカーネルは３次元である
- **畳み込み**の仕組み上、**同じ層の全ニューロン**はすべて、**同一のカーネルを共有**

元データ



- **畳み込み層**では，一度に**多数の畳み込み**を行う
- **同じ層の全ニューロン**はすべて、**同一のカーネルを共有し、パラメータの量**は**少ない**



学習の仕組み

- **学習**では、ニューラルネットワークの**パラメータが最適化**される。

【全結合層（一般のニューラルネットワークの層）の学習】

- **パラメータ：**ニューロンの**重み**や**バイアス**
- 全結合層のニューロンごと独自の**重み**や**バイアス**を持つ（
- **パラメータは大量**

【畳み込み層の学習】

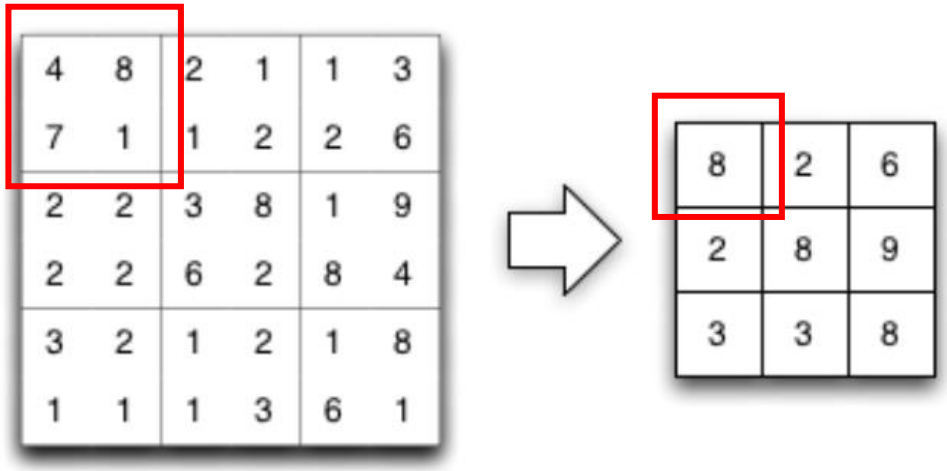
- パラメータ：**カーネル**
- **同じ層の全ニューロン**はすべて、**同一のカーネルを共有**
- **パラメータは少量**
- カーネルの共有によって、過学習を避けつつ高速な学習が実現される。

5-8. プーリング層

プーリングを行う Max Pooling 層

- ２次元のデータの**縮小**

  （例）サイズ 100 × 100 ⇒ 50 × 50 のように
- 一定領域内の**結果**を，**１つにまとめる**.
- 学習の対象ではない
- **Max Pooling は，縮小後に，範囲内の最大値が残る**



・4, 8, 7, 1 の最大値は 4

・「**4, 8, 7, 1**」**の 4マス**から，**最大値の 8 を選ぶ**.

出典: https://github.com/jeffheaton/t81_558_deep_learning/blob/084023876b6cf09c931b452584dbd44c56314a03/t81_558_class_06_2_cnn.ipynb

5-9. CNN Explainer のデモ

畳み込みニューラルネットワーク（CNN）の特徴

	畳み込みニューラルネットワーク	一般のニューラルネットワーク
局所的な特徴の抽出能力	画像の局所的な特徴をとらえる**畳み込み層を持つ**	畳み込み層はない
仕組み	**畳み込み層、プーリング層、全結合層**などから構成	主に全結合層
パラメータ	学習では、重み、バイアスなどのパラメータを調整。ただし、**畳み込み層**では、**カーネルを共有する**ため、学習が効率化	学習では、重み、バイアスなどのパラメータを調整

演習４

畳み込みニューラルネットワーク

CNN Explainer

- **CNN Explainer**　ジョージア工科大学 Polo Club
- 畳み込み層などの仕組みをビジュアルに学ぶことができるサイト

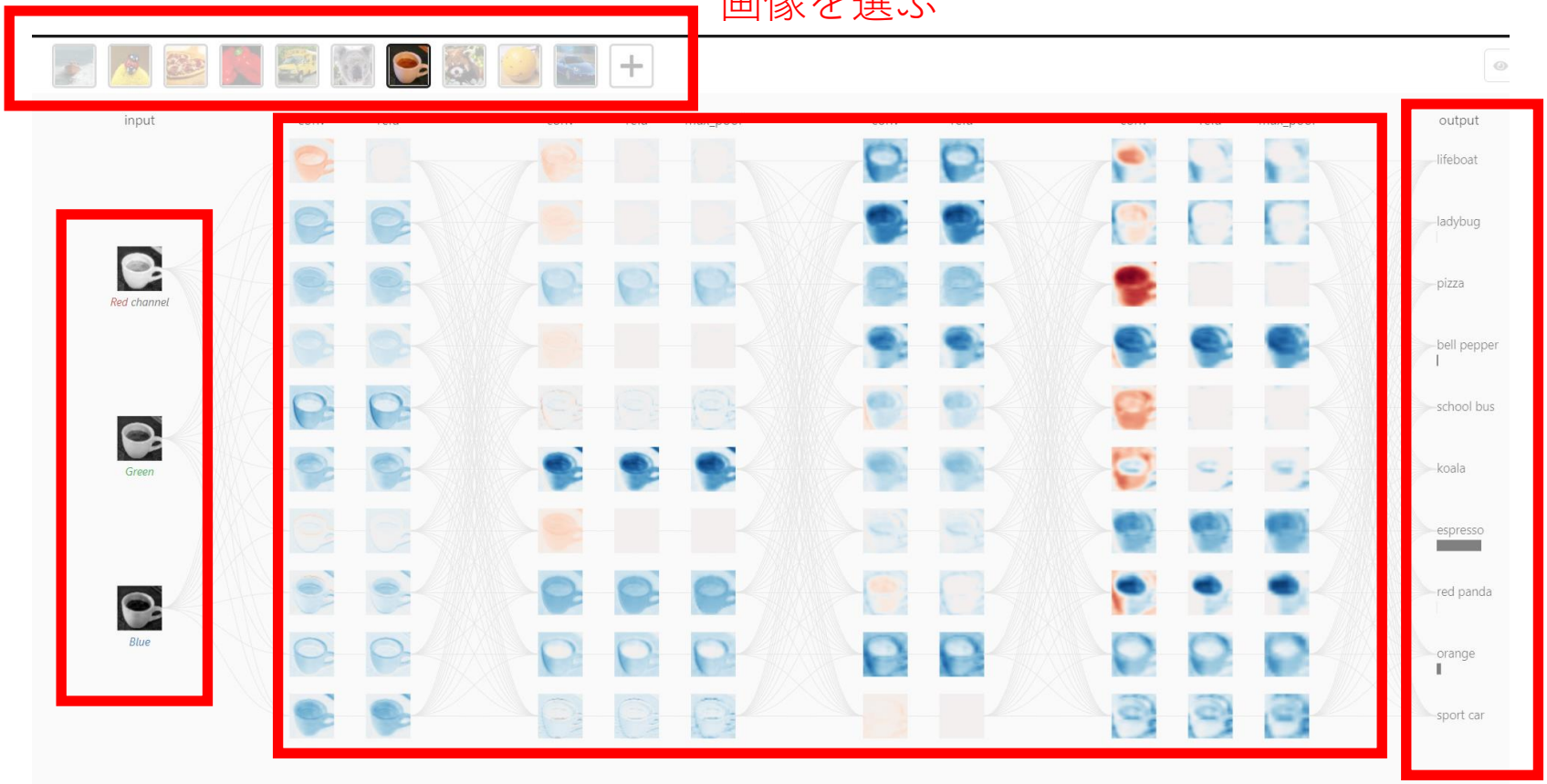
Webブラウザで次の URL を開く

https://poloclub.github.io/cnn-explainer/

① 画面の確認

このニューラルネットワークは，**画像分類**を行う

画像を選ぶ



元画像の
赤青緑の成分

ニューラルネットワーク

画像の分類結果.
ここでは espresso

② ニューラルネットワークの構成

畳み込み層と**プーリング層**を含む

conv relu　　conv relu max_pool　conv relu　conv relu max_pool

畳み込み層　畳み込み層　プーリング層　畳み込み層　畳み込み層　プーリング層

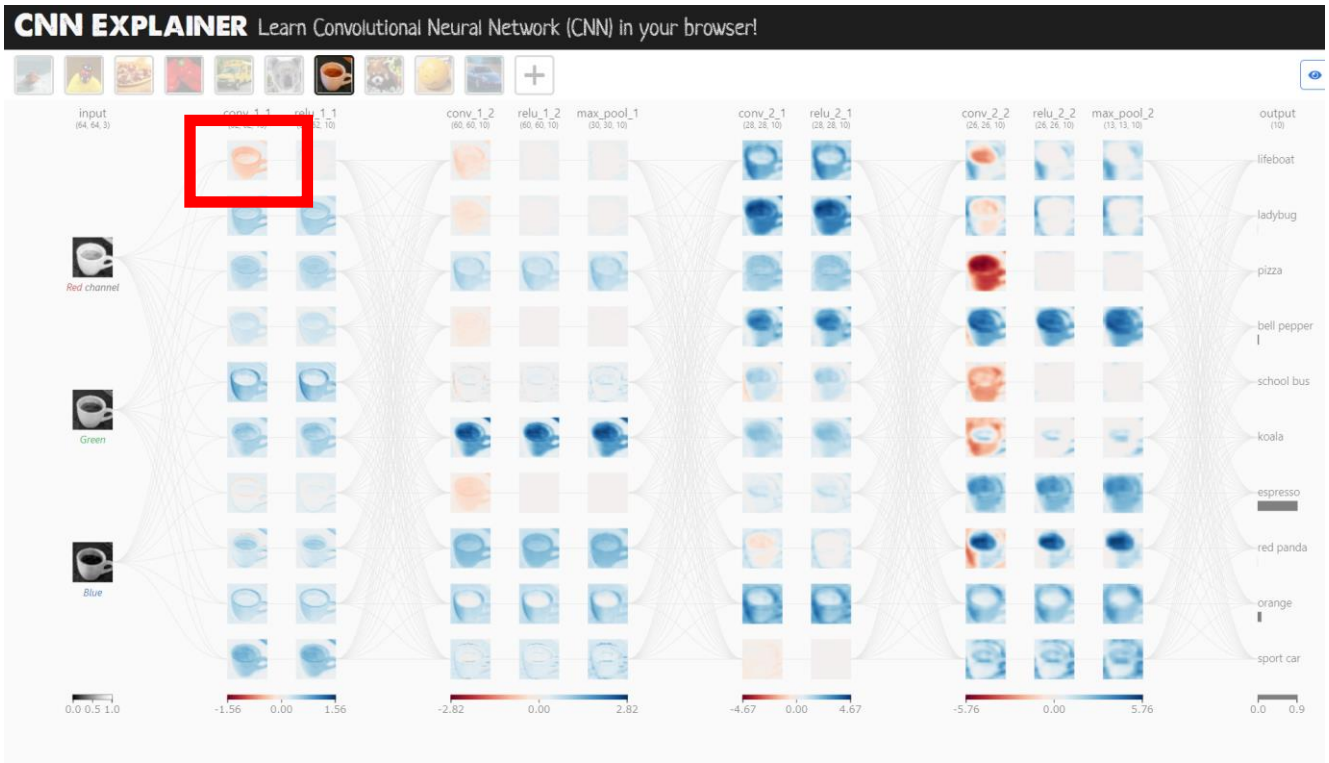
conv は畳み込み層で，max_pool はプーリング層

全結合層　全結合層　全結合層　全結合層

③ 左上の画像をクリック

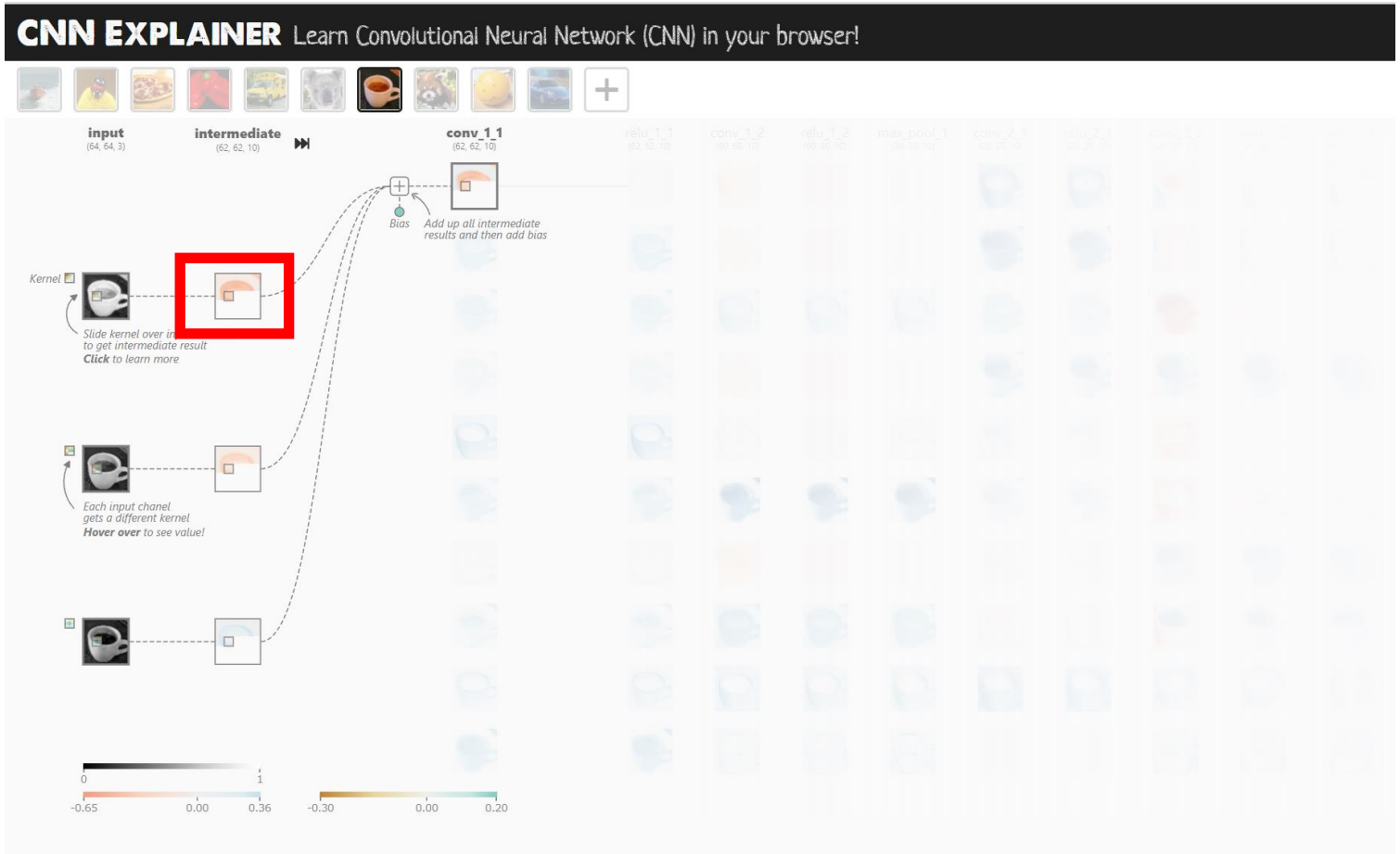
→　畳み込みの様子をアニメーションで確認できる

（この画像は，**各層での処理結果**である．画像１個がニューロン１つというわけではない）

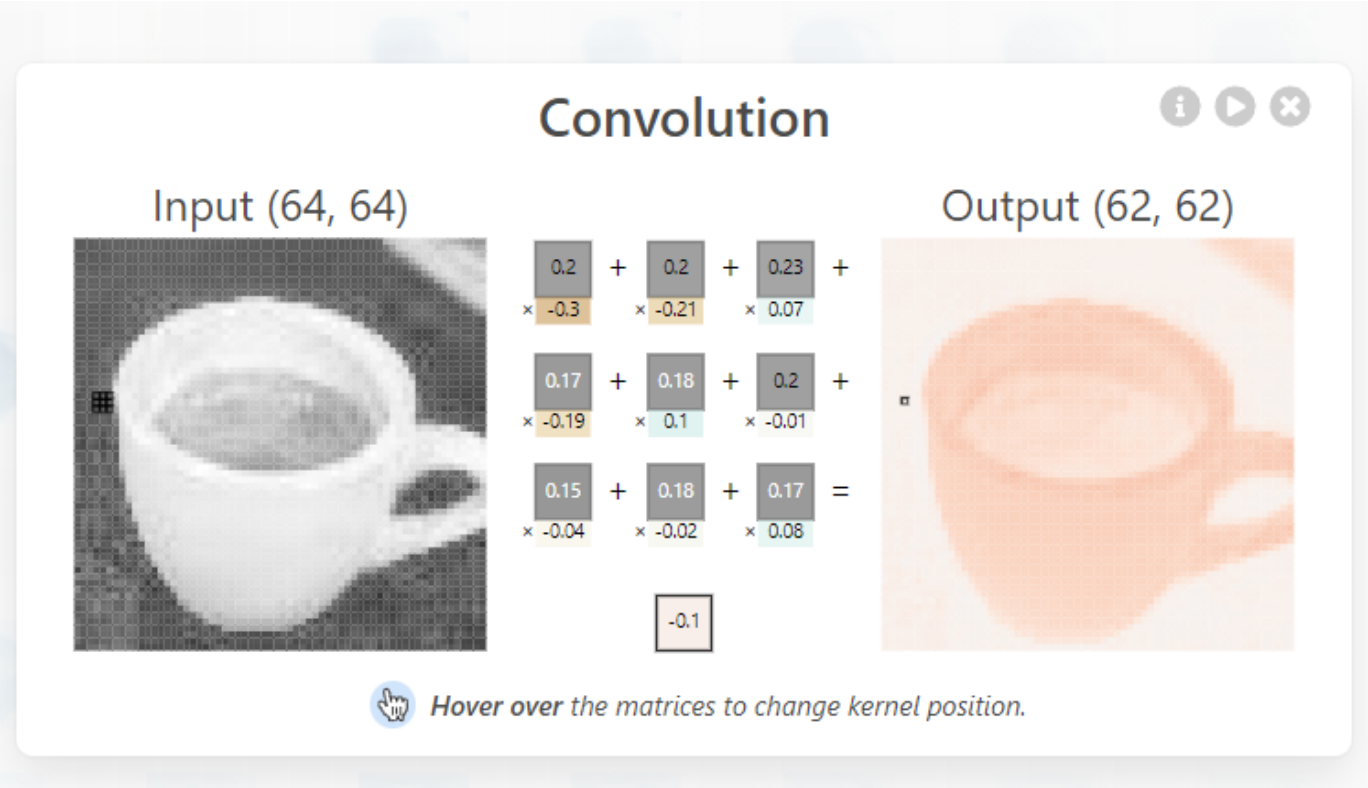


④ 出てきた画像をクリック

→　畳み込みの詳細をアニメーションで確認できる



⑤ 畳み込みの様子がアニメーションで表示される



その他の層についてもビジュアルに表示できる（いろいろ試すことは，各自の自主的な自習とする）

全体まとめ ①

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

- **物体検出**

物体の種類、場所、大きさを理解。

- **セグメンテーション**

画素単位で理解。

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

全体まとめ ②

- **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

学習の仕組み

- ニューラルネットワークの学習では、パラメータの最適化が行わる
- 畳み込み層のパラメータの量は少ない。これにより、過学習を回避