深層学習入門

2. 深層学習

深層学習

ニューラルネットワーク

畳み込みニューラルネットワーク

- 物体分類
- 物体検出・セグメンテーション

敵対的生成ネットワーク

深層学習

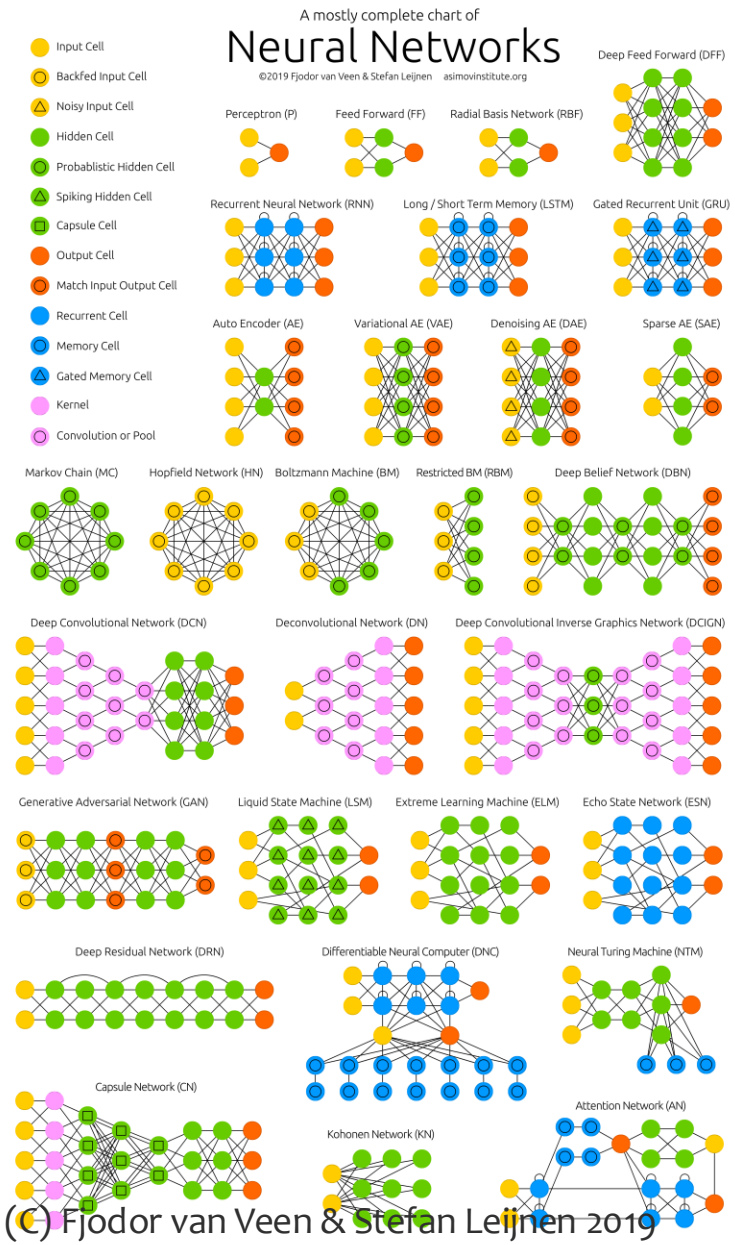
ニューラルネットワーク

畳み込みニューラルネットワーク

- 物体分類
- 物体検出・セグメンテーション

敵対的生成ネットワーク

ニューラルネットワーク



- 多層化パーセプトロン
  入力から出力まですべての層で完全に結合されているニューラルネットワーク。分類や回帰に用いられる。
- 畳み込みニューラルネットワーク（CNN）
  畳み込み演算（移動平均）により画像の特徴を抽出する層を加えたニューラルネットワーク。画像分類や物体検出に広く使われている。
- 再帰型ニューラルネットワーク（RNN）
  過去の情報を記録するパラメータを加えたニューラルネットワーク。時系列データを利用した予測などに使われる。LSTM や GRU などの改良型がある。
- 敵対的生成ネットワーク（GAN）
  敵対的生成ネットワークは、2 つのニューラルネットワーク（生成器と識別器）を対立的に学習させ、その後、生成器だけを利用してデータ生成させる。

深層学習

ニューラルネットワーク

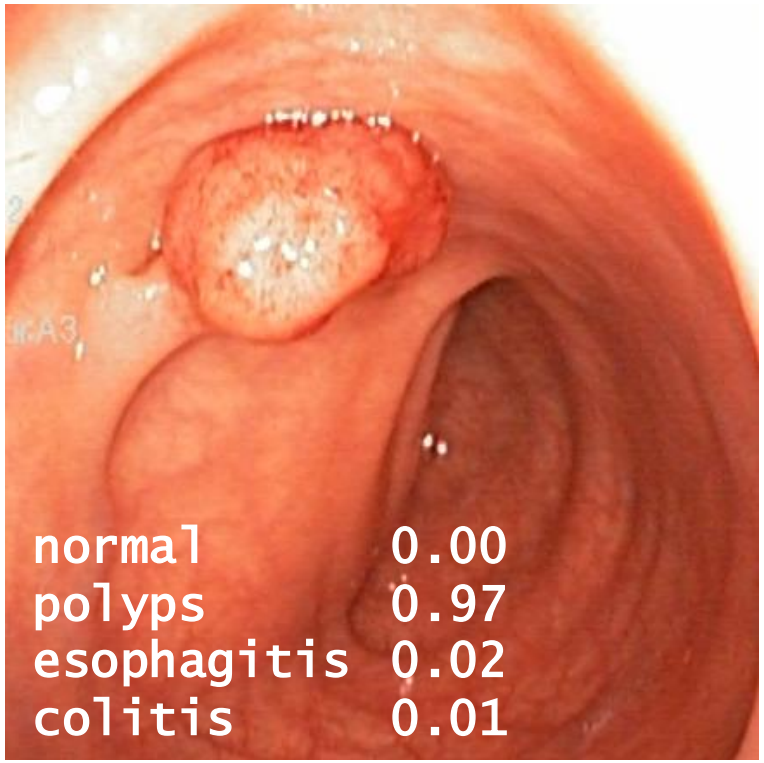
畳み込みニューラルネットワーク

物体分類

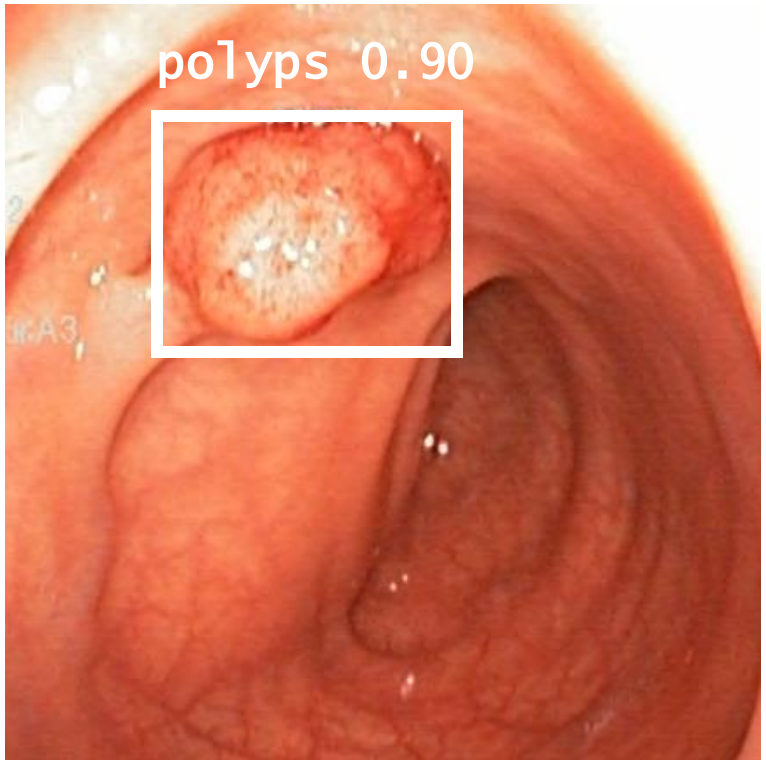
物体検出・セグメンテーション

敵対的生成ネットワーク

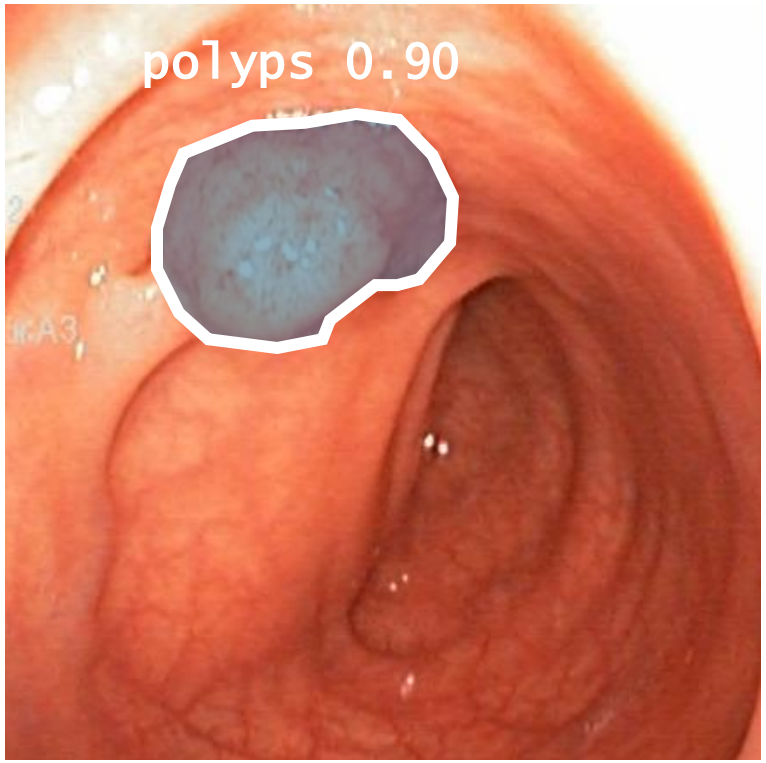
物体分類



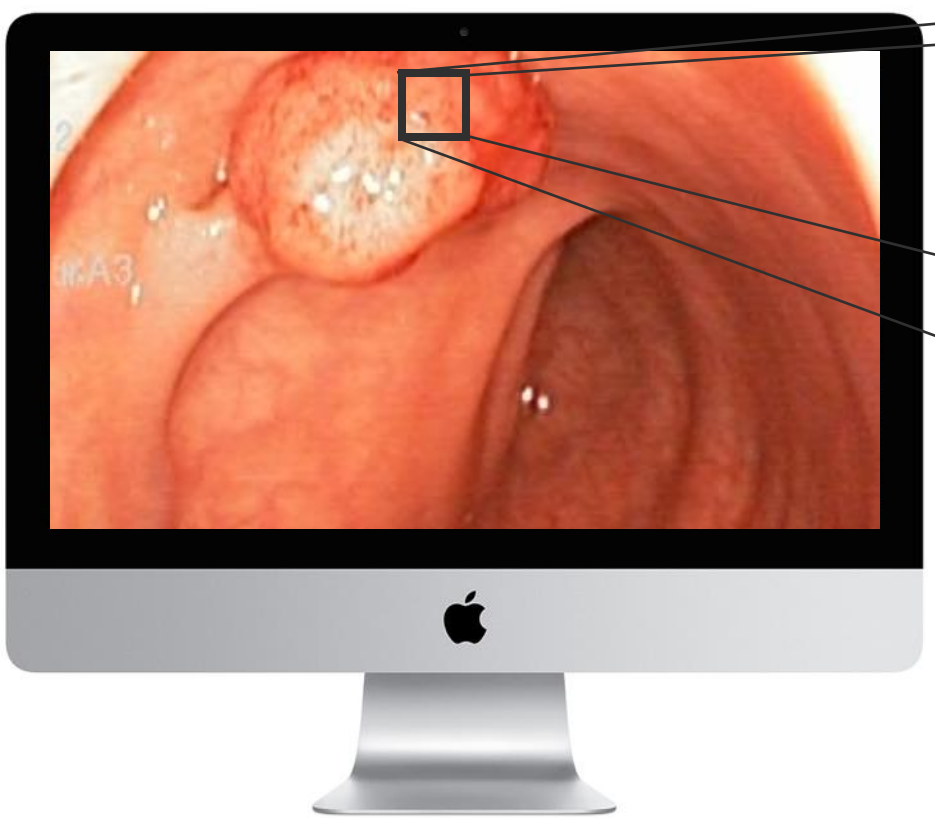
物体検出

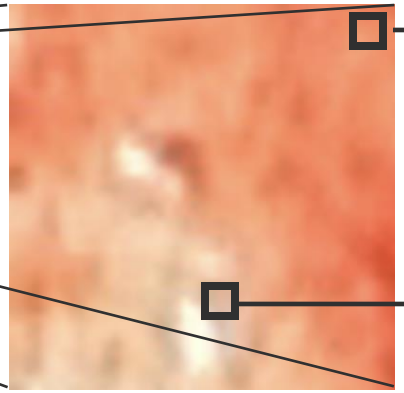


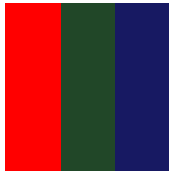
セグメンテーション



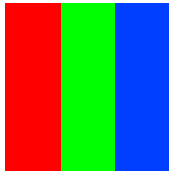
物体認識



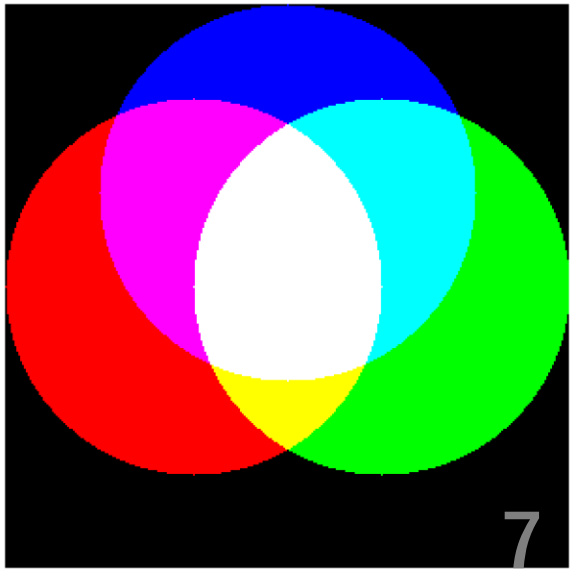




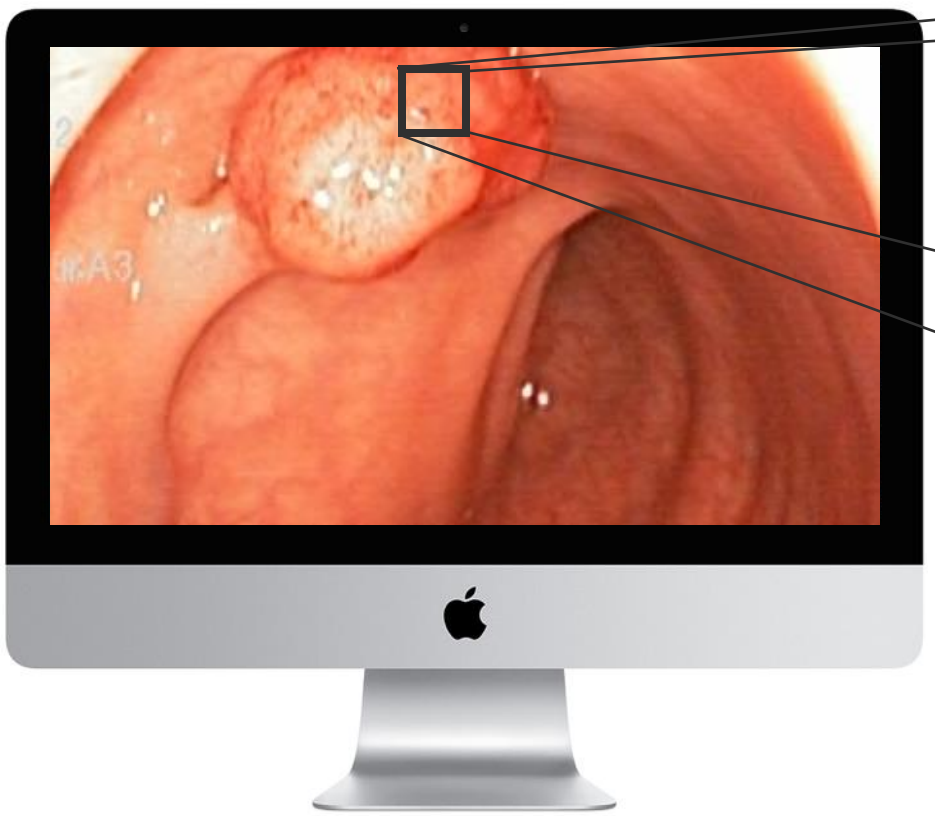
220 032 035

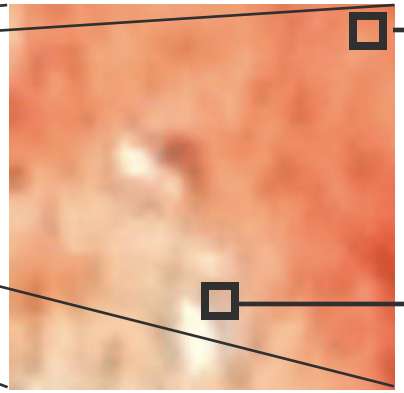


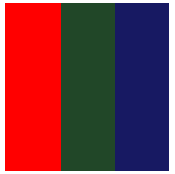
232 243 215



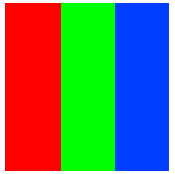
物体認識







220 032 035



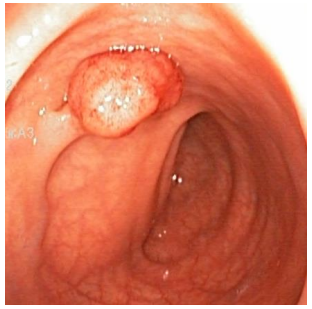
232 243 215

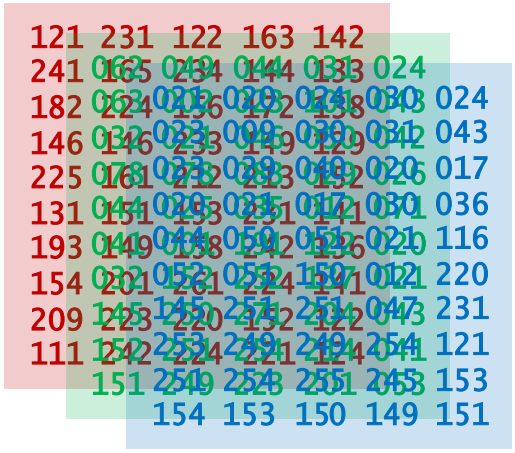
121	231	122	163	142
241	165	234	144	133
182	224	156	172	138
146	146	233	149	129
225	161	212	213	152
131	151	153	231	141
193	149	158	242	136
154	201	261	224	141
209	223	220	152	122
111	242	224	221	124

062	049	044	031	024
063	002	023	101	043
032	021	046	050	042
078	078	083	029	026
044	025	035	012	071
041	001	191	122	020
032	152	252	137	021
145	250	271	204	043
152	251	254	214	041
151	249	223	201	053

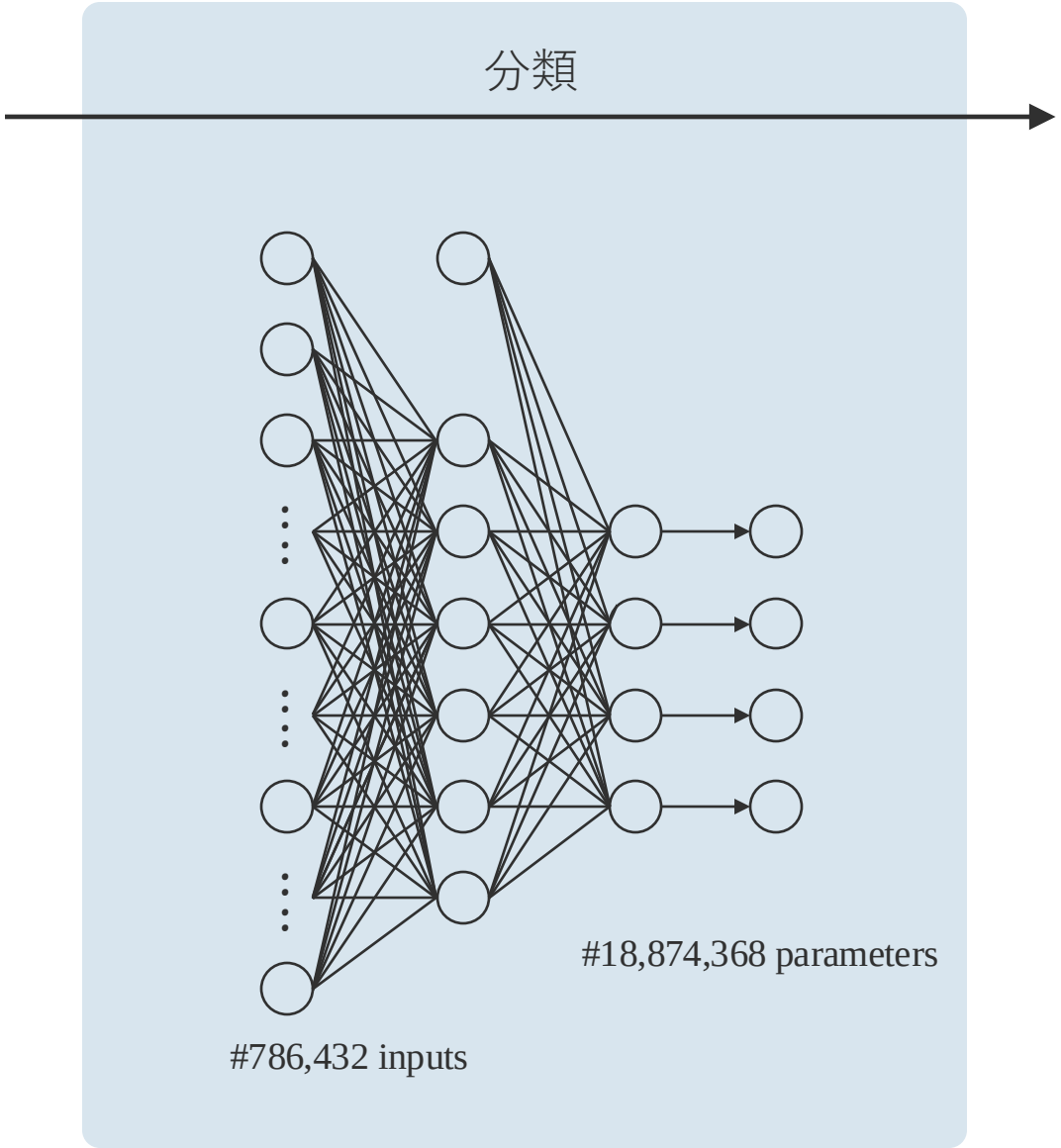
021	020	024	030	024
023	009	030	031	043
023	029	040	020	017
020	021	017	030	036
044	050	051	021	116
052	051	150	012	220
145	251	251	047	231
253	249	249	254	121
251	254	255	245	153
154	153	150	149	151

物体分類



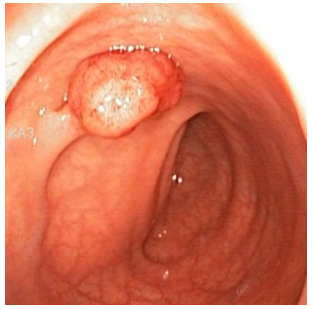


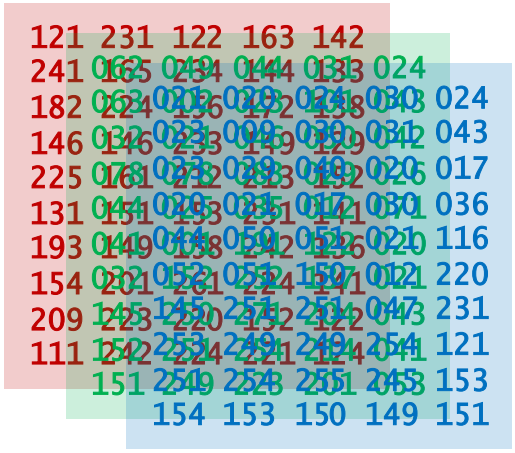
512 × 512 × 3 = 786,432



normal	0.00
polyps	0.97
esophagitis	0.02
colitis	0.01

物体分類





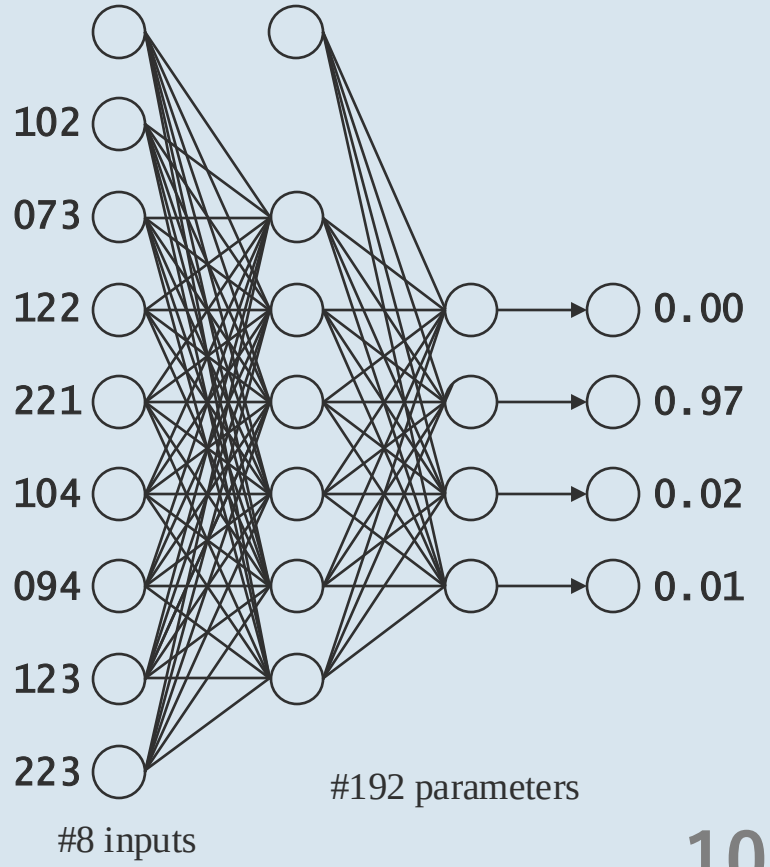
512 × 512 × 3 = 786,432

特徴抽出

102
073
122
221
104
094
123
223

分類

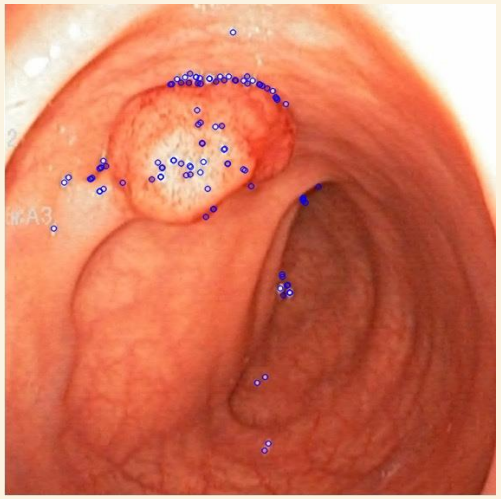
normal	0.00
polyps	0.97
esophagitis	0.02
colitis	0.01

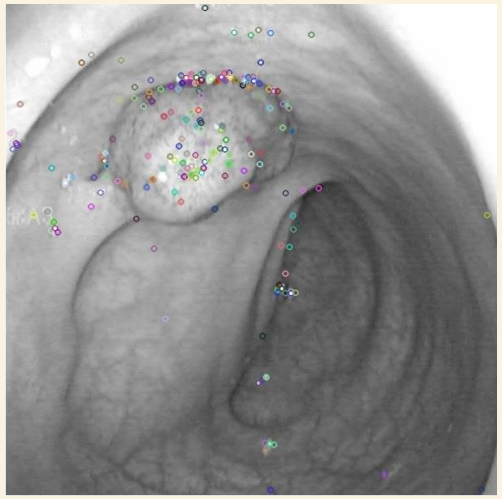


SIFT

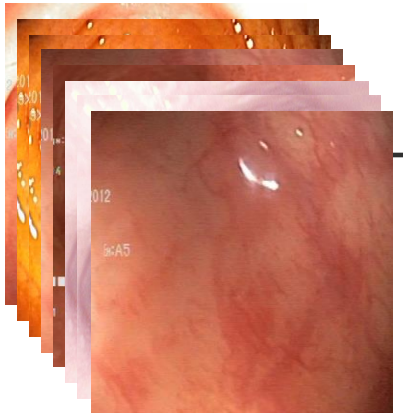
SURF

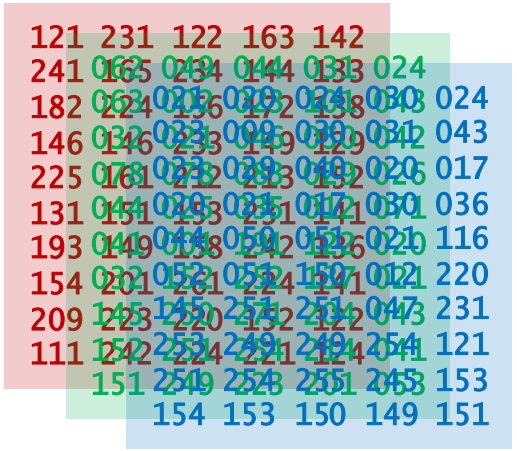
AKAZE





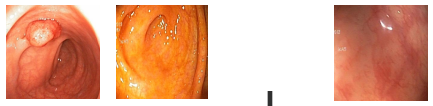
物体分類（学習）



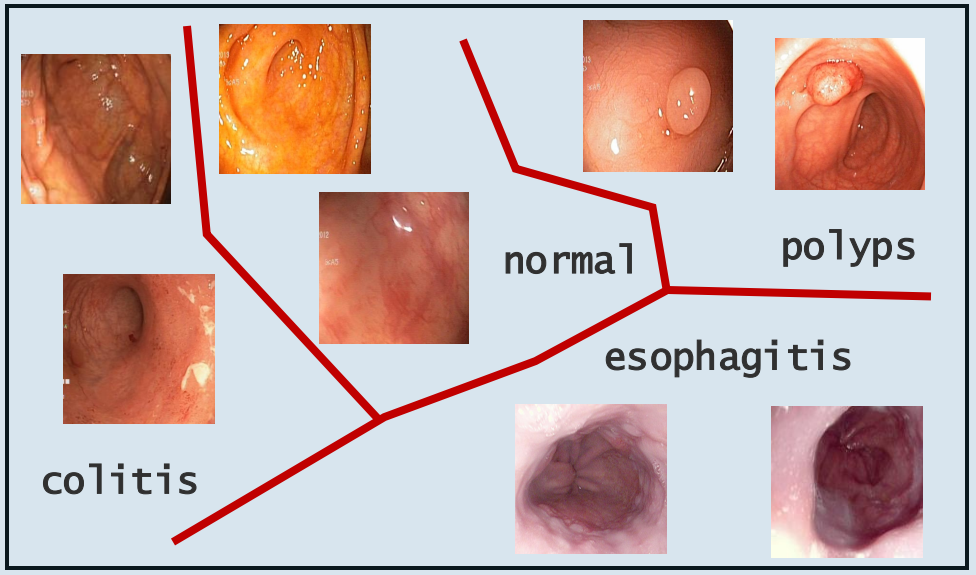


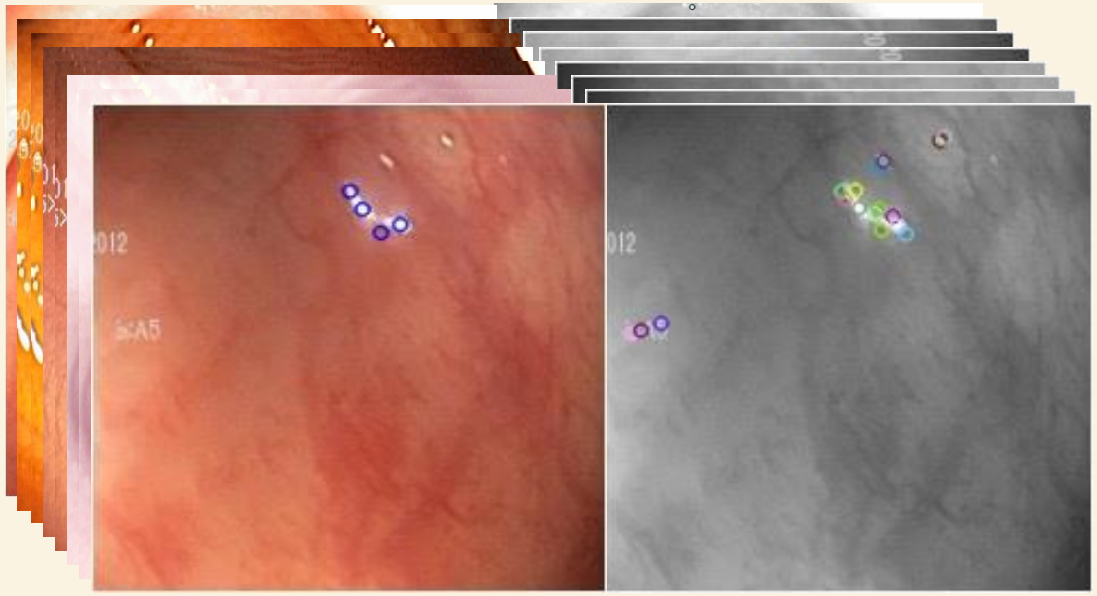
特徴抽出

102	212	....	122
073	131		211
122	032		212
221	103		042
104	093		043
094	143		241
123	221		223
223	031		201

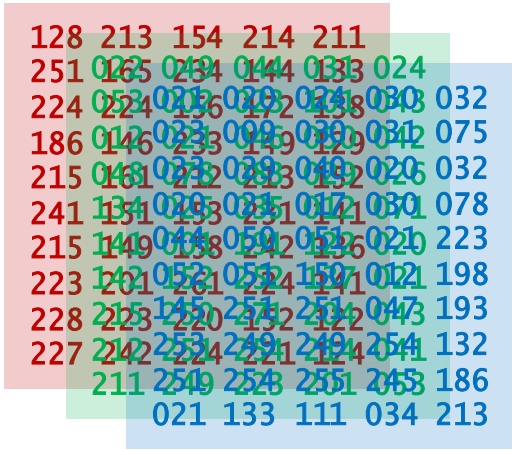


学習





物体分類（推論）

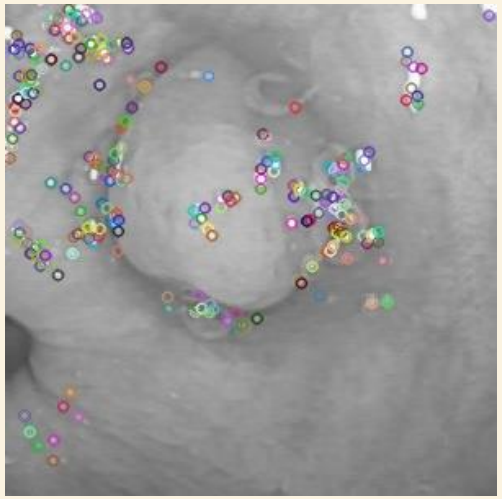


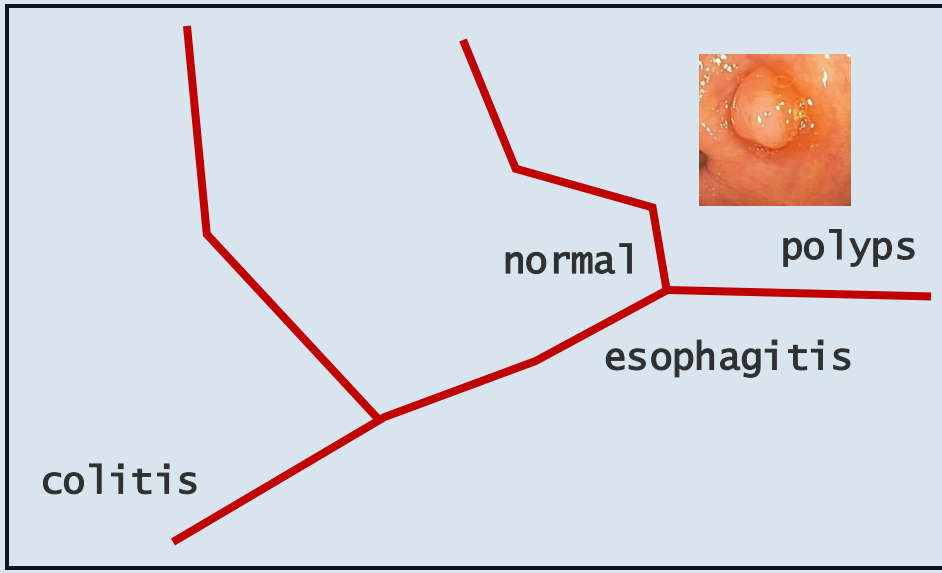
特徴抽出

220
023
019
233
098
212
023
167

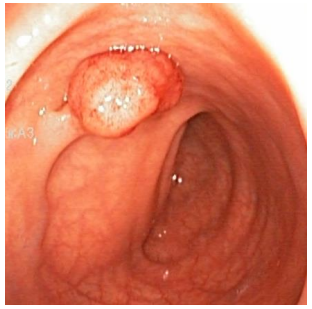
分類

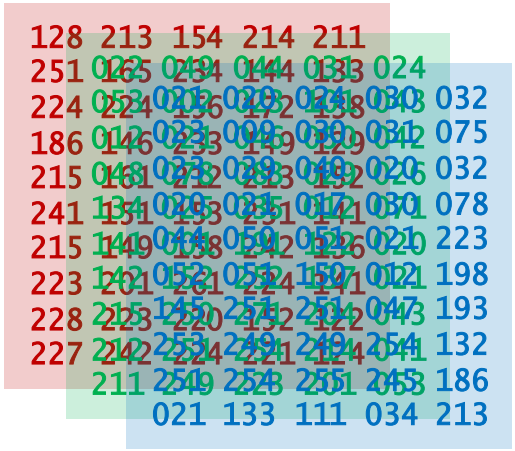
normal	0.01
polyps	0.99
esophagitis	0.00
colitis	0.00





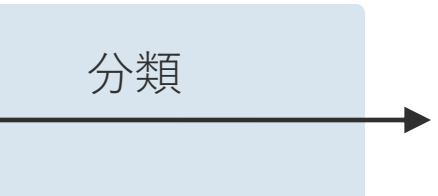
物体分類





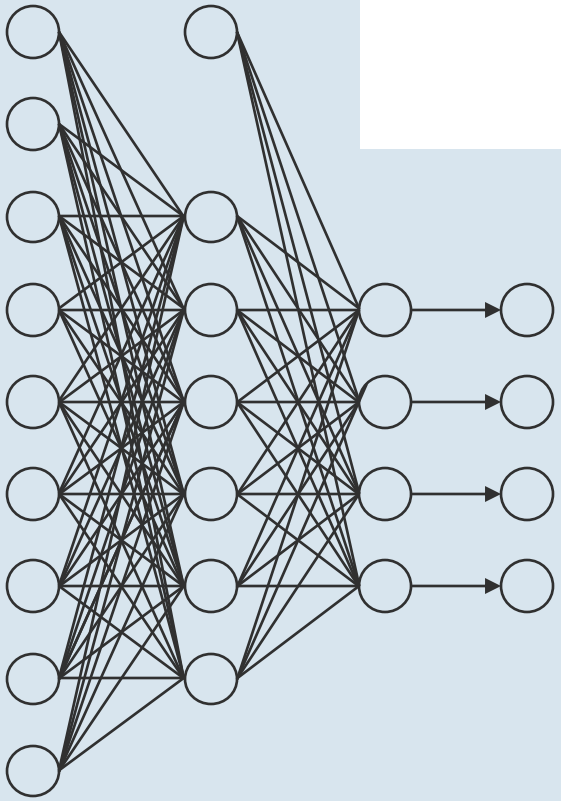
特徴抽出

分類

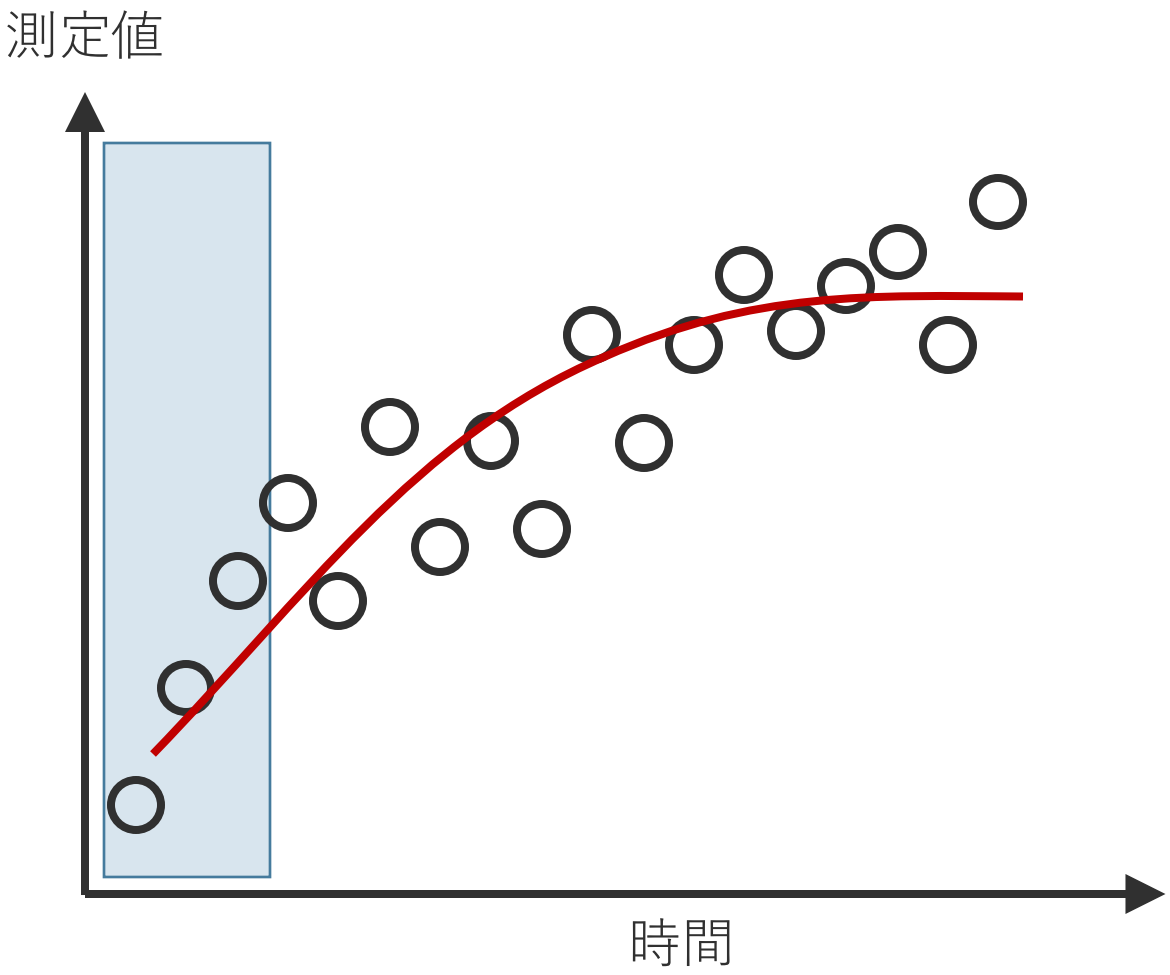


特徴抽出を、ニューラルネットワークのように自動的に学習する仕組みは実現できないだろうか？

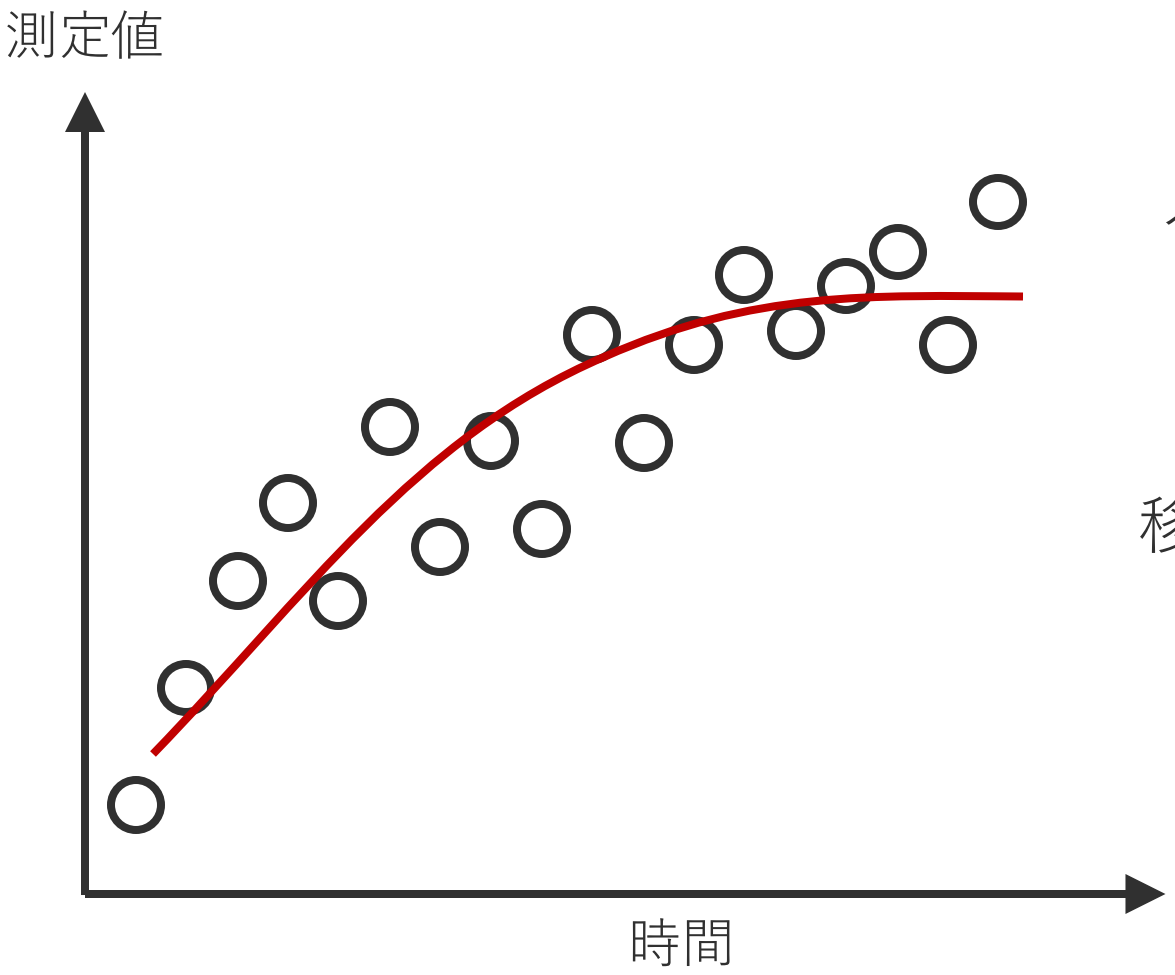
- 畳み込み演算
- プーリング演算

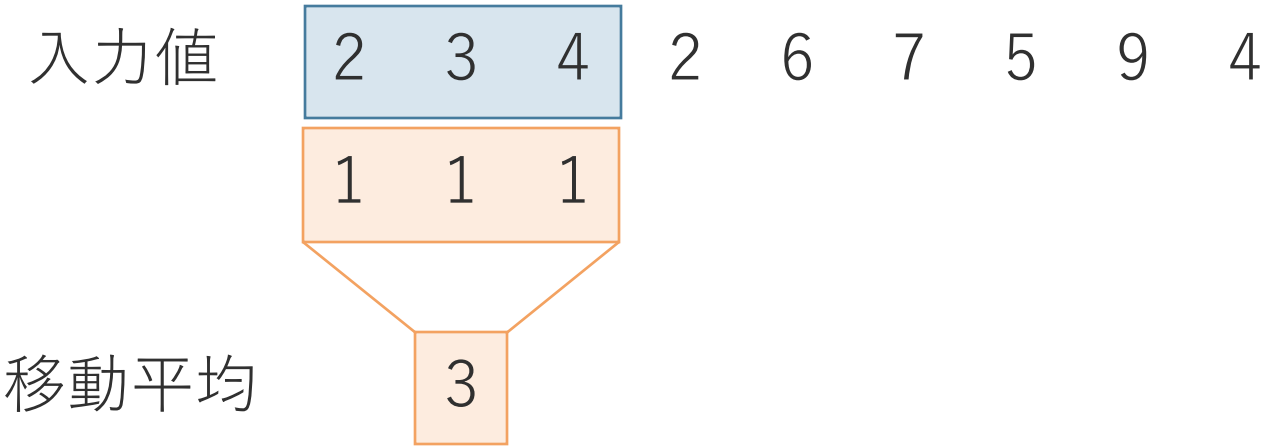


移動平均

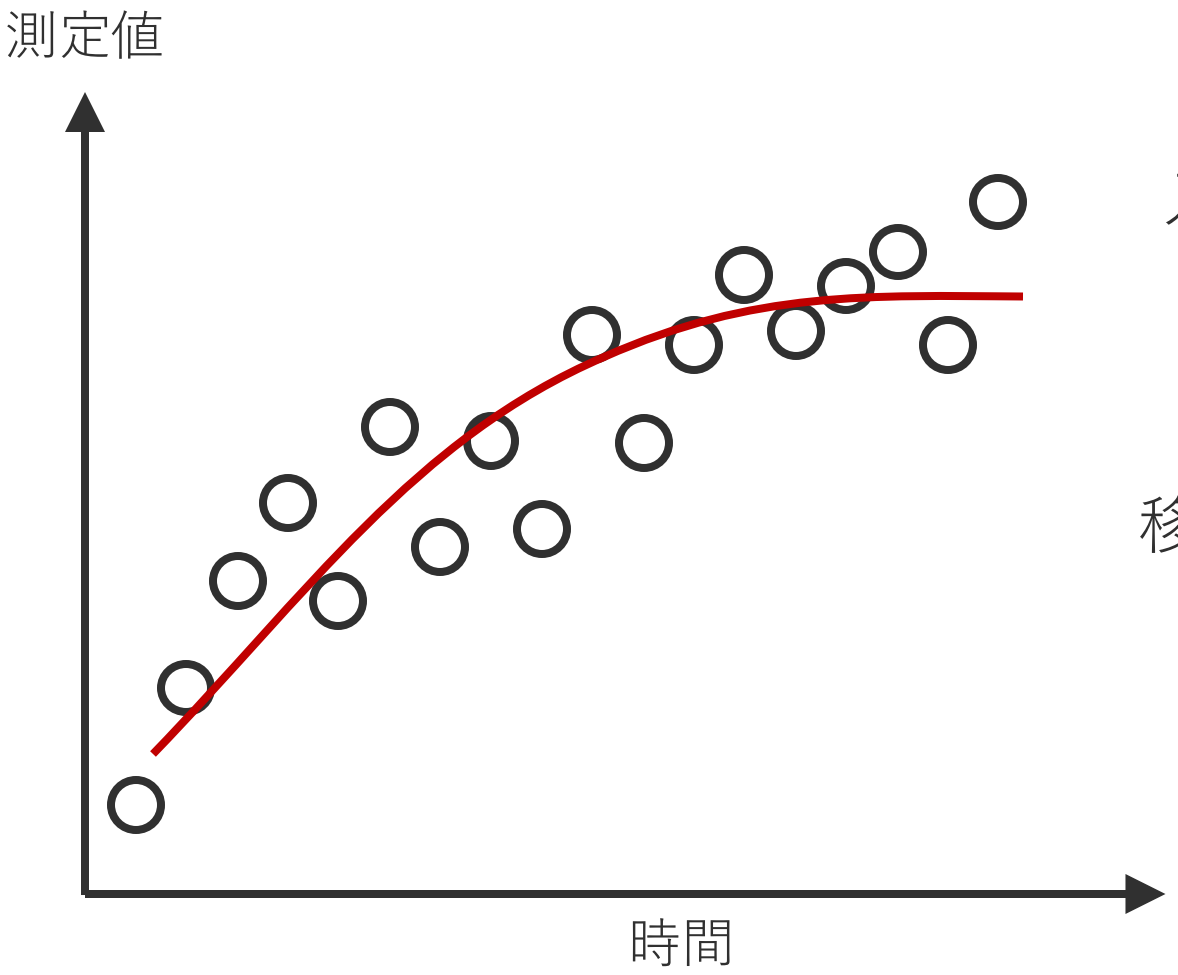


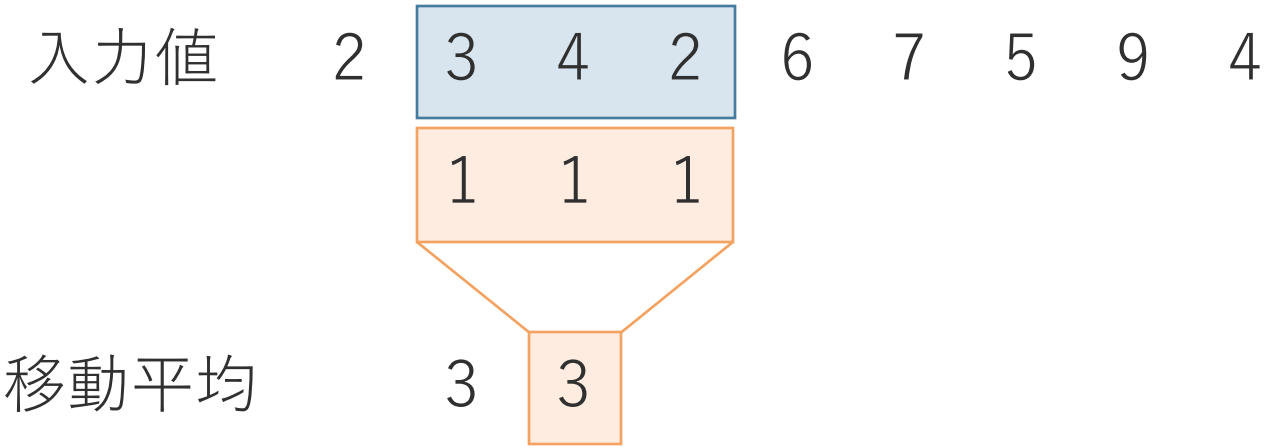
移動平均



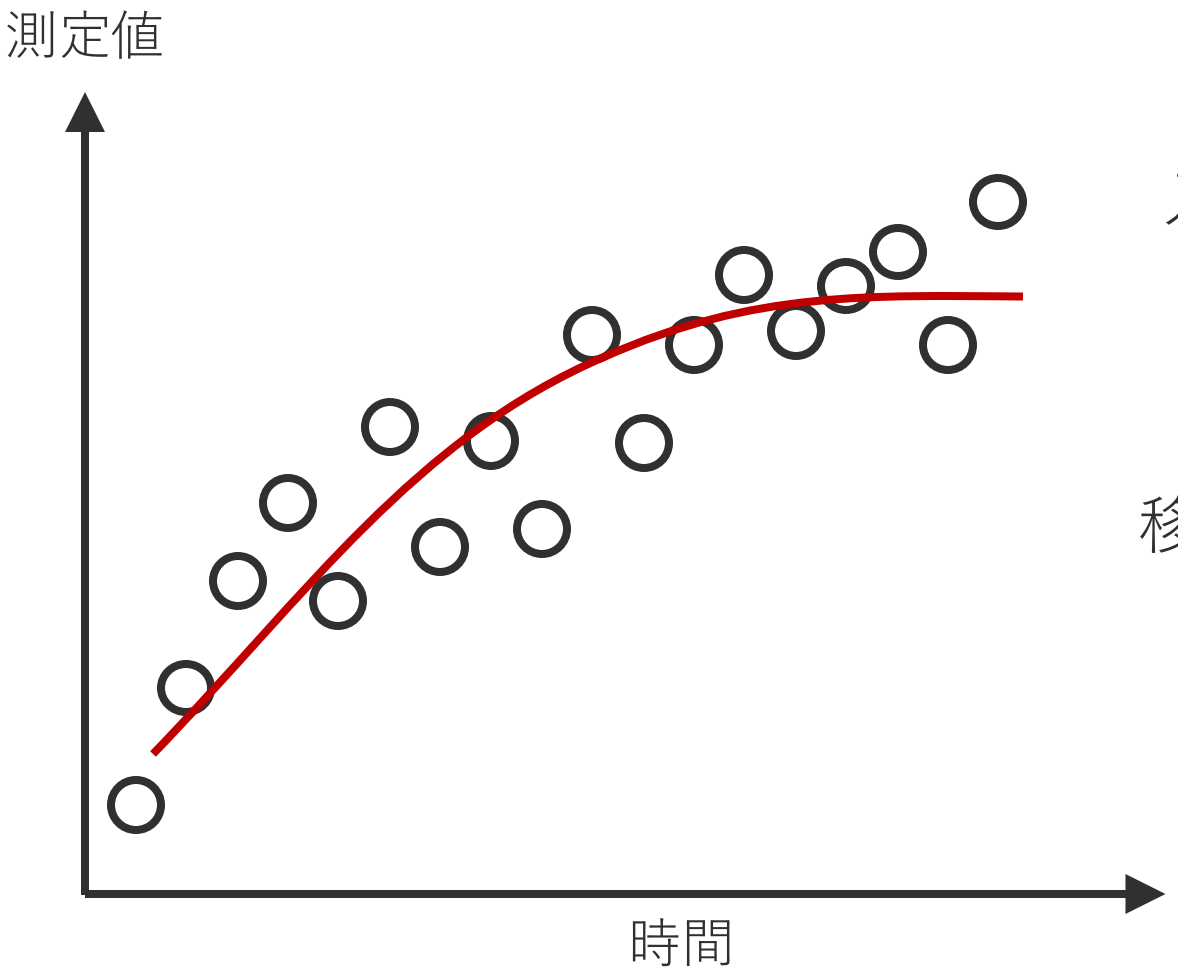


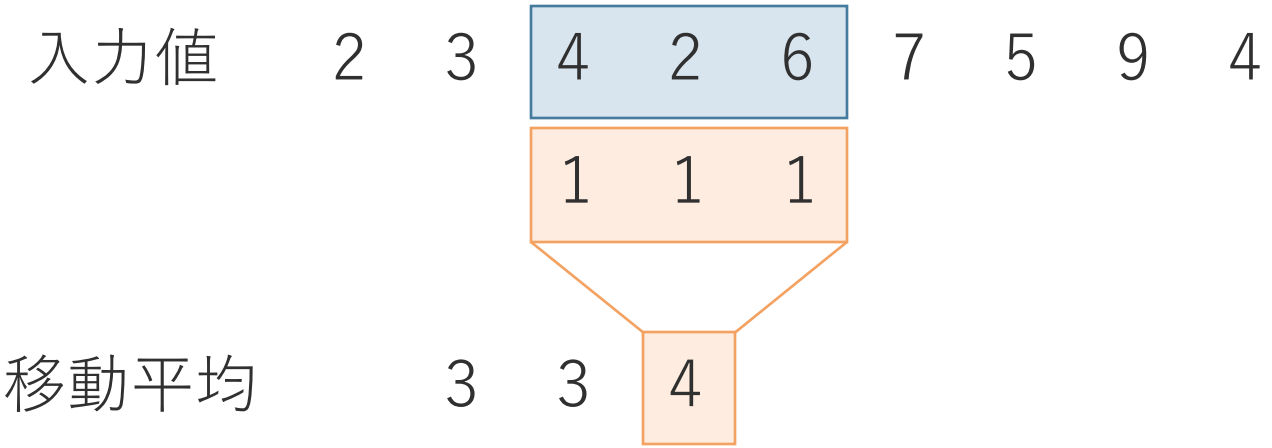
移動平均



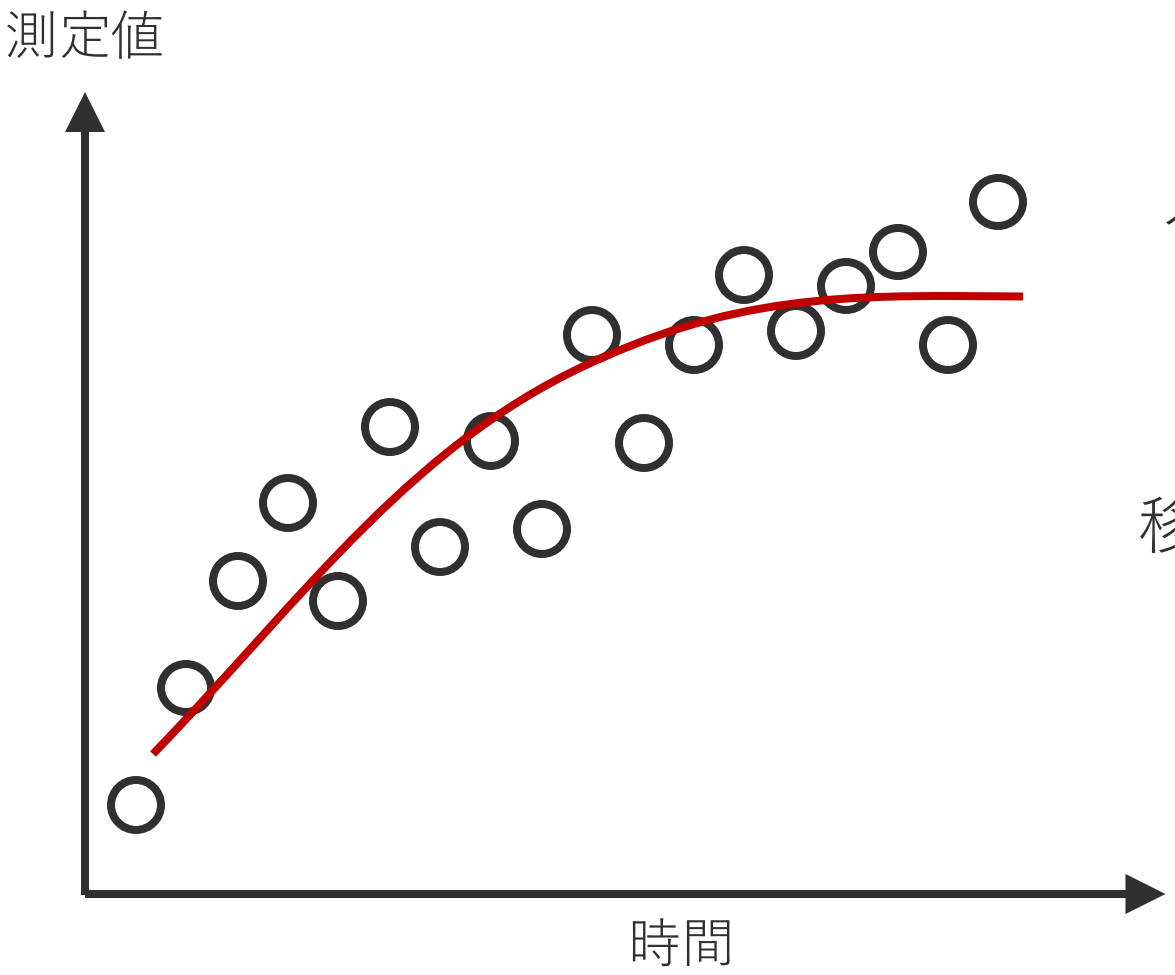


移動平均



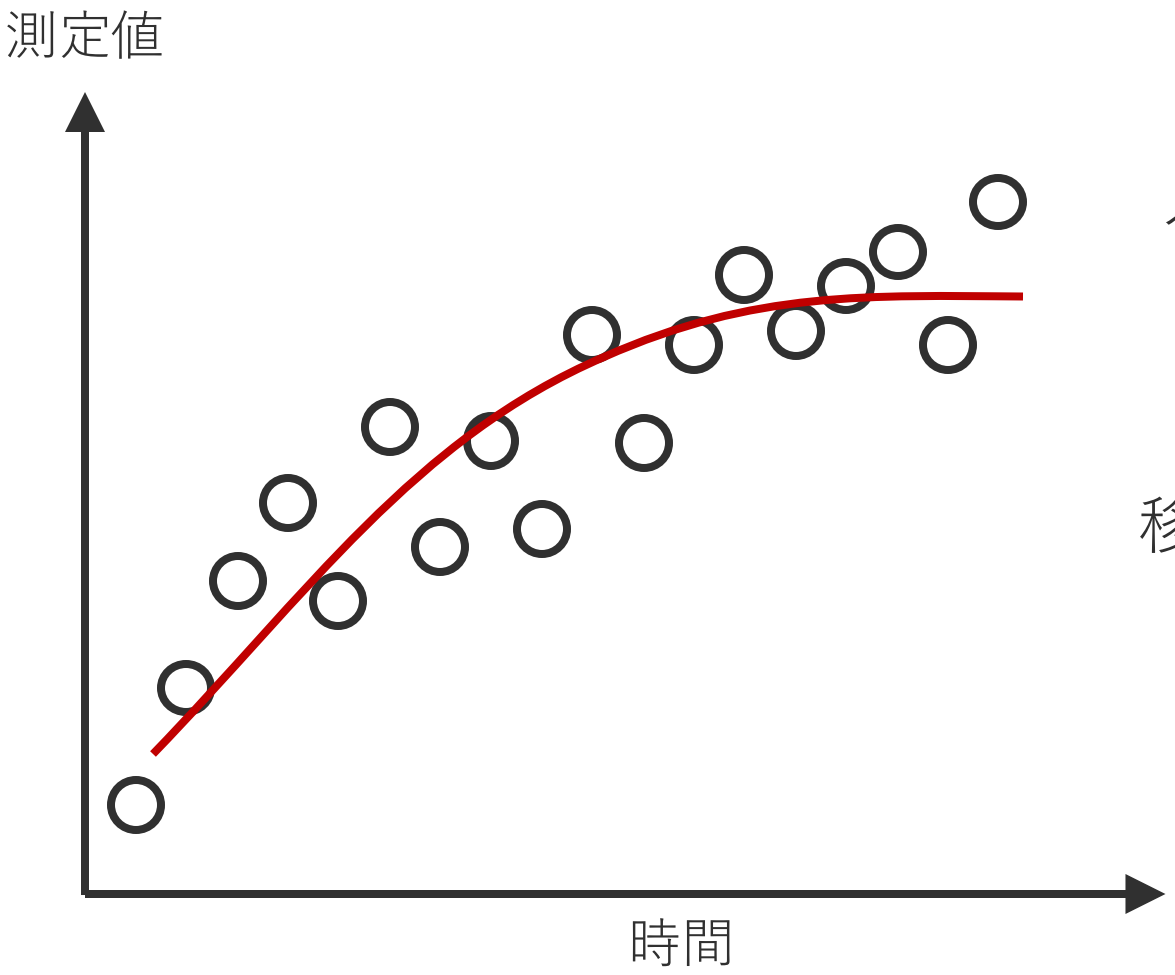


移動平均



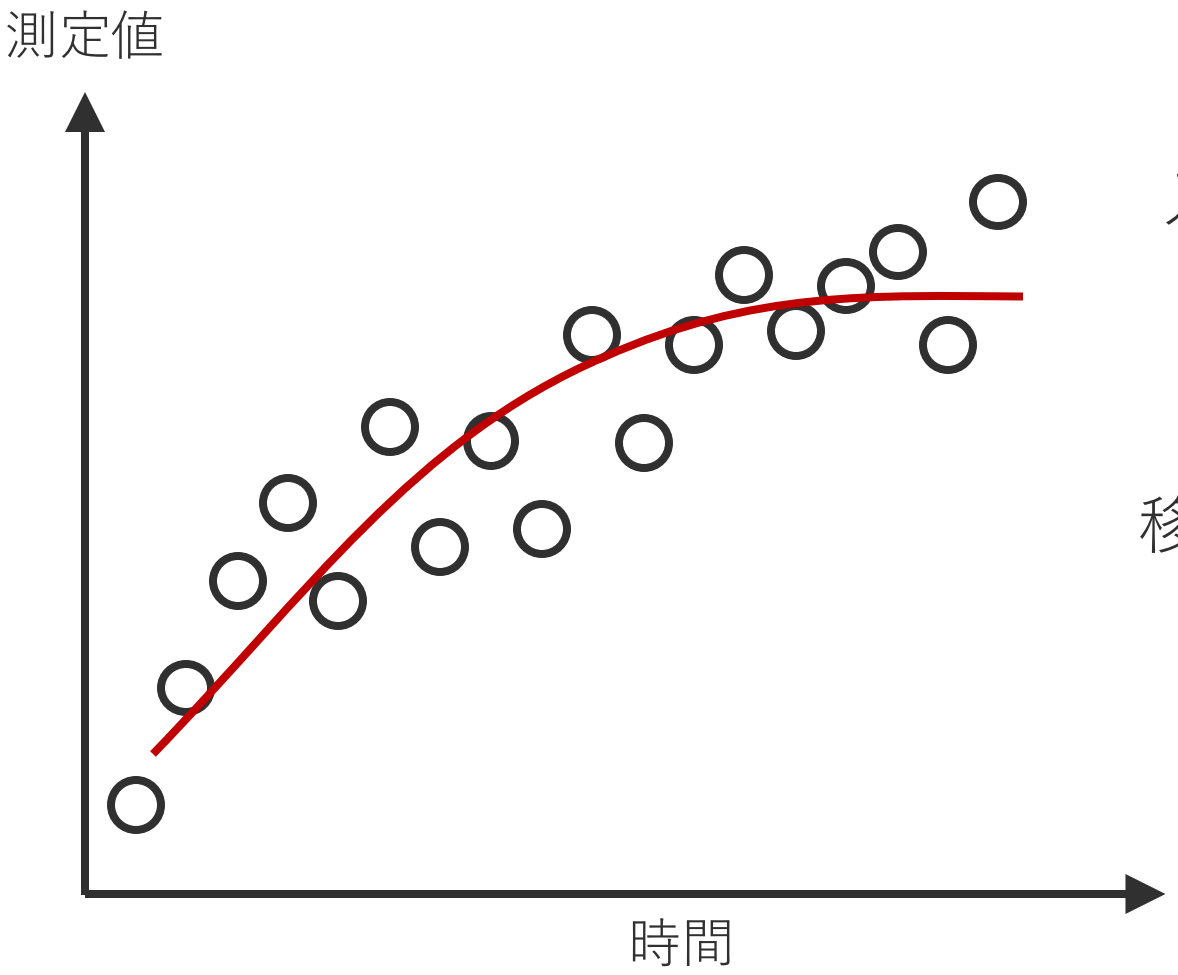
入力値	2	3	4	2	6	7	5	9	4
				1	1	1			
移動平均			3	3	4	5			

移動平均



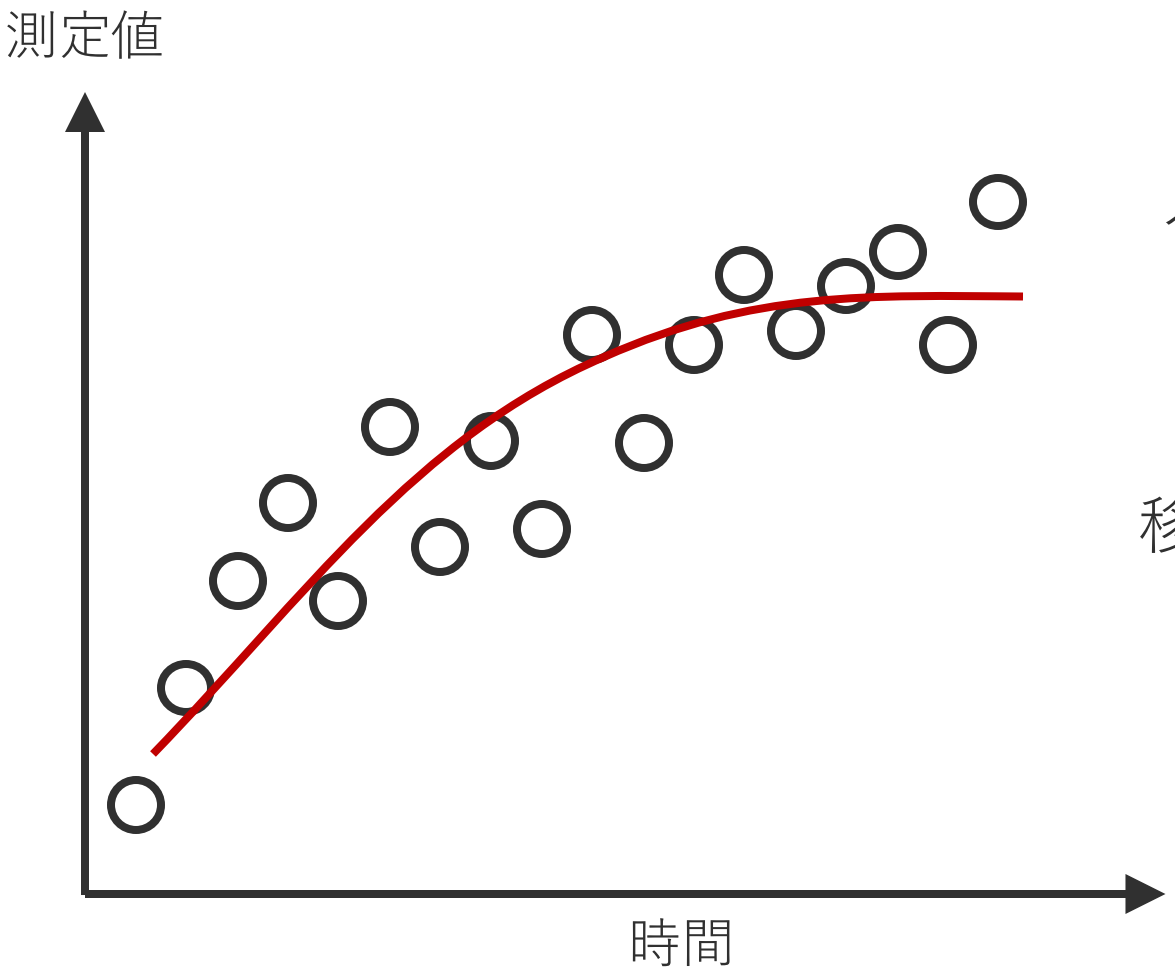
入力値	2	3	4	2	6	7	5	9	4
						1	1	1	
移動平均		3	3	4	5	6	7		

移動平均



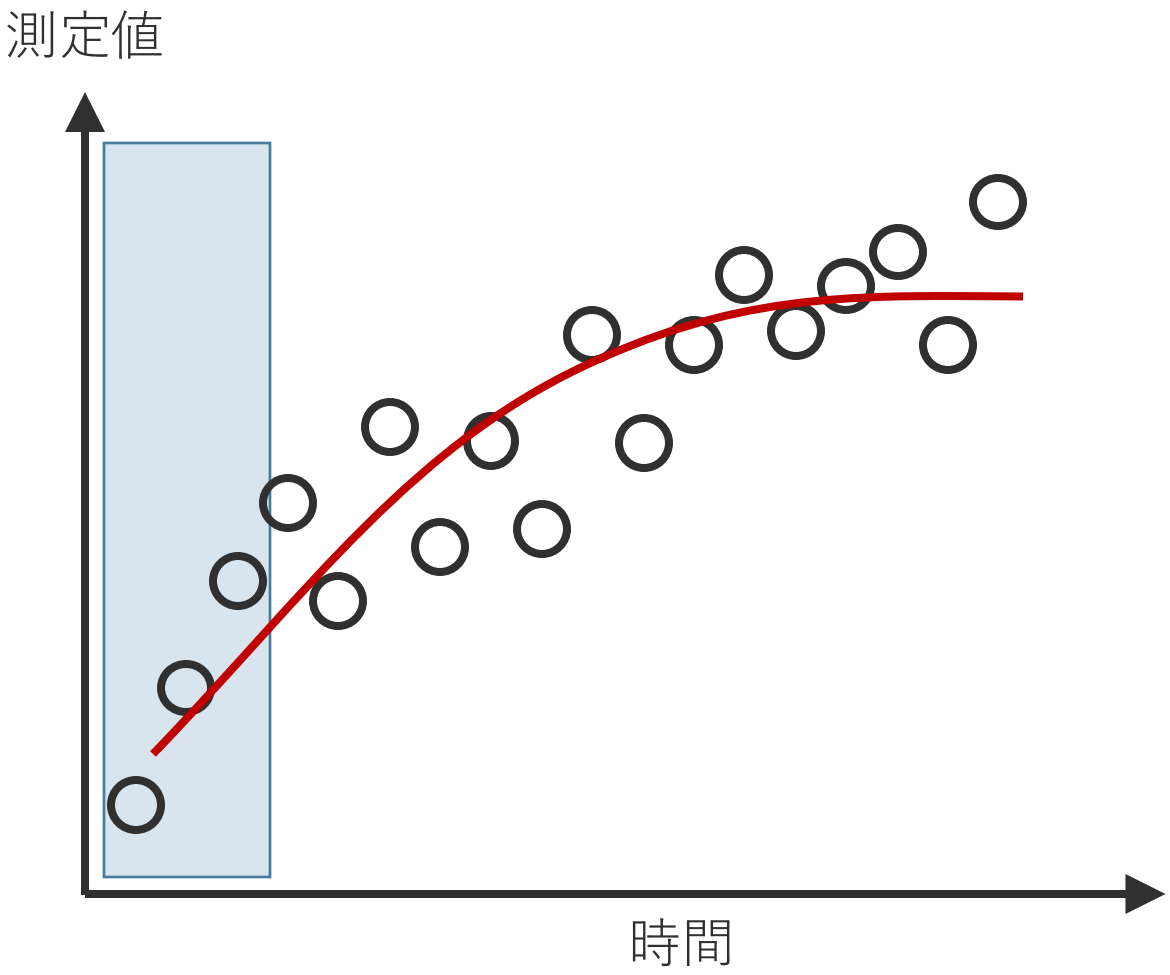
入力値	2	3	4	2	6	7	5	9	4
							1	1	1
移動平均		3	3	4	5	6	7	6	

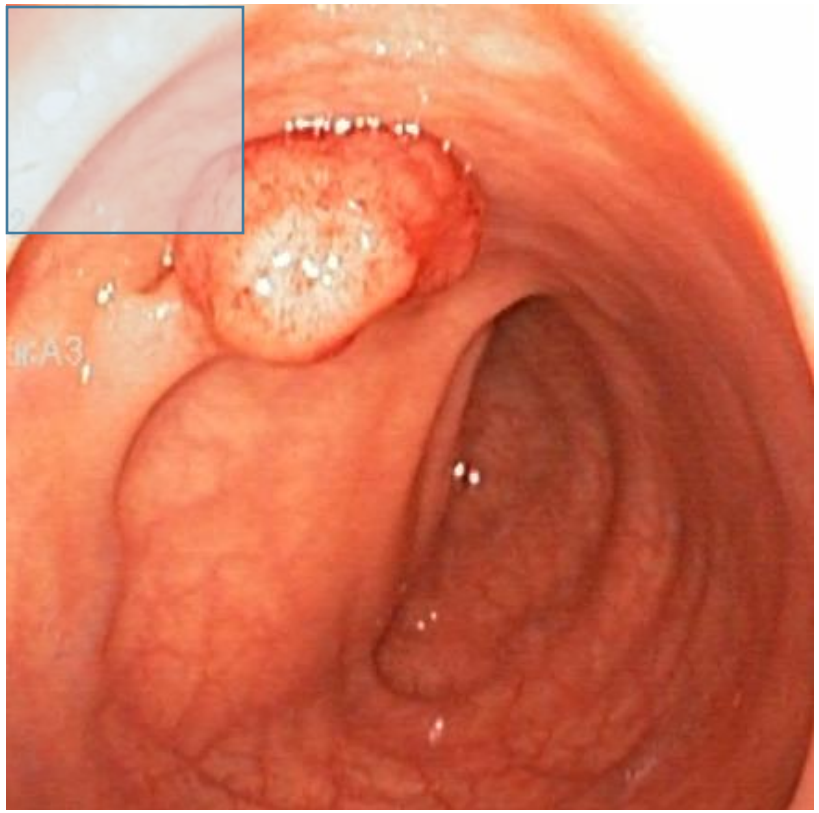
移動平均



入力値	2	3	4	2	6	7	5	9	4
							1	0	1
移動平均		3.0	2.5	5.0	4.5	5.5	8.0	4.5	

畳み込み（移動平均）





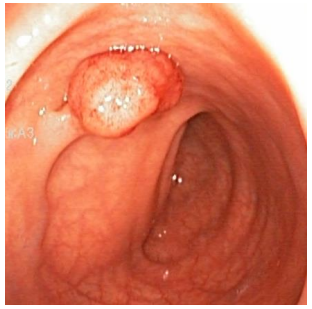
畳み込み演算

Grayscale Photo

162	149	144	131	124
163	202	123	101	143
132	121	146	150	142
178	178	183	129	126
144	125	135	112	171
241	201	191	122	120
132	152	152	137	121
145	150	171	104	143
152	151	154	114	141
151	149	123	101	153

1 channel

Color Photo



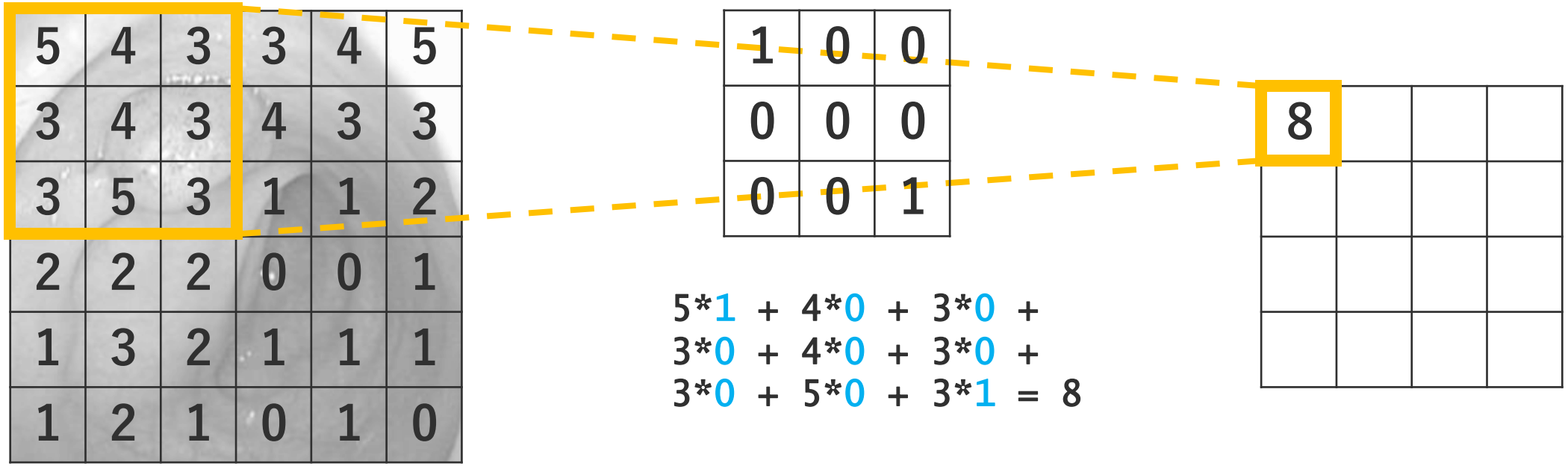
021	031	022	063	042
041	065	034	044	033
082	024	056	072	038
046	046	033	049	029
025	061	032	103	053
031	053	053	031	041
093	049	058	093	036
054	101	051	024	041
109	033	026	053	033
111	042	024	021	024

162	149	144	131	124
183	202	123	101	143
132	121	146	150	142
178	178	183	129	126
144	125	135	112	171
241	201	191	122	120
132	152	152	137	121
145	150	171	104	143
152	151	154	114	141
151	149	123	101	153

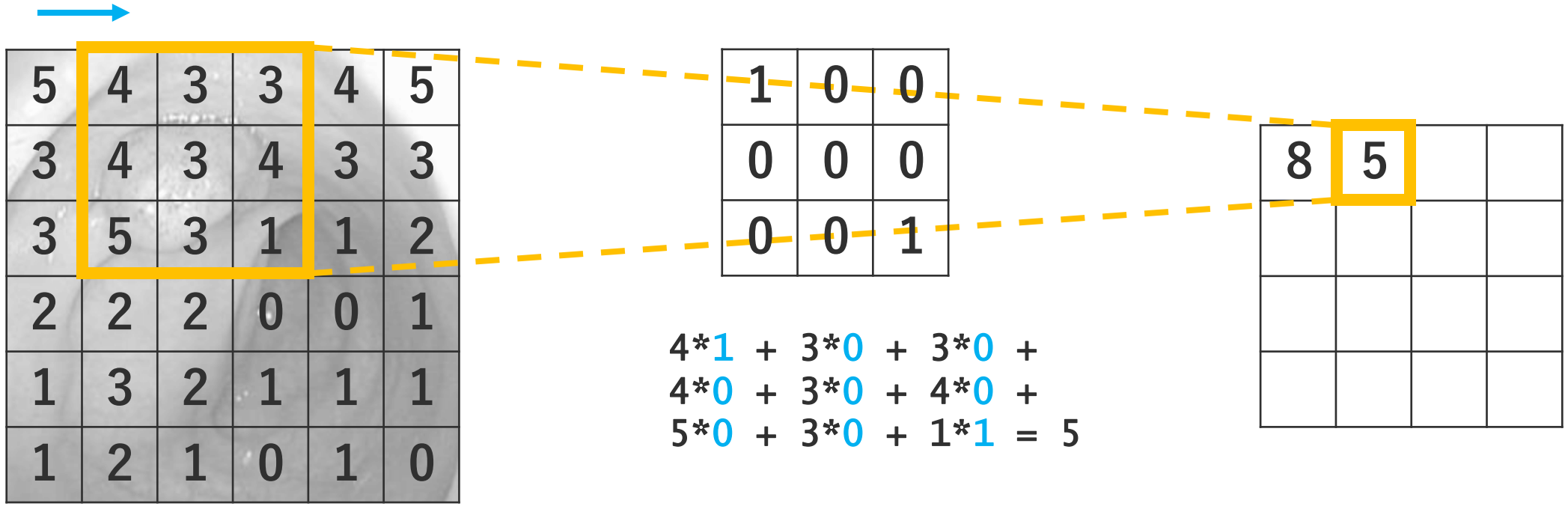
224	243	217	236	216
220	231	221	243	231
253	249	249	254	221
251	254	255	245	253
254	253	250	249	251

3 channels

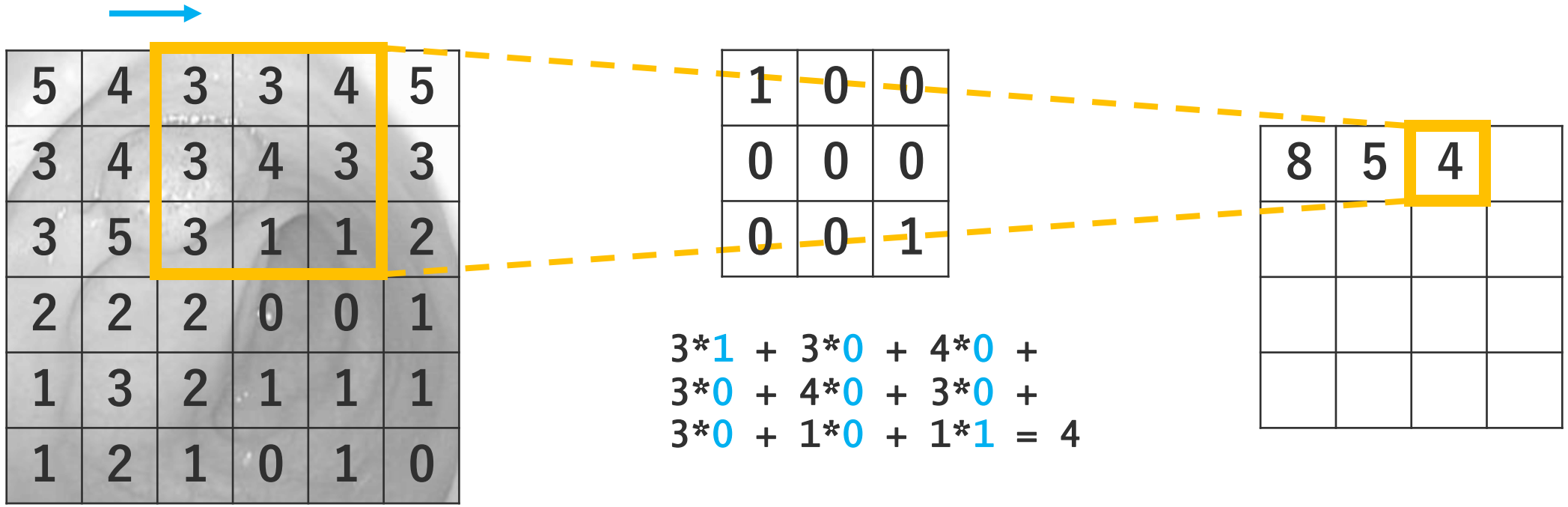
畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
1 0 0
0 0 0
0 0 1
5*1 + 4*0 + 3*0 +
3*0 + 4*0 + 3*0 +
3*0 + 5*0 + 3*1 = 8
8

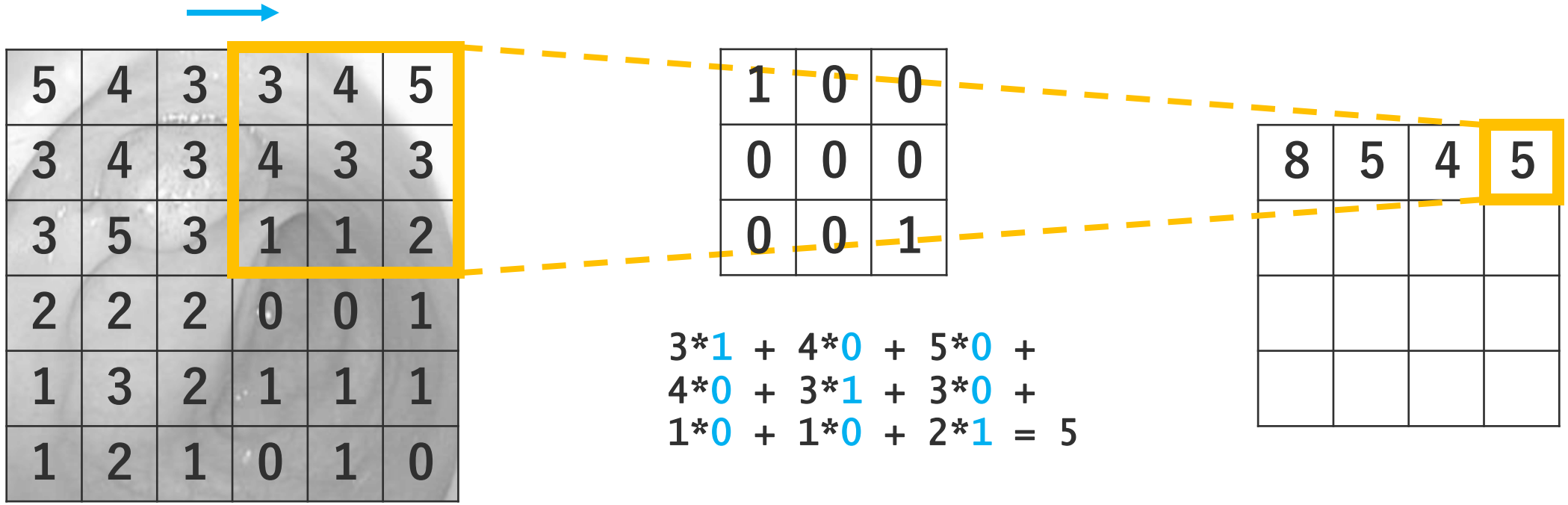
畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
1 0 0
0 0 0
0 0 1
4*1 + 3*0 + 3*0 +
4*0 + 3*0 + 4*0 +
5*0 + 3*0 + 1*1 = 5
8 5

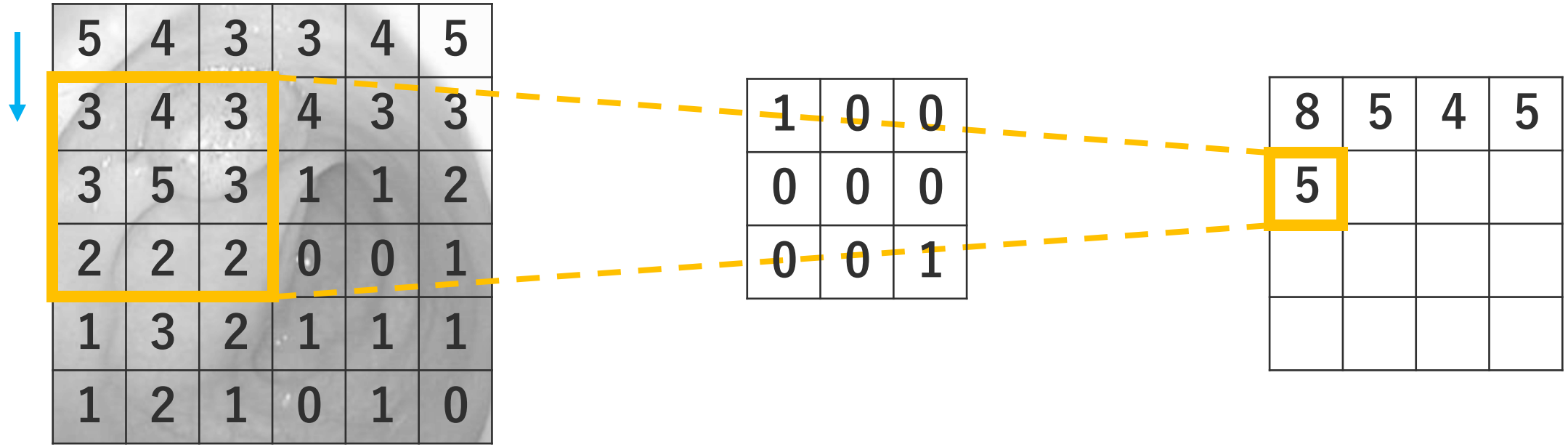
畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
1 0 0
0 0 0
0 0 1
3*1 + 3*0 + 4*0 +
3*0 + 4*0 + 3*0 +
3*0 + 1*0 + 1*1 = 4
8 5 4

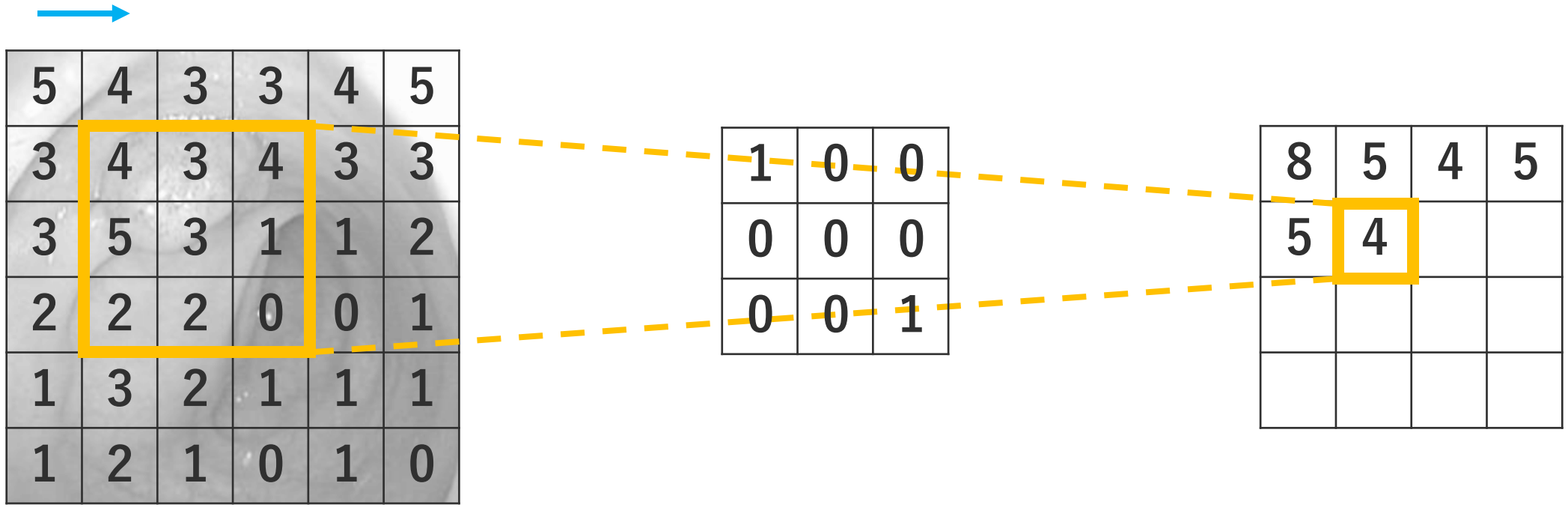
畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
1 0 0
0 0 0
0 0 1
3*1 + 4*0 + 5*0 +
4*0 + 3*1 + 3*0 +
1*0 + 1*0 + 2*1 = 5
8 5 4 5

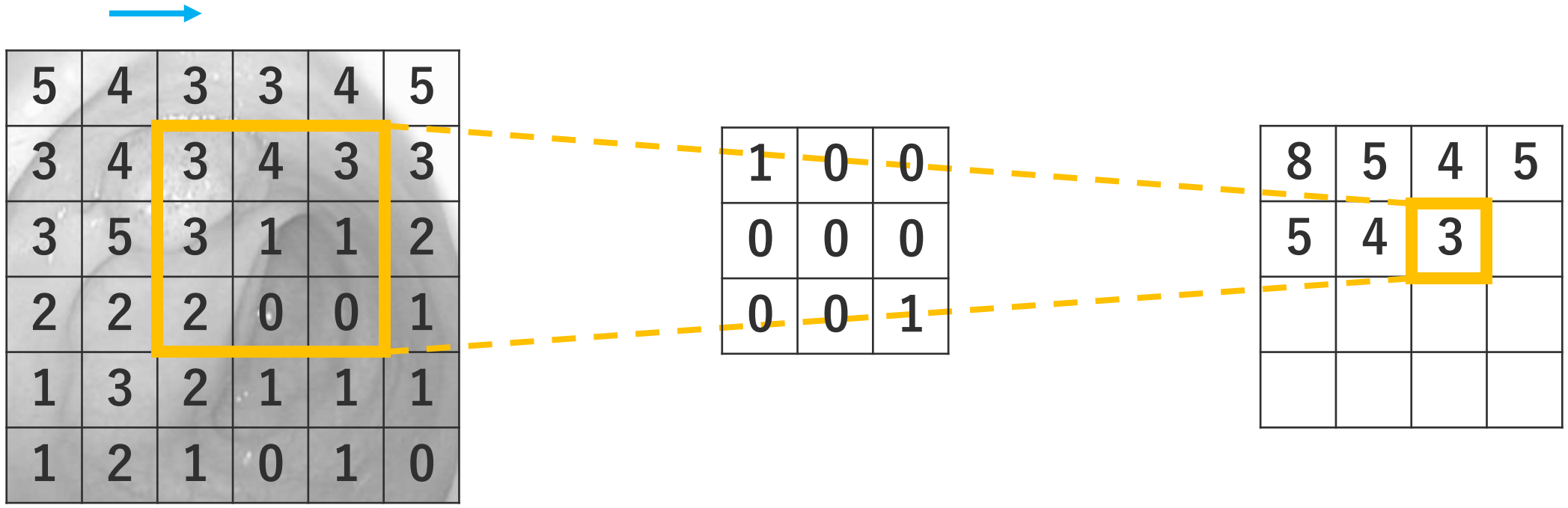
畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
1 0 0
0 0 0
0 0 1
8 5 4 5
5

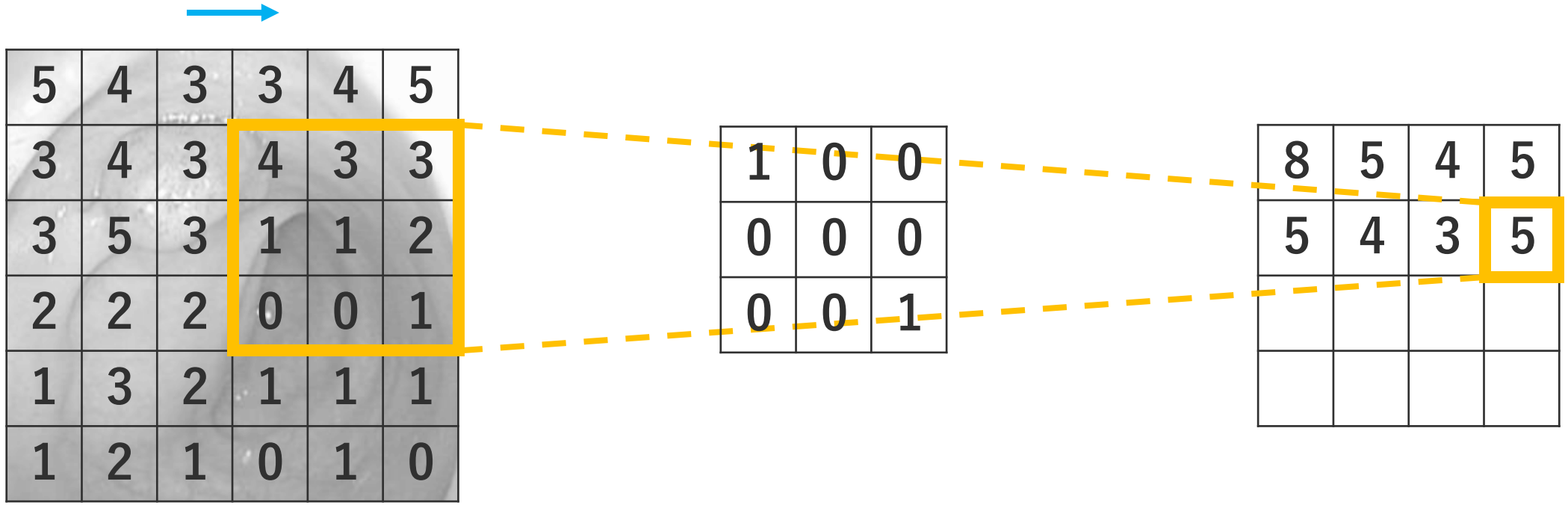
畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
1 0 0
0 0 0
0 0 1
8 5 4 5
5 4

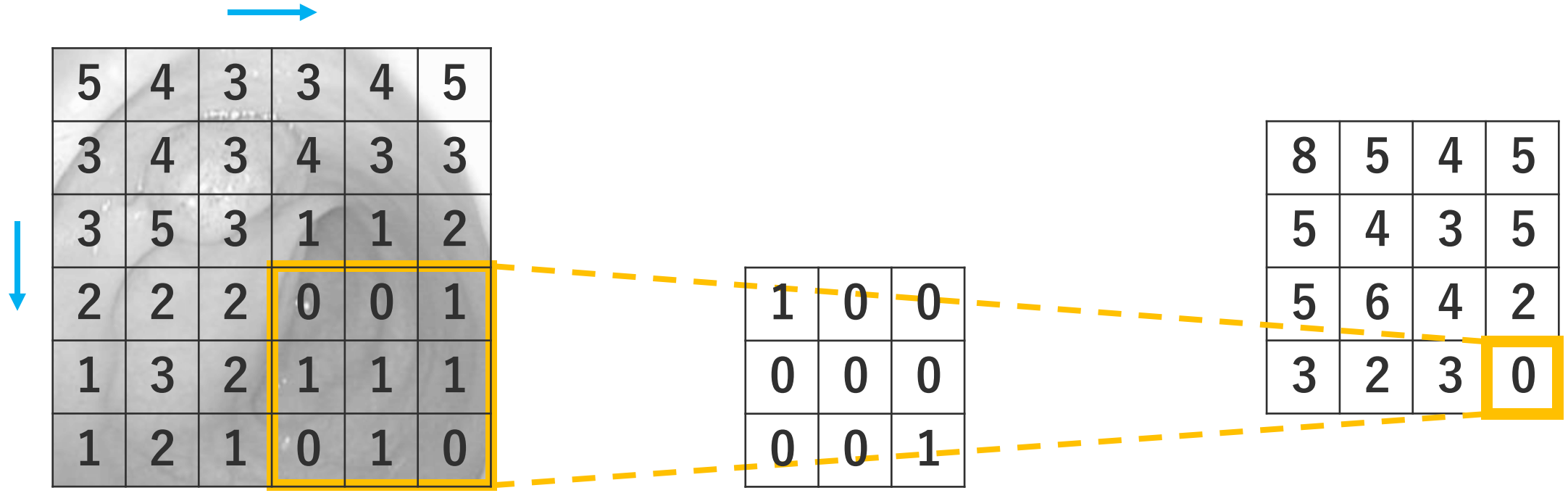
畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
1 0 0
0 0 0
0 0 1
8 5 4 5
5 4 3

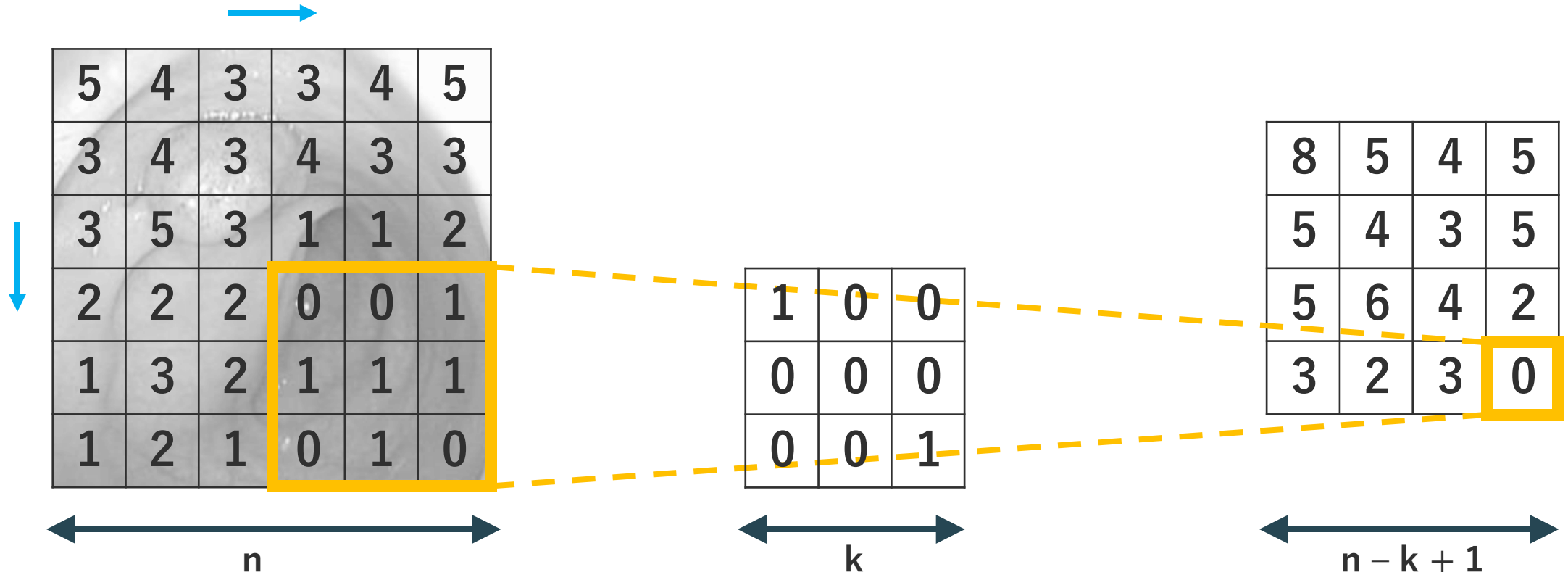
畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
1 0 0
0 0 0
0 0 1
8 5 4 5
5 4 3 5

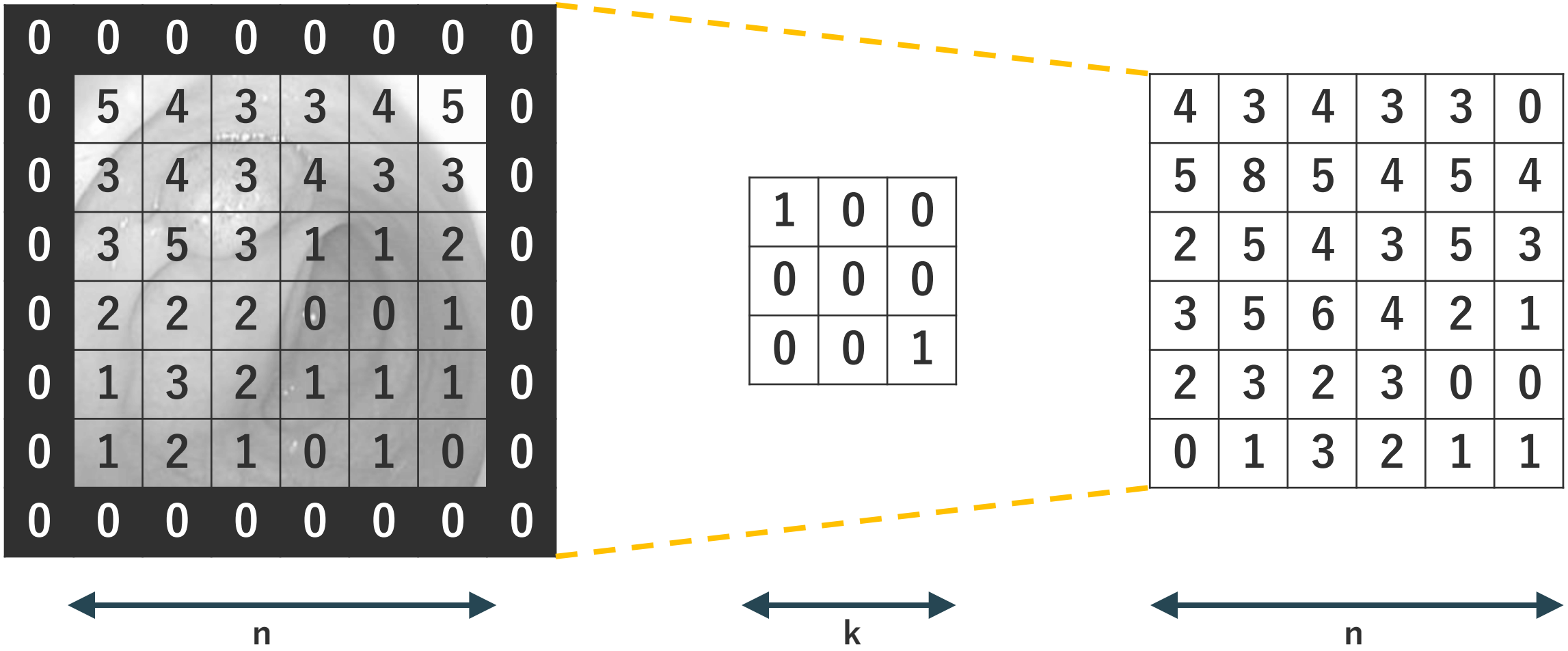
畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
1 0 0
0 0 0
0 0 1
8 5 4 5
5 4 3 5
5 6 4 2
3 2 3 0

畳み込み演算


5 4 3 3 4 5
3 4 3 4 3 3
3 5 3 1 1 2
2 2 2 0 0 1
1 3 2 1 1 1
1 2 1 0 1 0
n
1 0 0
0 0 0
0 0 1
k
8 5 4 5
5 4 3 5
5 6 4 2
3 2 3 0
n − k + 1

畳み込み演算（zero-padding）



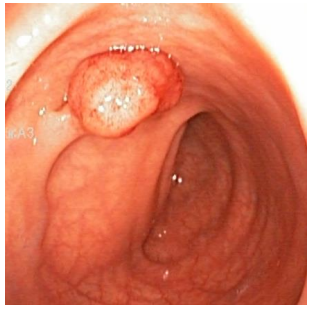
畳み込み演算

Grayscale Photo

162	149	144	131	124
163	202	123	101	143
132	121	146	150	142
178	178	183	129	126
144	125	135	112	171
241	201	191	122	120
132	152	152	137	121
145	150	171	104	143
152	151	154	114	141
151	149	123	101	153

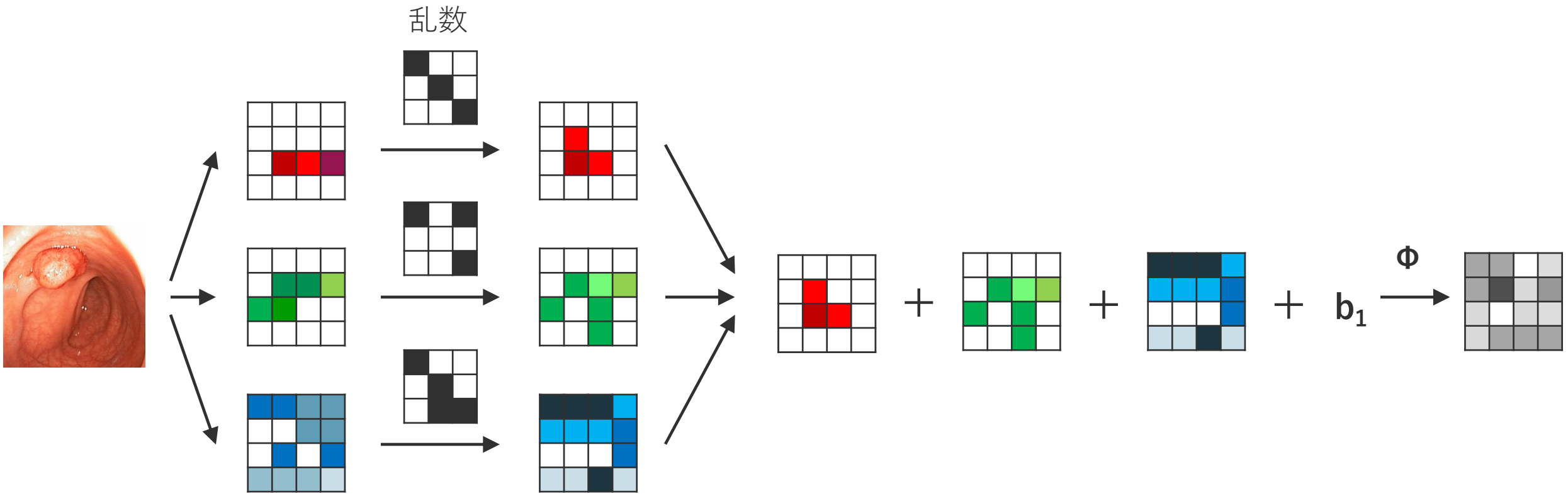
1 channel

Color Photo

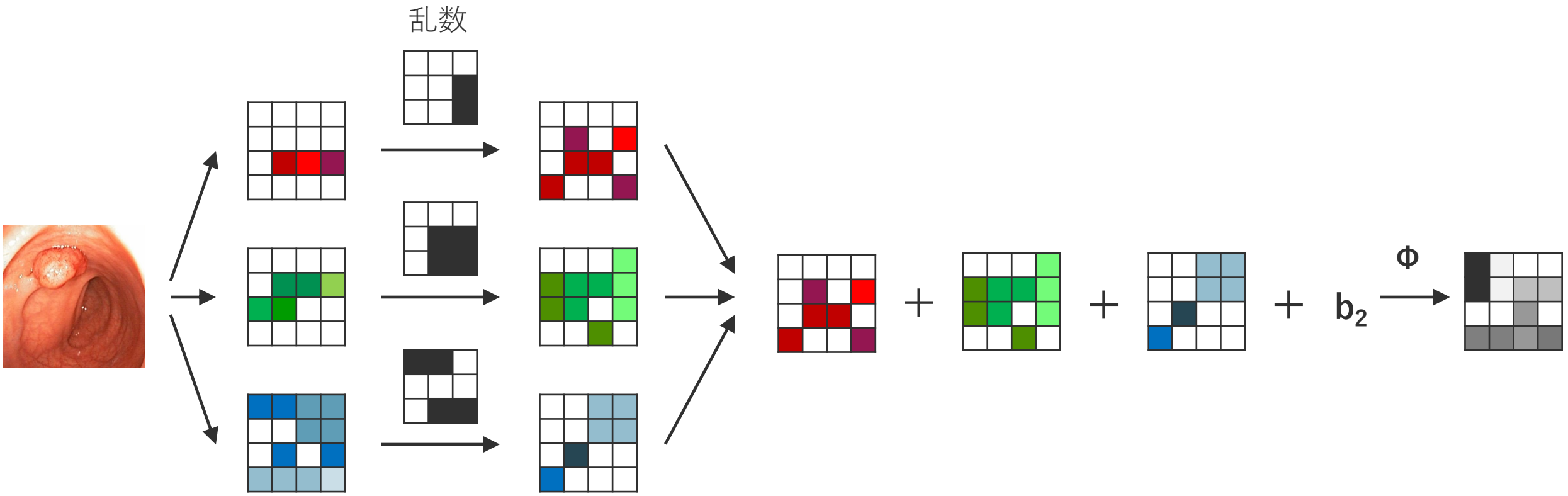


3 channels

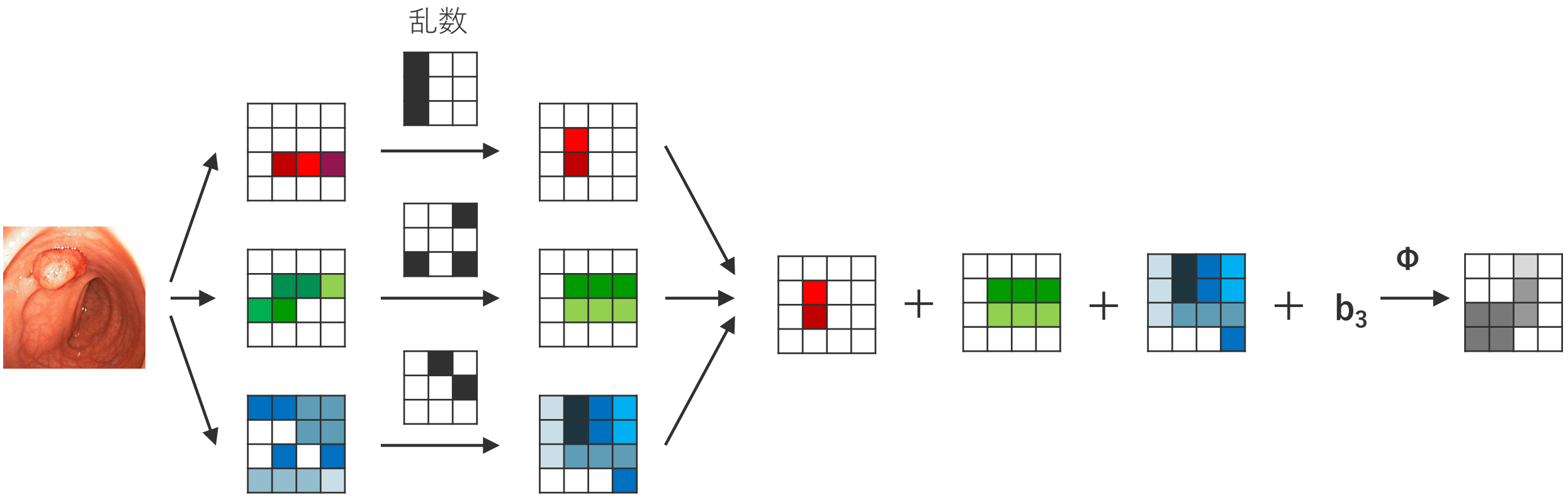
畳み込み演算



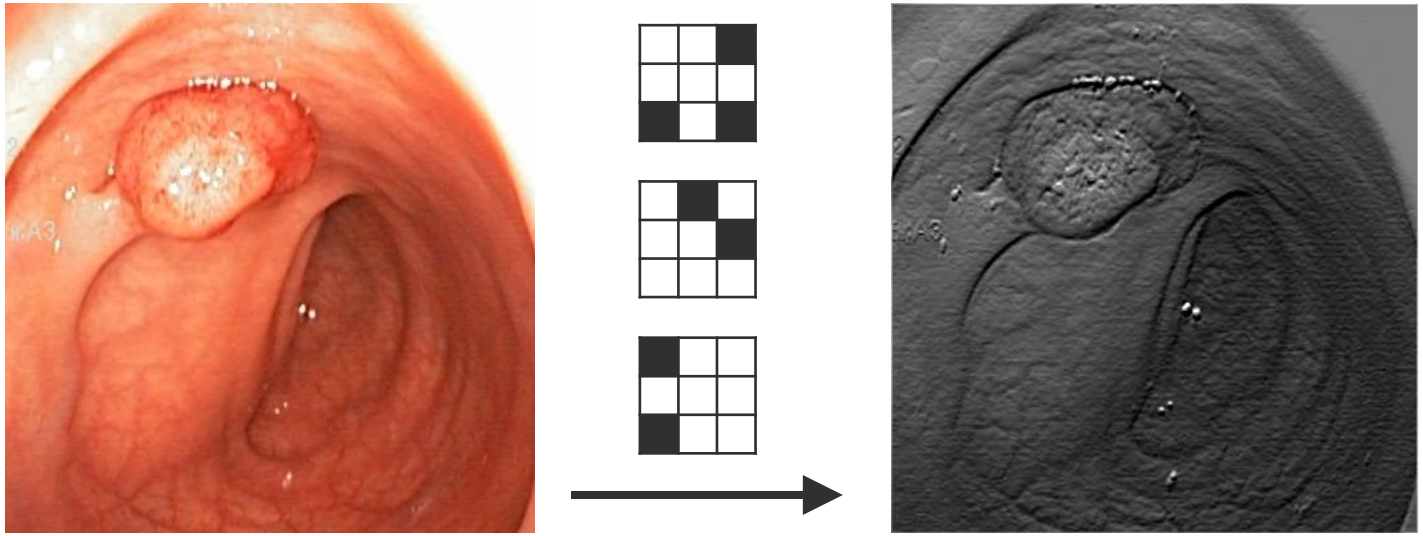
畳み込み演算



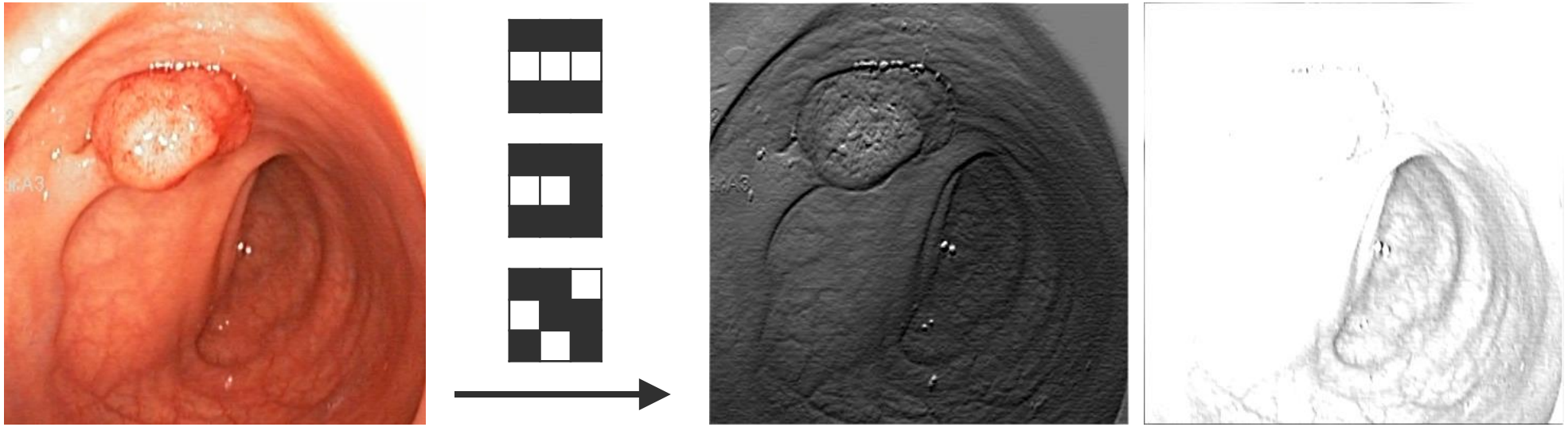
畳み込み演算



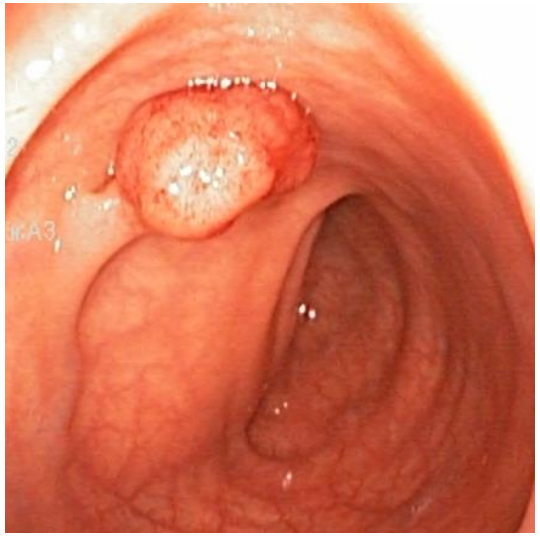
畳み込み演算

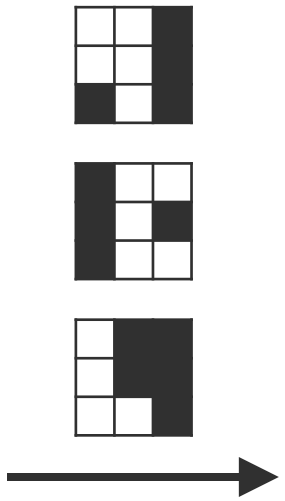


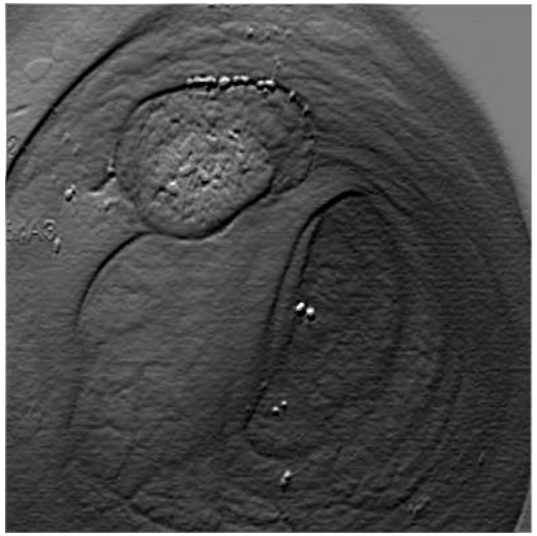
畳み込み演算

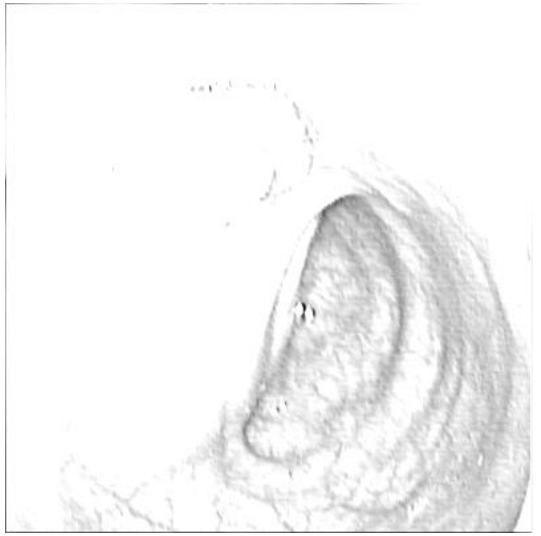


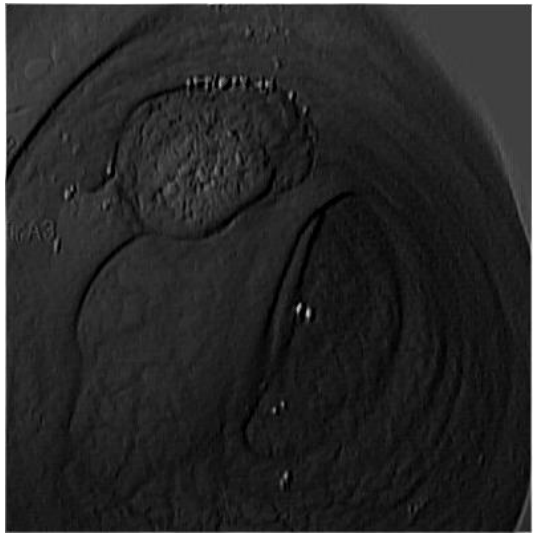
畳み込み演算



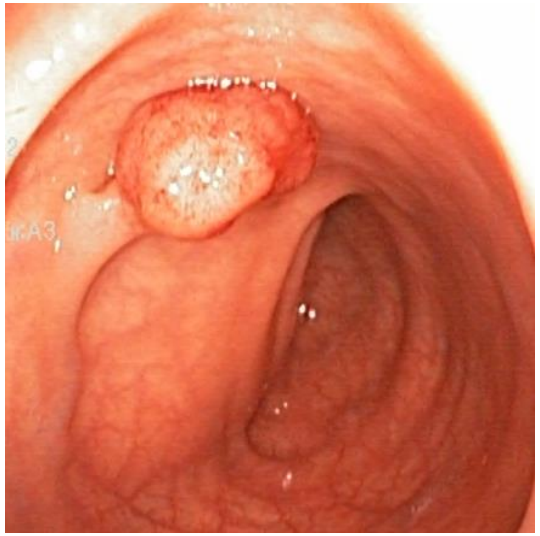


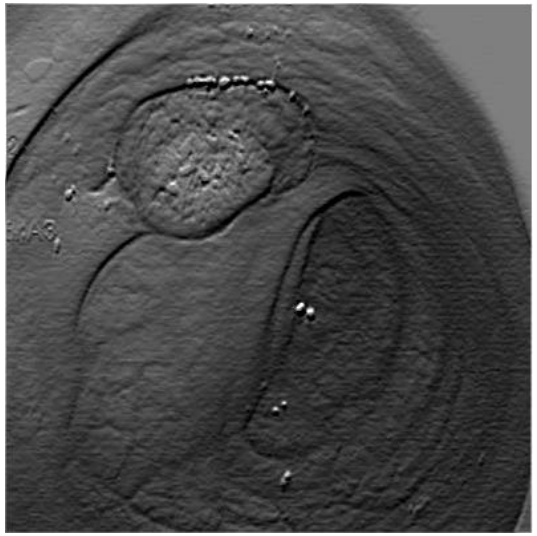


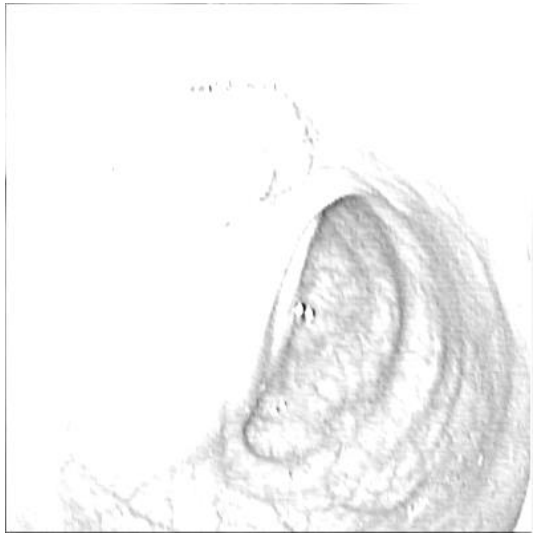


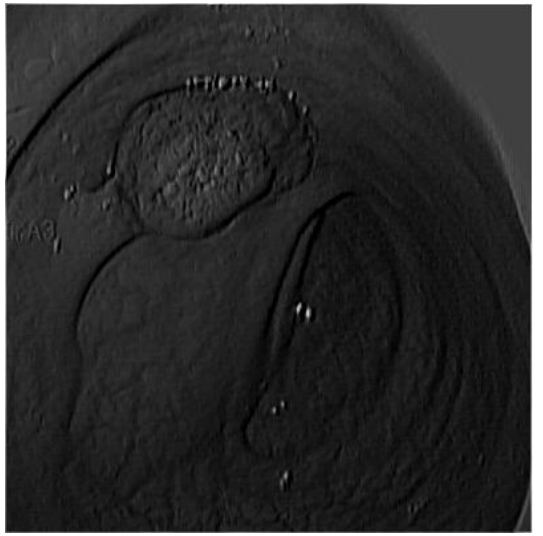


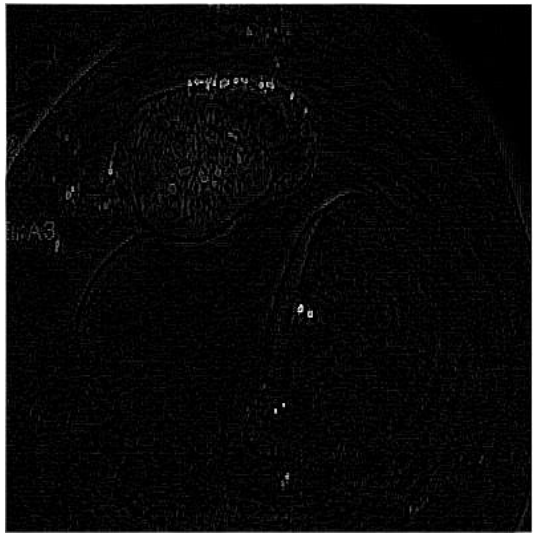
畳み込み演算



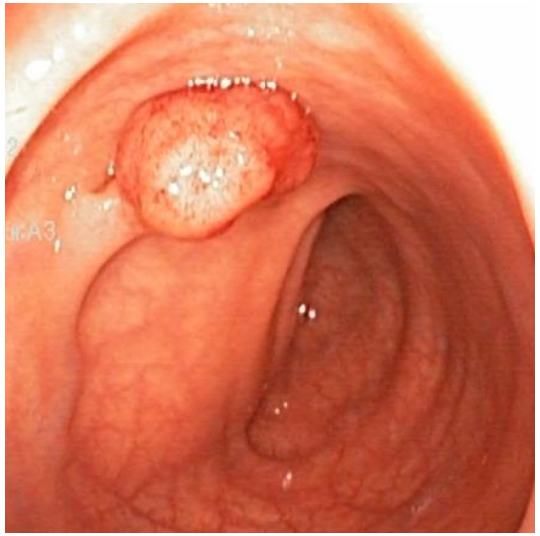


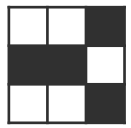


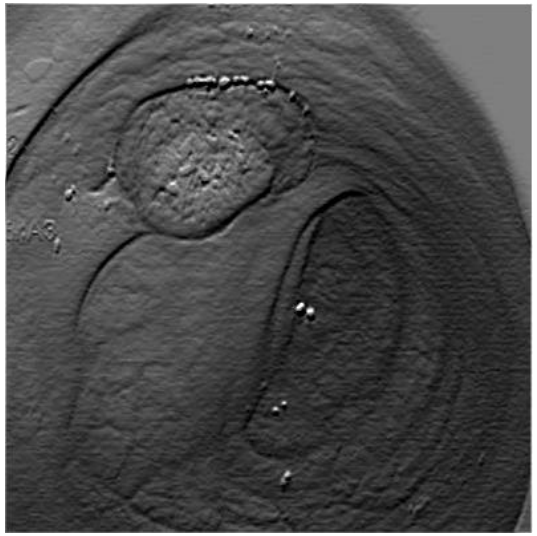


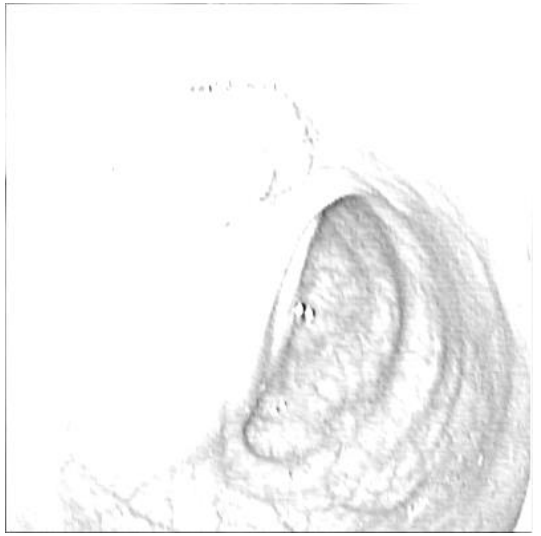


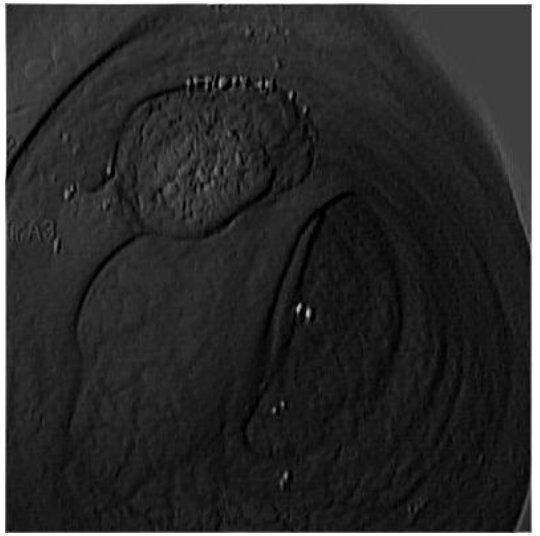
畳み込み演算

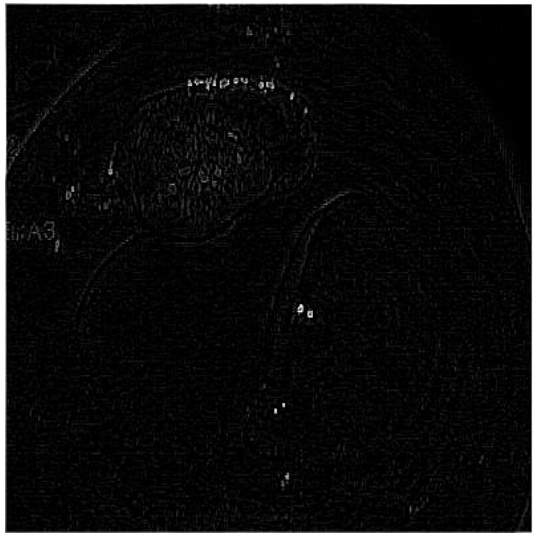


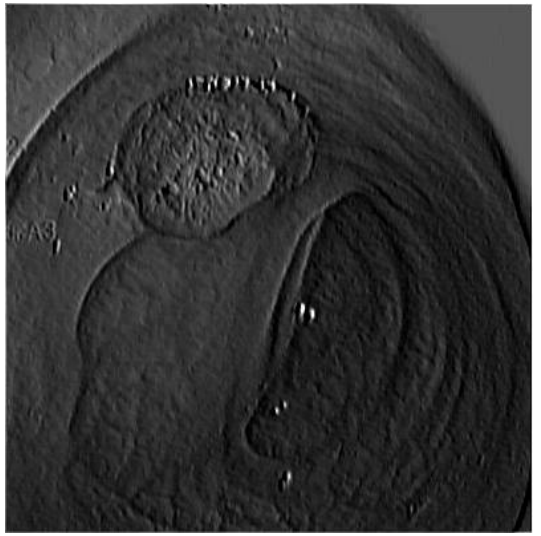




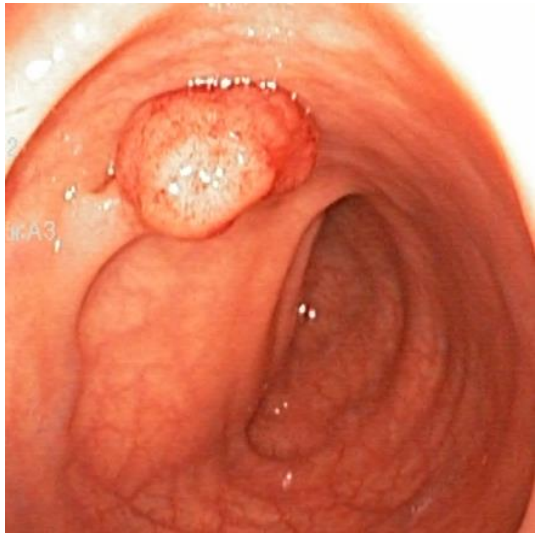


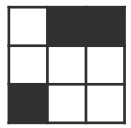


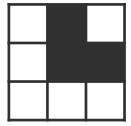


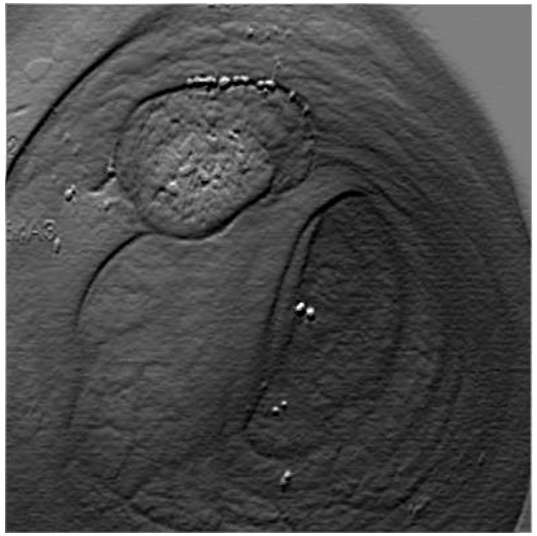


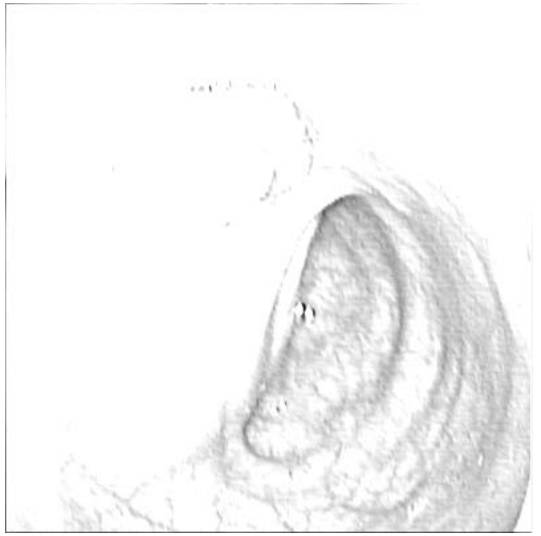
畳み込み演算

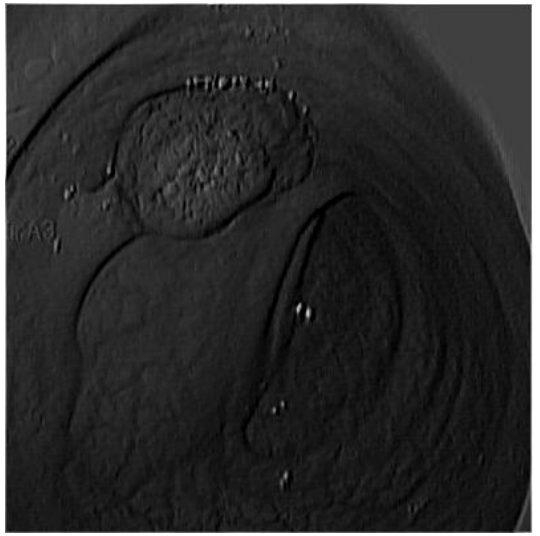


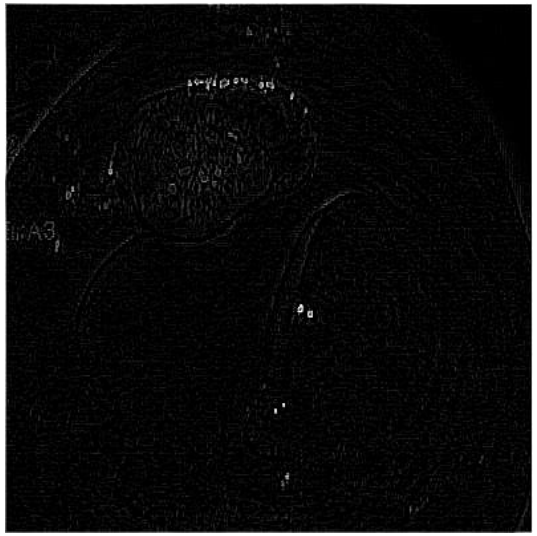


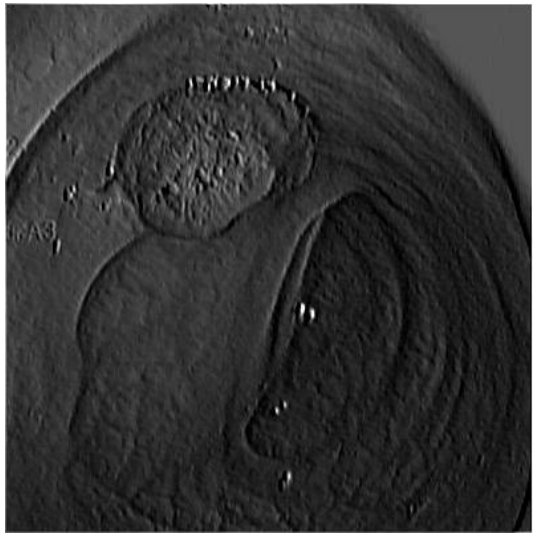


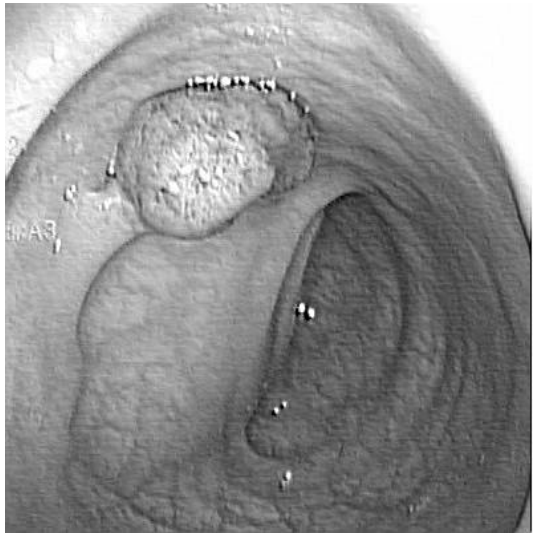




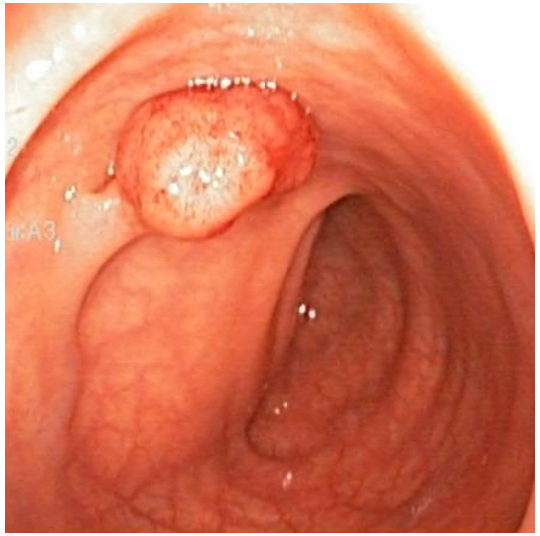








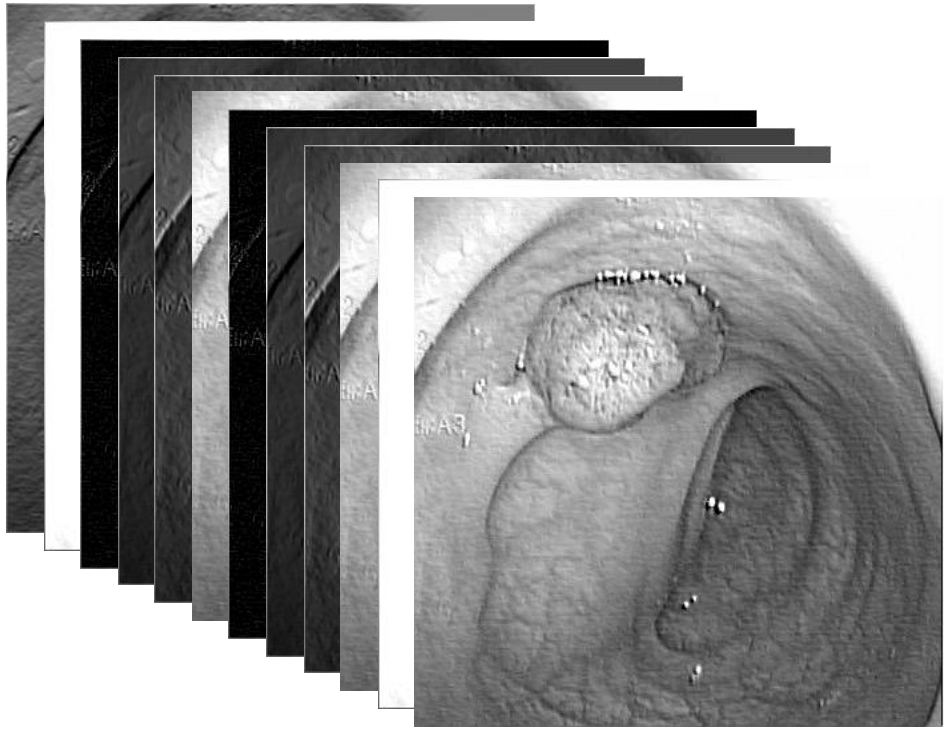
畳み込み演算



入力画像

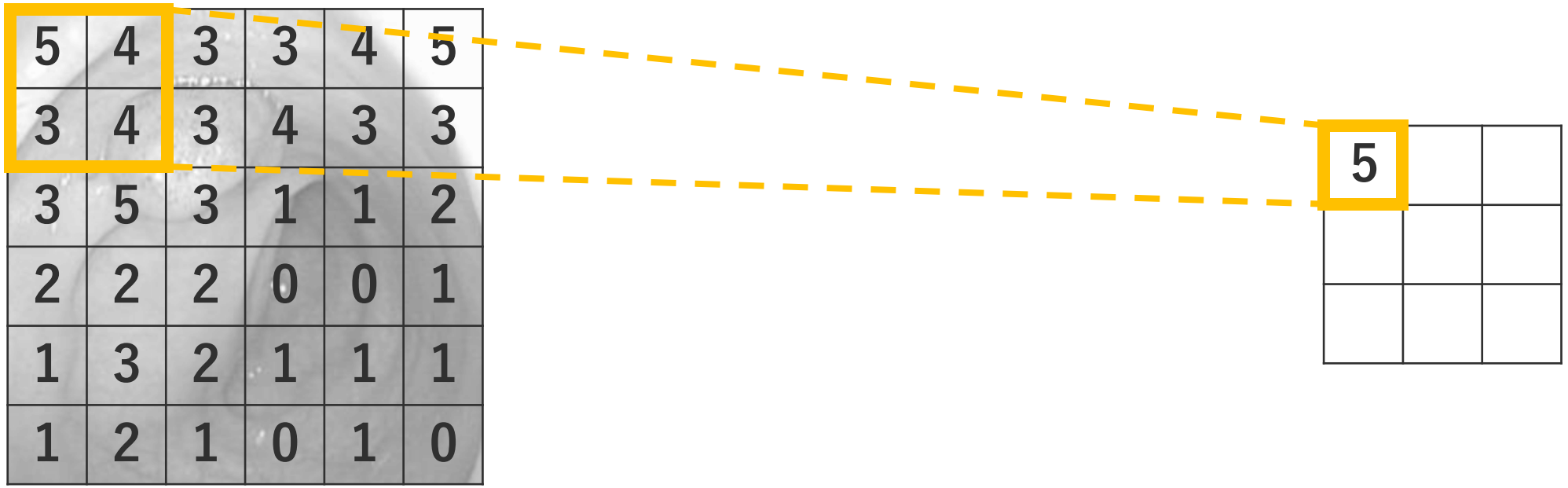
畳み込み演算

- 特徴の抽出
  - 輪郭・模様など
  - 抽象的な概念など
- 空間情報の保存
- パラメータの削減

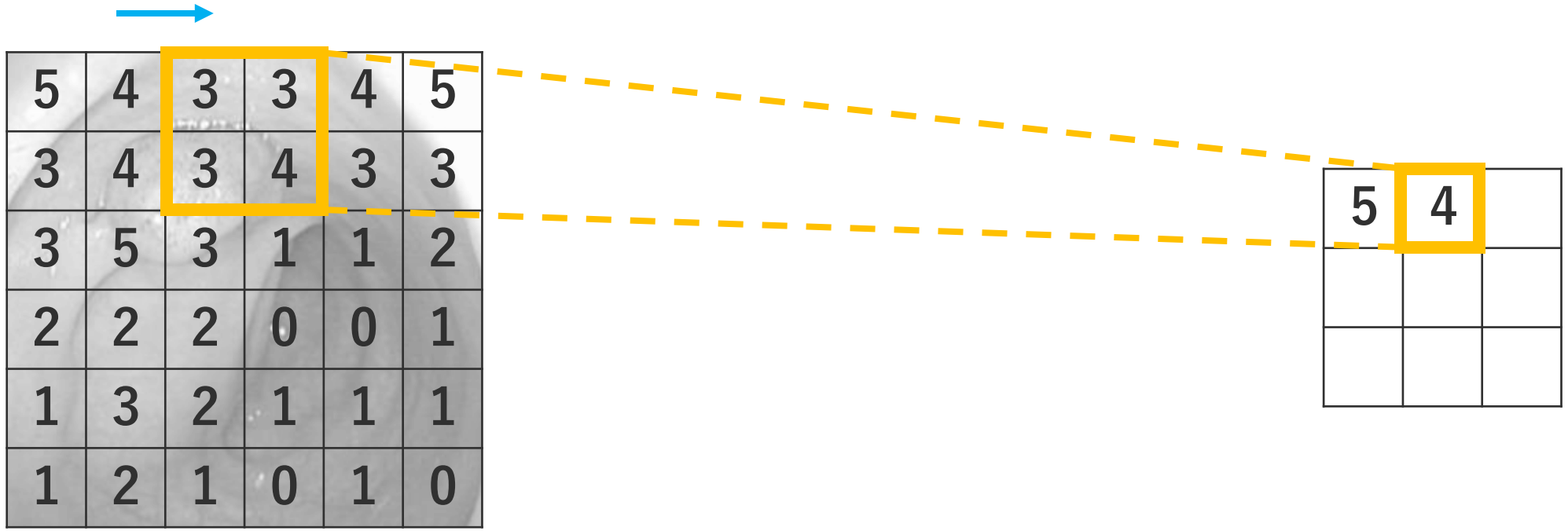


出力画像
(特徴マップ)

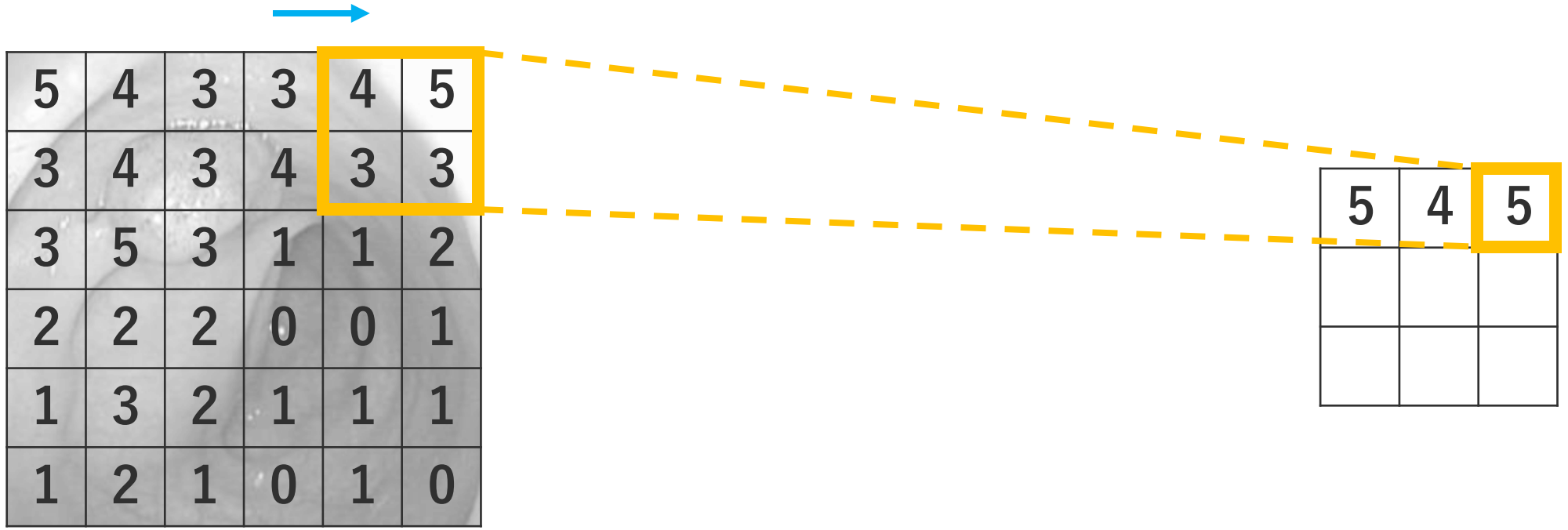
最大プーリング演算



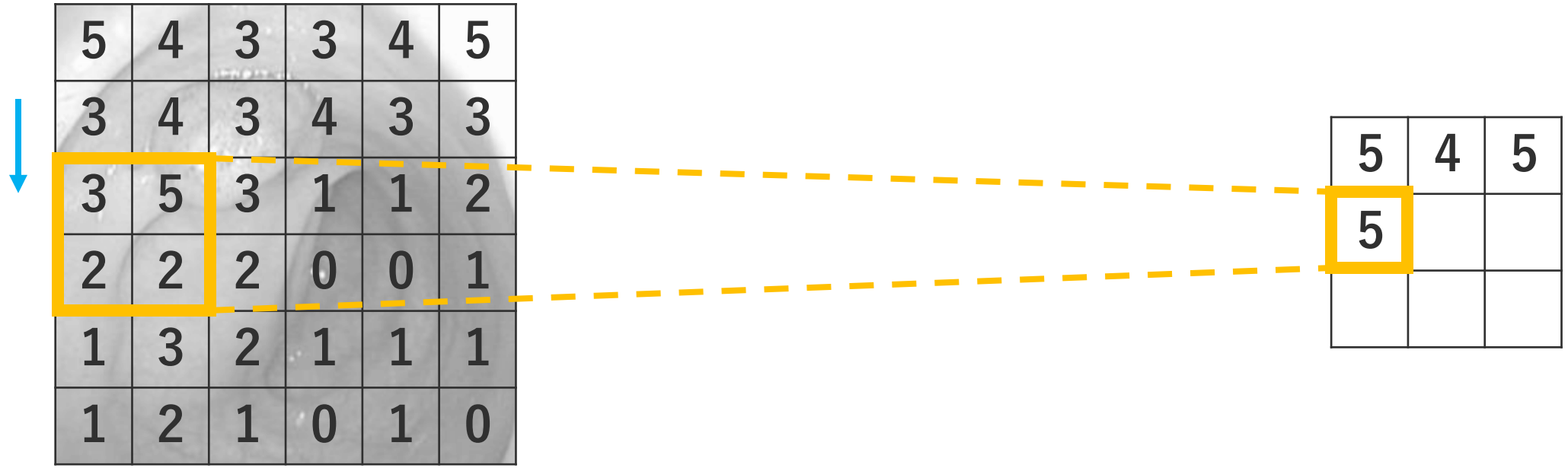
最大プーリング演算



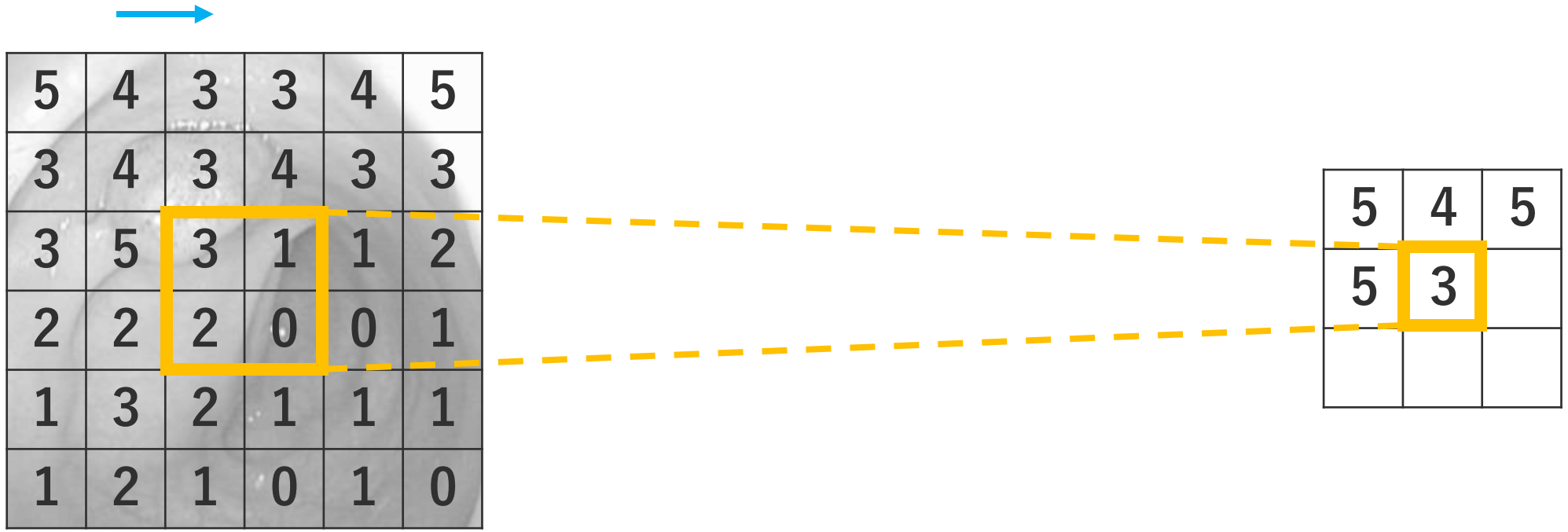
最大プーリング演算



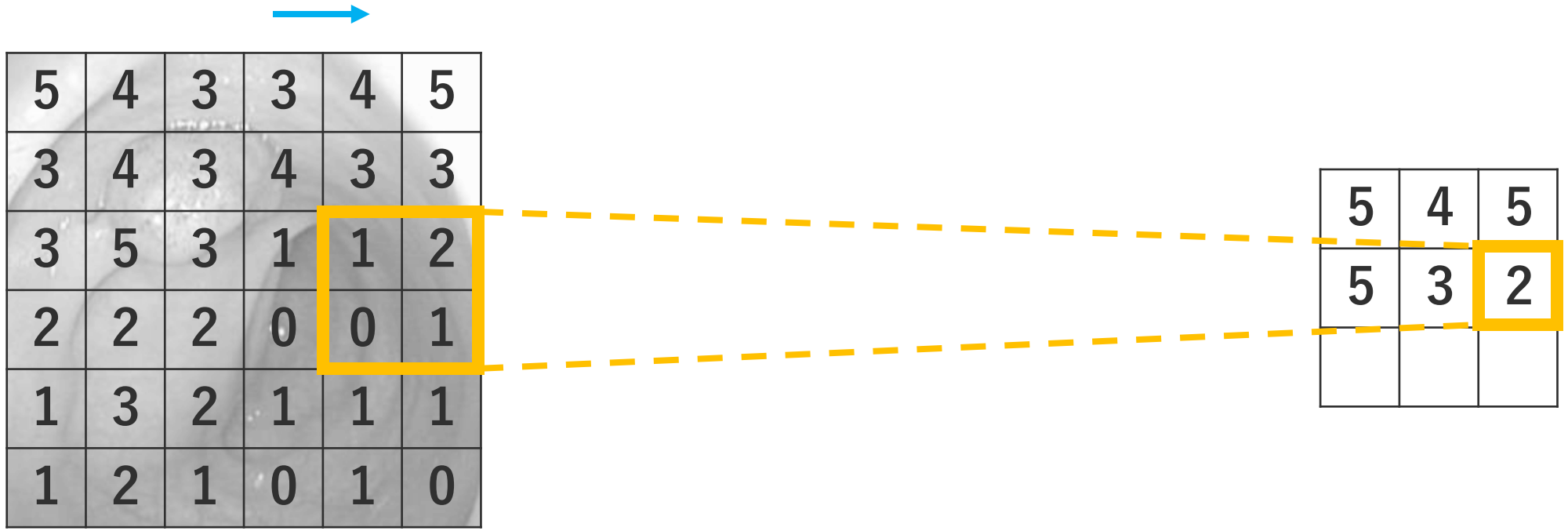
最大プーリング演算



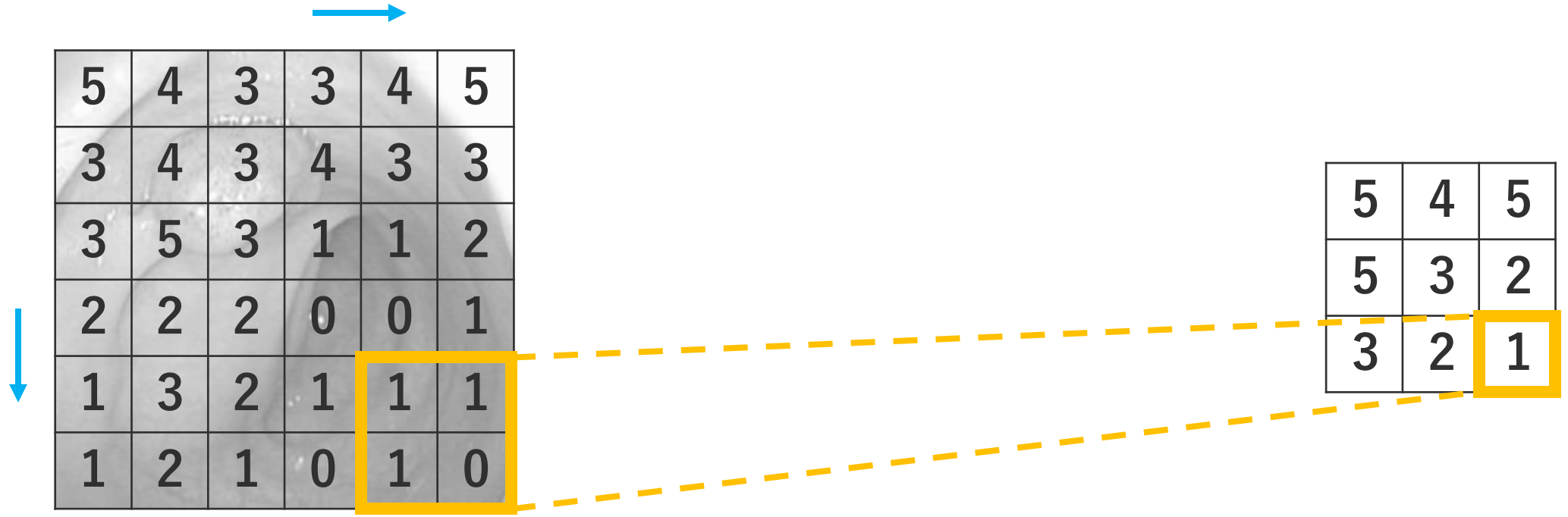
最大プーリング演算



最大プーリング演算



最大プーリング演算

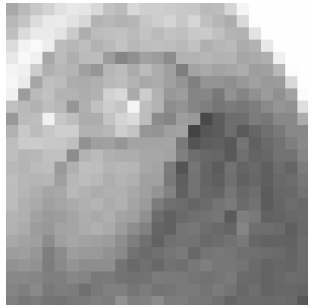


最大プーリング演算

入力画像

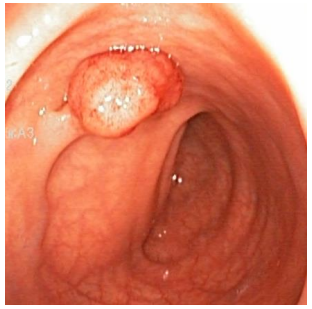
最大プーリング演算

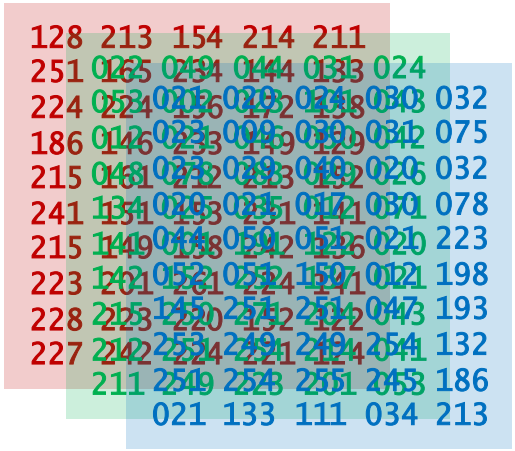
- 特徴マップの圧縮
- パラメータの削減
- 平均プーリング演算、グローバルプーリング演算などもある



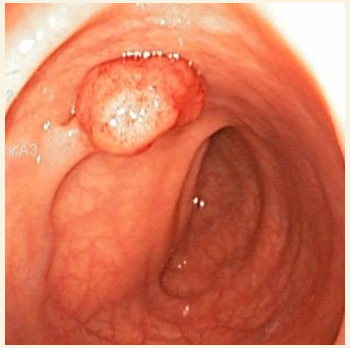
出力画像

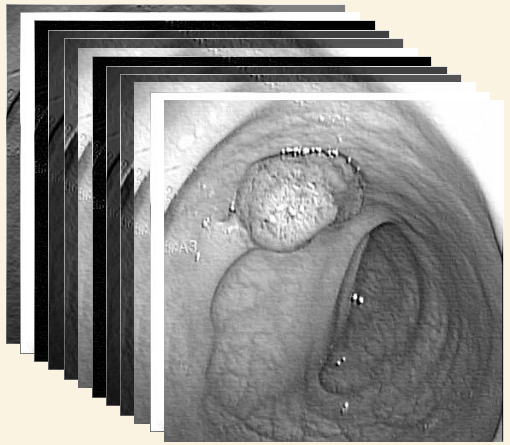
物体分類

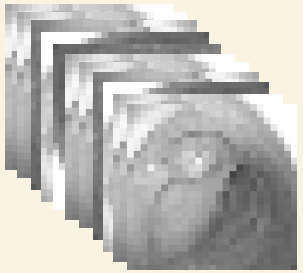




特徴抽出





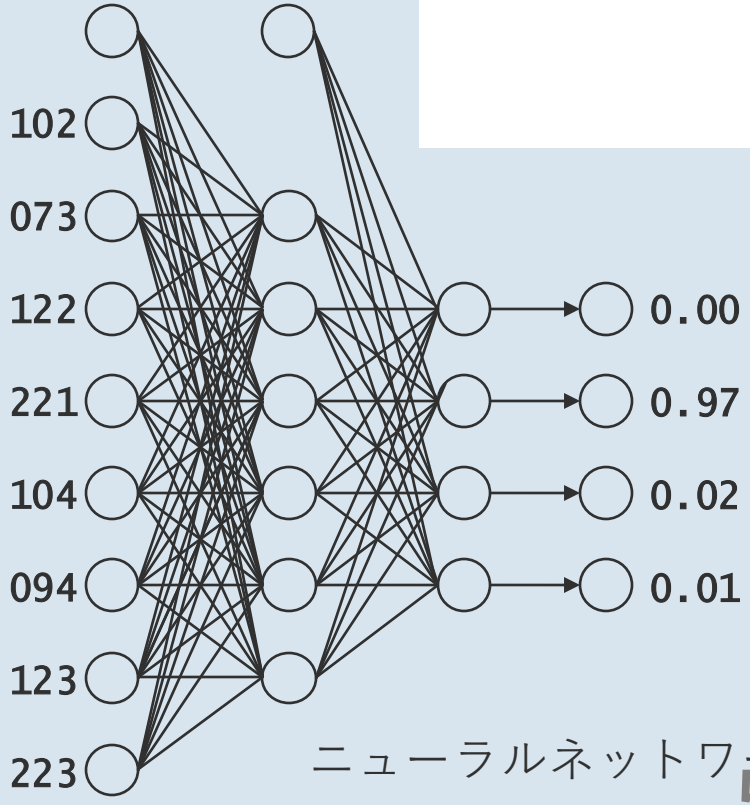


畳み込み

102
073
122
221
104
094
123
223

分類

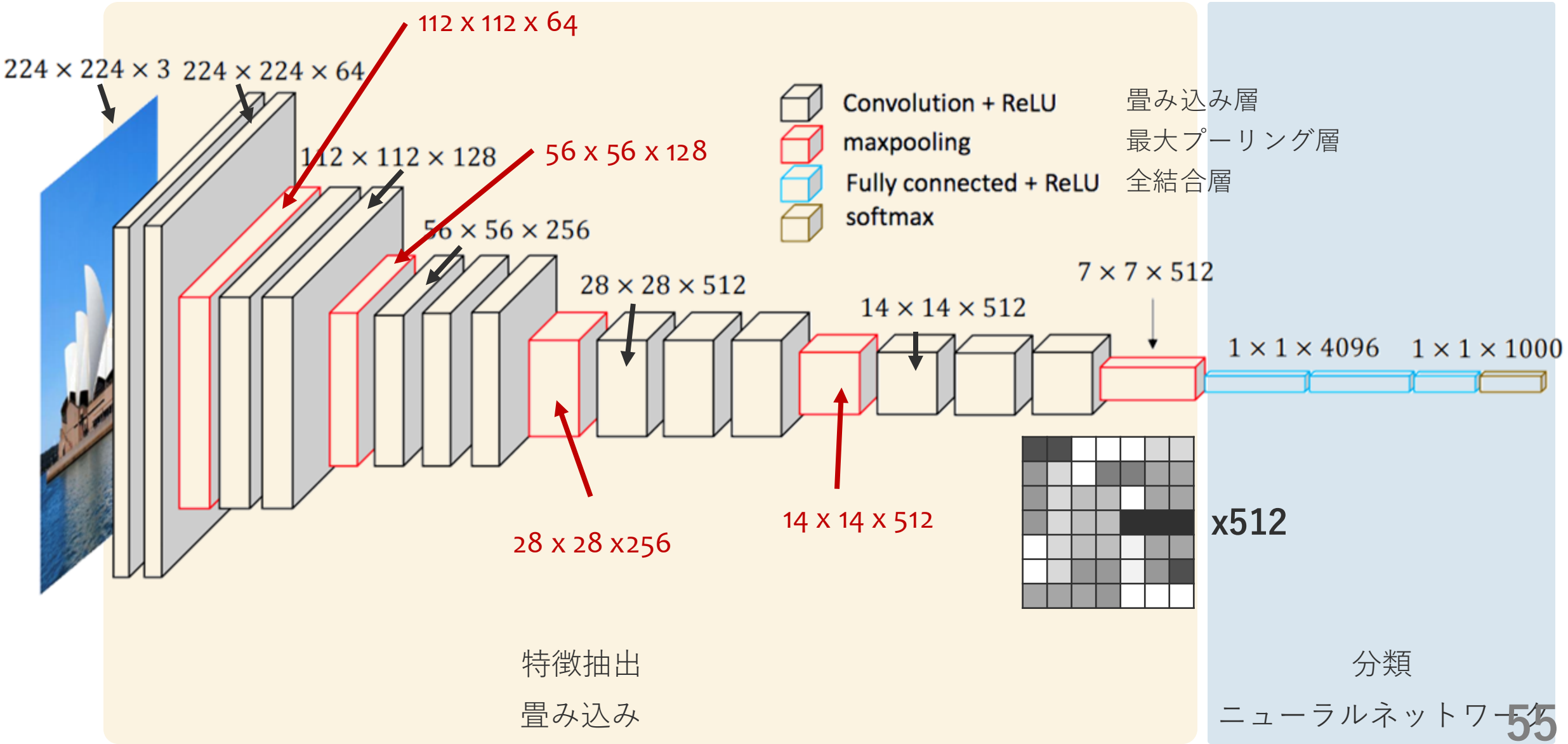
normal	0.00
polyps	0.97
esophagitis	0.02
colitis	0.01



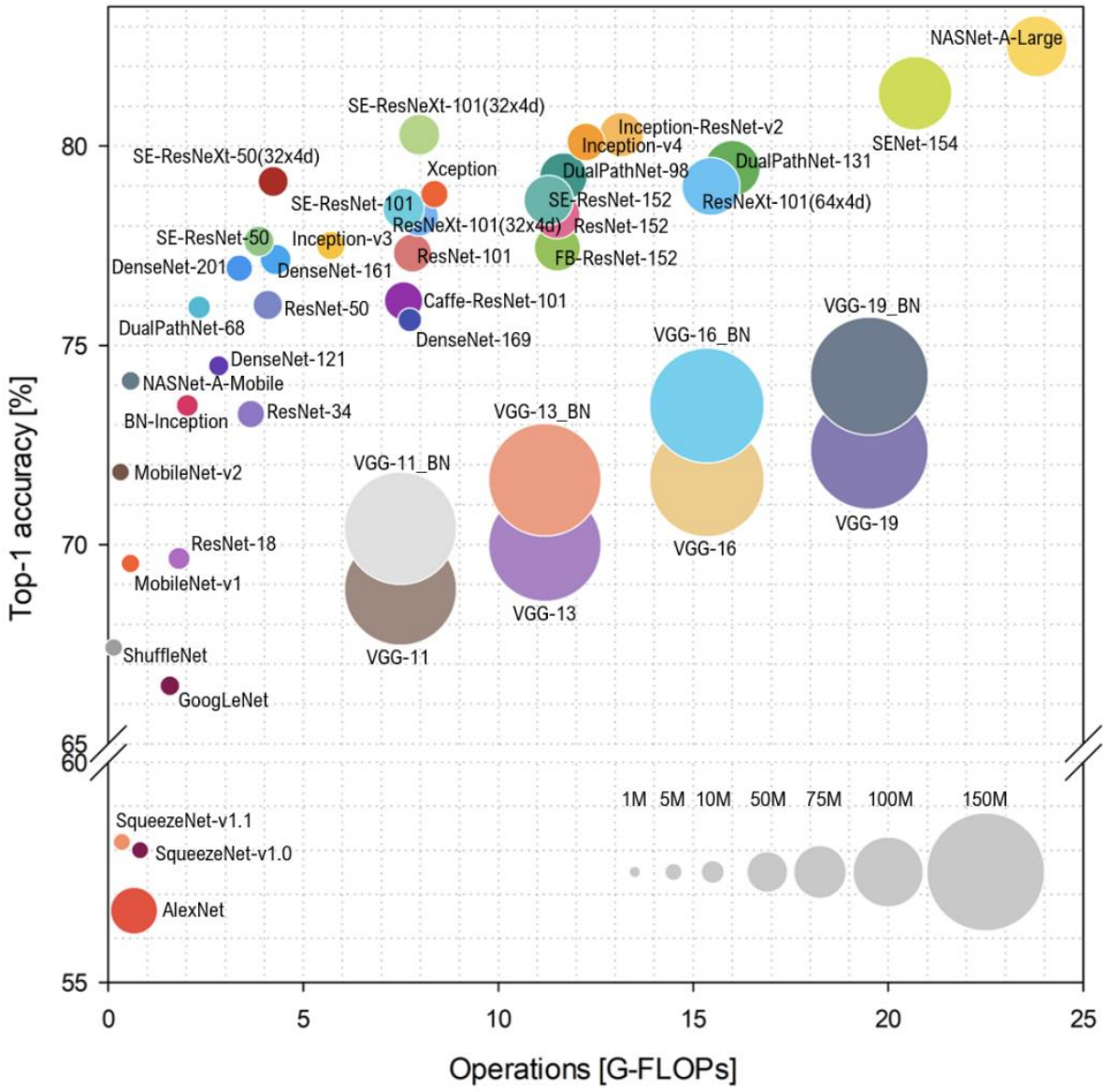
ニューラルネットワーク

VGG16

https://doi.org/10.1145/3038912.3052638

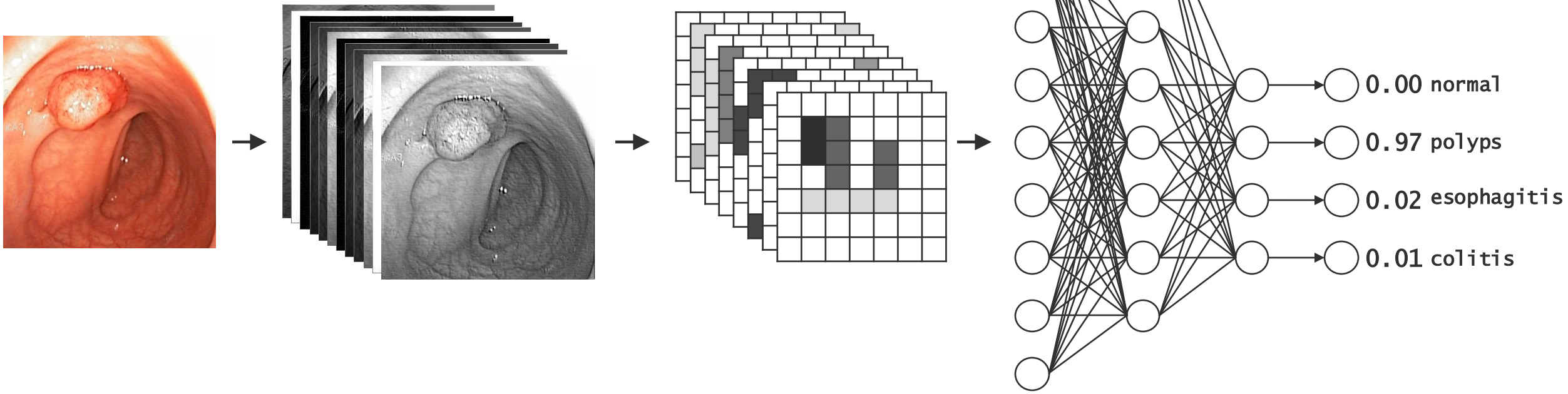


CNNアーキテクチャ

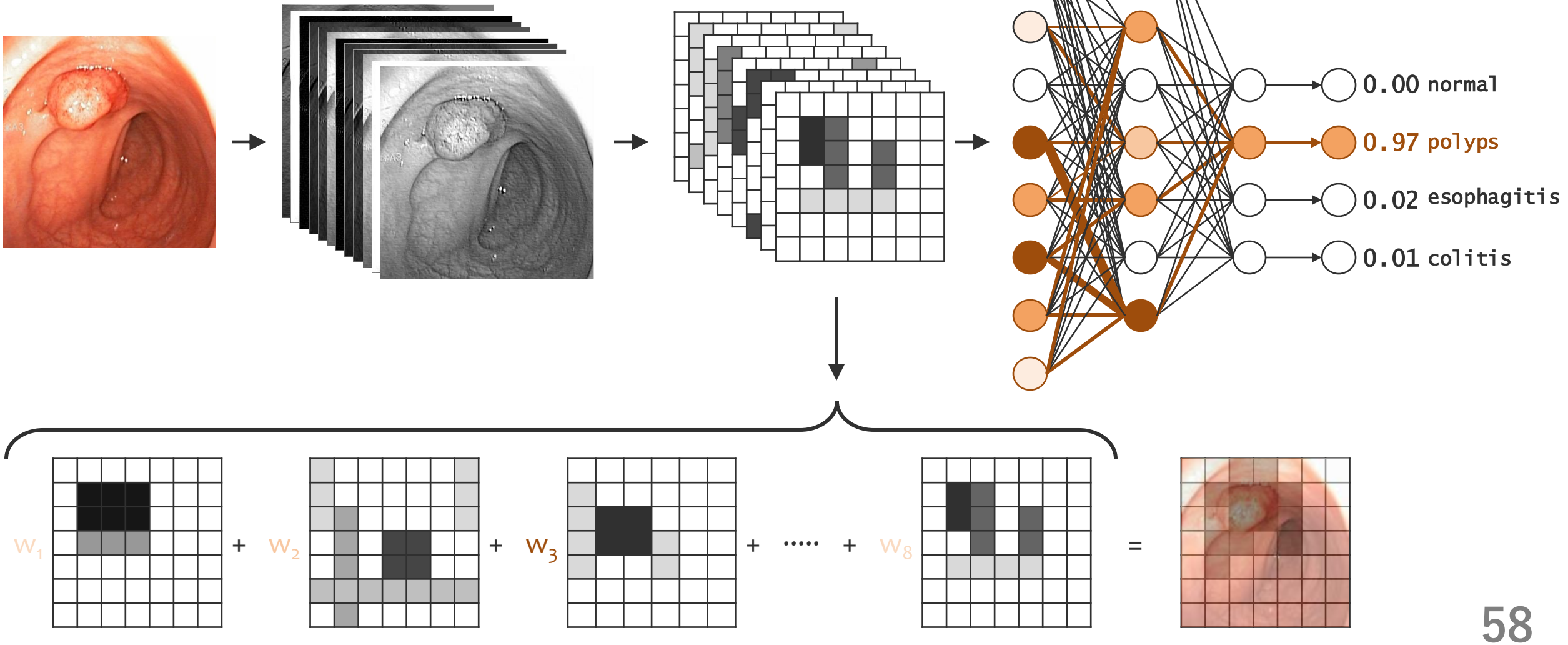


Bianco et al., 2018 https://doi.org/10.48550/arXiv.1810.00736

判定根拠の可視化

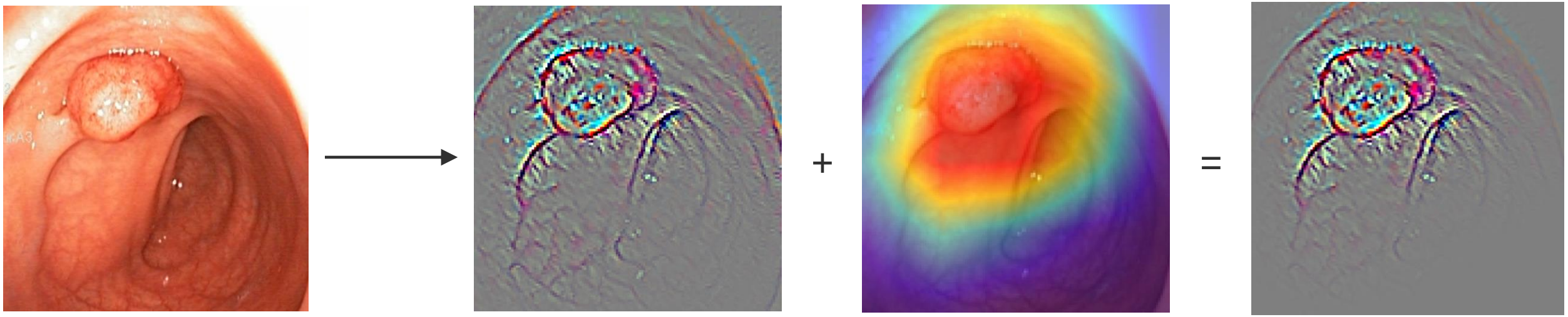


判定根拠の可視化


0.00 normal
0.97 polyps
0.02 esophagitis
0.01 colitis
w1
+
w2
+
w3
+
.....
+
w8
=

判定根拠の可視化

guided backpropagation　　Grad-CAM　　Guided Grad-CAM



深層学習

ニューラルネットワーク

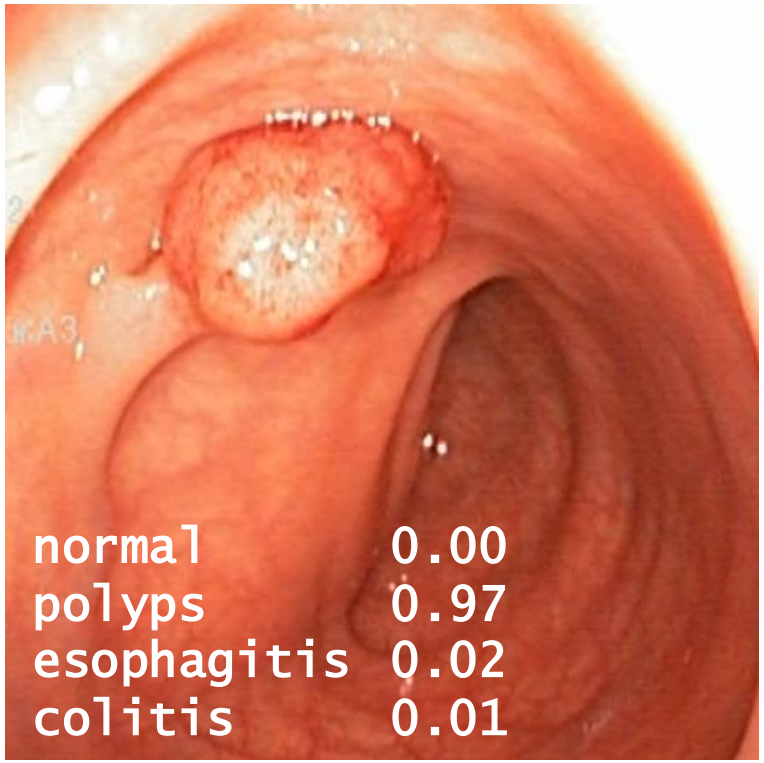
畳み込みニューラルネットワーク

物体分類

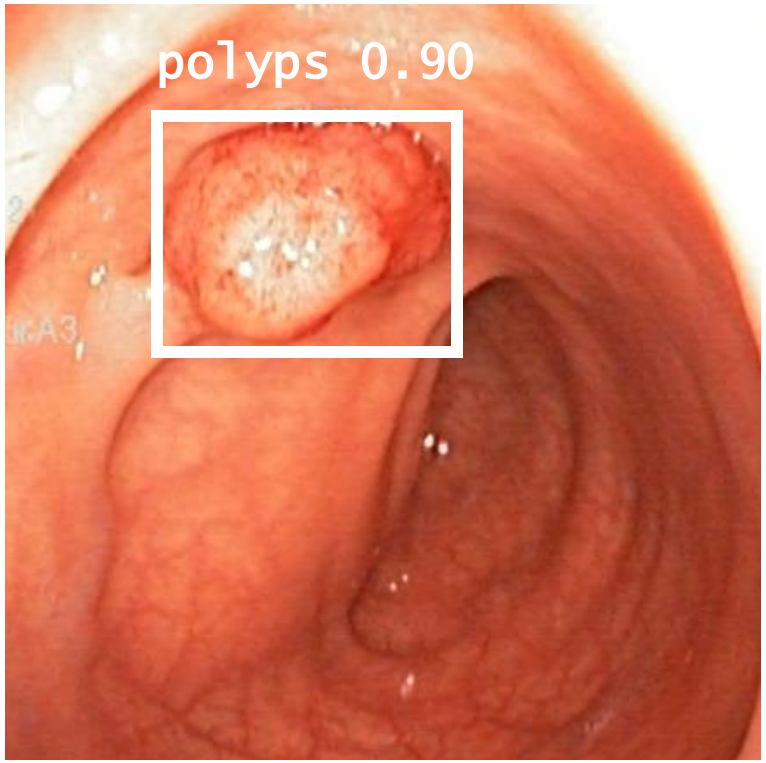
物体検出・セグメンテーション

敵対的生成ネットワーク

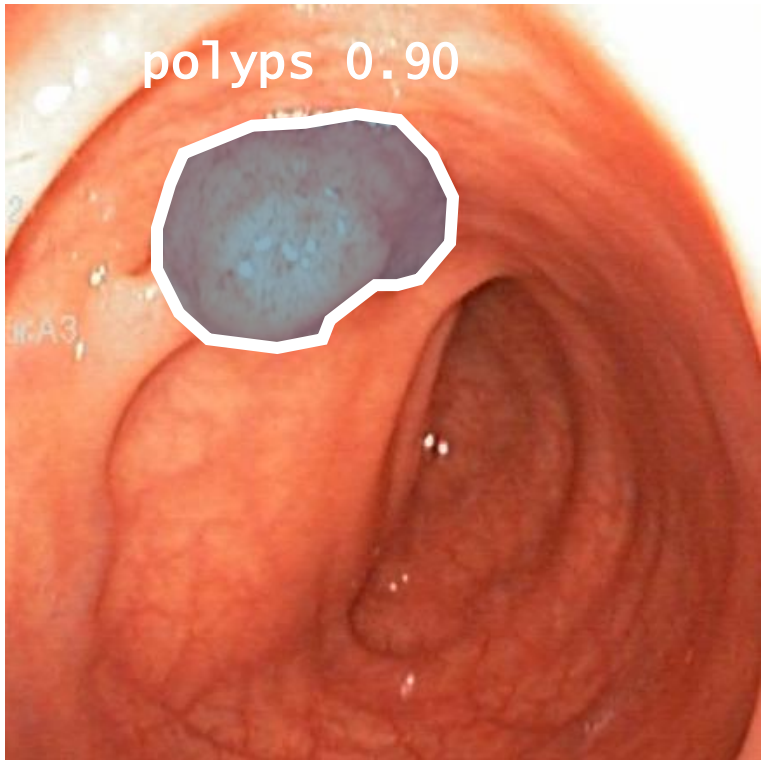
物体分類



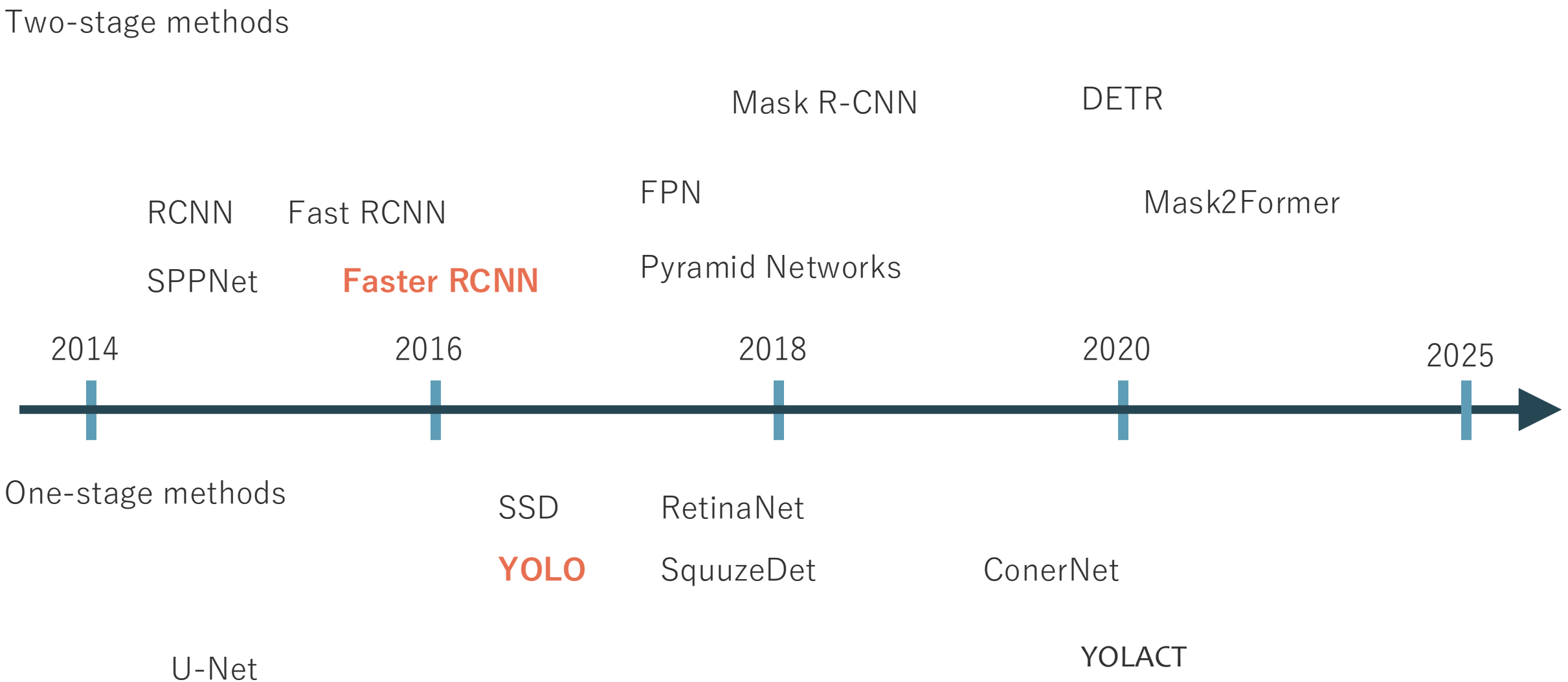
物体検出



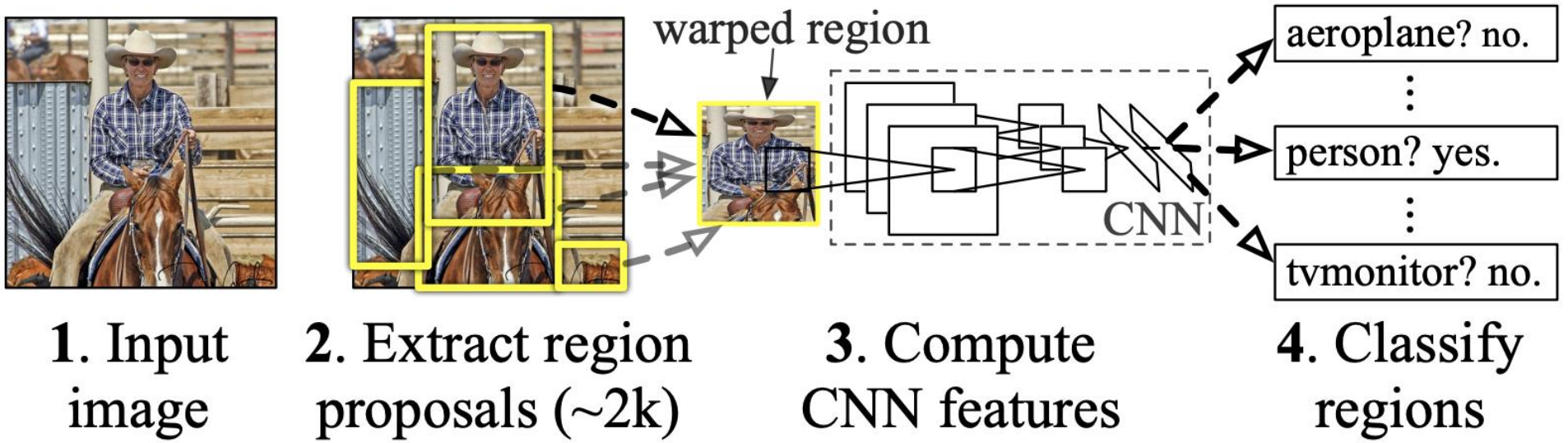
セグメンテーション



物体検出アルゴリズム


Two-stage methods
Mask R-CNN
DETR
FPN
RCNN
Fast RCNN
Mask2Former
Pyramid Networks
SPPNet
Faster RCNN
2014
2016
2018
2020
2025
One-stage methods
SSD
RetinaNet
YOLO
SquuzeDet
ConerNet
U-Net
YOLACT

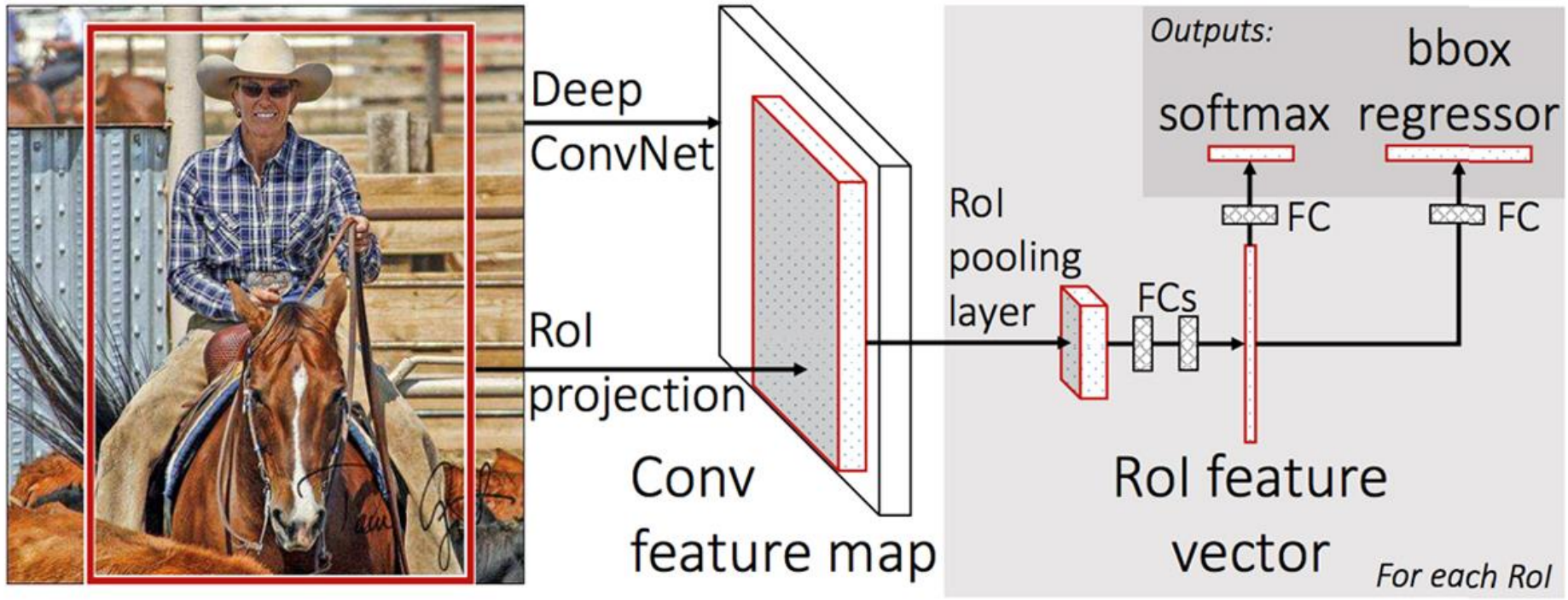
R-CNN



- Select search アルゴリズムを利用し、入力画像にあるオブジェクトが存在していそうな領域候補（Region of Interest; RoI）を決定する。
- 各候補領域を CNN に入力し特徴抽出を行う。
- CNN から出力される特徴を SVM に入力し分類を行う。

https://arxiv.org/abs/1311.2524

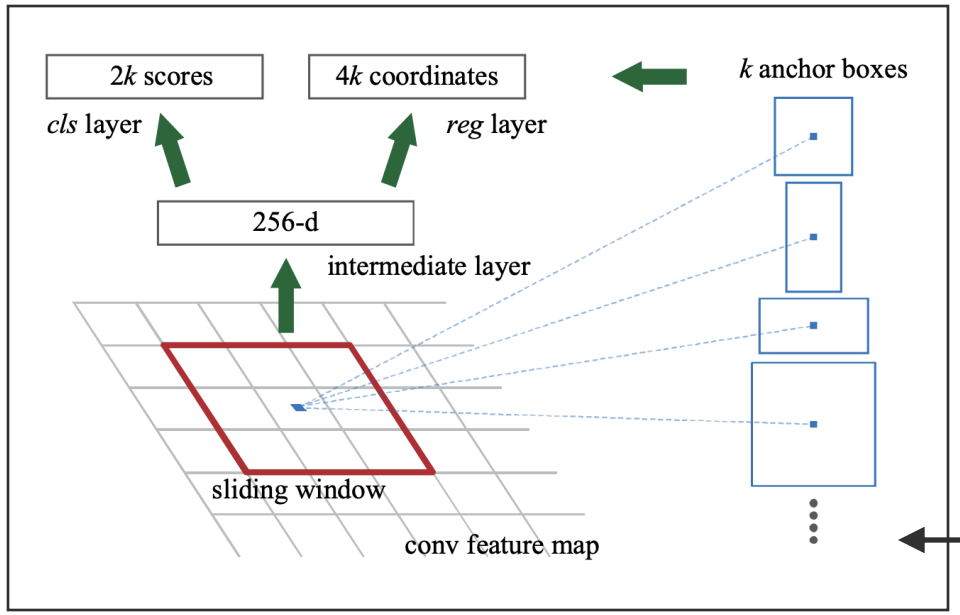
Fast R-CNN

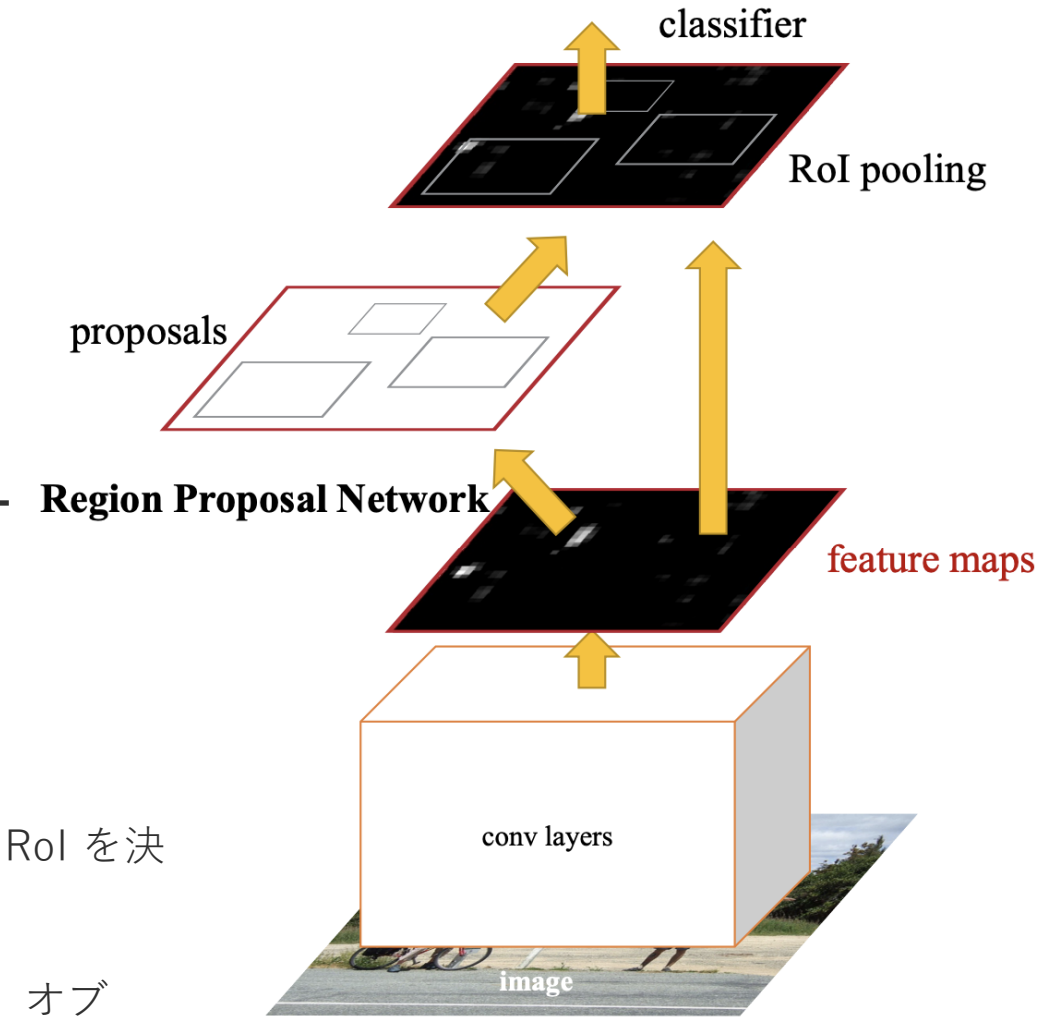


- 入力画像を CNN に入力し特徴マップを計算する。
- 入力画像をもとに決定した RoI を特徴マップに射影する。
- 特徴マップ上の RoI を処理し、物体分類とバウンディングボックスの座標を予測する。

https://arxiv.org/abs/1504.08083

Faster R-CNN

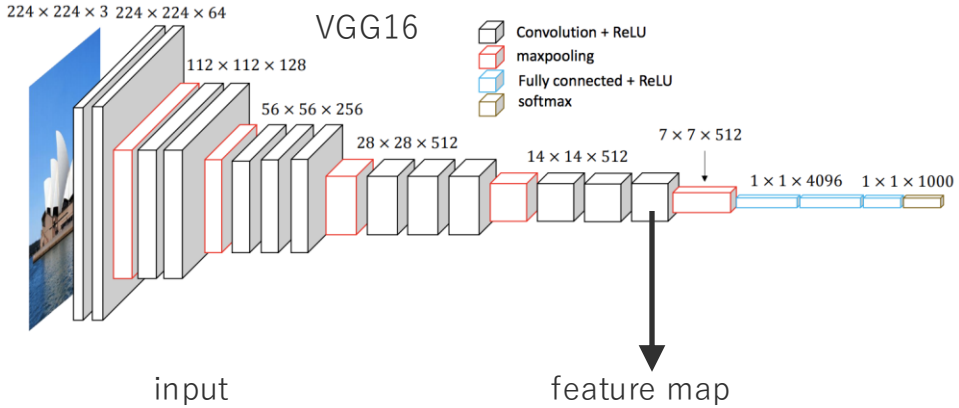


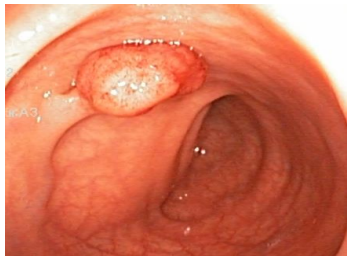


- 入力画像を CNN に入力し特徴マップを計算する。
- 特徴マップを Region Proposal Network に入力し、RoI を決定する。
- 特徴マップと RPN から出力される RoI を利用して、オブジェクトの分類を行う。

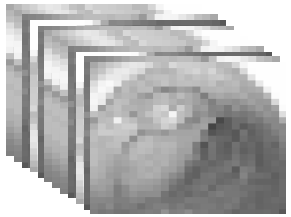
https://arxiv.org/abs/1506.01497

Resion Proposal Network

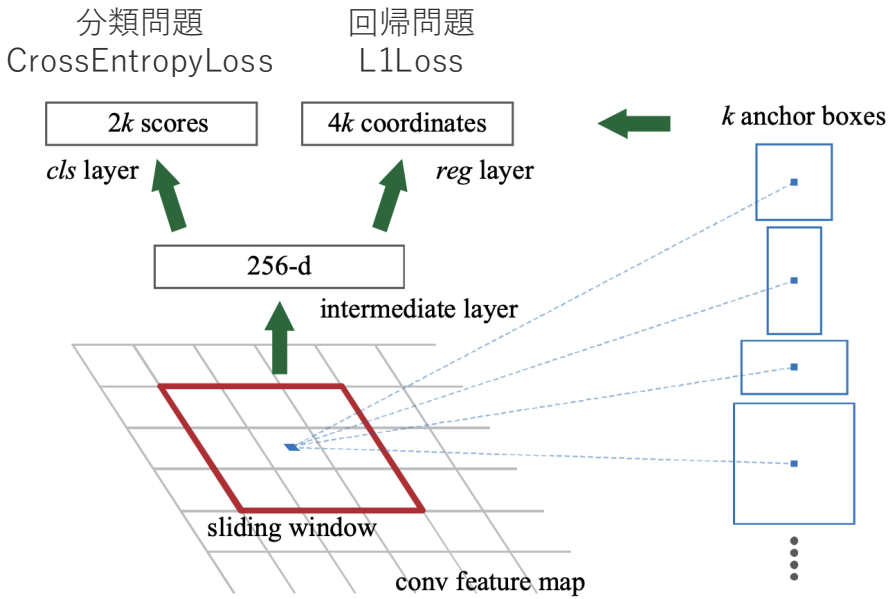


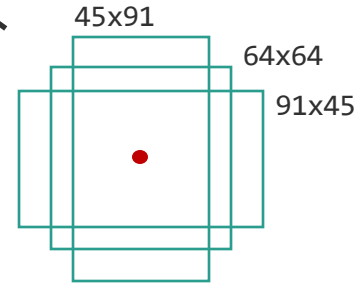


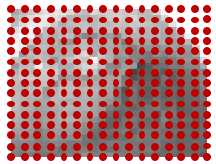
1024x768x3



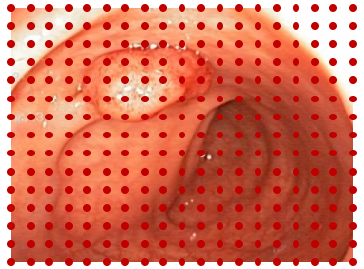
64x48x512



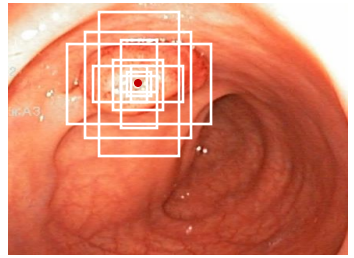




64x48 anchors



64x48 anchors

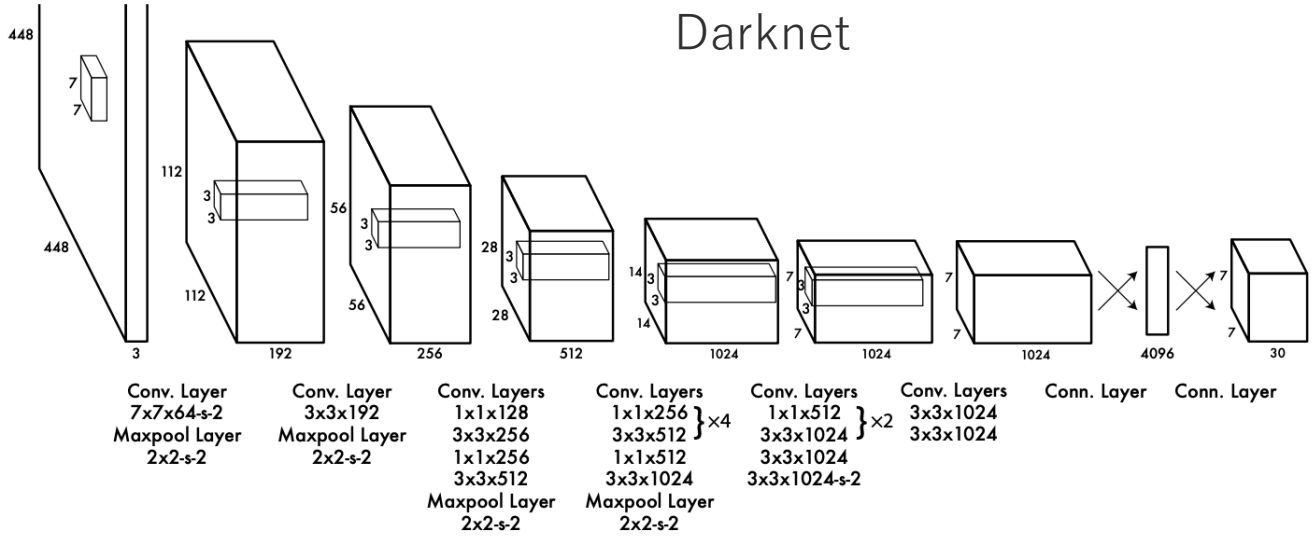


9 anchor boxes / anchor

parameters
base width: 64, 128, 256
ratio: 1:1, 1:2, 2:1

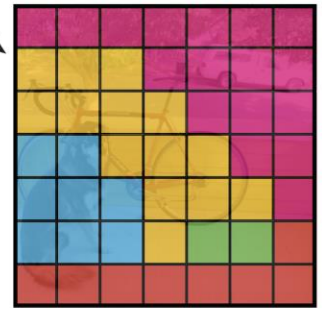
https://arxiv.org/abs/1506.01497
https://doi.org/10.1145/3038912.3052638

You Only Look Once (YOLO)

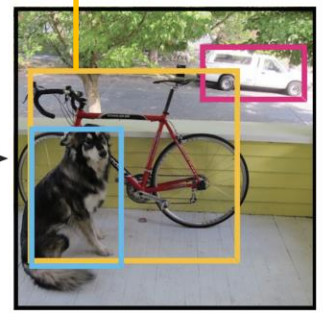



S × S grid on input


Bounding boxes + confidence


Class probability map

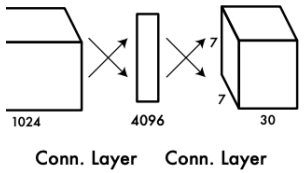
$$\begin{pmatrix} x \\ y \\ w \\ h \\ conf \\ p(c_1) \\ p(c_2) \\ \vdots \\ p(c_m) \end{pmatrix}$$

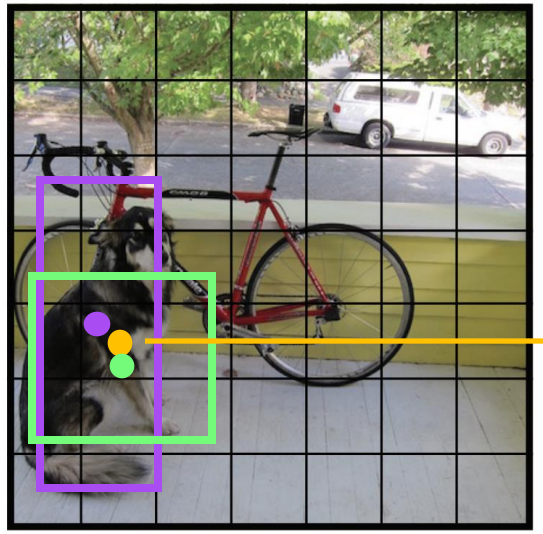

Final detections

https://arxiv.org/abs/1506.02640

YOLO – モデル訓練

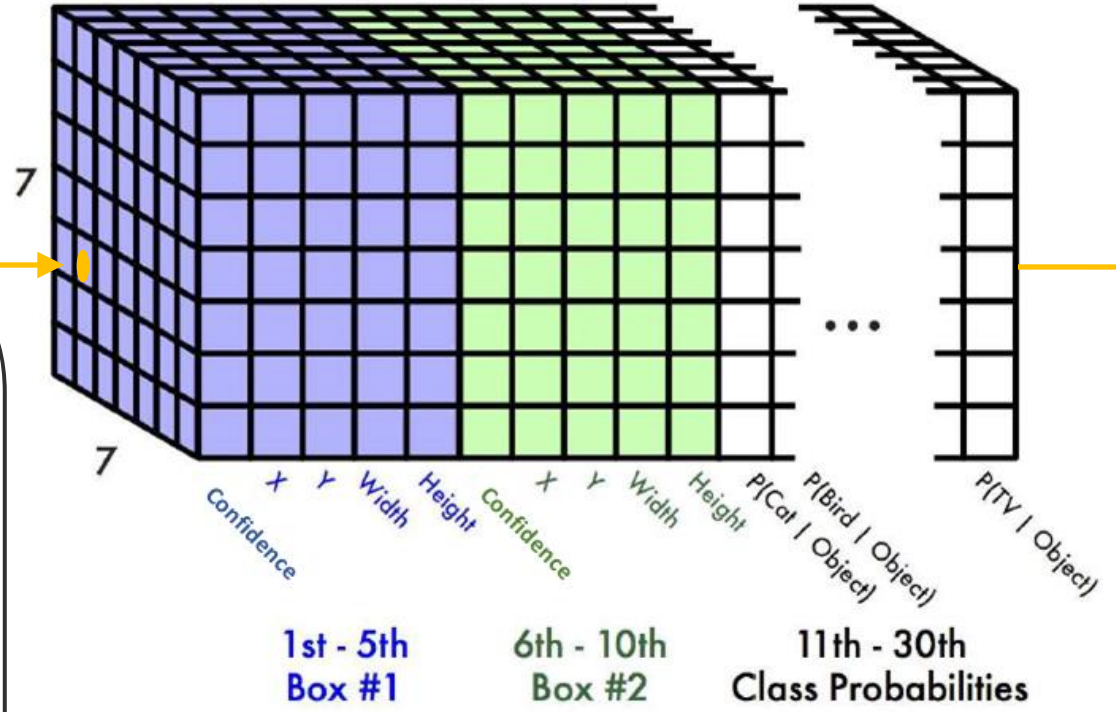
$$conf = \underset{\hat{b}\in\{b_1,b_2\}}{\text{argmax}}\{p(object)\cdot \text{IoU}_{\hat{b}}^{truth}\}$$





S × S grid on input

$$\begin{pmatrix} x \\ y \\ w \\ h \\ conf \\ x \\ y \\ w \\ h \\ conf \\ p(c_1) \\ p(c_2) \\ \vdots \\ p(c_m) \end{pmatrix}$$



$$\begin{pmatrix} 0.286 \\ 0.714 \\ 0.332 \\ 0.584 \\ 0.845 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad p(\text{dog})$$

教師ラベル

https://arxiv.org/abs/1506.02640

YOLO – 損失関数

訓練過程で、オブジェクトのあるセルではバウンディングボックスの中心位置および幅と高さが学習され、信頼度も増加する。背景のセルでは学習されずに出鱈目なバウンディングボックスを低信頼度で生成される。

1: オブジェクト
0: 背景

bbox 中心座標の予測誤差

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right]$$

bbox の幅と高さの予測誤差

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

予測誤差が非常に小さいので、定数倍 $\lambda_{\text{coord}} = 5$ をかけて大きくする。

confidence の予測誤差

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2$$

自信を持って生成した bbox が偶然に教師ラベルと重なった場合、損失がゼロになる。

confidence の予測誤差（背景セルならば $C_i = 0$）

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

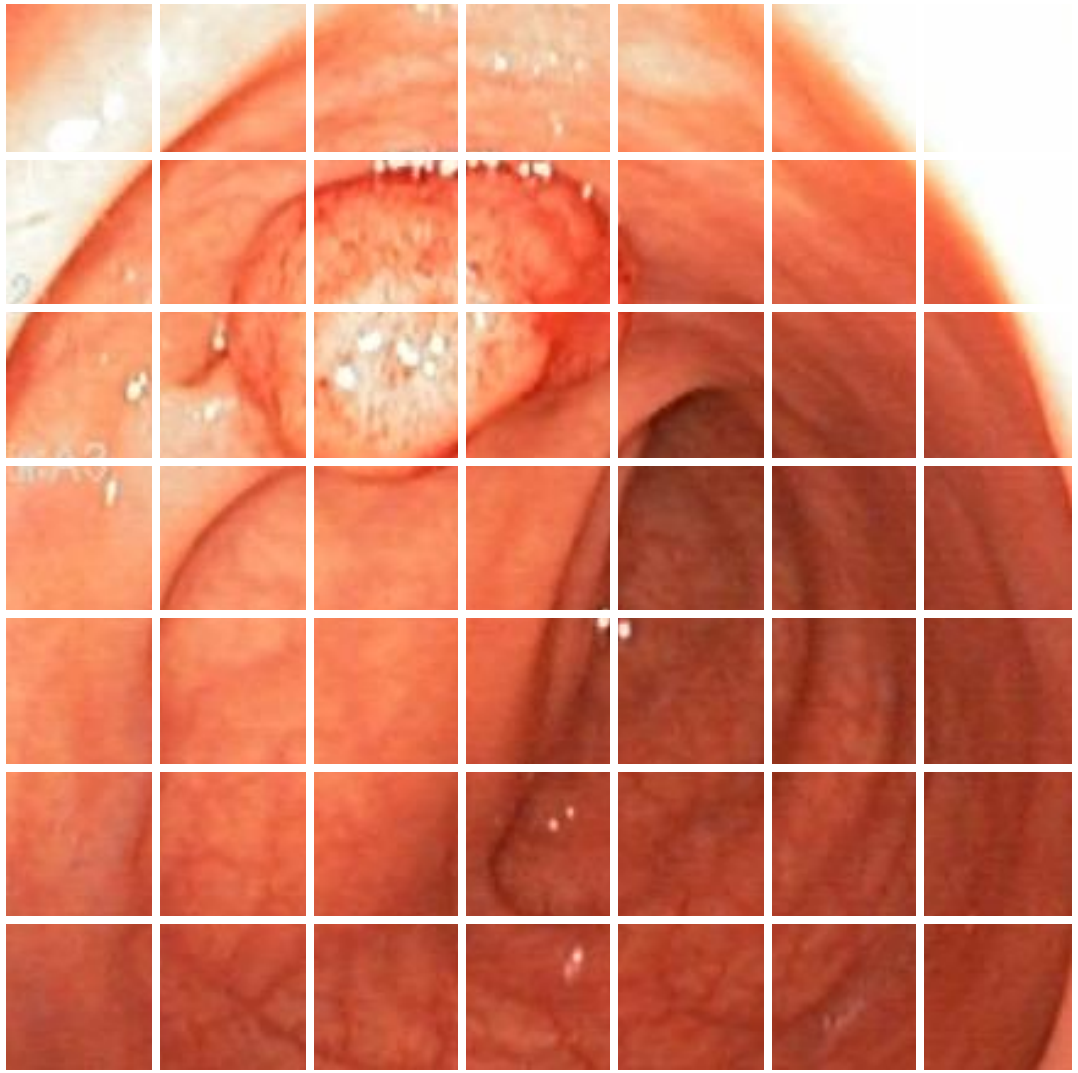
自信を持って生成した bbox にオブジェクトが含まれていなければ、その分だけ損失を計上する。

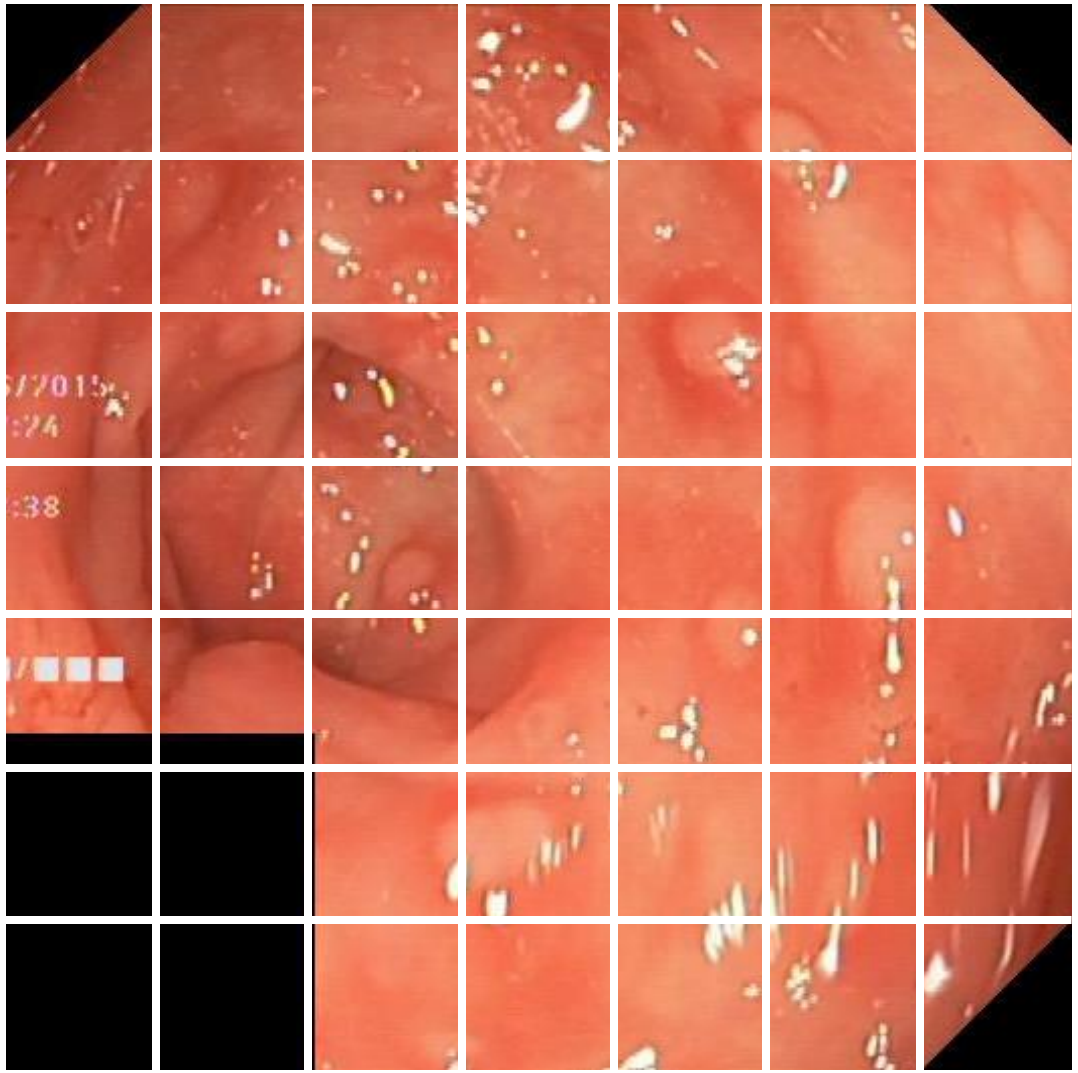
画像の大部分が背景であるため、この項が大きな値を取りやすく、他の項の値が相対的に小さくなり、学習が進まない。この効果を弱めるために、定数倍 $\lambda_{\text{noobj}} = 0.5$ をかける。

クラスの予測誤差

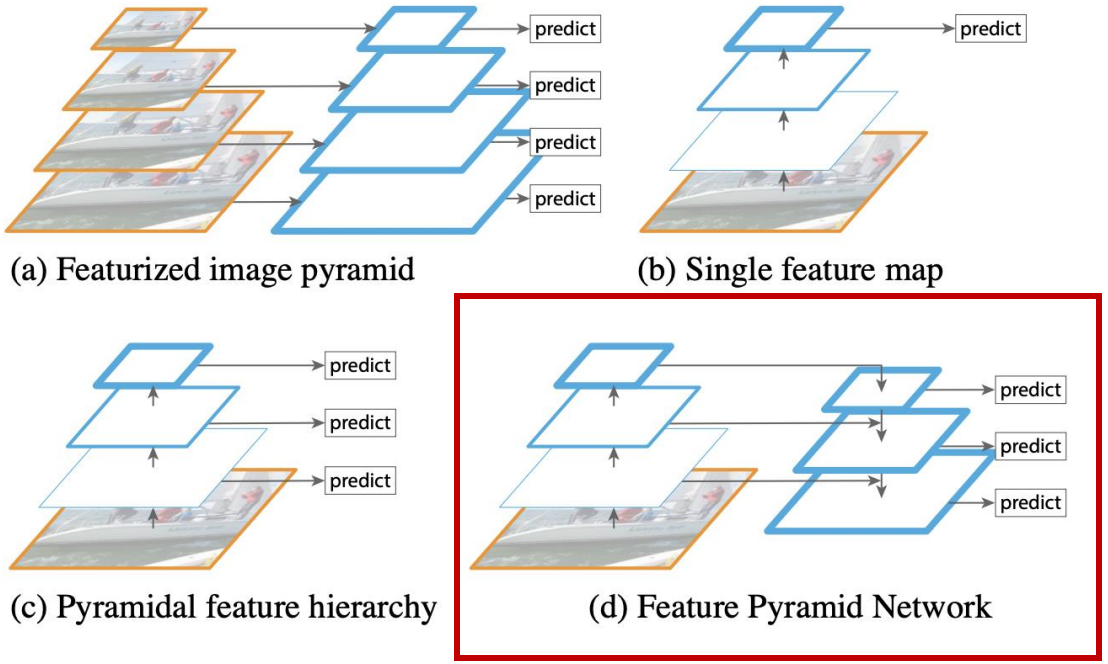
$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \hat{p}_i(c) \right)^2$$

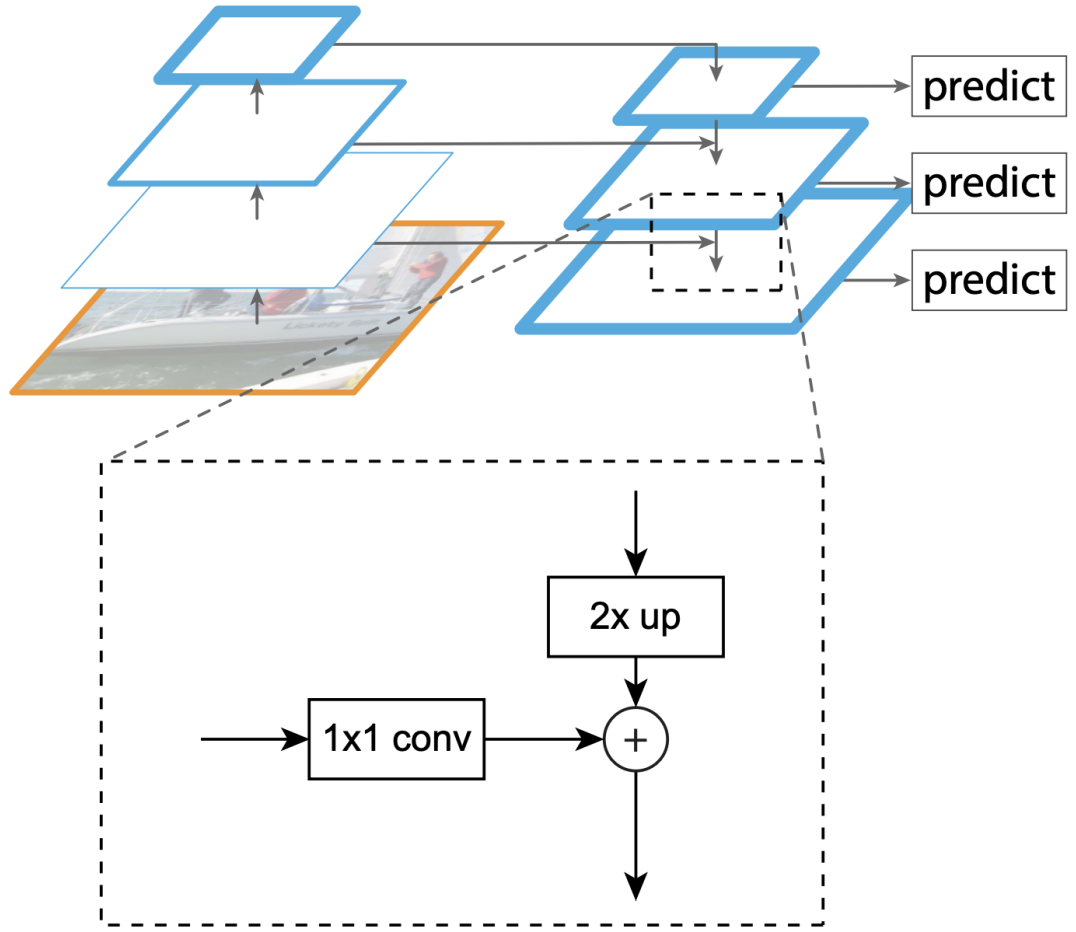
YOLO





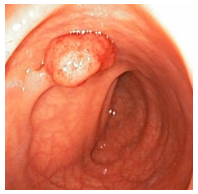
Feature Pyramid Network (FPN)





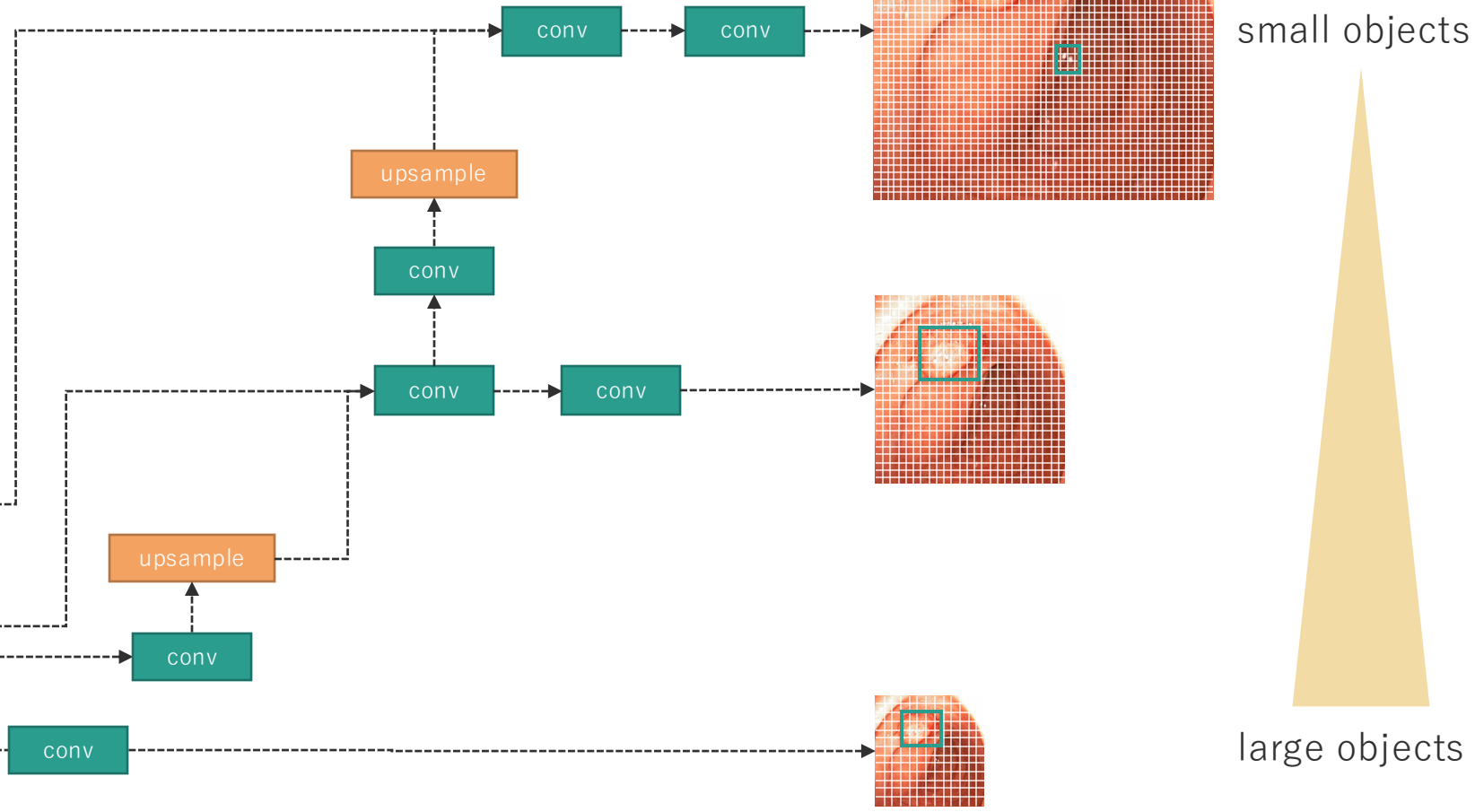
https://arxiv.org/abs/1612.03144

YOLO3



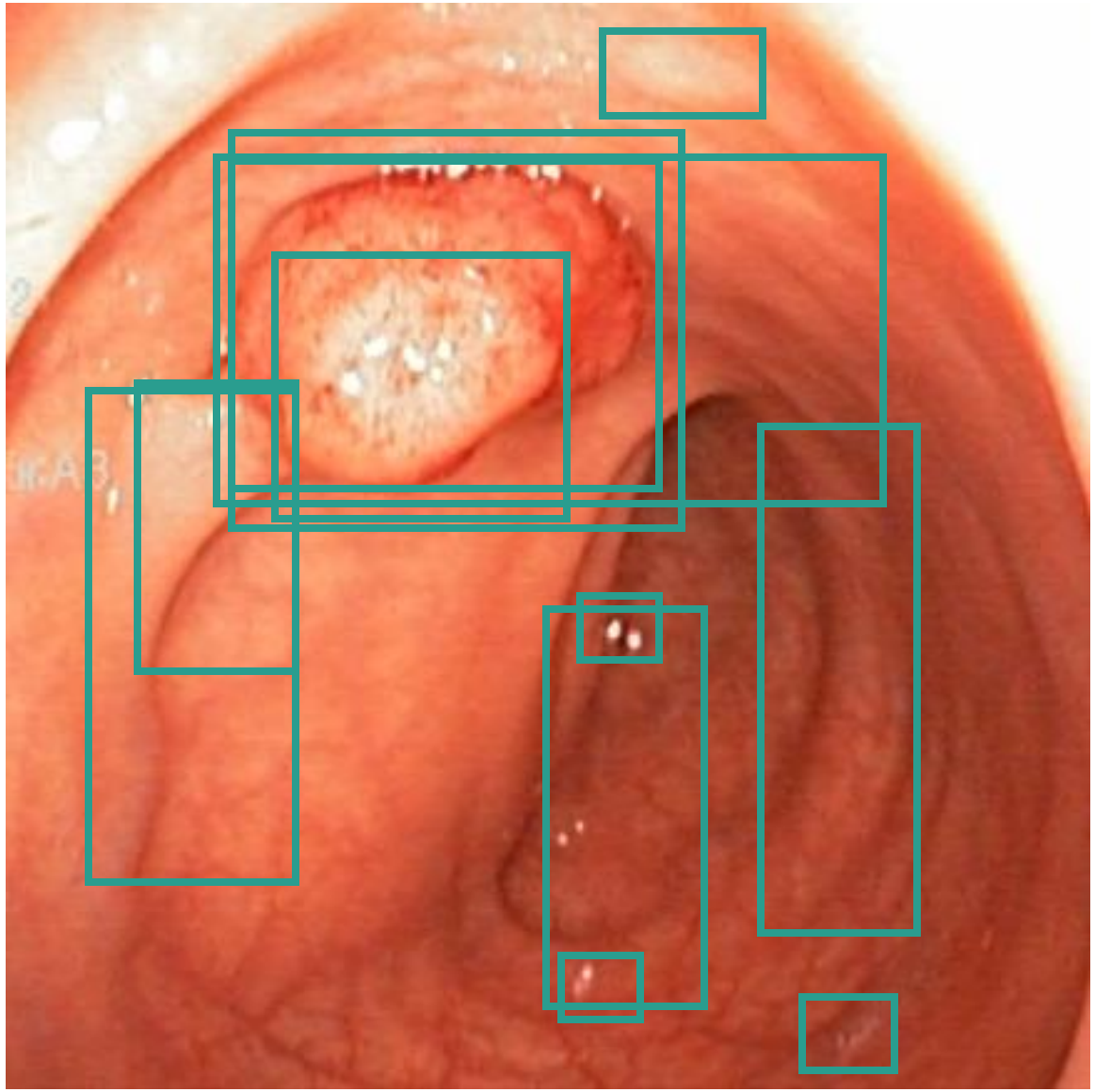
Darknet-53

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			



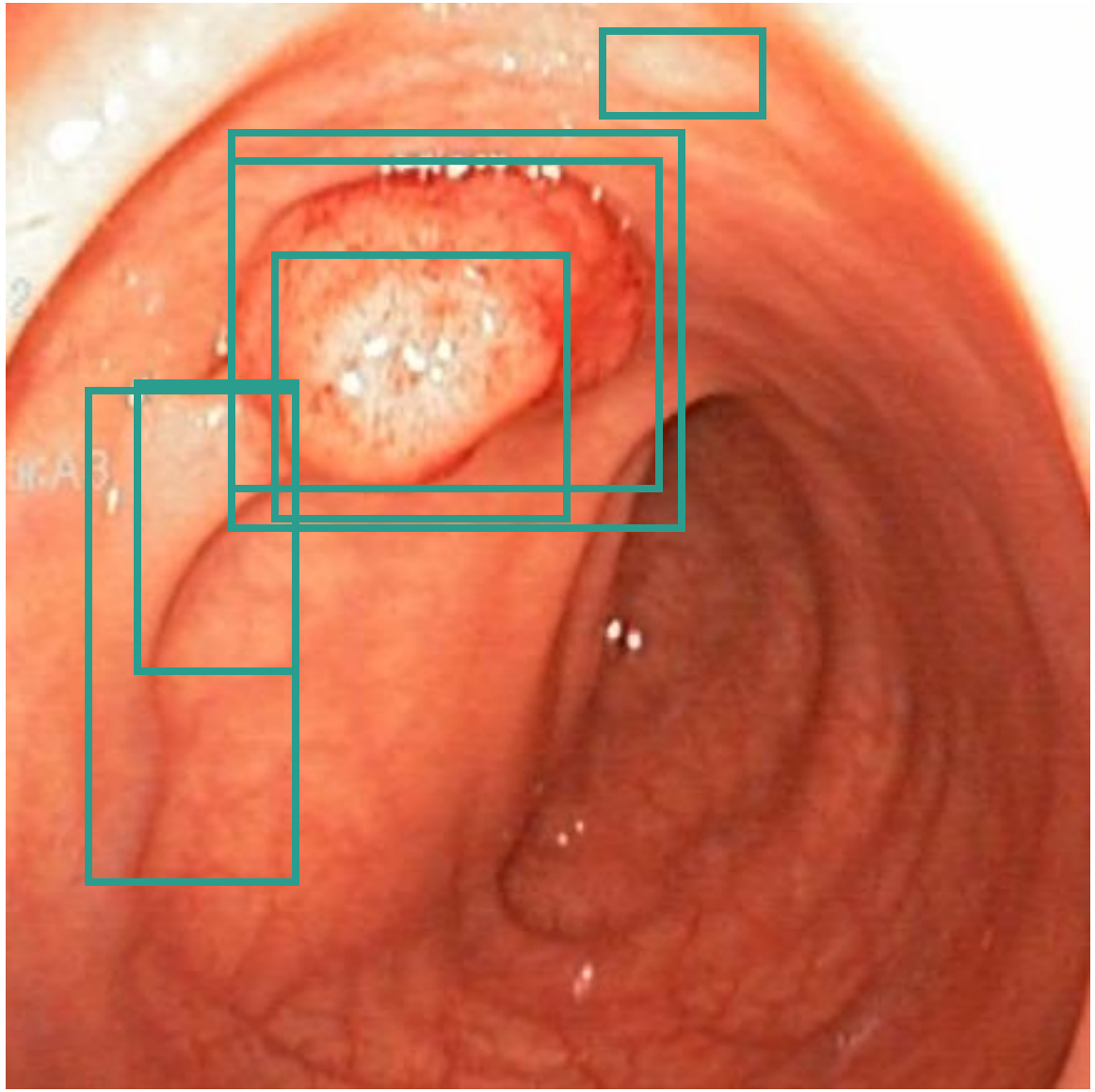
https://arxiv.org/abs/1804.02767

Non-maximum Suppression (NMS)



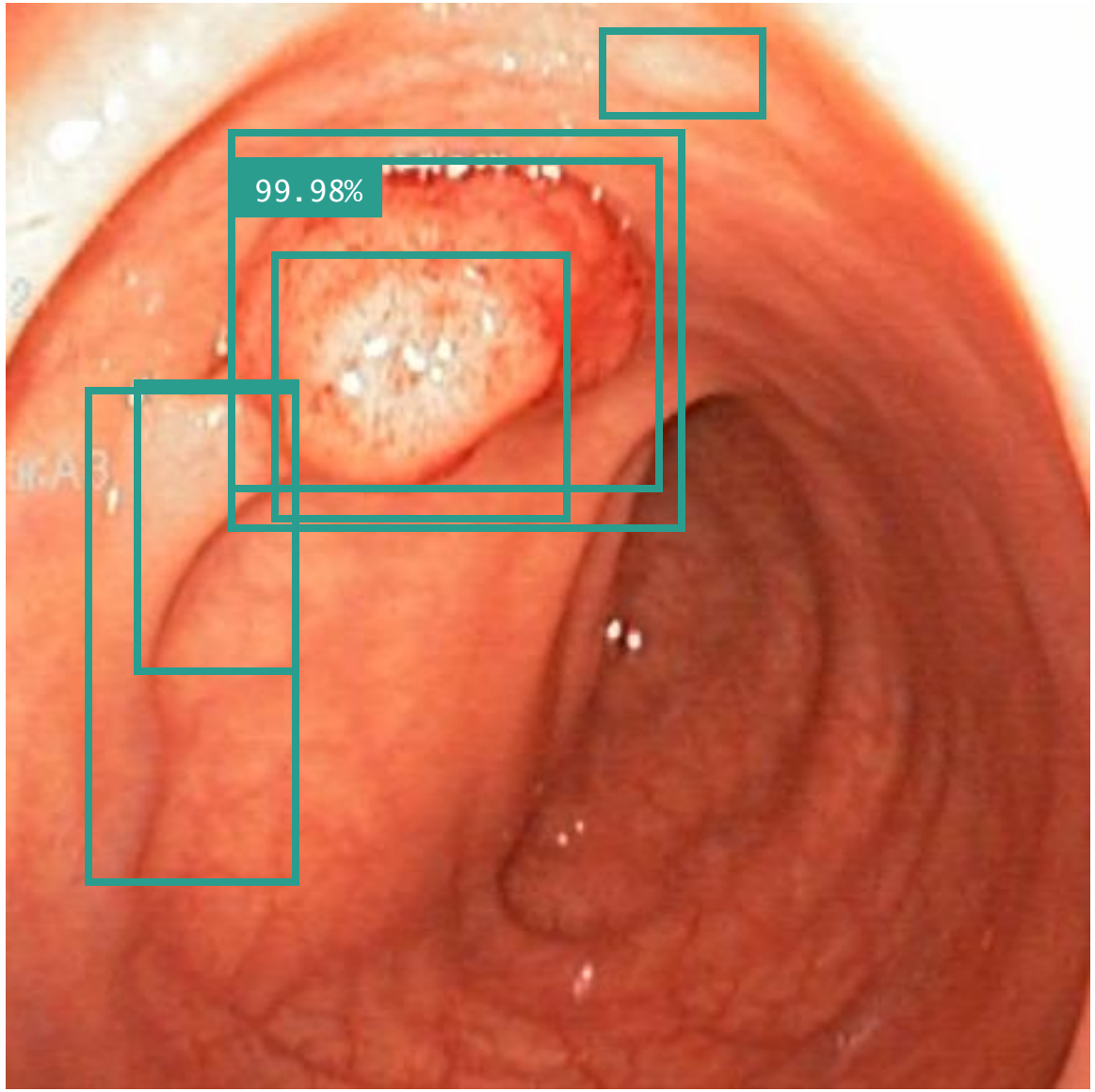
1. 予測において数多くの bbox が出力される。

Non-maximum Suppression (NMS)



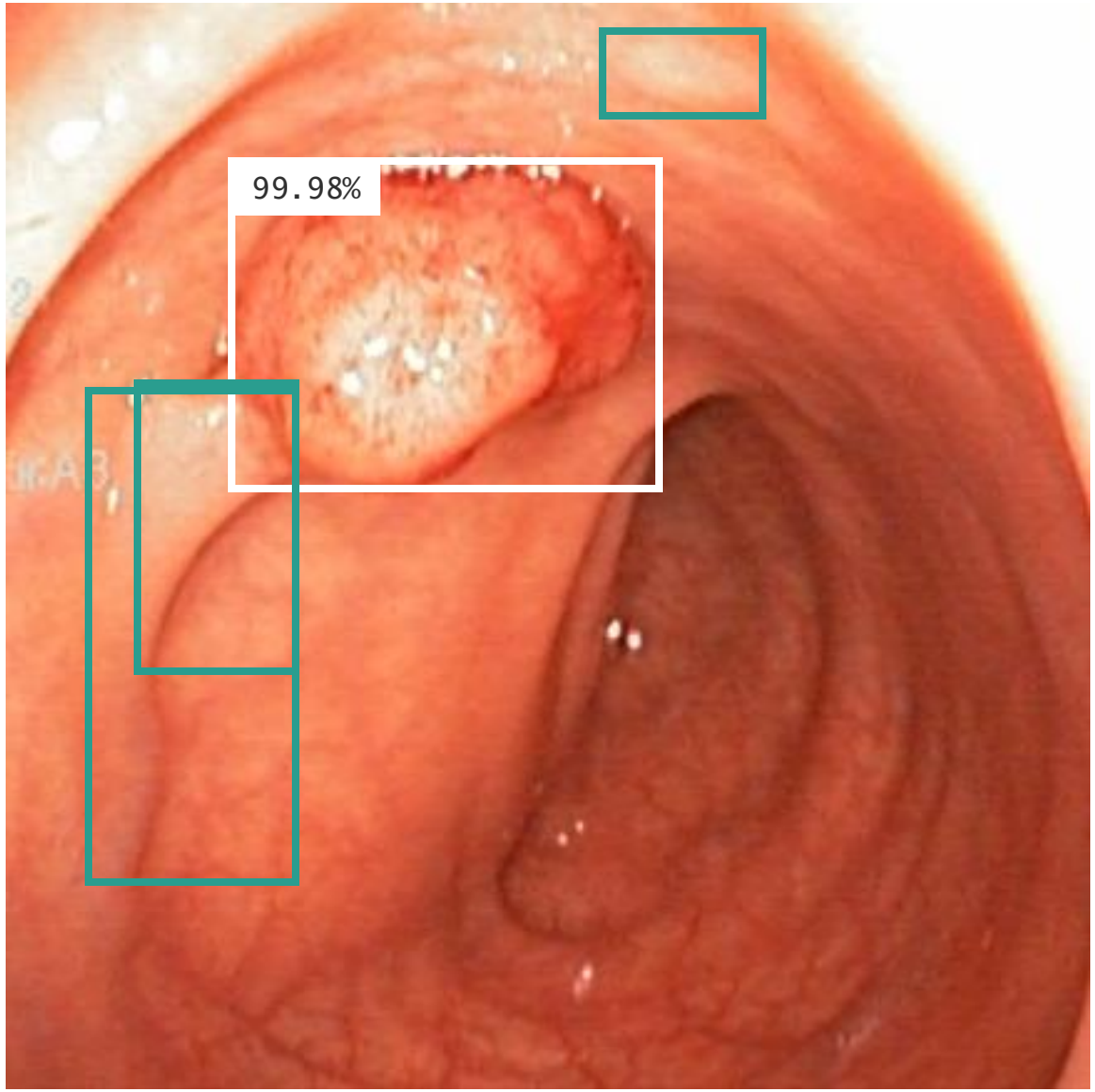
1. 予測において数多くの bbox が出力される。
2. 信頼度が最も高い k 個の bbox を残す。

Non-maximum Suppression (NMS)



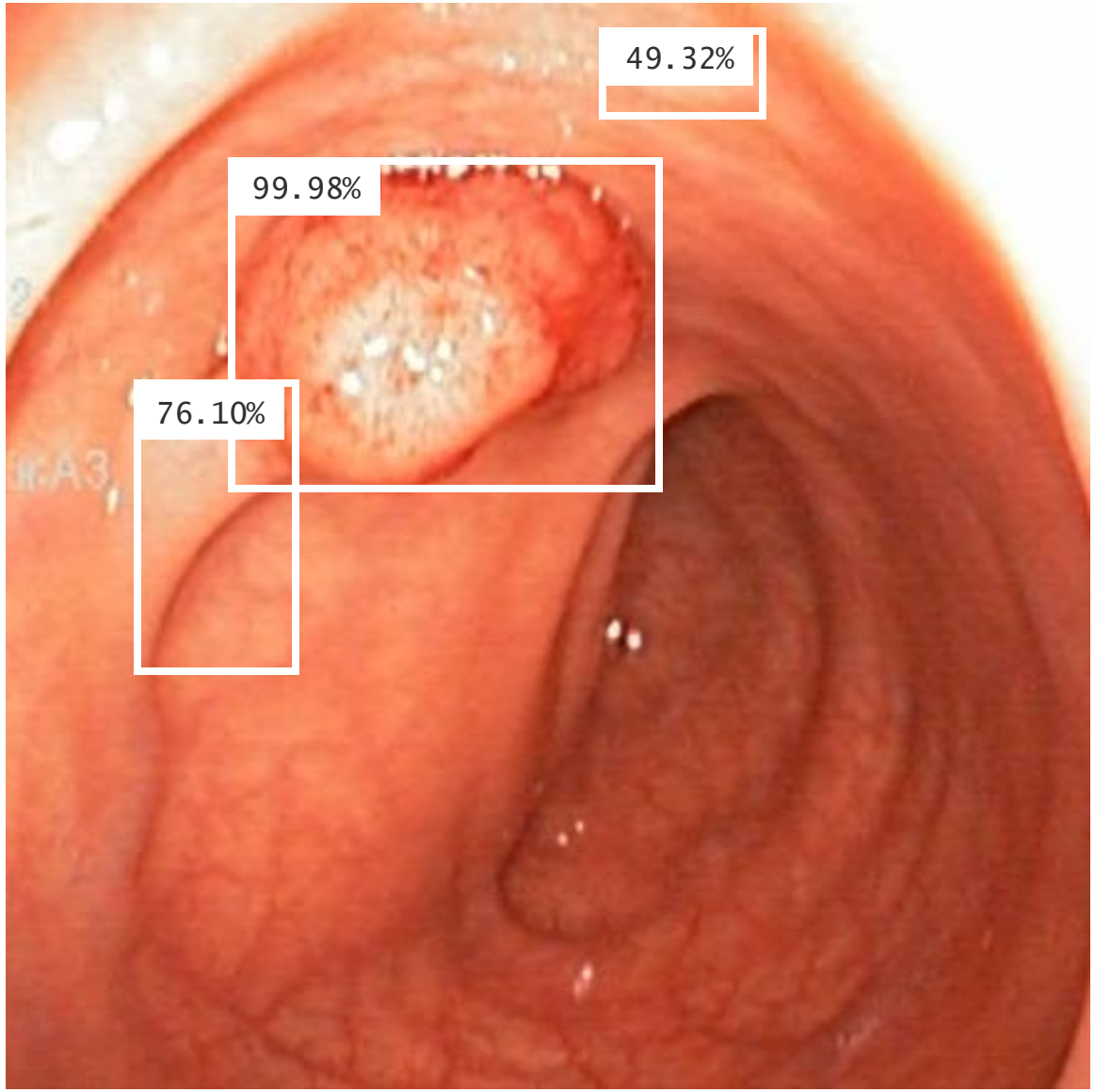
1. 予測において数多くの bbox が出力される。
2. 信頼度が最も高い k 個の bbox を残す。
3. 信頼度が最も高い bbox を基準として、他の bbox との IoU を計算する。

Non-maximum Suppression (NMS)



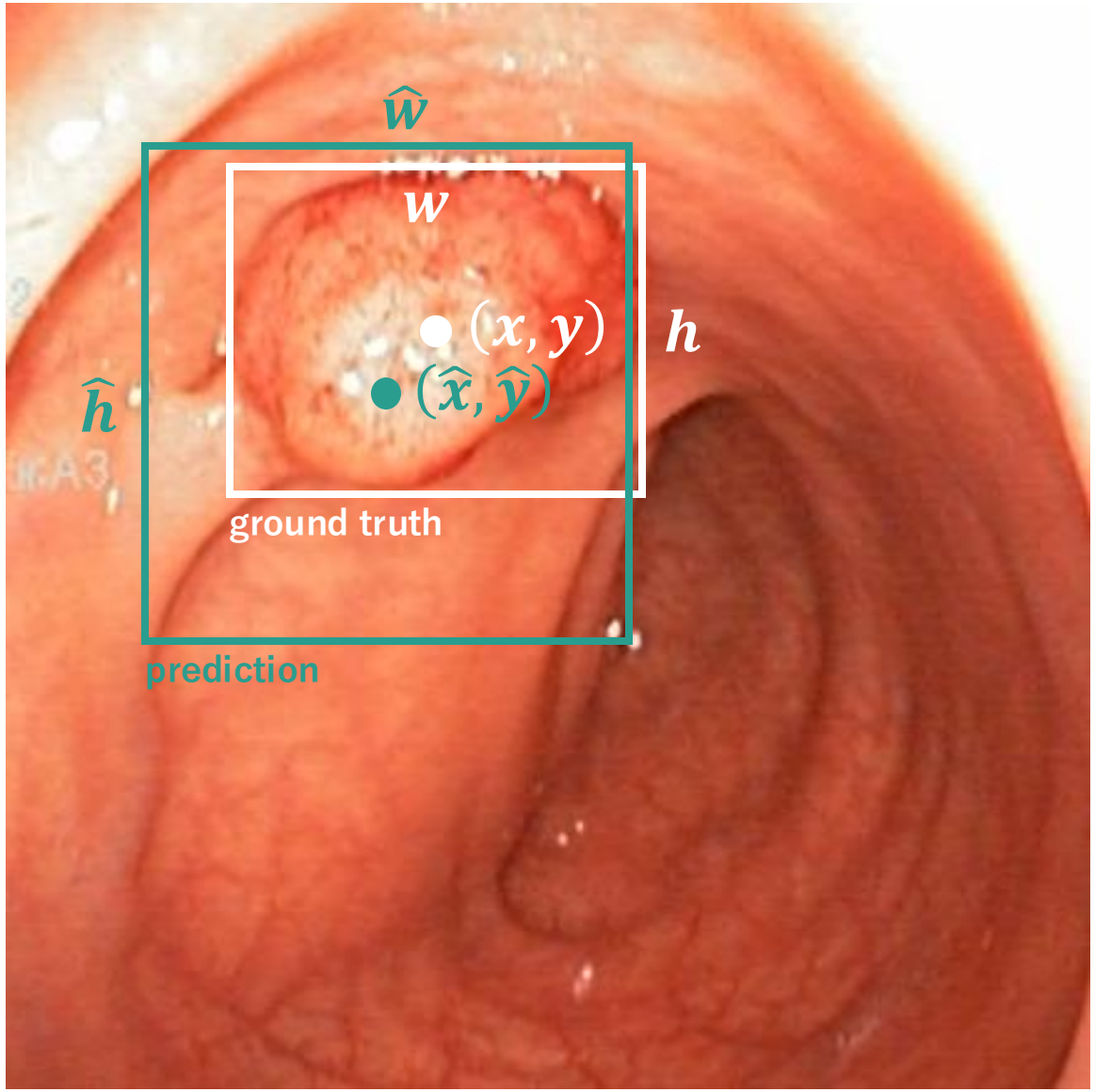
1. 予測において数多くの bbox が出力される。
2. 信頼度が最も高い k 個の bbox を残す。
3. 信頼度が最も高い bbox を基準として、他の bbox との IoU を計算する。
4. IoU > 0.5 ならばその bbox を除去し、基準となった bbox を予測結果として出力する。

Non-maximum Suppression (NMS)

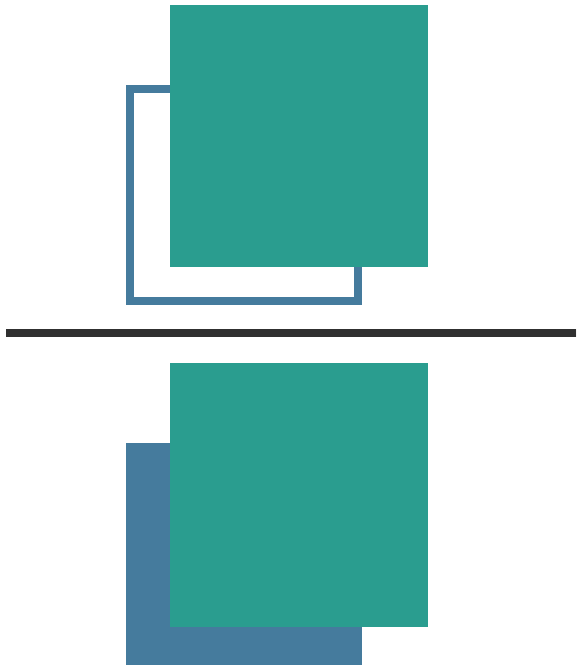


1. 予測において数多くの bbox が出力される。
2. 信頼度が最も高い k 個の bbox を残す。
3. 信頼度が最も高い bbox を基準として、他の bbox との IoU を計算する。
4. IoU > 0.5 ならばその bbox を除去し、基準となった bbox を予測結果として出力する。
5. 残った bbox に対して、手順 3-4 を繰り返す。

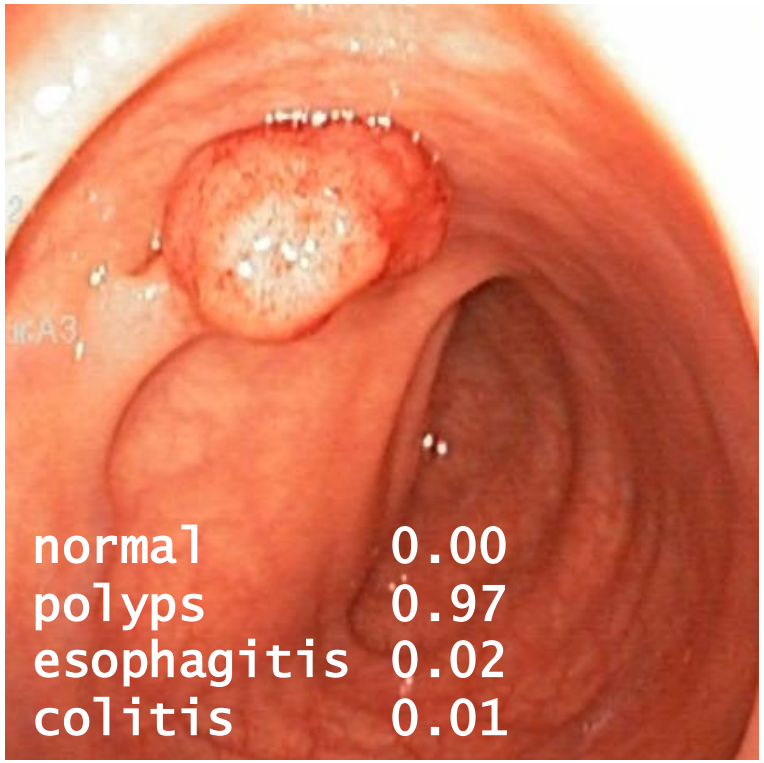
Intersection over Union (IoU)



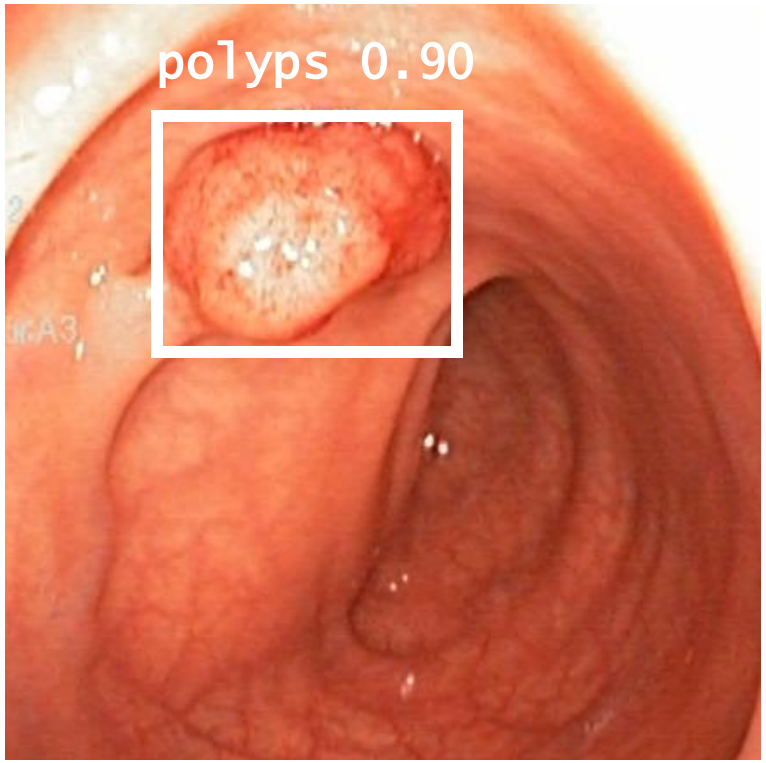
$$\text{IoU} = \frac{\text{intersection}}{\text{union}}$$



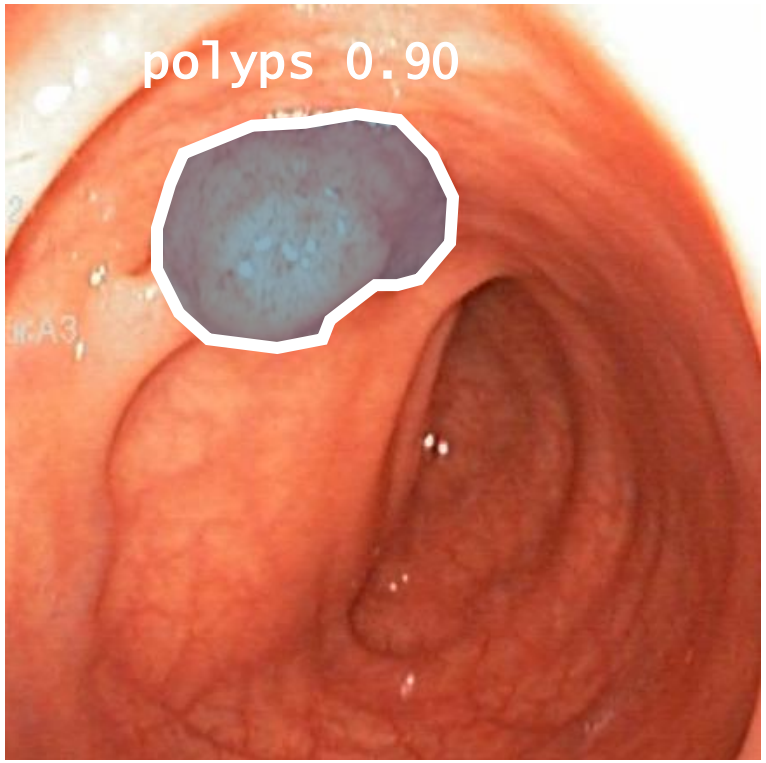
物体分類



物体検出

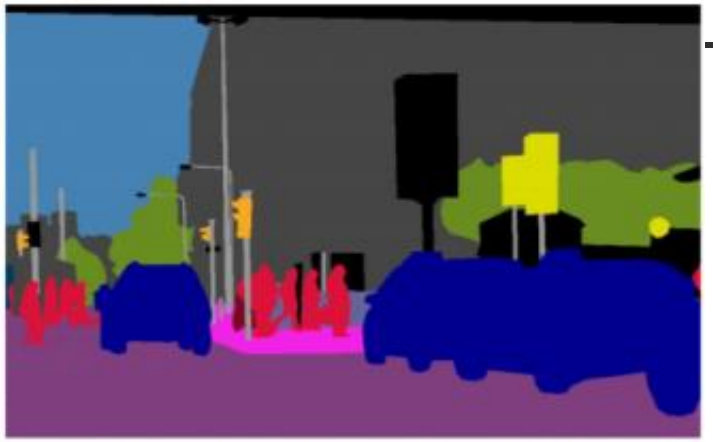


セグメンテーション



セグメンテーション

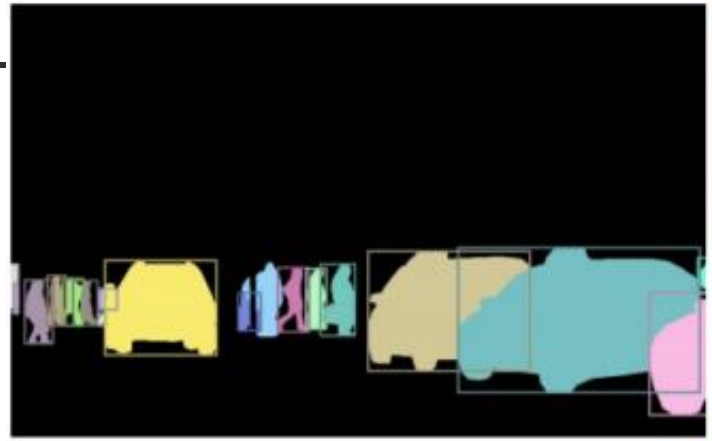
(a) image



(b) semantic segmentation

すべてのピクセルをクラスレベルで検出。

個々のオブジェクトをピクセルレベルで検出し、オブジェクトごとに ID を付与。



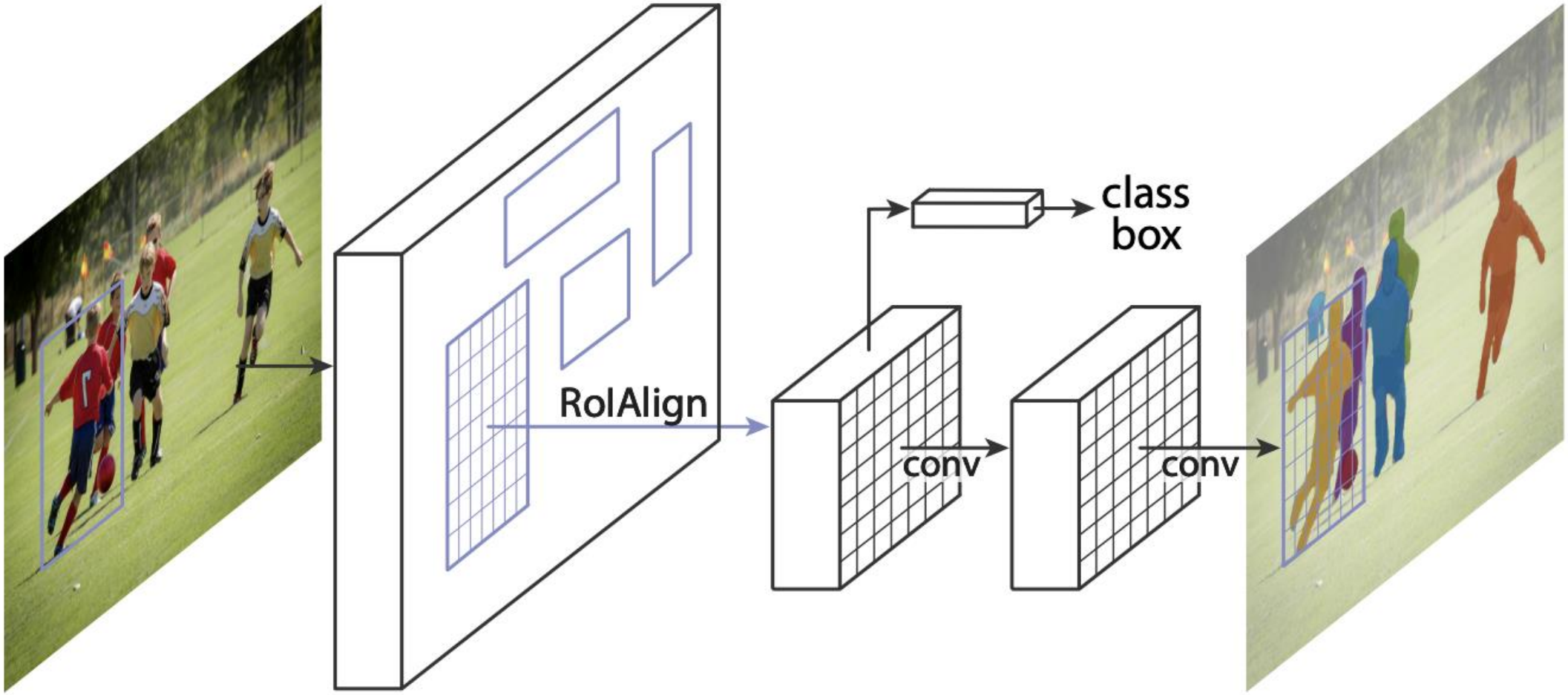
(c) instance segmentation

(d) panoptic segmentation

すべてのピクセルをクラスレベルで検出し、オブジェクトごとに ID を付与。

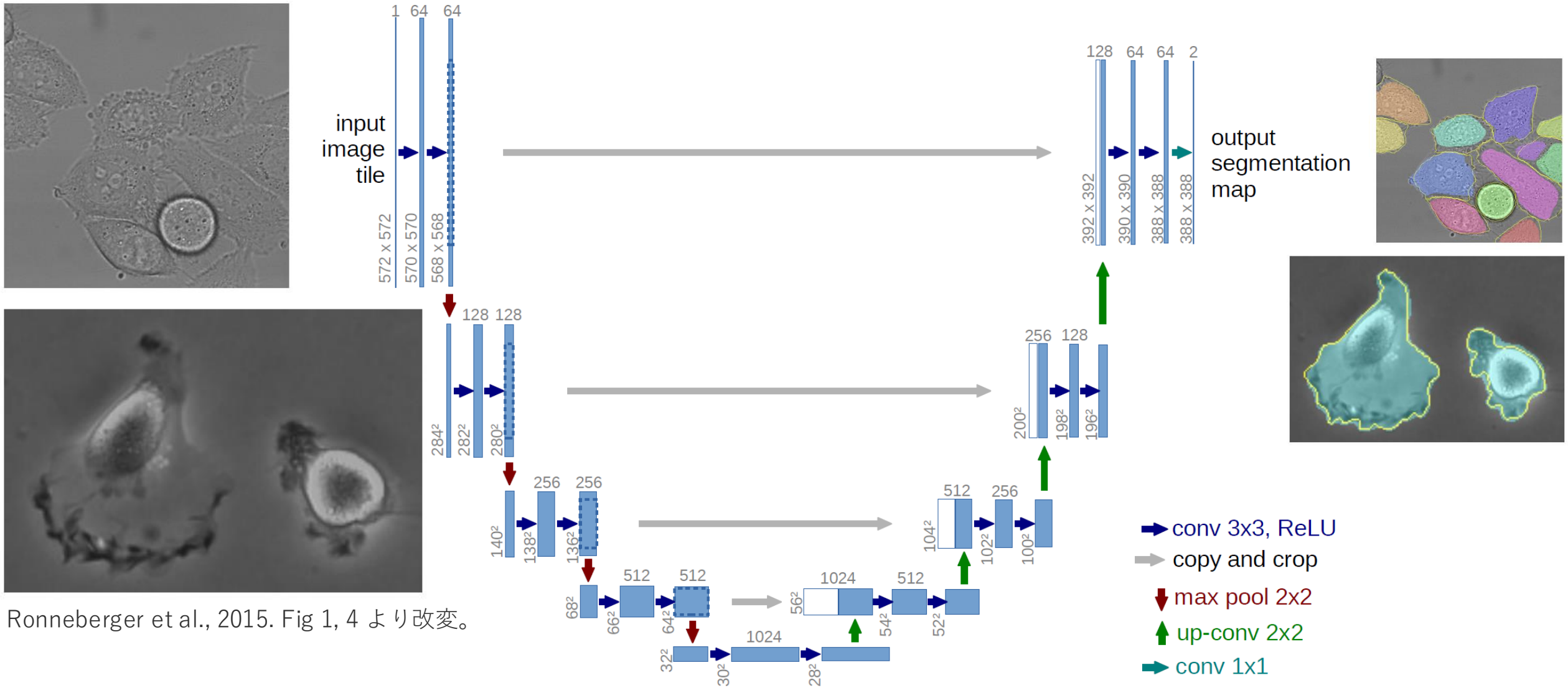
https://arxiv.org/abs/1801.00868

Mask R-CNN



https://arxiv.org/abs/1703.06870

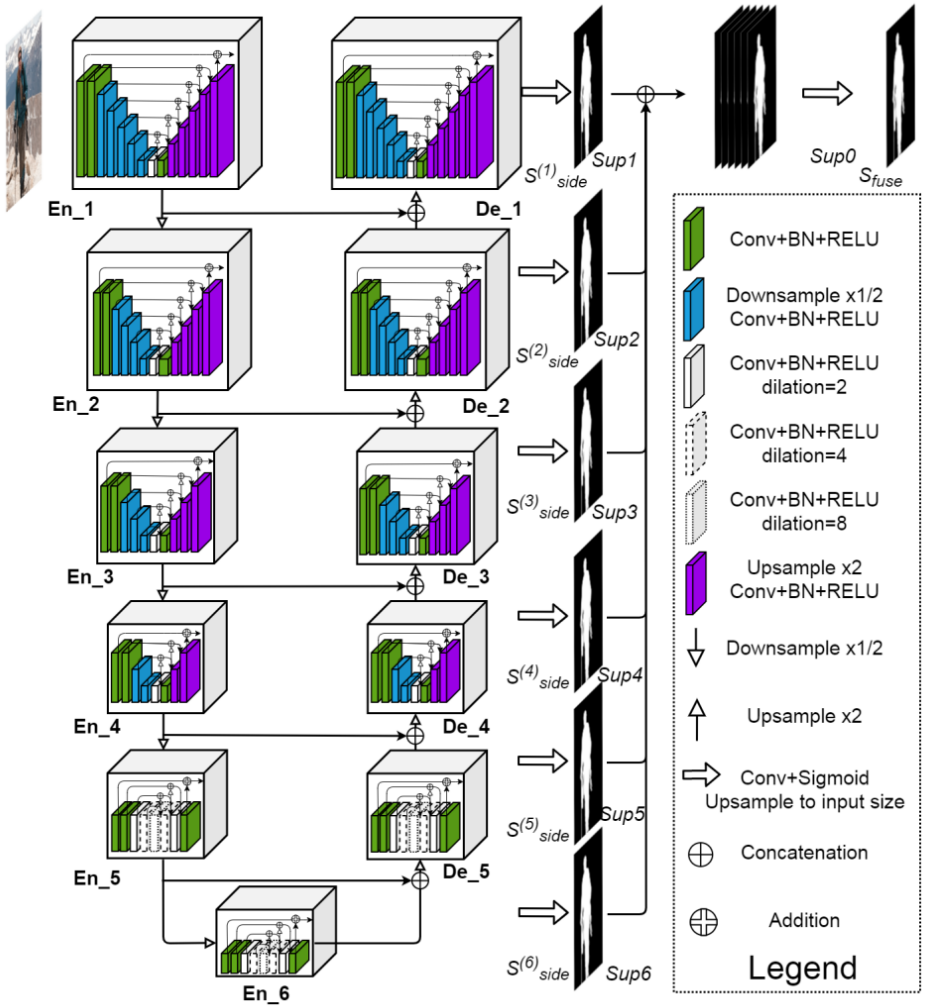
U-Net



Ronneberger et al., 2015. Fig 1, 4 より改変。

https://arxiv.org/abs/1505.04597

Salient Object Detection

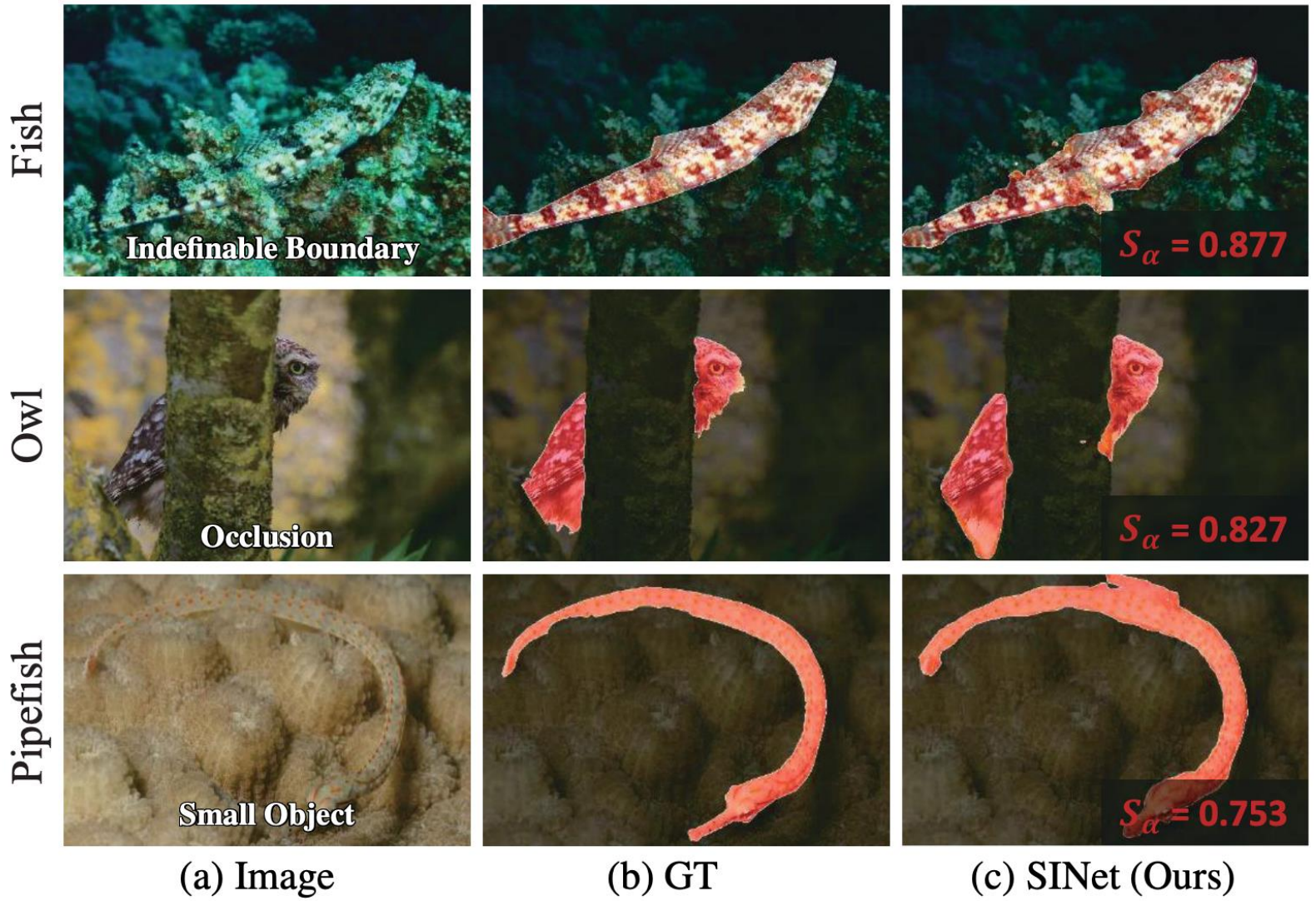


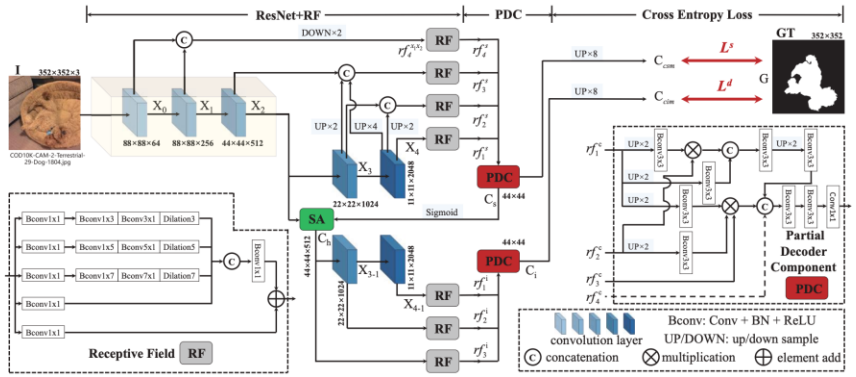
Qin et al., 2020. Fig 5

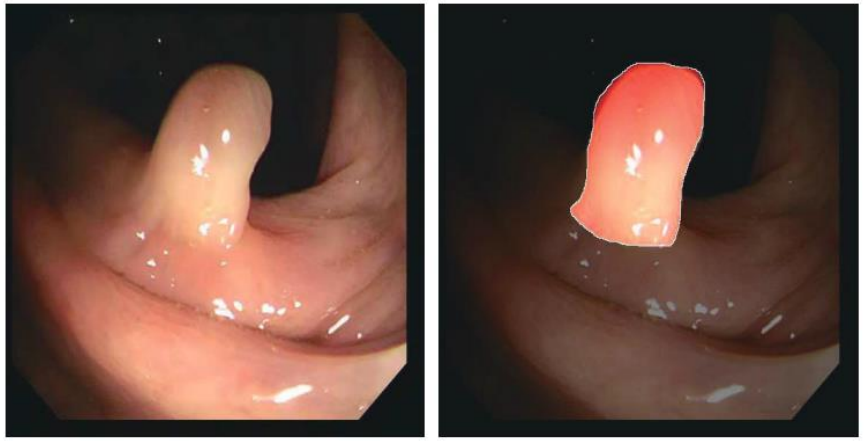
https://github.com/xuebinqin/U-2-Net

https://arxiv.org/abs/2005.09007

Camouflaged Object Detection







Fan et al., 2020. Fig 8, 9, 10 より改変

https://doi.org/10.1109/CVPR42600.2020.00285

深層学習

ニューラルネットワーク

畳み込みニューラルネットワーク

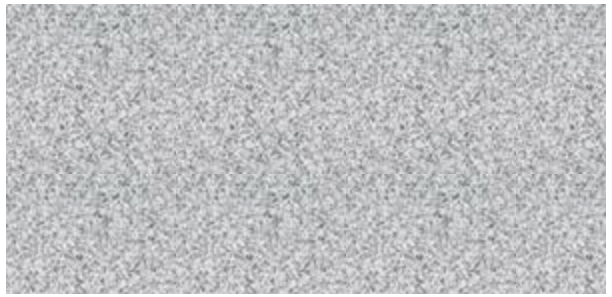
　物体分類

　物体検出・セグメンテーション

敵対的生成ネットワーク

敵対的生成ネットワーク

本物データ



生成データ

Discriminator

real

fake

Generator

ノイズ

敵対的生成ネットワーク

本物データ

生成データ

Discriminator

real

fake

Generator

ノイズ

敵対的生成ネットワーク

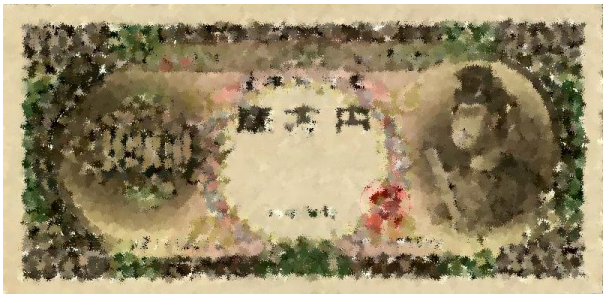
本物データ

Discriminator

real

fake

Generator



生成データ

ノイズ

敵対的生成ネットワーク

本物データ

ノイズ

Generator

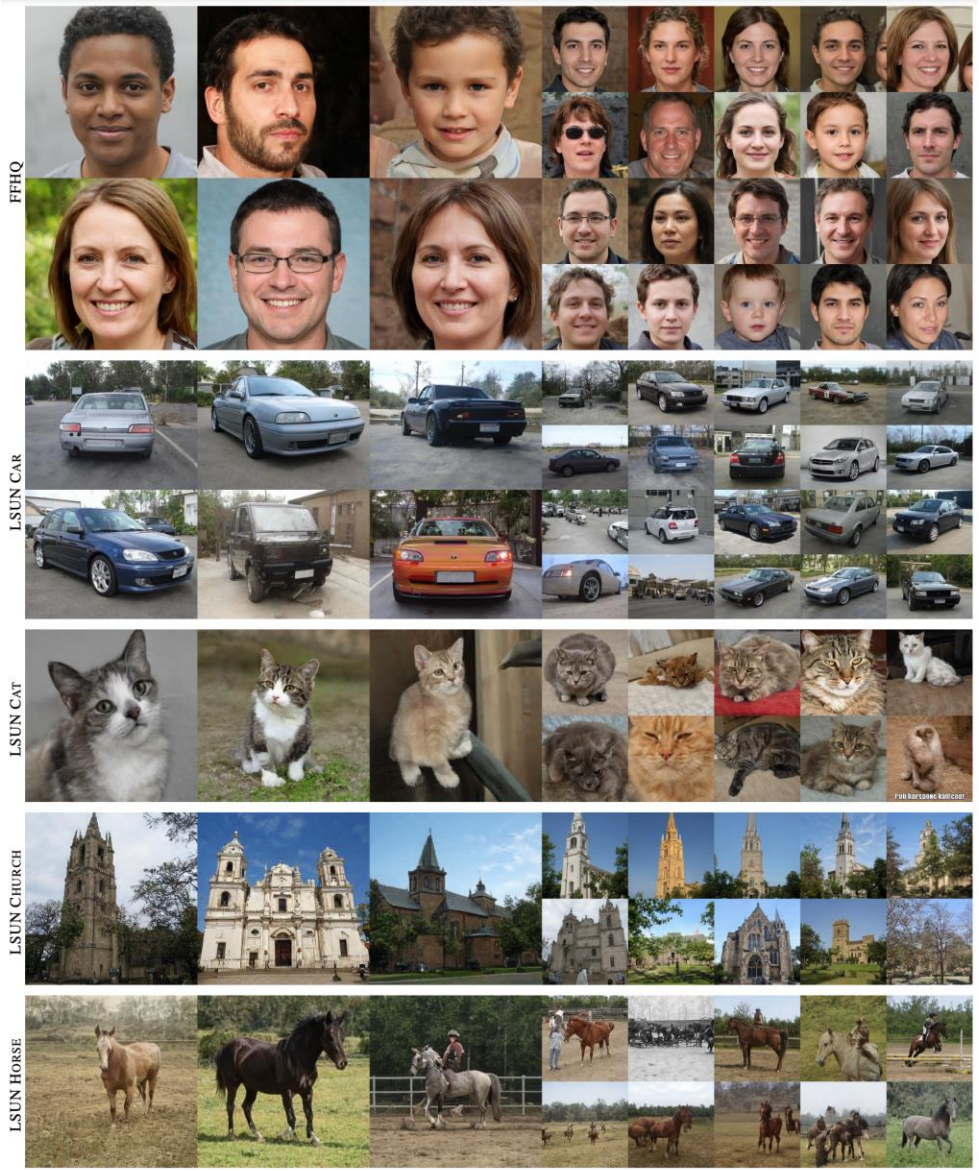
生成データ

Discriminator

real

fake

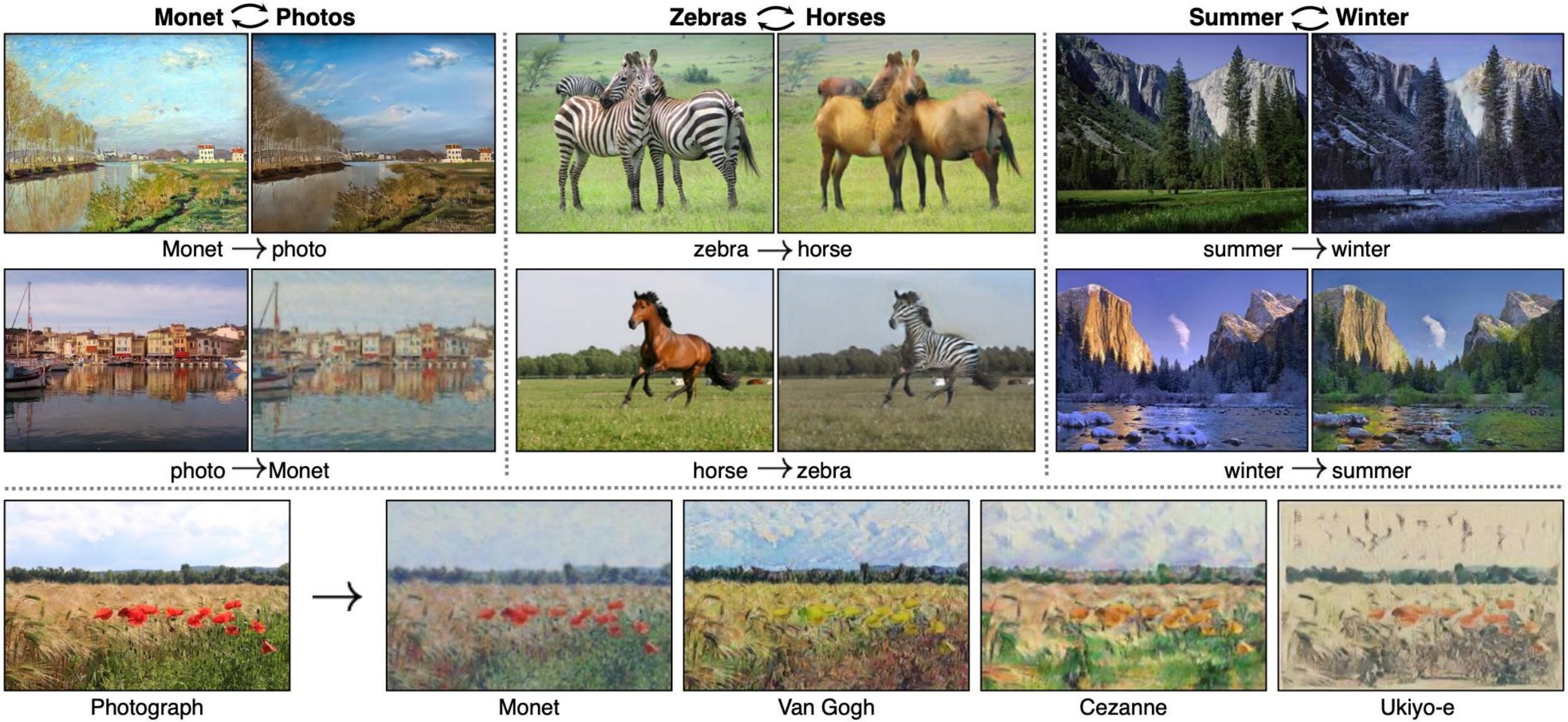
StyleGAN2



Karras et al., 2020. Fig 12

https://arxiv.org/abs/1912.04958

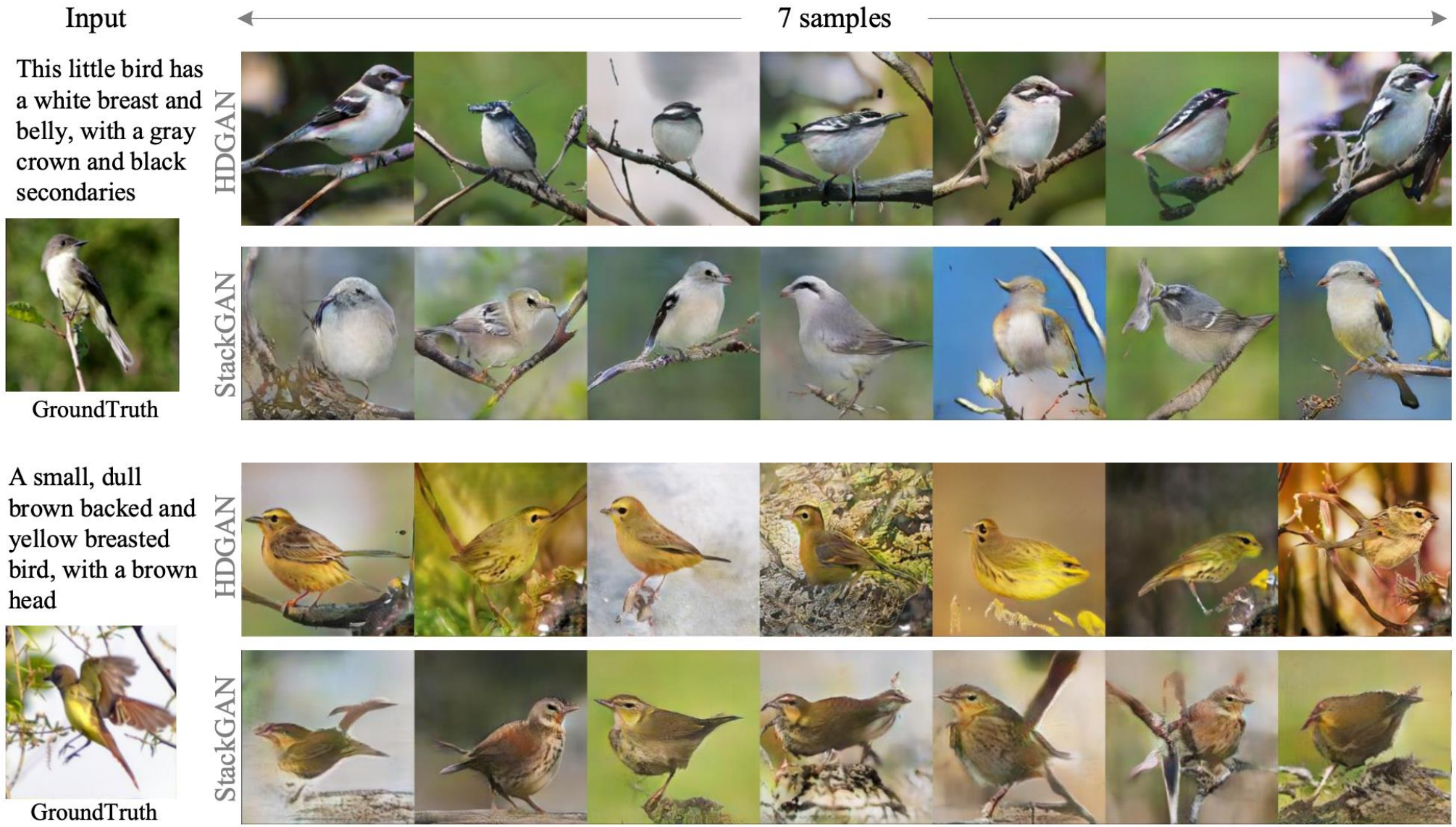
CycleGAN



Zhu et al., 2020. Fig 1

https://arxiv.org/abs/1703.10593

HDGAN



Zhang et al., 2018. Fig 7 改変

https://arxiv.org/abs/1802.09178