深層学習の実際

現在の中心的な技術は

多層ニューラル

ネットワーク

しかも２次元で

多層ニューラル
ネットワーク

Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 1980;36(4):193-202. doi: 10.1007/BF00344251. PMID: 7370364.

ネオコグニトロン

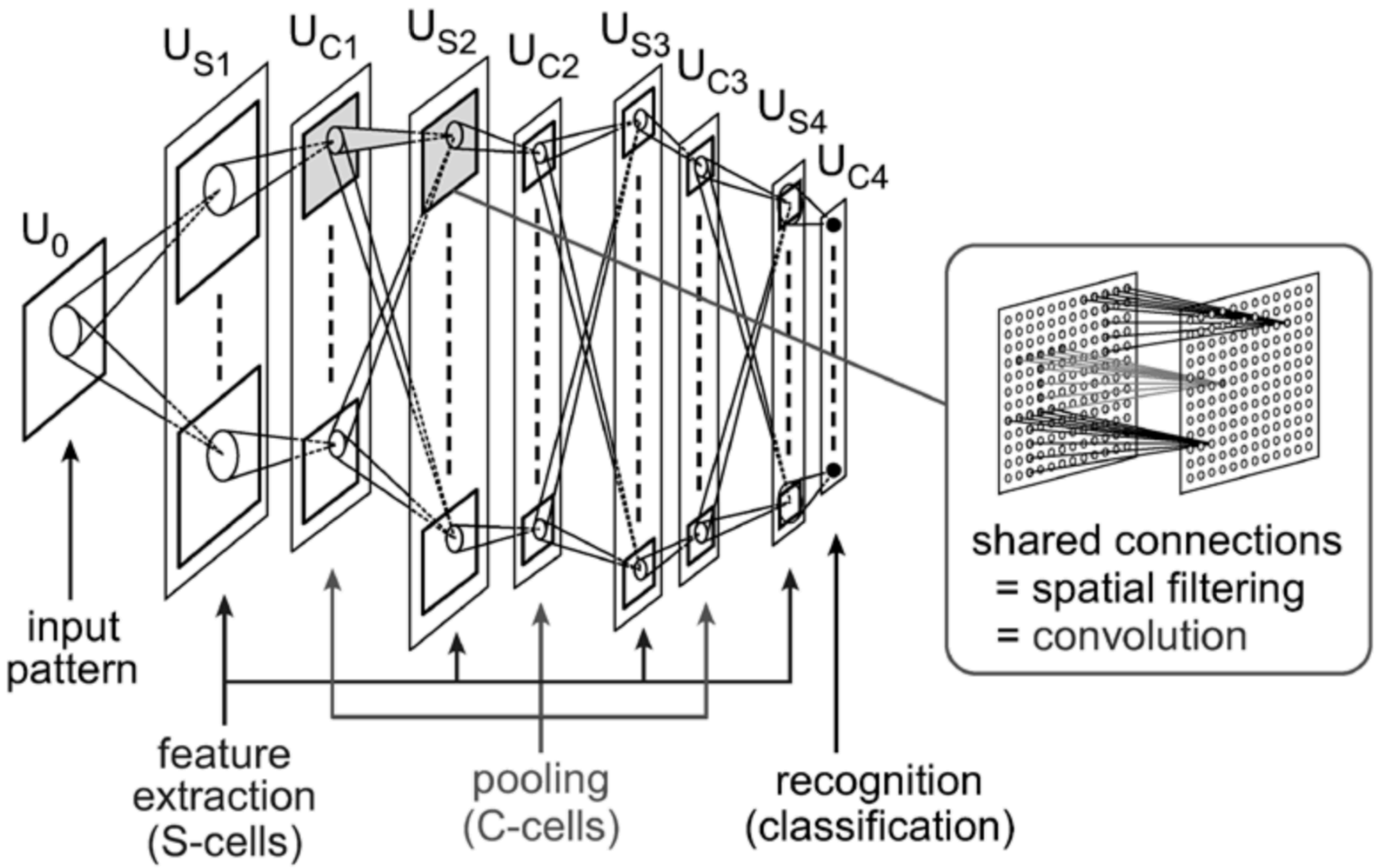


図 1　ネオコグニトロンの回路構造.

(VISION Vol. 29, No. 1, 1–5, 2017)

Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 1980;36(4):193-202. doi: 10.1007/BF00344251. PMID: 7370364.

ネオコグニトロン

画像を直接入力

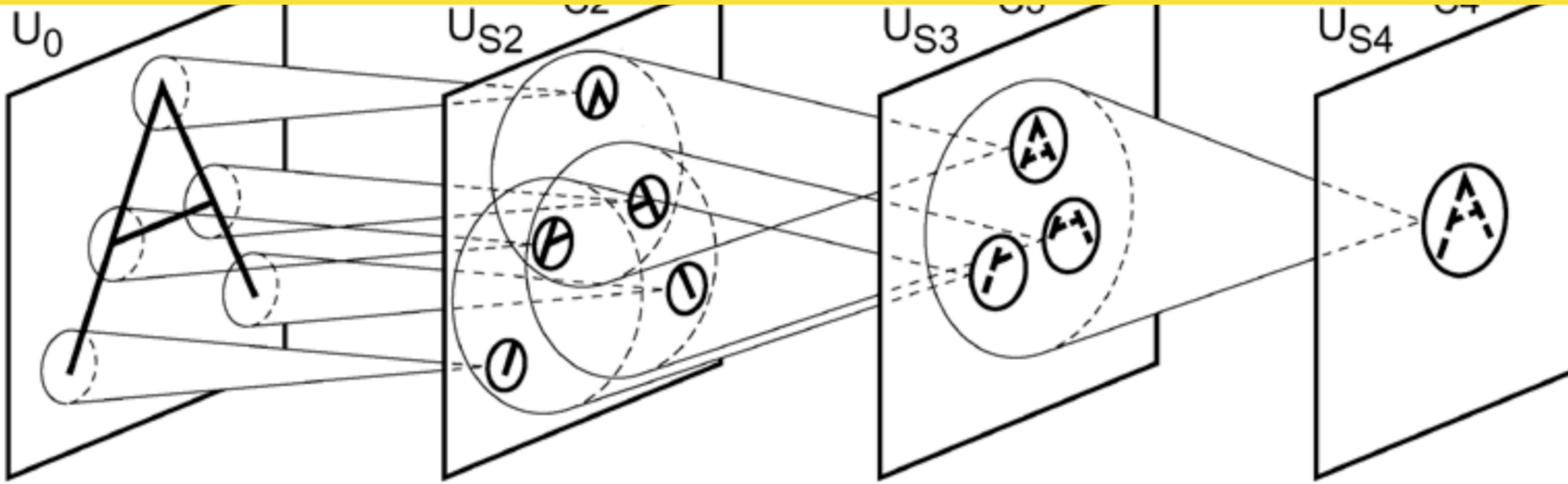


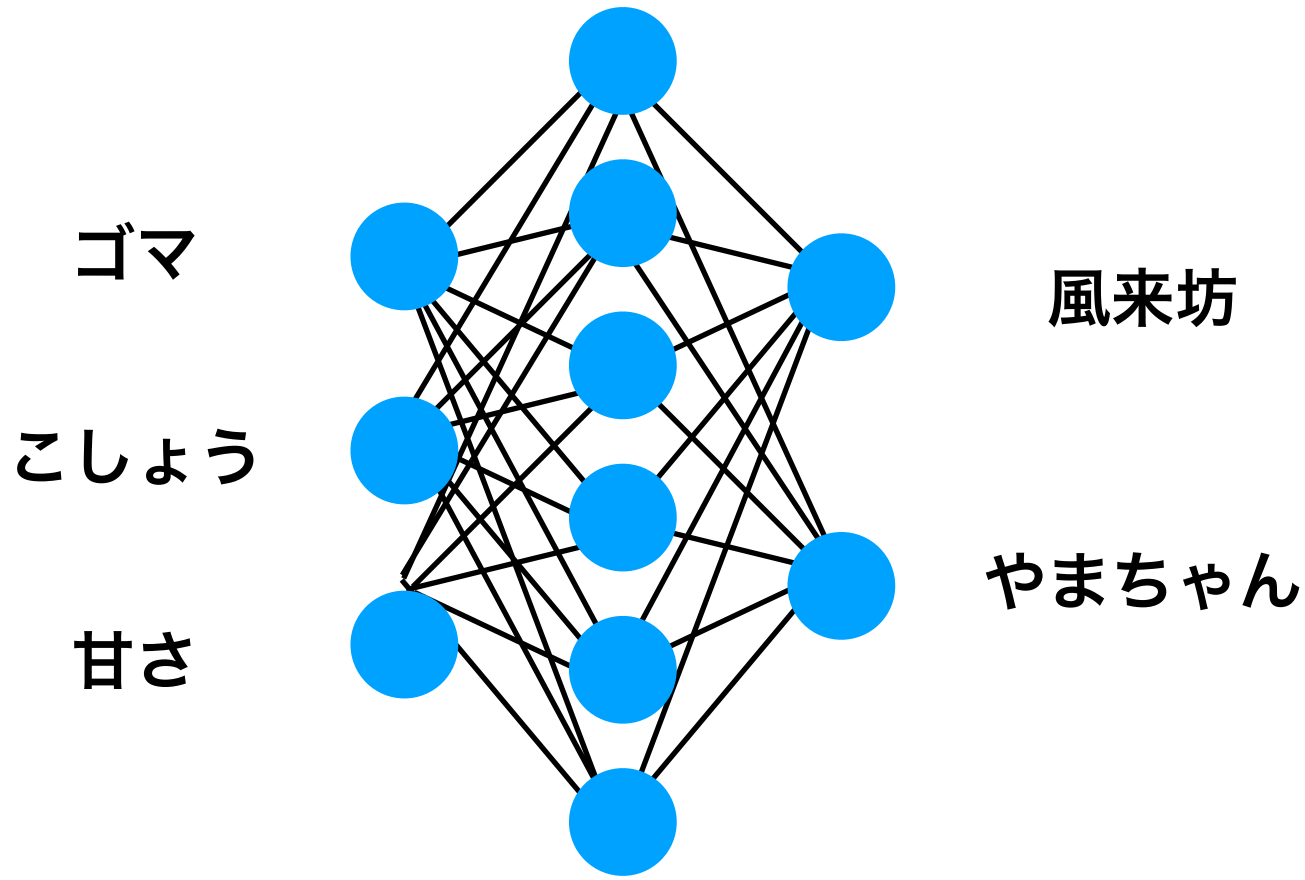
図2　ネオコグニトロンにおけるパターン認識の原理[1)]

(VISION Vol. 29, No. 1, 1–5, 2017)

名古屋は「手羽先」だがや

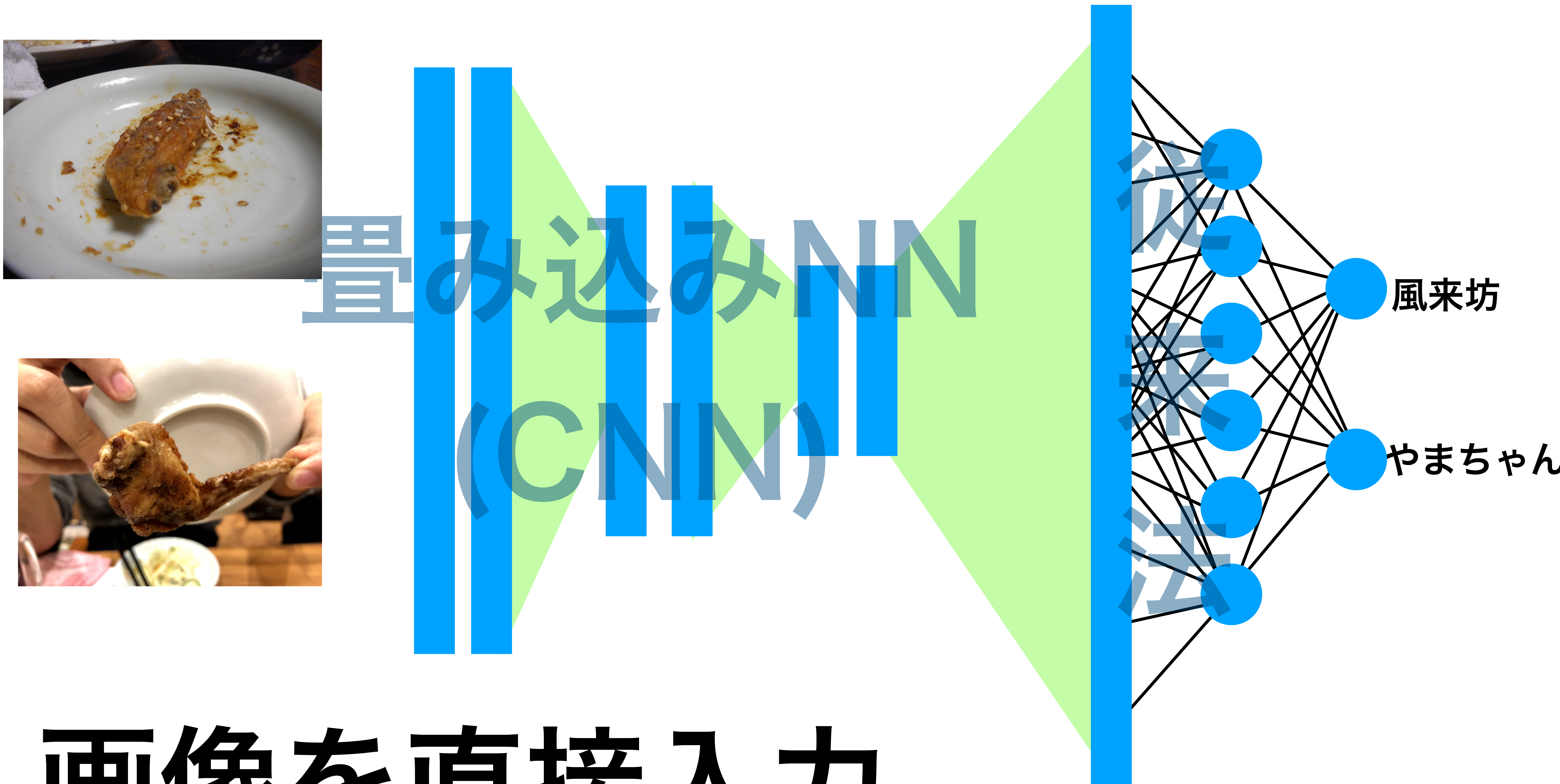
山ちゃん or 風来坊 自動判別

従来技術



特徴を手動で抽出

どう変わったか？



画像を直接入力

深層学習
考え方


数値
深層
学習
説明変数
x
f
関数
目的変数
y
★
I
or
II
or
III

x, y で関数を決める

学習とはコレを決定すること

機械学習で獲得

$$y = f(\boldsymbol{x})$$

目的変数
（群／数値）

説明変数
（ベクトル）

教師あり学習

関数っていろいろあったなぁ〜

$$f(x) = ax^2 + bx + c$$ **モデル**

ちょっとだけ一般化

$$f(x, \theta), \quad \theta = \{a, b, c\}$$ **パラメタ**

学習データと評価データ

- 性能評価のためには検証が必要

学習とはコレを
決定すること

機械学習で獲得

$$y = f(\boldsymbol{x})$$

目的変数
（群／数値）

説明変数
（ベクトル）

教師あり学習

fを構築するためのデータ：学習データ／Trainingデータ
その後に評価するためのデータ：評価データ／Testデータ

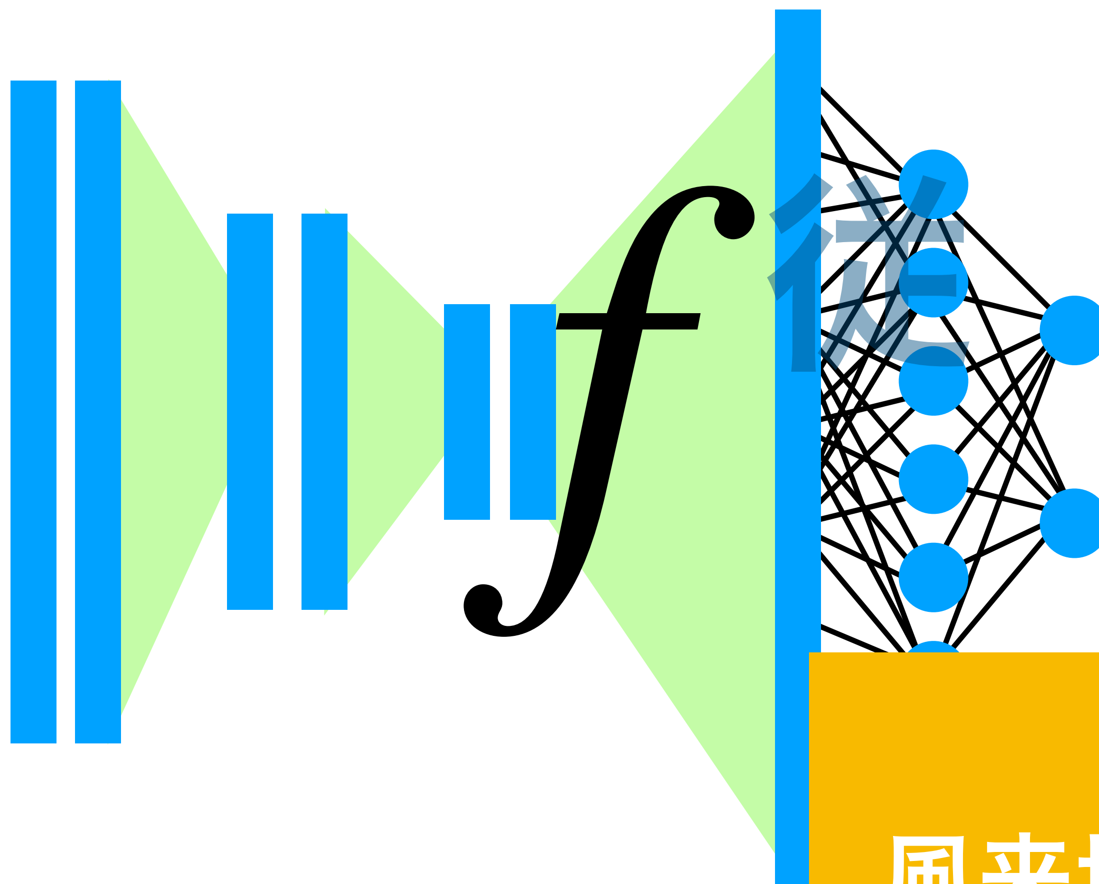
学習データと評価データが混ざるといわゆる「カンニング」

混ざらないように　かつ　少ないデータを最大限活用する工夫
Cross-validation/Leave-one-outなど

x =

画像も音声も文字もすべて数値データ

x



f

y

(1, 0)
風来坊を表すベクトル

$y = \{風来坊\}$

ディープラーニングを使う

x, y に何を持ってくるか?

ディープラーニングは
与えられた x, y
に従って関数を決定する

ディープラーニングを使う

$$x \longrightarrow y$$

x	y
文字	良性悪性（２分類）
音声	ステージ（多分類）
画像	数値
映像	検出（箱型）
	領域抽出（精密抽出）
	画像
	映像

課題を組み合わせに落とし込む

画像分野での代表的な問題

検出 四角で検出
領域分割 精密に抽出
分類 ２クラス／多クラス
推定 回帰
画質改善 ノイズ除去，被ばく低減
画像生成 フェイク画像生成
データ抽出 異常検知（教師なし学習）

データと正解があれば
自動で学習

じゃどのくらい？

JACC: CARDIOVASCULAR IMAGING VOL. 13, NO. 2, 2020

PUBLISHED BY ELSEVIER

ORIGINAL RESEARCH

A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images

Kenya Kusunose, MD, PhD,[a] Takashi Abe, MD, PhD,[b] Akihiro Haga, PhD,[c] Daiju Fukuda, MD, PhD,[a] Hirotsugu Yamada, MD, PhD,[a] Masafumi Harada, MD, PhD,[b] Masataka Sata, MD, PhD[a]

ABSTRACT

OBJECTIVES This study investigated whether a deep convolutional neural network (DCNN) could provide improved detection of regional wall motion abnormalities (RWMAs) and differentiate among groups of coronary infarction territories from conventional 2-dimensional echocardiographic images compared with that of cardiologists, sonographers, and resident readers.

BACKGROUND An effective intervention for reduction of misreading of RWMAs is needed. The hypothesis was that a DCNN trained using echocardiographic images would provide improved detection of RWMAs in the clinical setting.

METHODS A total of 300 patients with a history of myocardial infarction were enrolled. From this cohort, 3 groups of 100 patients each had infarctions of the left anterior descending (LAD) artery, the left circumflex (LCX) branch, and the right coronary artery (RCA). A total of 100 age-matched control patients with normal wall motion were selected from a database. Each case contained cardiac ultrasonographs from short-axis views at end-diastolic, mid-systolic, and end-systolic phases. After the DCNN underwent 100 steps of training, diagnostic accuracies were calculated from the test set. Independently, 10 versions of the same model were trained, and ensemble predictions were performed using those versions.

x

超音波画像３枚

y

局所壁運動

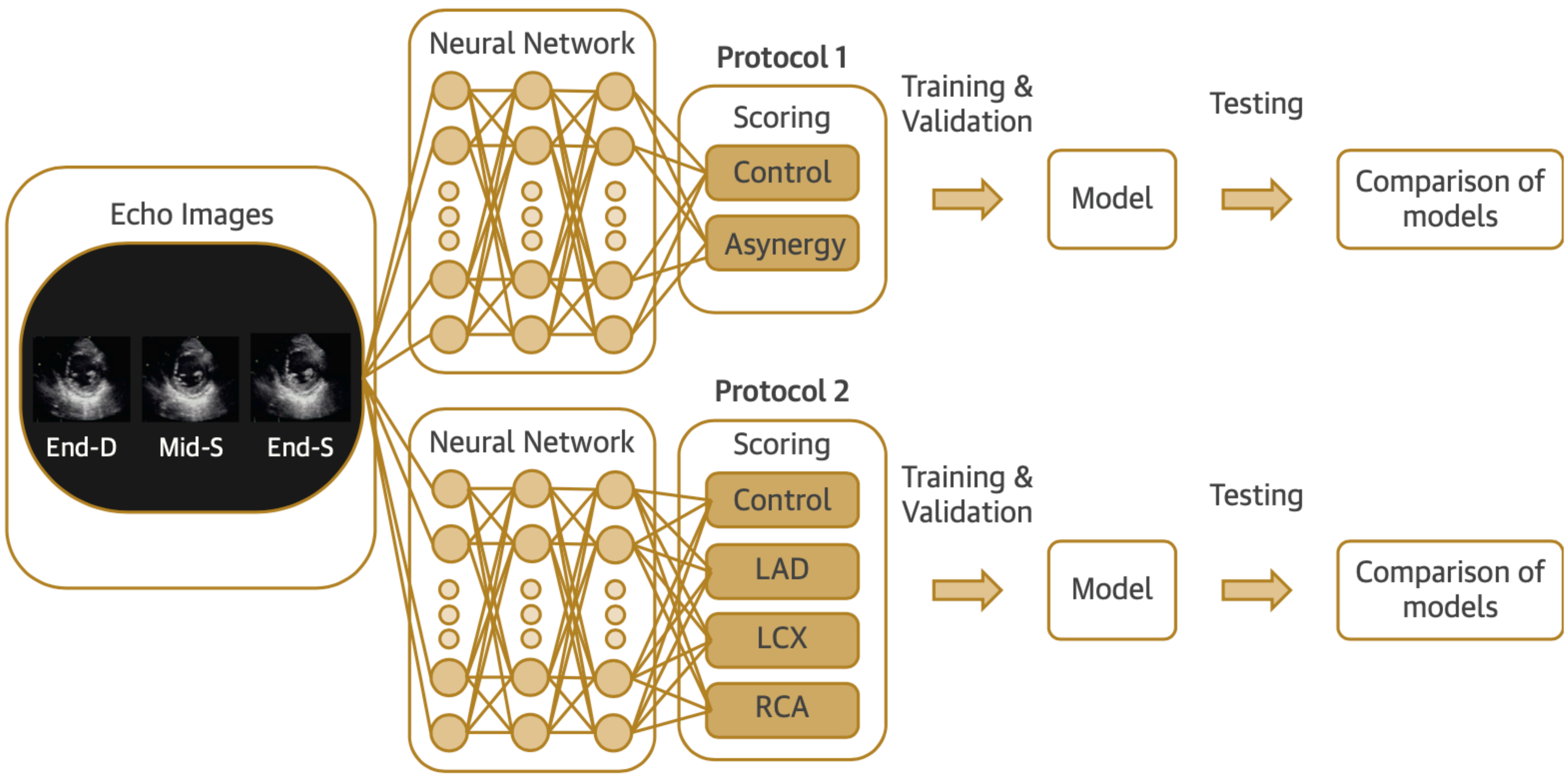
異常の有無

（２分類）

$n = 300$

Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, Sata M. A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images. JACC Cardiovasc Imaging. 2020 Feb;13(2 Pt 1):374-381. doi: 10.1016/j.jcmg.2019.02.024. Epub 2019 May 15. PMID: 31103590.

CENTRAL ILLUSTRATION Neural Networks for the Presence of RWMAs and the Territory of RWMAs

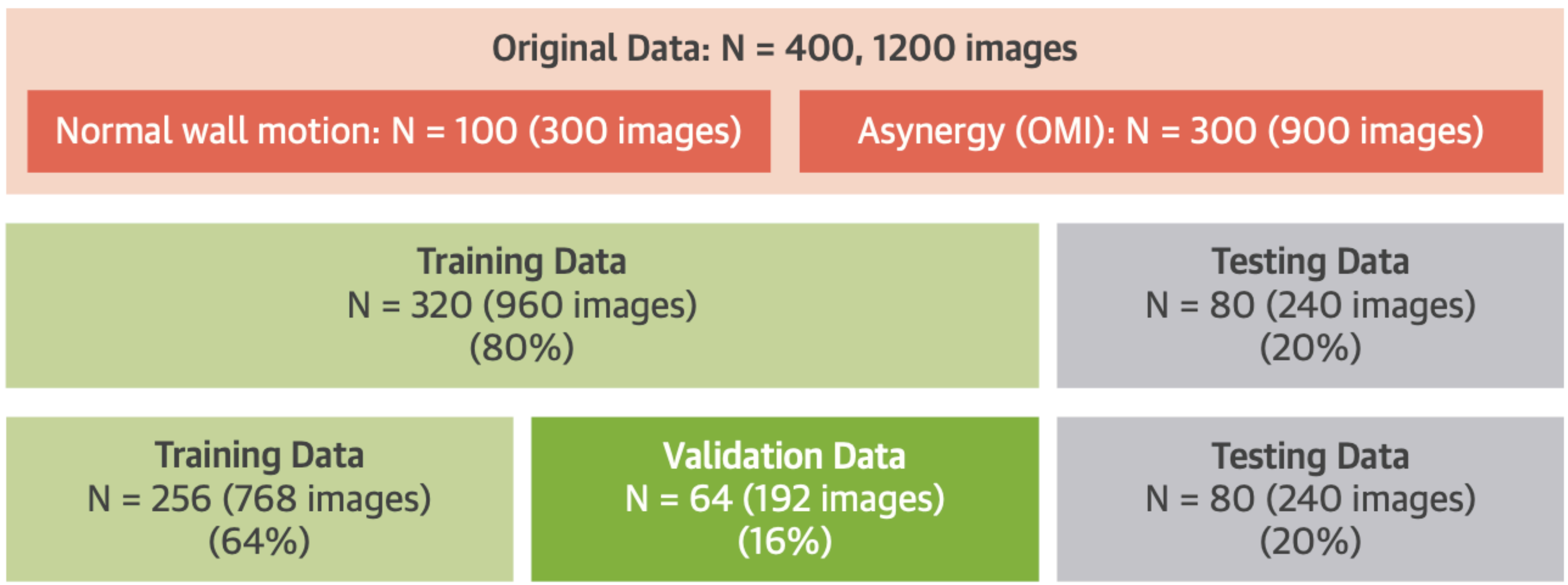


Kusunose, K. et al. J Am Coll Cardiol Img. 2020;13(2):374-81.

The fully connected layers transform the image features into the final scores by adjusting weights for neuron activations during training. Echo = echocardiography; End-D = end diastolic (phase); LAD = left anterior descending artery; LCX = left circumflex artery; Mid/End-S = mid-/end systolic (phases); RCA = right coronary artery; RWMA = regional wall motion abnormality.

Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, Sata M. A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images. JACC Cardiovasc Imaging. 2020 Feb;13(2 Pt 1):374-381.
doi: 10.1016/j.jcmg.2019.02.024. Epub 2019 May 15. PMID: 31103590.

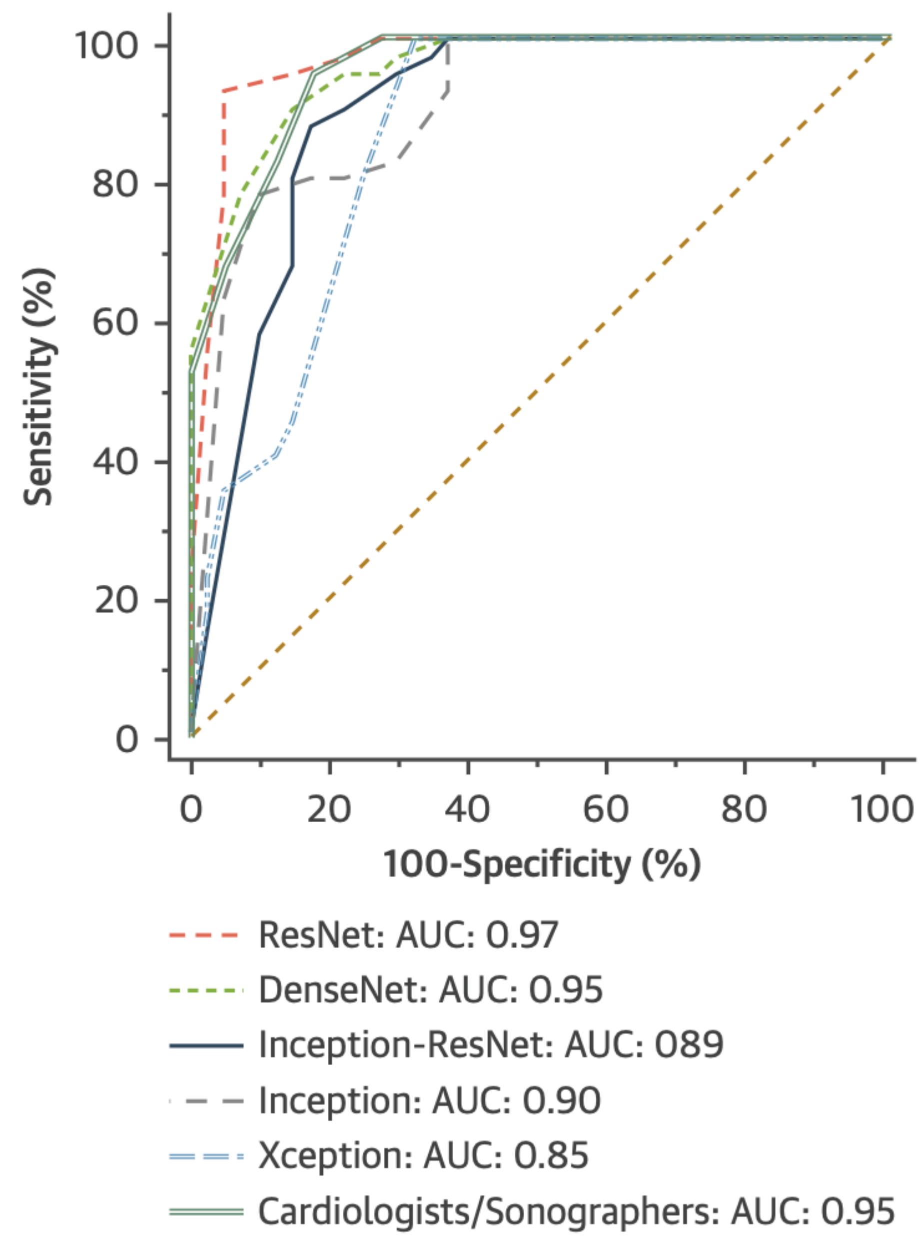
FIGURE 1 **Import Data**



There were a total of 400 cases from which 1,200 images were split with 256 cases (786 images) as the training set, 64 cases (192 images) as the validation set, and 80 cases (240 images) as the test set. OMI = old myocardial infarction.

Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, Sata M. A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images. JACC Cardiovasc Imaging. 2020 Feb;13(2 Pt 1):374-381.
doi: 10.1016/j.jcmg.2019.02.024. Epub 2019 May 15. PMID: 31103590.

FIGURE 3 **Diagnostic Ability to Detect the Presence of RWMAs**



The area under the curves by several deep learning algorithms for detection of territories of wall motion abnormality were good. AUC = area under the curve; ResNet = residual neural network; other abbreviation as in **Figure 2**.

Kusunose K, Abe T, Haga A, Fukuda D, Yamada H, Harada M, Sata M. A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images. JACC Cardiovasc Imaging. 2020 Feb;13(2 Pt 1):374-381. doi: 10.1016/j.jcmg.2019.02.024. Epub 2019 May 15. PMID: 31103590.

Circulation

ORIGINAL RESEARCH ARTICLE

Fully Automated Echocardiogram Interpretation in Clinical Practice

Feasibility and Diagnostic Accuracy

Editorials, see p 1636 and p 1639

BACKGROUND: Automated cardiac image interpretation has the potential to transform clinical practice in multiple ways, including enabling serial assessment of cardiac function by nonexperts in primary care and rural settings. We hypothesized that advances in computer vision could enable building a fully automated, scalable analysis pipeline for echocardiogram interpretation, including (1) view identification, (2) image segmentation, (3) quantification of structure and function, and (4) disease detection.

METHODS: Using 14035 echocardiograms spanning a 10-year period, we trained and evaluated convolutional neural network models for multiple tasks, including automated identification of 23 viewpoints and segmentation of cardiac chambers across 5 common views. The segmentation output was used to quantify chamber volumes and left ventricular mass, determine ejection fraction, and facilitate automated determination of longitudinal strain through speckle tracking. Results were evaluated through comparison to manual segmentation and measurements from 8666 echocardiograms obtained during the routine clinical workflow. Finally, we developed models to detect 3 diseases: hypertrophic cardiomyopathy, cardiac amyloid, and pulmonary arterial hypertension.

Jeffrey Zhang, BA
Sravani Gajjala, MBBS
Pulkit Agrawal, PhD
Geoffrey H. Tison, MD, MPH
Laura A. Hallock, BS
Lauren Beussink-Nelson, RDCS
Mats H. Lassen, BM
Eugene Fan, MD
Mandar A. Aras, MD, PhD
ChaRandle Jordan, MD, PhD
Kirsten E. Fleischmann, MD, MPH
Michelle Melisko, MD
Atif Qasim, MD, MSCE
Sanjiv J. Shah, MD
Ruzena Bajcsy, PhD
Rahul C. Deo, MD, PhD

x

超音波画像 1 枚

y

分割画像

多クラス分類

$n = 14035$

Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. 2018 Oct 16;138(16):1623-1635. doi: 10.1161/CIRCULATIONAHA.118.034338. PMID: 30354459; PMCID: PMC6200386.

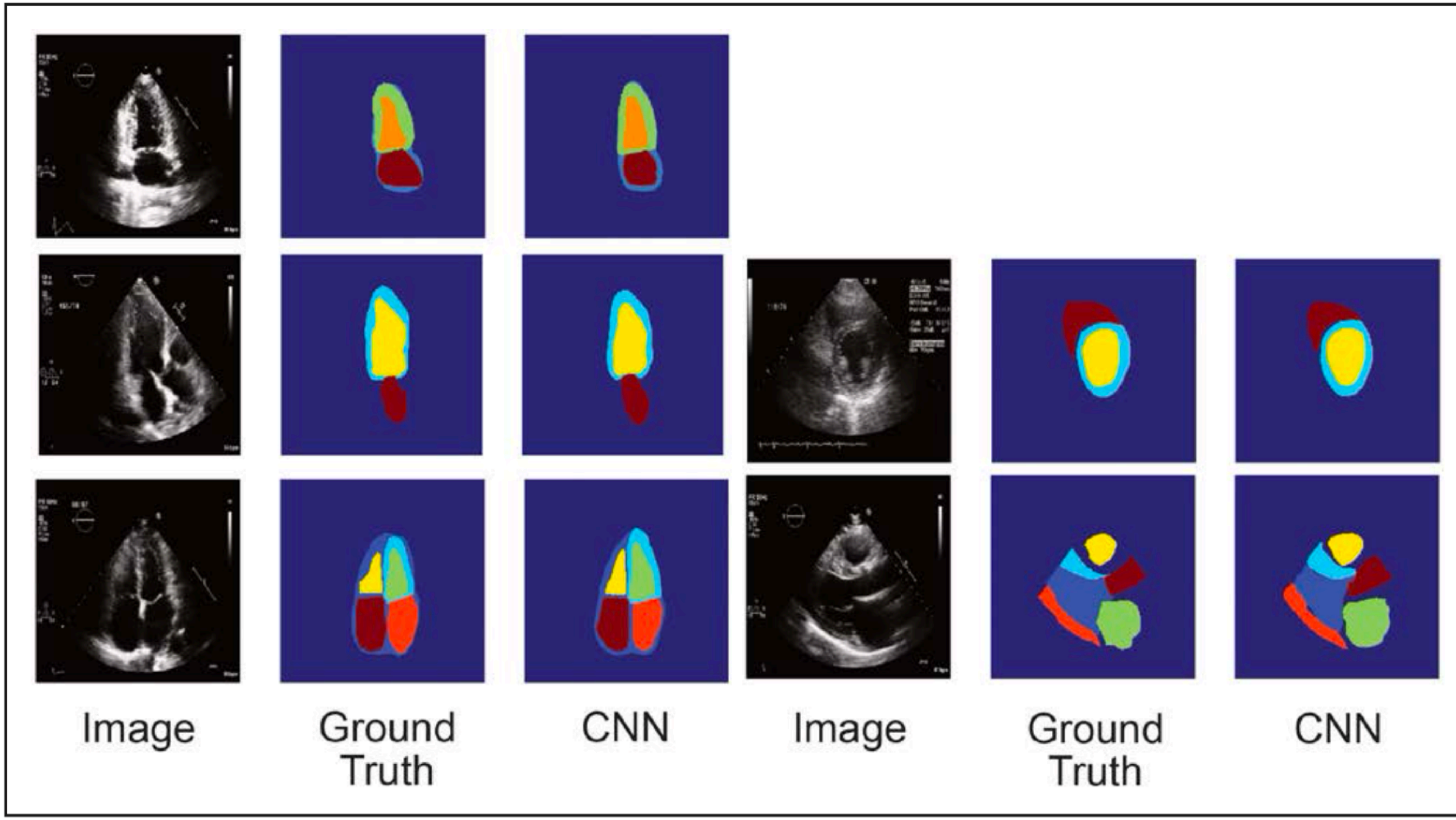


Figure 3. Convolutional neural networks successfully segment cardiac chambers.
We used the U-net algorithm to derive segmentation models for 5 views: A2c, A3c, A4c (left: top, middle, and bottom, respectively), parasternal short axis at the level of the papillary muscle (right, middle), and PLAX (right, bottom). For each view, the trio of images, from left to right, corresponds to the original image, the manually traced image used in training (ground truth), and the automated segmented image (determined as part of the cross-validation process). A2c indicates apical 2-chamber; A3c, apical 3-chamber; A4c, apical 4-chamber; CNN, convolutional neural network; and PLAX, parasternal long axis.

Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. 2018 Oct 16;138(16):1623-1635. doi: 10.1161/CIRCULATIONAHA.118.034338. PMID: 30354459; PMCID: PMC6200386.

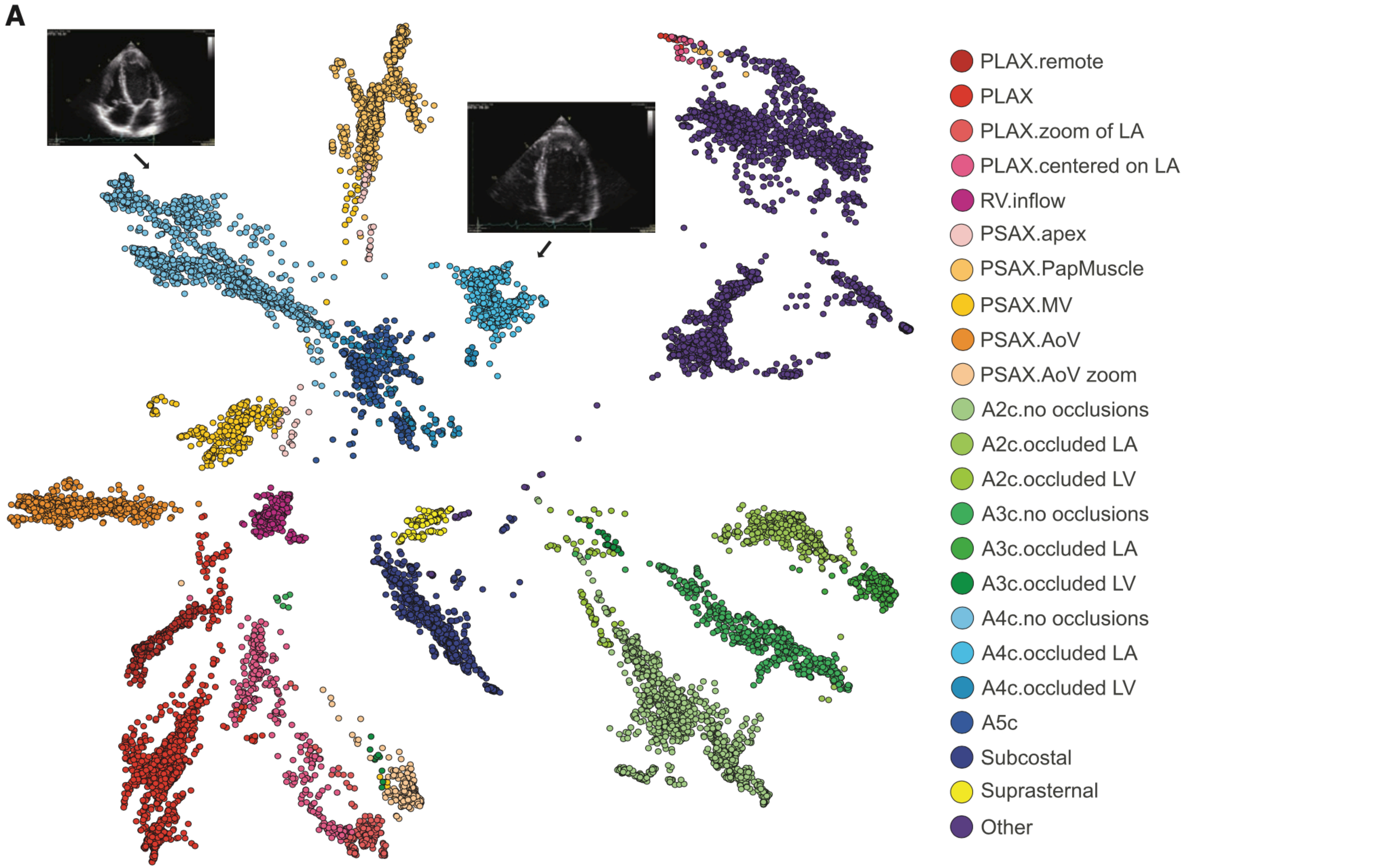


Figure 2. Convolutional neural networks successfully discriminate echocardiographic views.
A, t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization of view classification. t-SNE is an algorithm used to visualize high-dimensional data in lower dimensions. It depicts the successful grouping of test images corresponding to 23 different echocardiographic views. Echocardiographic still images indicate the distinct clustering of images of A4c views without occlusions and those with occlusion of the left atrium. **B**, Confusion matrix demonstrating successful and unsuccessful view classifications within the test data set. Numbers along the diagonal represent successful classifications, whereas off-diagonal entries are misclassifications. A2c indicates apical 2-chamber; A3c, apical 3-chamber; A4c, apical 4-chamber; echo, echocardiogram; LV, left ventricular; and PLAX, parasternal long axis.

Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. 2018 Oct 16;138(16):1623-1635. doi: 10.1161/CIRCULATIONAHA.118.034338. PMID: 30354459; PMCID: PMC6200386.

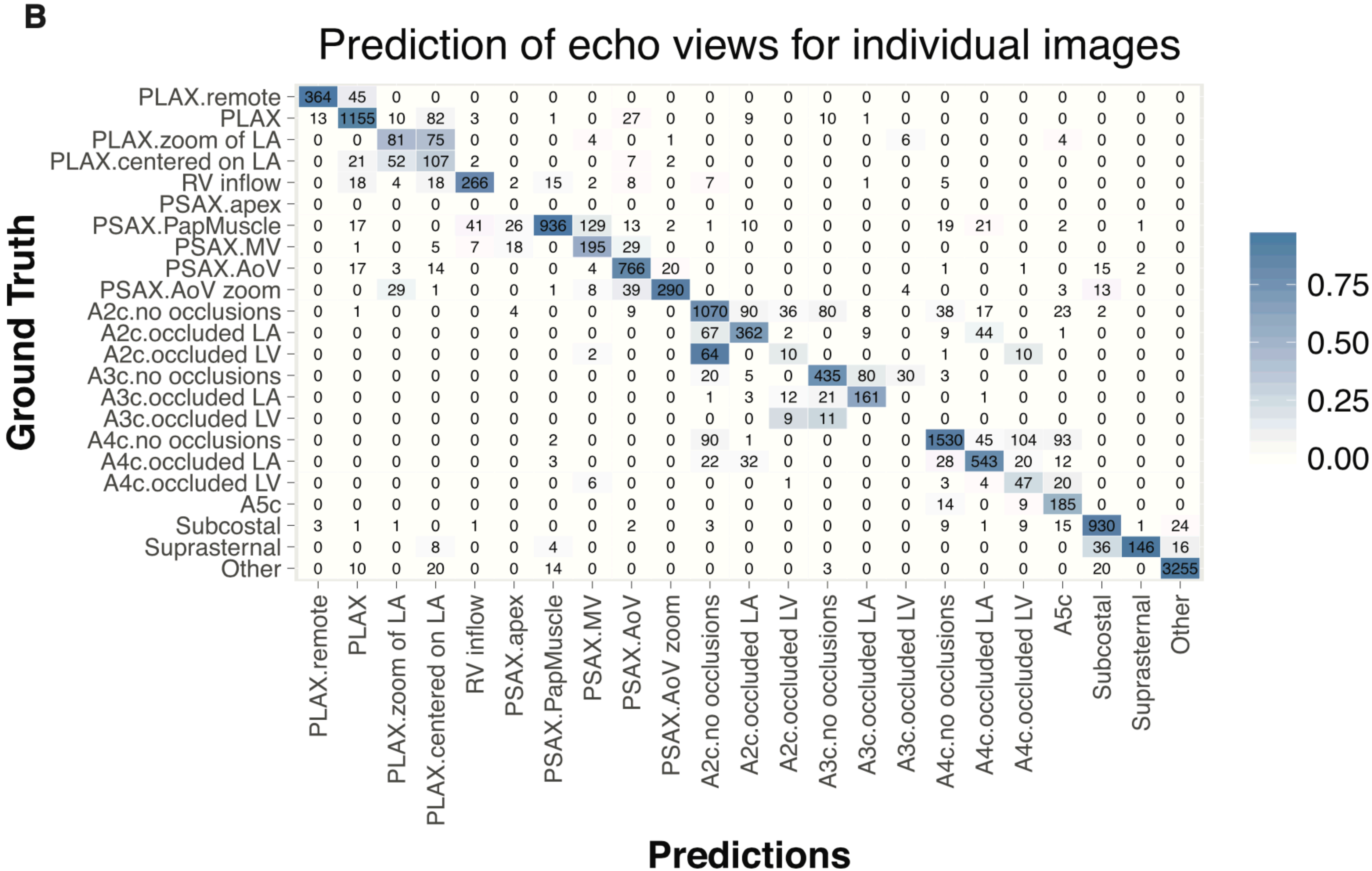


Figure 2. Convolutional neural networks successfully discriminate echocardiographic views.
A, t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization of view classification. t-SNE is an algorithm used to visualize high-dimensional data in lower dimensions. It depicts the successful grouping of test images corresponding to 23 different echocardiographic views. Echocardiographic still images indicate the distinct clustering of images of A4c views without occlusions and those with occlusion of the left atrium. **B**, Confusion matrix demonstrating successful and unsuccessful view classifications within the test data set. Numbers along the diagonal represent successful classifications, whereas off-diagonal entries are misclassifications. A2c indicates apical 2-chamber; A3c, apical 3-chamber; A4c, apical 4-chamber; echo, echocardiogram; LV, left ventricular; and PLAX, parasternal long axis.

Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH, Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo RC. Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. 2018 Oct 16;138(16):1623-1635. doi: 10.1161/CIRCULATIONAHA.118.034338. PMID: 30354459; PMCID: PMC6200386.

European Journal of Radiology 139 (2021) 109717

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Artificial intelligence in ultrasound

Yu-Ting Shen [a], Liang Chen [b], Wen-Wen Yue [a,*], Hui-Xiong Xu [a,*]

[a] Department of Medical Ultrasound, Shanghai Tenth People's Hospital, Ultrasound Research and Education Institute, Tongji University School of Medicine, Tongji University Cancer Center, Shanghai Engineering Research Center of Ultrasound Diagnosis and Treatment, National Clnical Research Center of Interventional Medicine, Shanghai, 200072, PR China

[b] Department of Gastroenterology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, 200072, PR China

ARTICLE INFO

Keywords:
Ultrasound
Artificial intelligence
Deep learning
Medical imaging

ABSTRACT

Ultrasound (US), a flexible green imaging modality, is expanding globally as a first-line imaging technique in various clinical fields following with the continual emergence of advanced ultrasonic technologies and the well-established US-based digital health system. Actually, in US practice, qualified physicians should manually collect and visually evaluate images for the detection, identification and monitoring of diseases. The diagnostic performance is inevitably reduced due to the intrinsic property of high operator-dependence from US. In contrast, artificial intelligence (AI) excels at automatically recognizing complex patterns and providing quantitative assessment for imaging data, showing high potential to assist physicians in acquiring more accurate and reproducible results. In this article, we will provide a general understanding of AI, machine learning (ML) and deep learning (DL) technologies; We then review the rapidly growing applications of AI-especially DL technology in the field of US-based on the following anatomical regions: thyroid, breast, abdomen and pelvis, obstetrics heart and blood vessels, musculoskeletal system and other organs by covering image quality control, anatomy localization, object detection, lesion segmentation, and computer-aided diagnosis and prognosis evaluation; Finally, we offer our perspective on the challenges and opportunities for the clinical practice of biomedical AI systems in US.

Shen YT, Chen L, Yue WW, Xu HX. Artificial intelligence in ultrasound. Eur J Radiol. 2021;139:109717. Epub 2021/05/08. doi: 10.1016/j.ejrad.2021.109717. PubMed PMID: 33962110.

European Journal of Radiology 139 (2021) 109717

Contents lists available at ScienceDirect

European Journal of Radiology

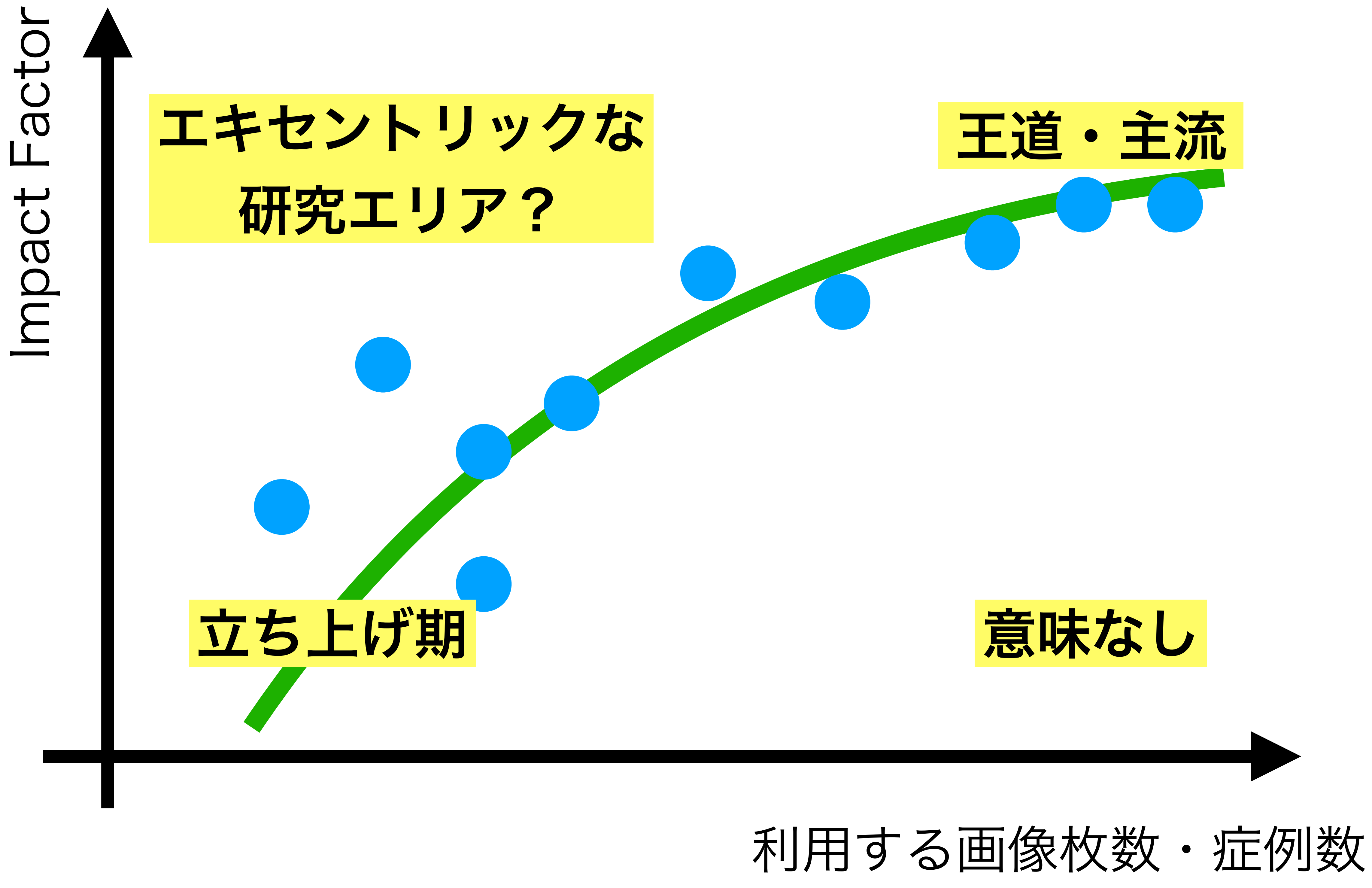
journal homepage: www.elsevier.com/locate/ejrad

Table 4
Summarize the machine learning applications in heart and blood vessels and musculoskeletal system as well as other organ systems ultrasonic image analysis in the papers surveyed.

Organ system and body location	Disease classification	Object detection	Image segmentation	Prognosis evaluation
Heart	View classification of echocardiograms [99,100] Classification of myocardial wall motion [101,102] Diagnosis of ventricular volume [161]	Detection of heart disease [98,103,104]	Segmentation of the ventricle of the heart [105] Segmentation of left ventricle and left atrium [106]	
Blood vessels	Classification of carotid artery intima-media thickness [107] Characterization of plaque composition in vascular [108]	Detection of vascular lumen [109]	Segmentation of lumen-intima and media-adventitia [110,111,112] Segmentation of vascular structure [113]	
Musculoskeletal system	Diagnosis of myositis from muscle US [114] Estimation of skeletal muscle status [115]	US-assisted vertebral body positioning [116] Detection and identification of spine level [117]	Segmentation of rectus femoris muscle [118] Segmentation of puborectalis muscle and urogenital hiatus [119]	
Other organ systems or body location	Classification of pediatric pneumonia [120] Assessment and diagnosis of lung US [122,123]	Improve US imaging contrast and detection rate [124]	Segmentation of subpleural pulmonary lesions [121]	

Note. US: ultrasound.

Shen YT, Chen L, Yue WW, Xu HX. Artificial intelligence in ultrasound. Eur J Radiol. 2021;139:109717. Epub 2021/05/08. doi: 10.1016/j.ejrad.2021.109717. PubMed PMID: 33962110.

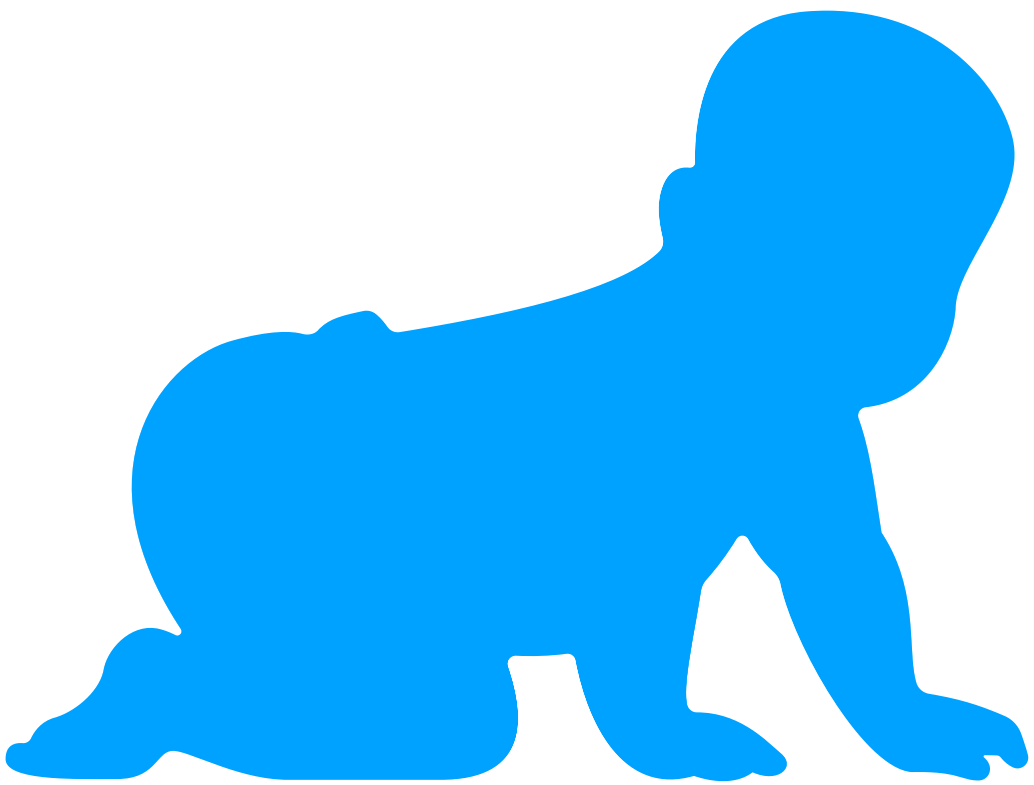


Impact Factor
エキセントリックな
研究エリア？
王道・主流
立ち上げ期
意味なし
利用する画像枚数・症例数

教師なし学習
異常検知

いつも吉野家

たまに 「すき家



いつも吉野家

吉野家好き

吉野家が甘さ 吉野家を

入力　醤油　想像

お酒

つゆ

たまねぎ

一致していると大満足！

吉野家が $\vec{a}$ 吉野家を

x 入力　醤油　想像 $\hat{x}$

たまねぎ

$$E = \|x - \hat{x}\|$$

一致していると大満足！

$$\cancel{x = f(x)}$$

Encode

画像をベクトルに変換して

画素数＝次元数

$$\vec{a} = f(x)$$

画像よりも低い次元数

画像をベクトルに符号化

Decode

変換したベクトルから
再び画像に戻す

$$\hat{x} = g(\vec{a})$$

ベクトルから復元した画像

ベクトルを画像に復号化

その差がゼロに近ければ
ベクトルは画像を効率よく
表現している　と言える

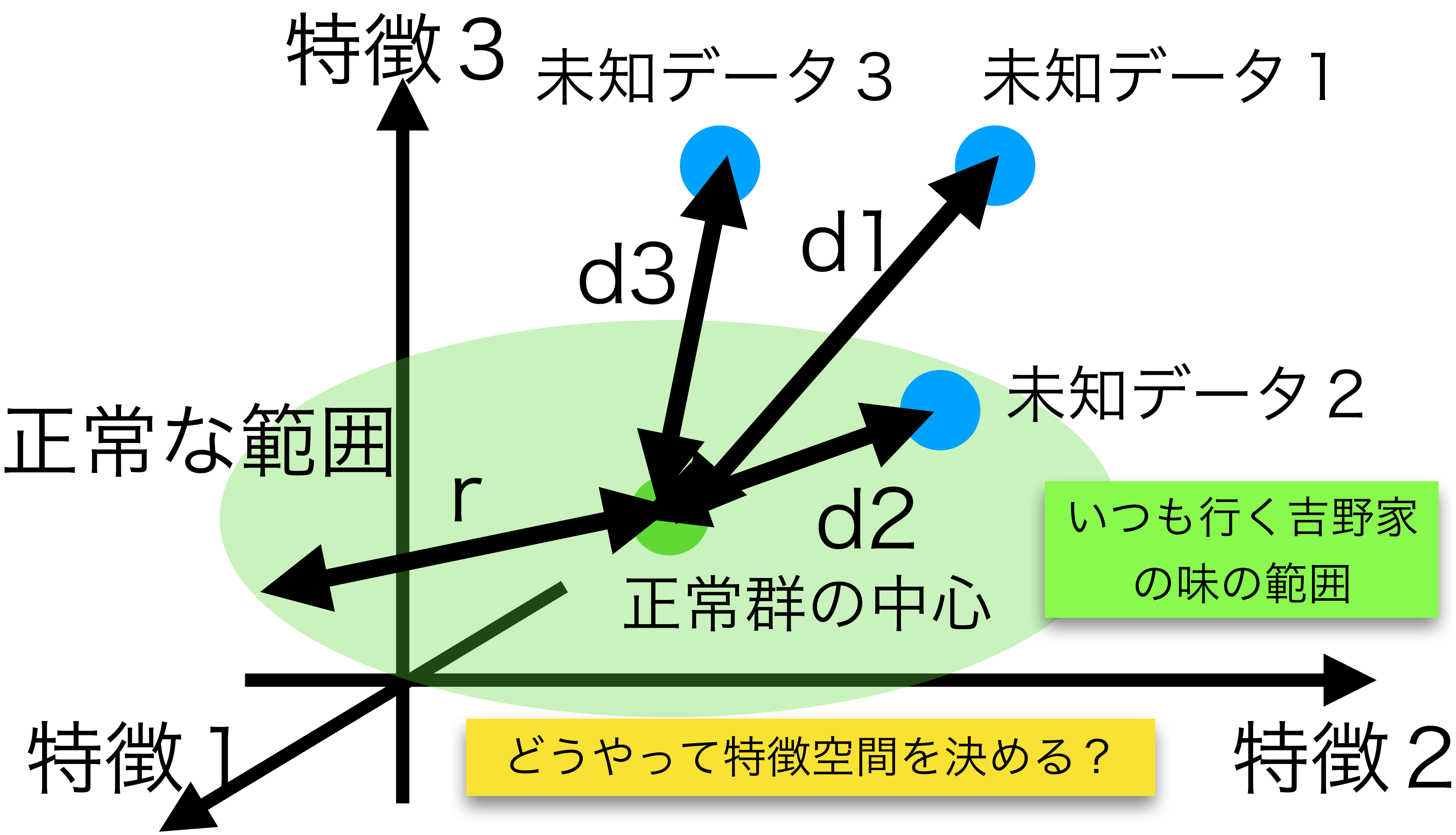
$$E = \|x - \hat{x}\|$$

元画像と復元画像の差

自己複製の自己＝Auto

異常検知の考え方

範囲が決まれば　そこからの逸脱で判定



r<d3<d1なのでデータ１と３は異常である
d2<rなのでデータ２は正常である

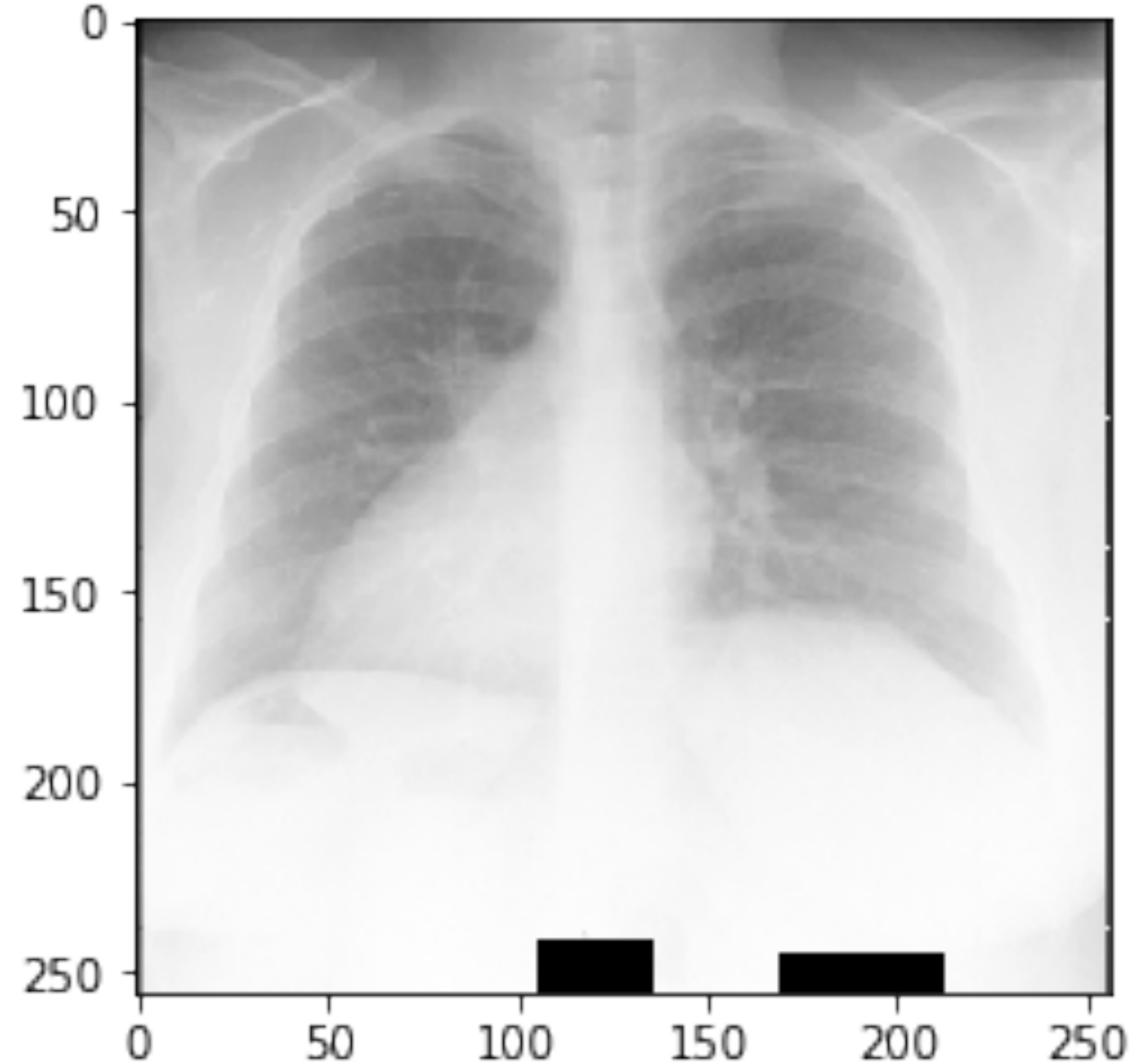
正常群から
中心と範囲を決める

未知データと
正常群の中心の
距離を計算する

その距離が正常な
範囲であるか
判定する


0
50
100
150
200
250
0
50
100
150
200
250

正常像


0
50
100
150
200
250
0
50
100
150
200
250

反転像

65536次元

画像を入力

256x256画素=**65536**

256x256x16

128x128x32

64x64x64

32x32x128

8x8x256

4096次元

4x256=**4096**

8x8x256

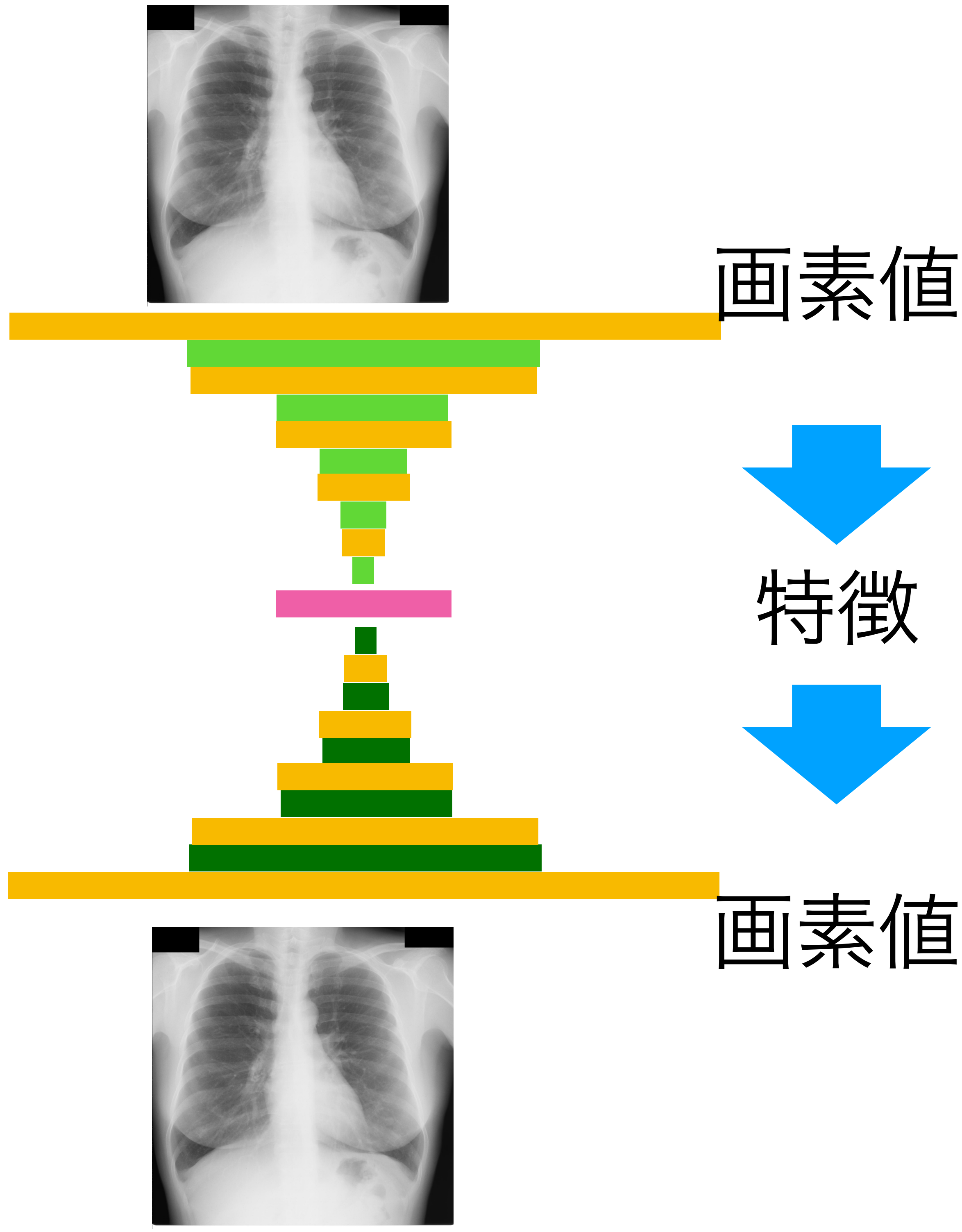
16x16x128

32x32x64

64x64x32

128x128x16

256x256x1



Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 256, 256, 16)	160
max_pooling2d_1 (MaxPooling2	(None, 128, 128, 16)	0
conv2d_2 (Conv2D)	(None, 128, 128, 32)	4640
max_pooling2d_2 (MaxPooling2	(None, 64, 64, 32)	0
conv2d_3 (Conv2D)	(None, 64, 64, 64)	18496
max_pooling2d_3 (MaxPooling2	(None, 32, 32, 64)	0
conv2d_4 (Conv2D)	(None, 32, 32, 128)	73856
max_pooling2d_4 (MaxPooling2	(None, 16, 16, 128)	0
conv2d_5 (Conv2D)	(None, 8, 8, 256)	295168
max_pooling2d_5 (MaxPooling2	(None, 4, 4, 256)	0
flatten_1 (Flatten)	(None, 4096)	0
reshape_1 (Reshape)	(None, 4, 4, 256)	0
conv2d_6 (Conv2D)	(None, 4, 4, 256)	590080
up_sampling2d_1 (UpSampling2	(None, 8, 8, 256)	0
conv2d_7 (Conv2D)	(None, 8, 8, 128)	295040
up_sampling2d_2 (UpSampling2	(None, 16, 16, 128)	0
conv2d_8 (Conv2D)	(None, 16, 16, 64)	73792
up_sampling2d_3 (UpSampling2	(None, 32, 32, 64)	0
conv2d_9 (Conv2D)	(None, 32, 32, 32)	18464
up_sampling2d_4 (UpSampling2	(None, 64, 64, 32)	0
conv2d_10 (Conv2D)	(None, 64, 64, 16)	4624
up_sampling2d_5 (UpSampling2	(None, 128, 128, 16)	0
up_sampling2d_6 (UpSampling2	(None, 256, 256, 16)	0
conv2d_11 (Conv2D)	(None, 256, 256, 1)	145

Total params: 1,374,465
Trainable params: 1,374,465
Non-trainable params: 0

同じ画像を出力

(同じ=自己=Auto)

65536次元

画像を入力

256x256画素=**65536**

256x256x16

128x128x32

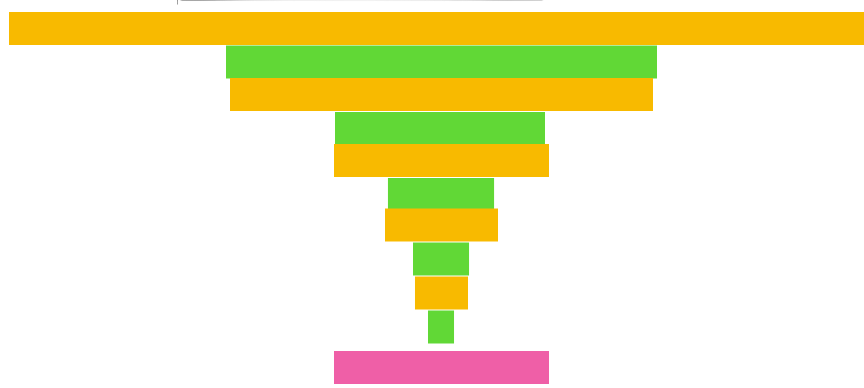
64x64x64

32x32x128

8x8x256

4096次元

4x4x256=**4096**

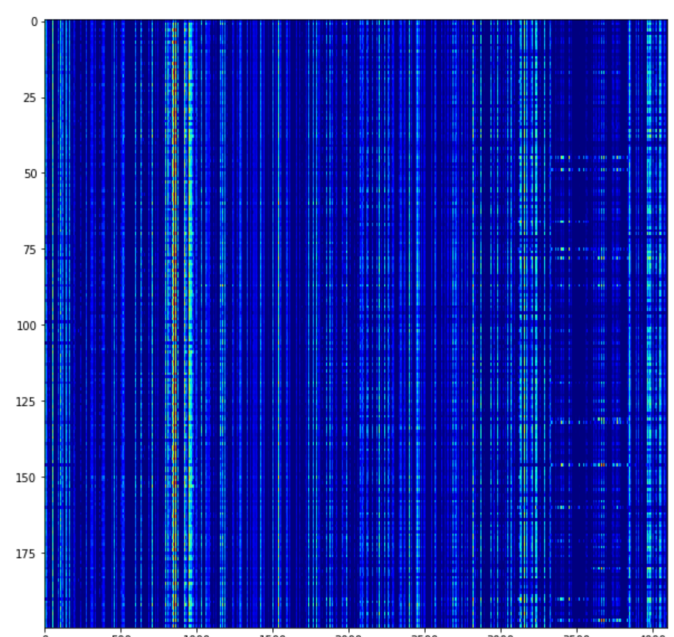


画素値

特徴

Layer (type)	Output Shape	Param #
conv2d_1_input (InputLayer)	(None, 256, 256, 1)	0
conv2d_1 (Conv2D)	(None, 256, 256, 16)	160
max_pooling2d_1 (MaxPooling2	(None, 128, 128, 16)	0
conv2d_2 (Conv2D)	(None, 128, 128, 32)	4640
max_pooling2d_2 (MaxPooling2	(None, 64, 64, 32)	0
conv2d_3 (Conv2D)	(None, 64, 64, 64)	18496
max_pooling2d_3 (MaxPooling2	(None, 32, 32, 64)	0
conv2d_4 (Conv2D)	(None, 32, 32, 128)	73856
max_pooling2d_4 (MaxPooling2	(None, 16, 16, 128)	0
conv2d_5 (Conv2D)	(None, 8, 8, 256)	295168
max_pooling2d_5 (MaxPooling2	(None, 4, 4, 256)	0
flatten_1 (Flatten)	(None, 4096)	0

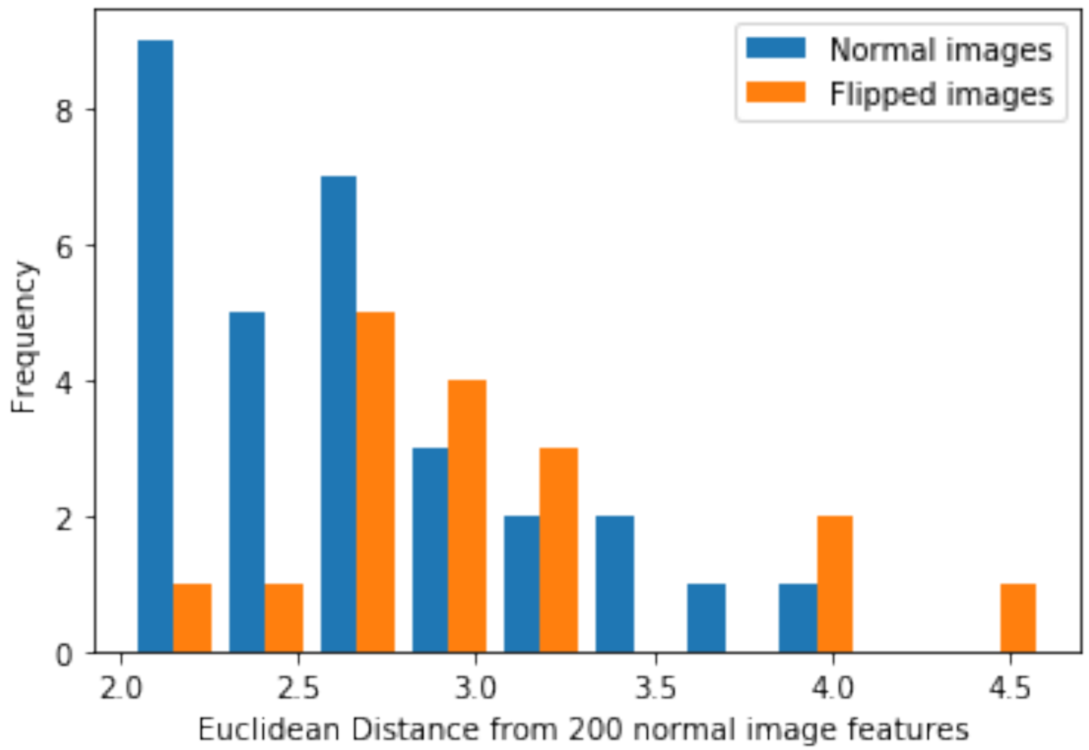
Total params: 392,320
Trainable params: 392,320
Non-trainable params: 0



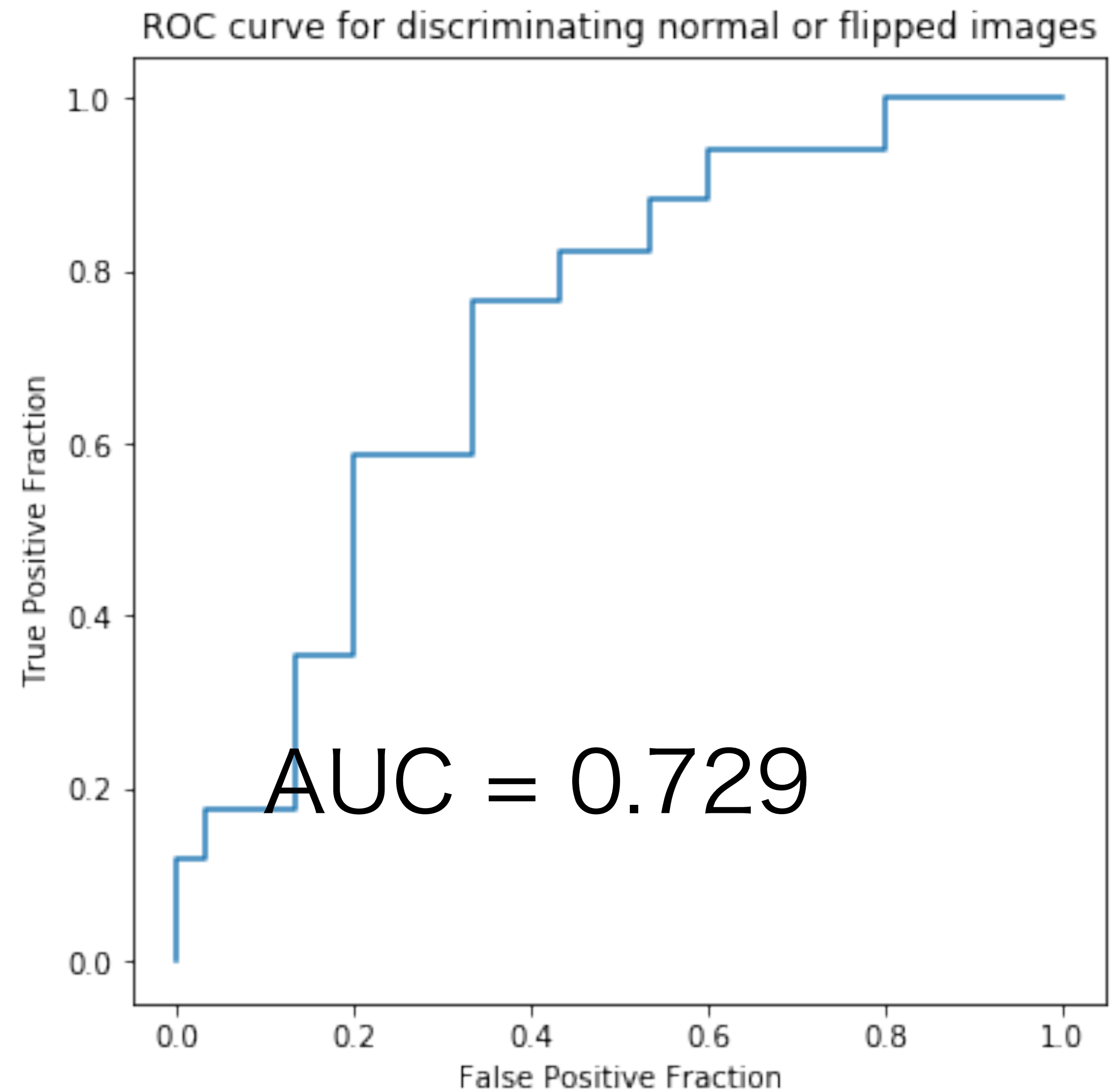
同じ画像を出力

(同じ=自己

=Auto)

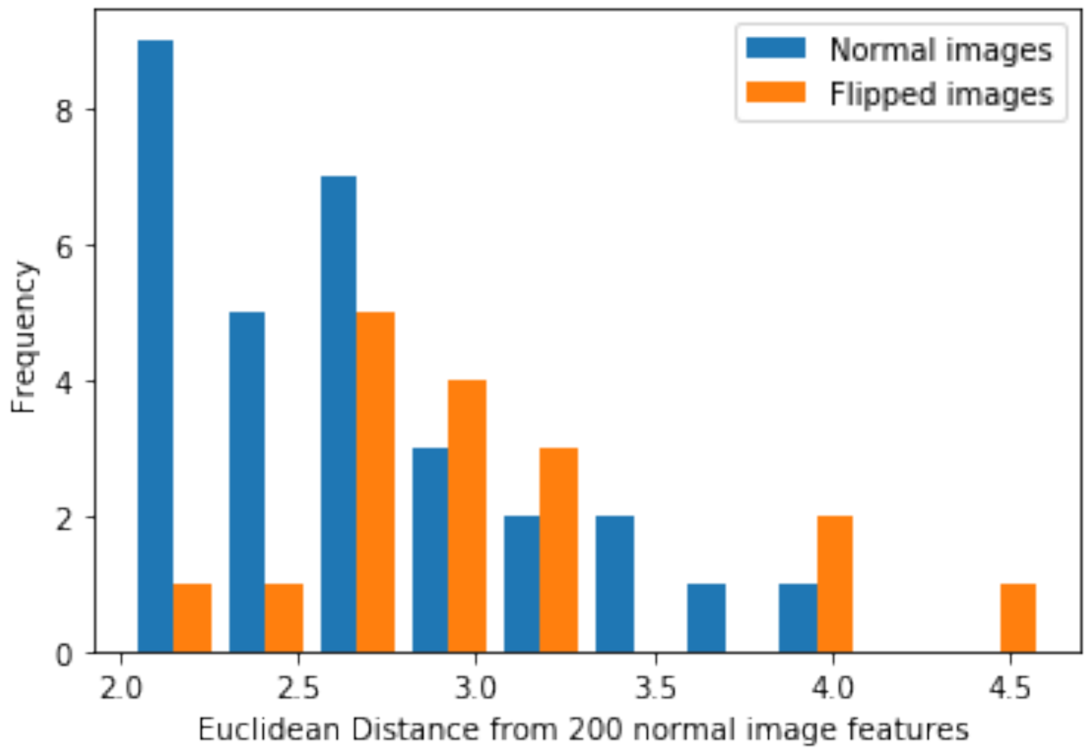


ユークリッド距離のヒストグラム

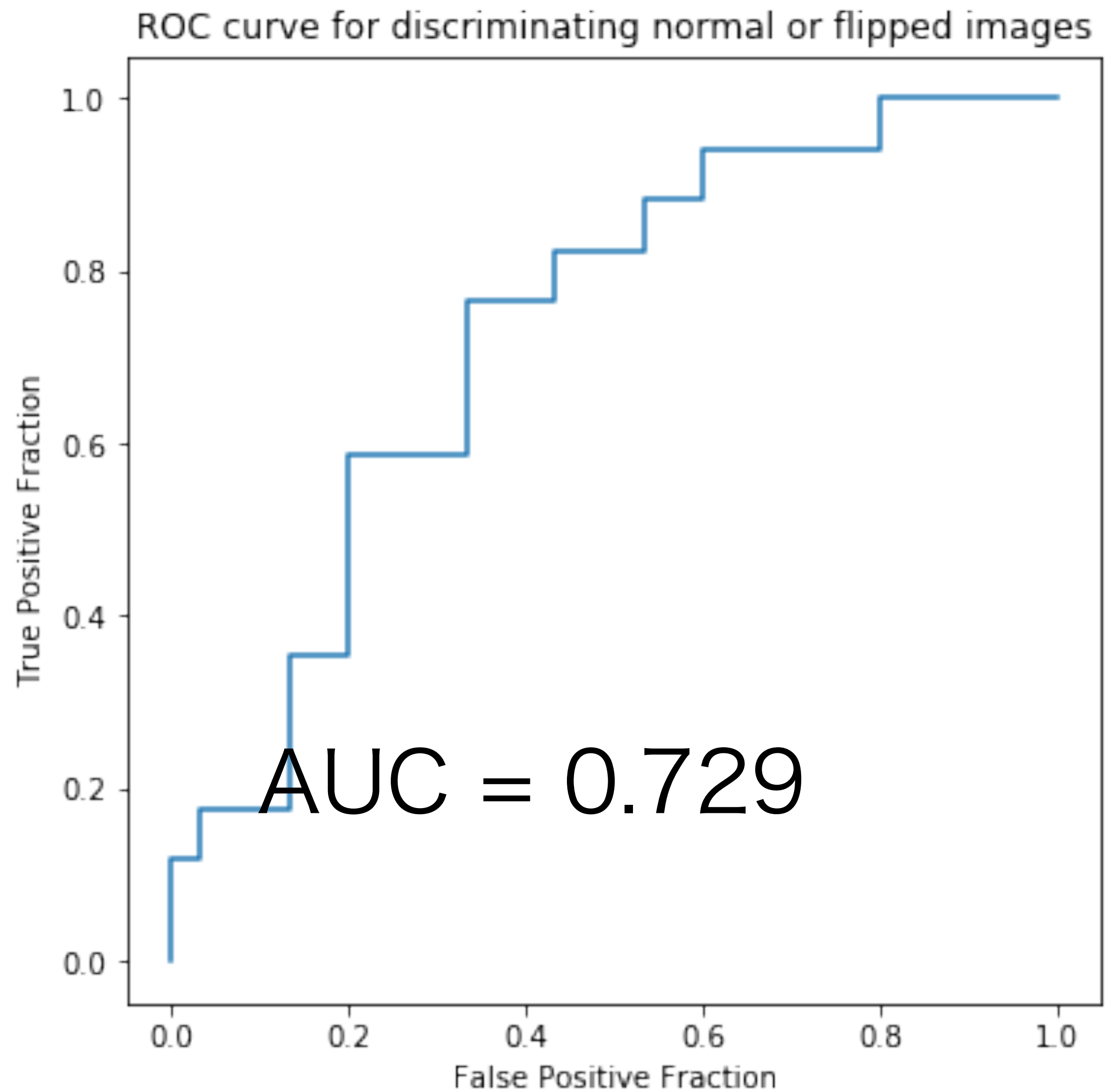


距離を確信度としたときのROC曲線

データだけ集めて手法開発



ユークリッド距離のヒストグラム



距離を確信度としたときのROC曲線

AutoEncoder＆異常検知

教師あり学習

代表例：画像分類

(問題点)
バランスよく学習データを
集める必要がある

10例　　1000例

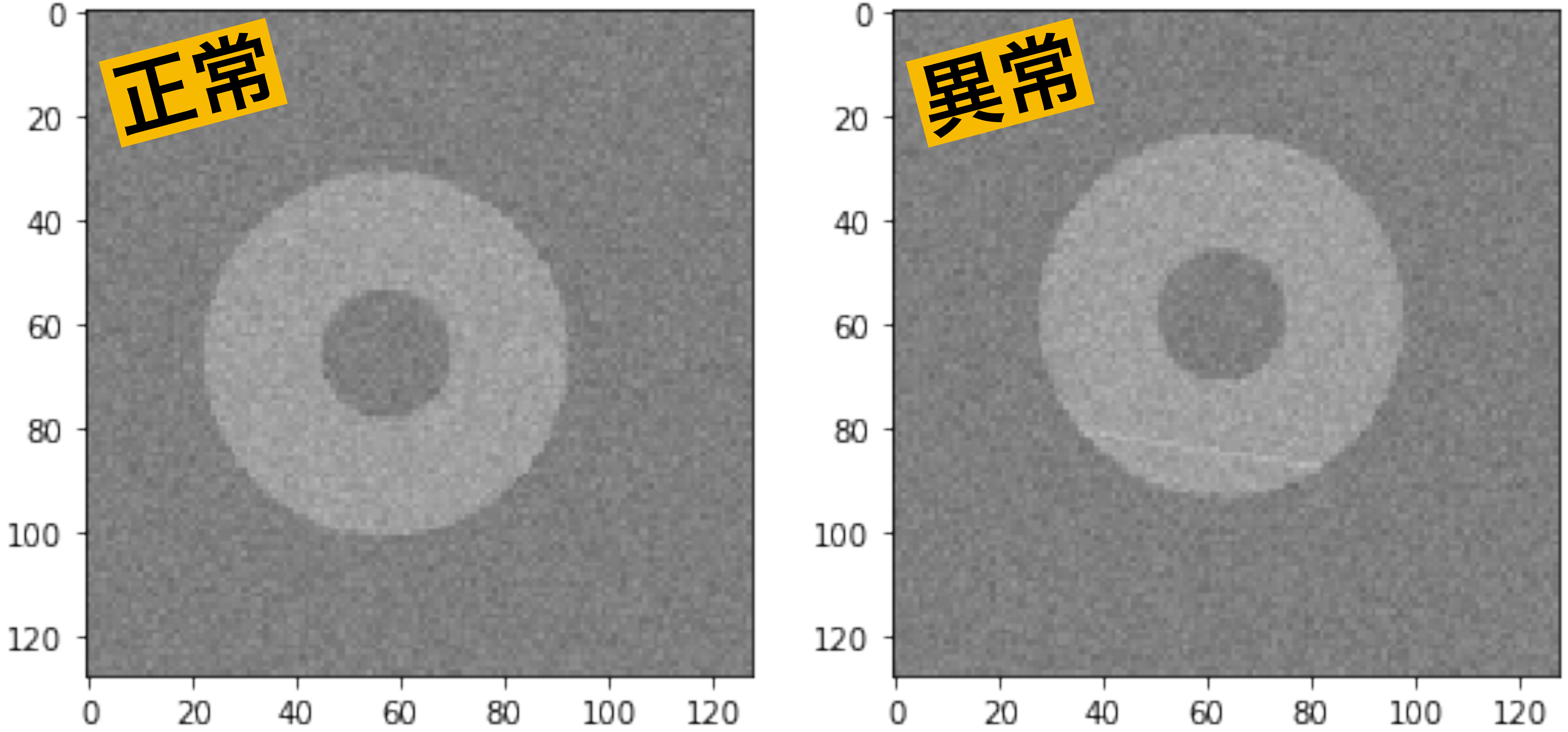
通常：異常例は少なく，正常例は多い
アンバランスが一般的

教師なし学習

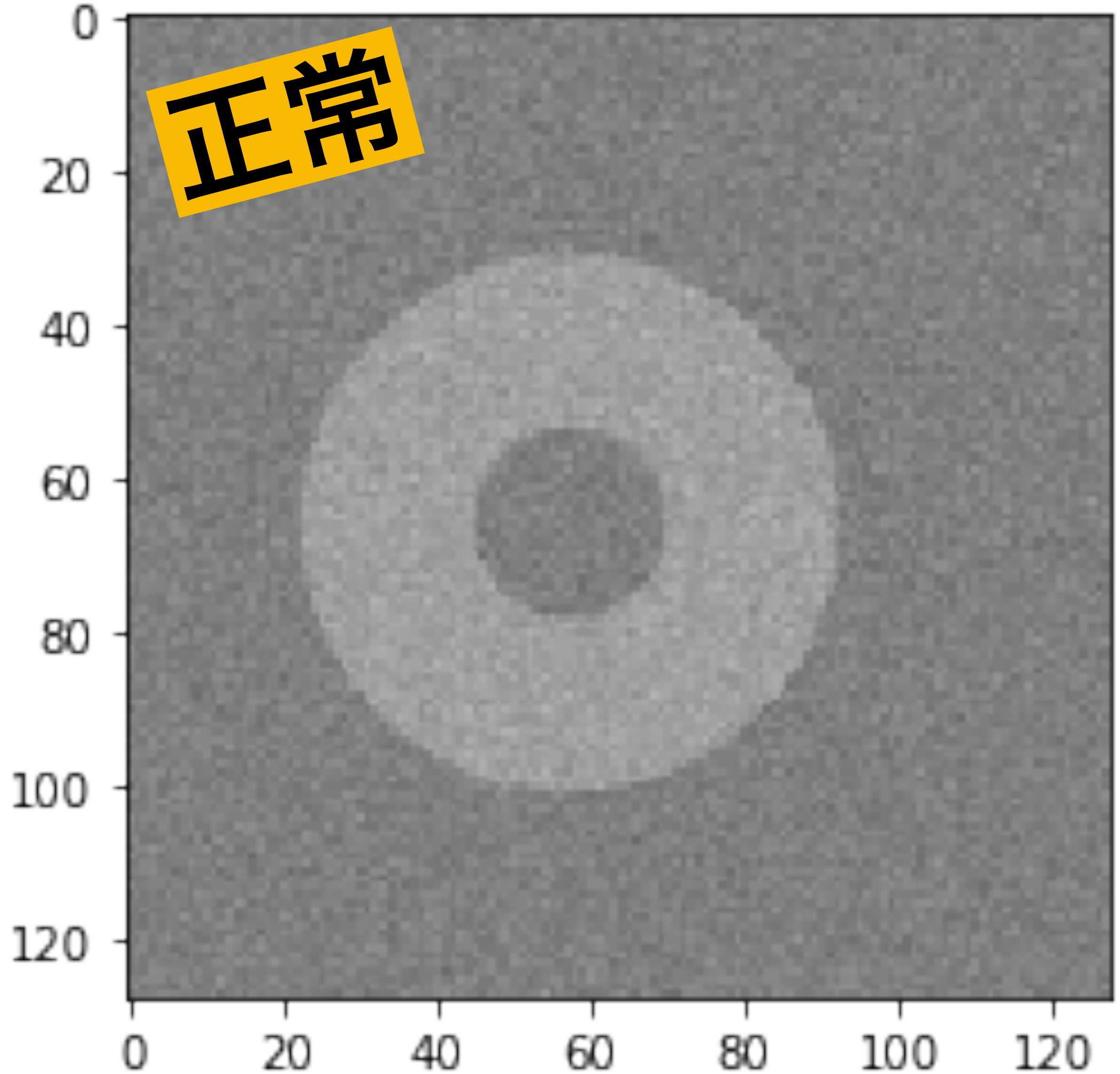
正常なデータのみから特徴抽出
そのからの逸脱を定量化

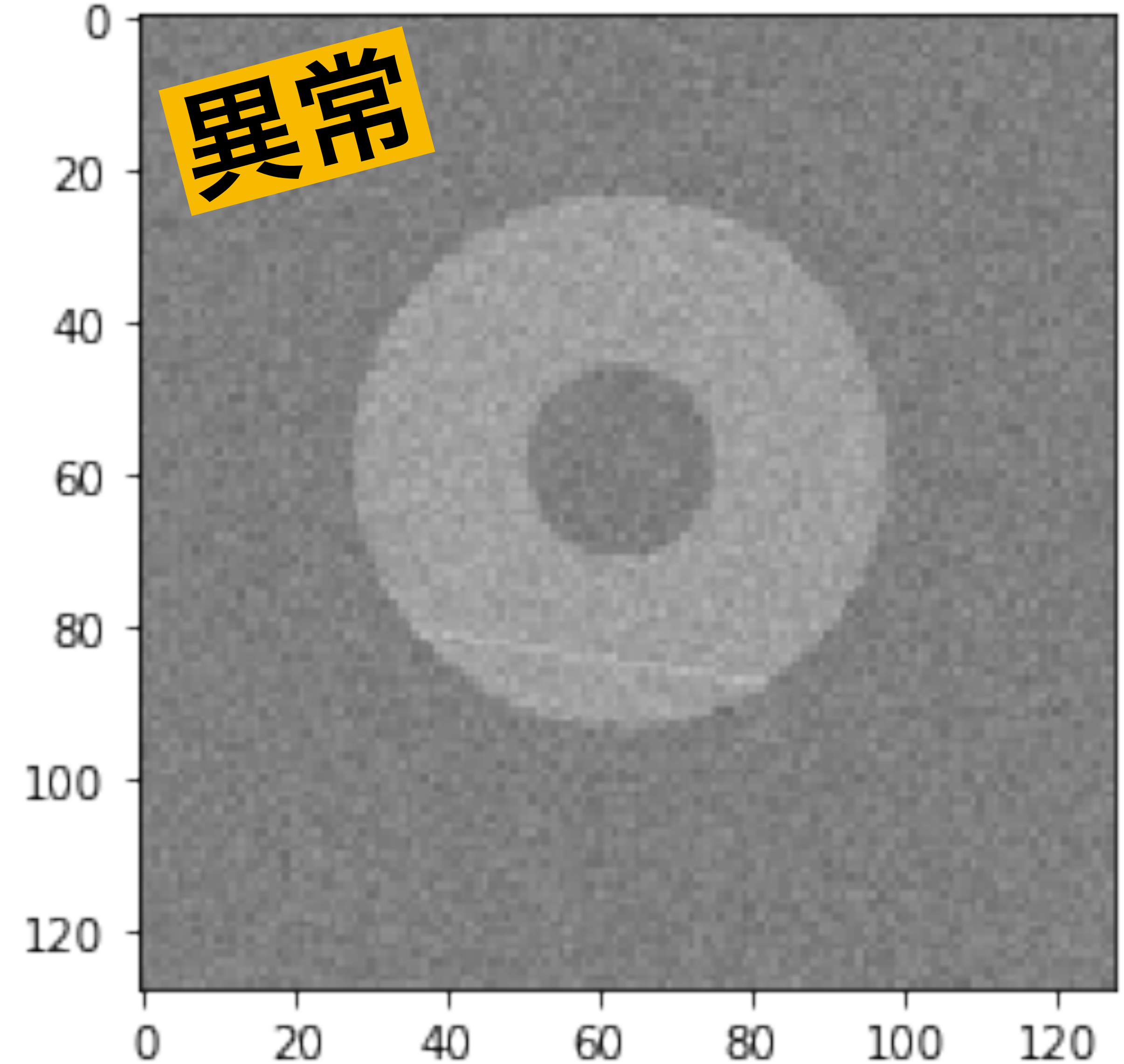
逸脱の大きい場合を
異常として検知

アンバランスなデータでも対応可能

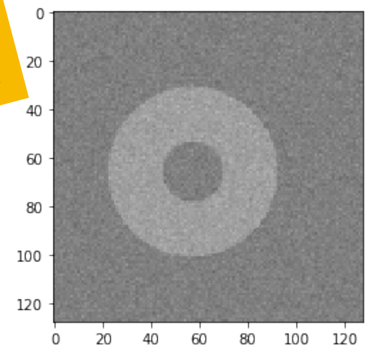

正常
異常
0
20
40
60
80
100
120

データだけ集めて手法開発

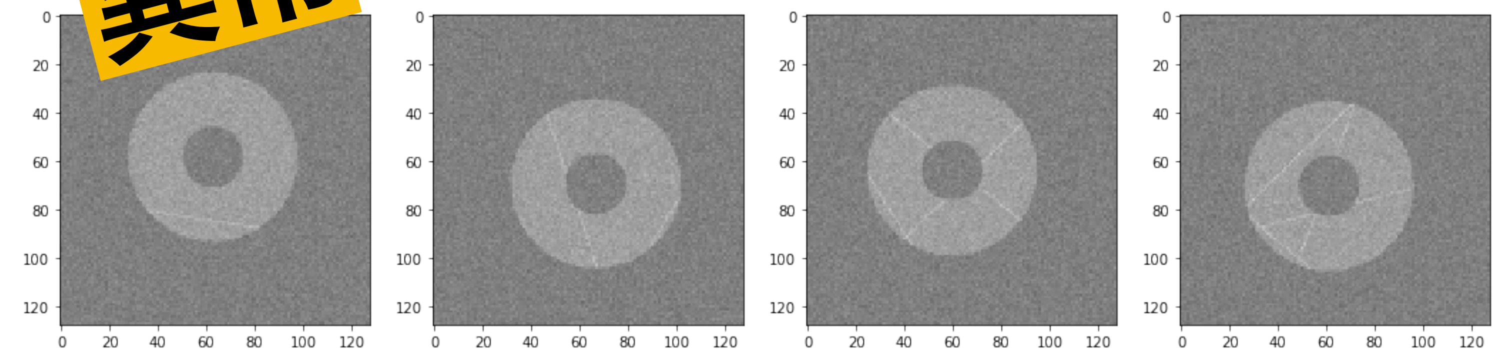


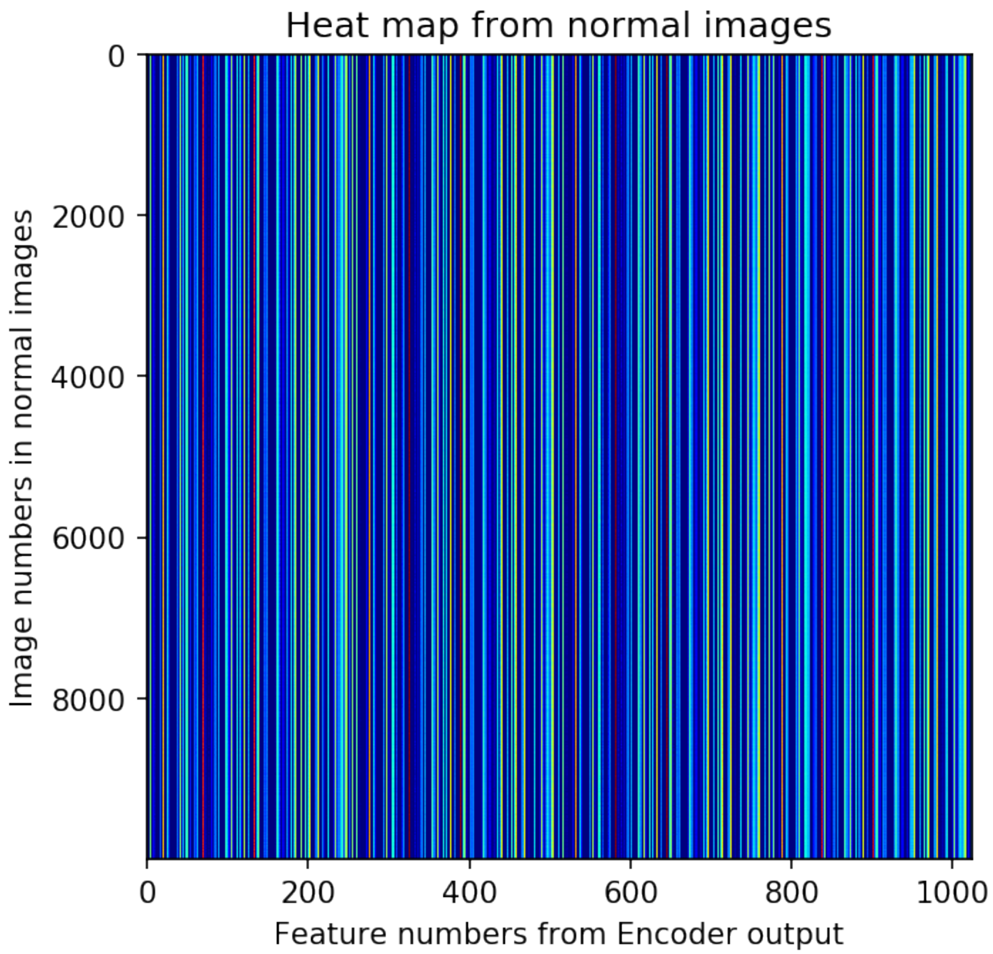


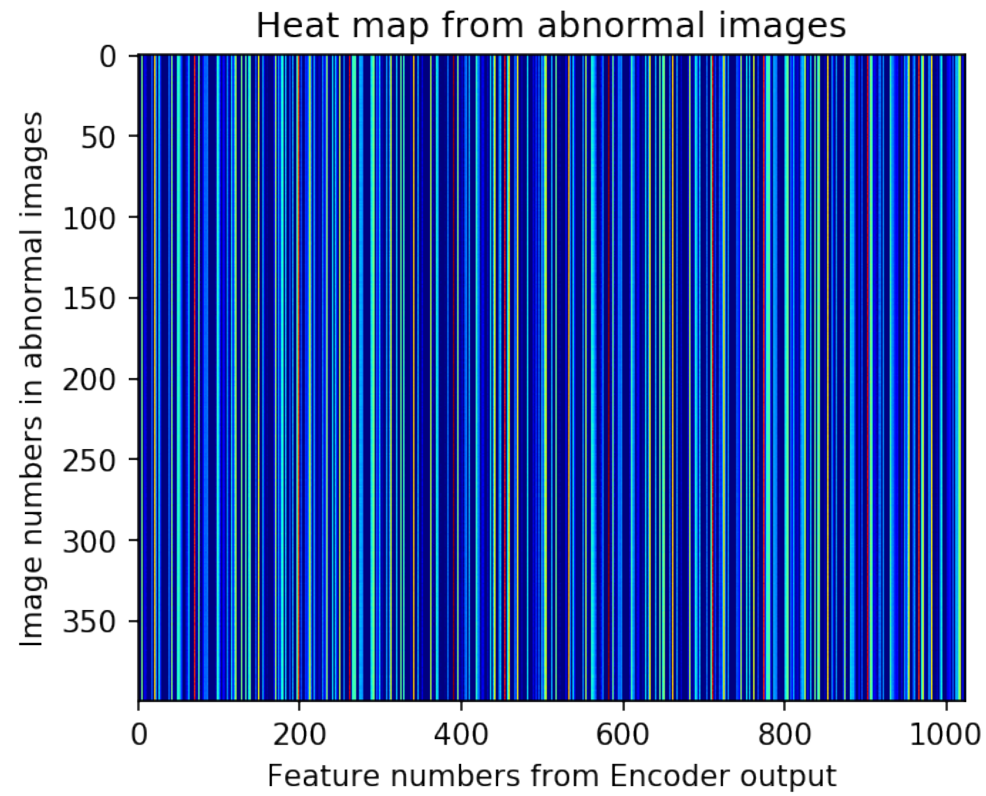
正常

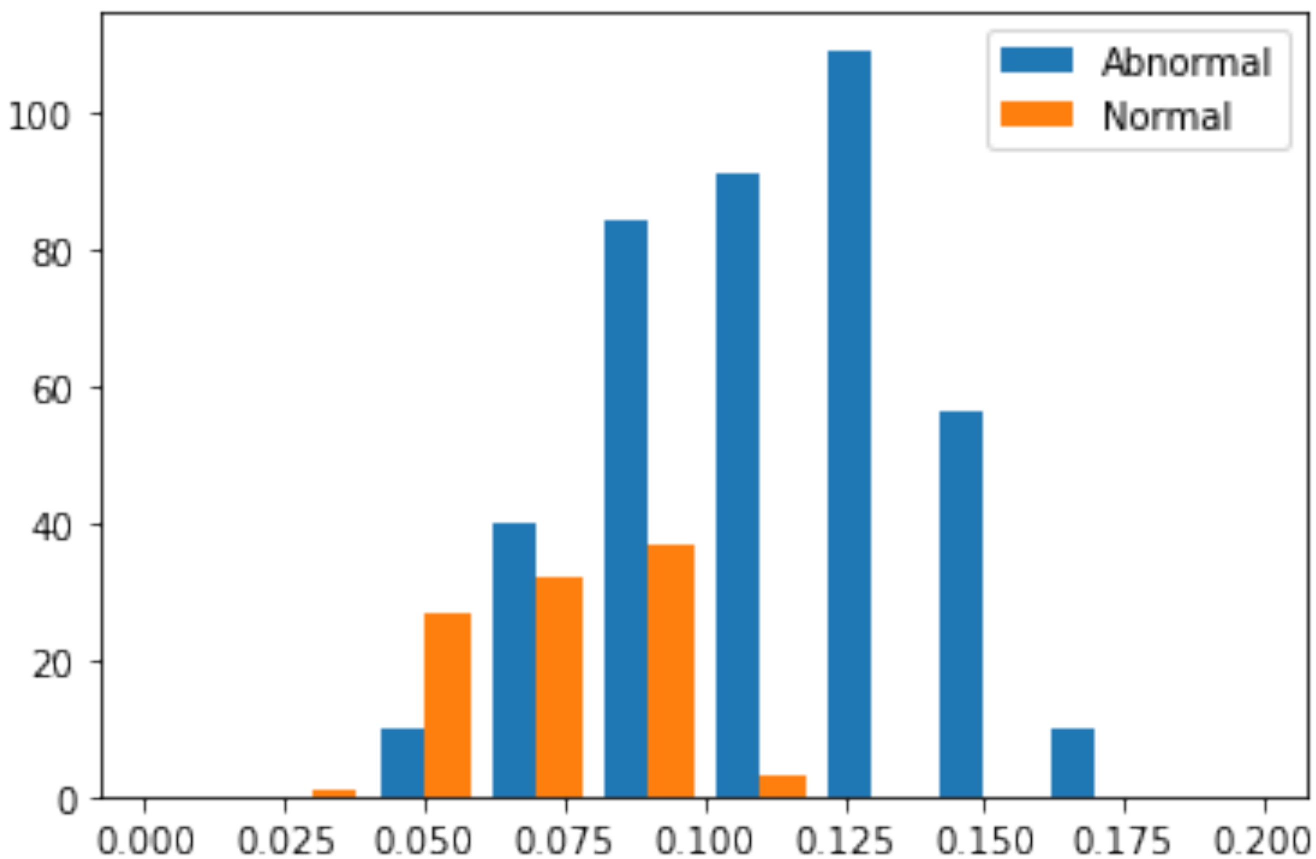


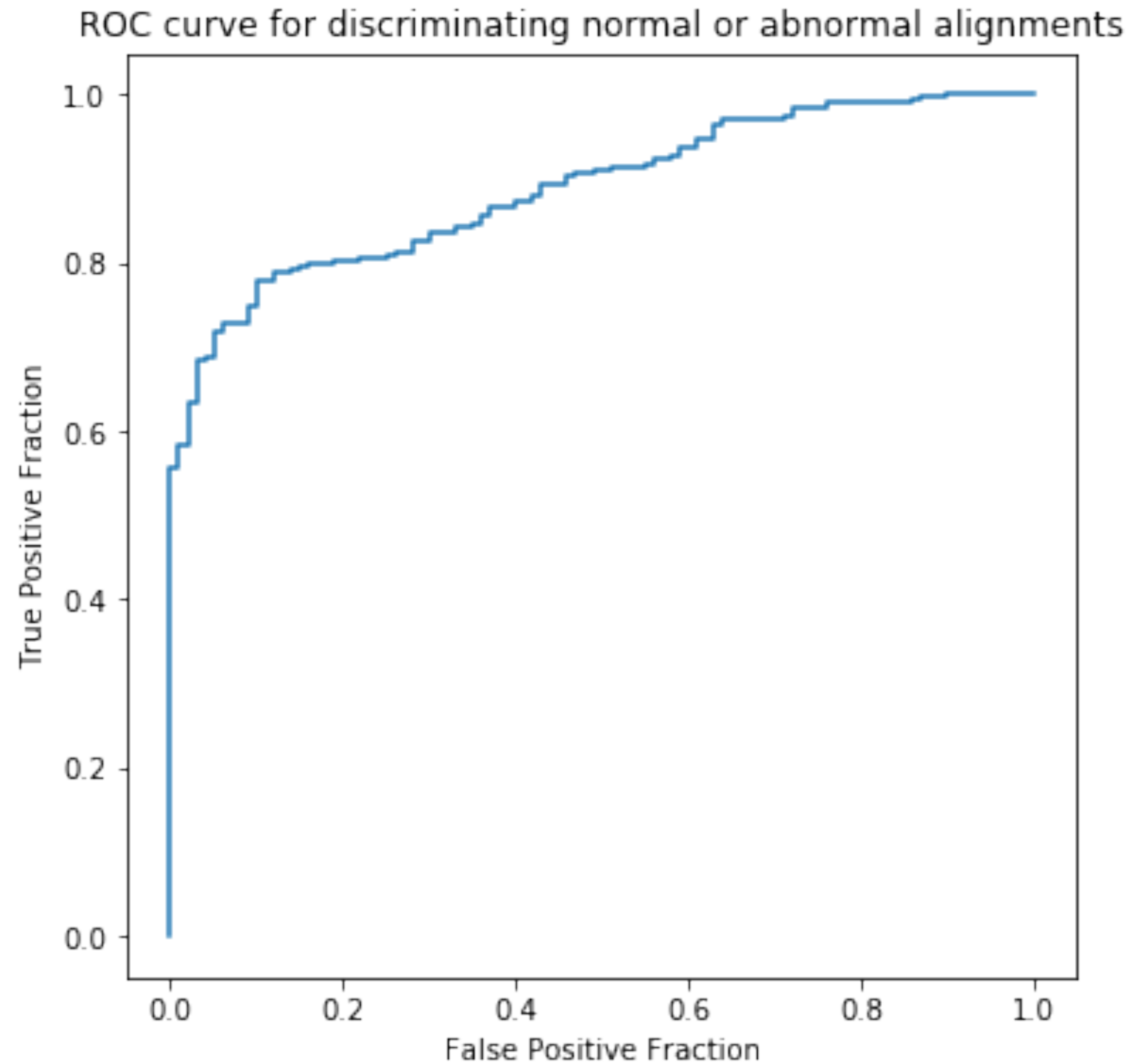
異常



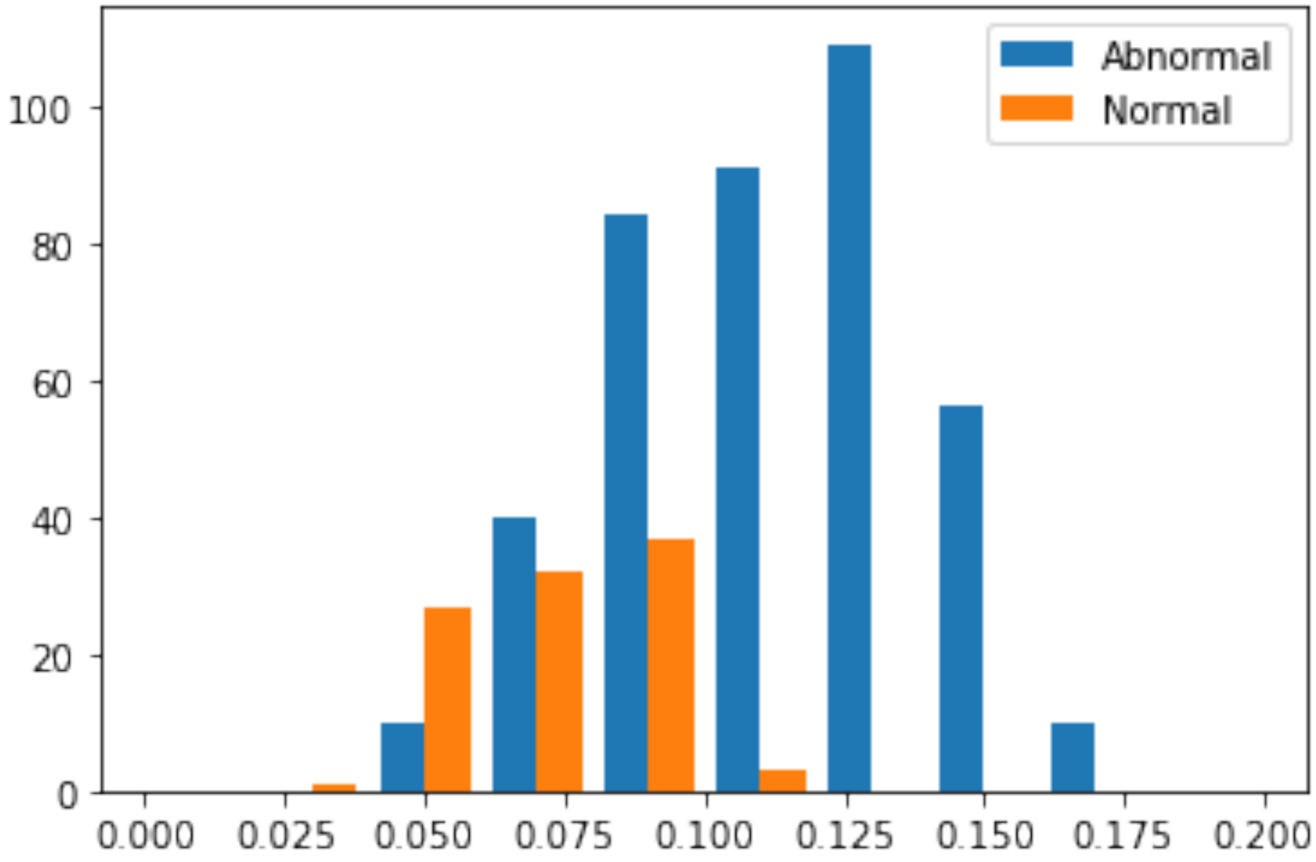

Heat map from normal images
Image numbers in normal images
Feature numbers from Encoder output

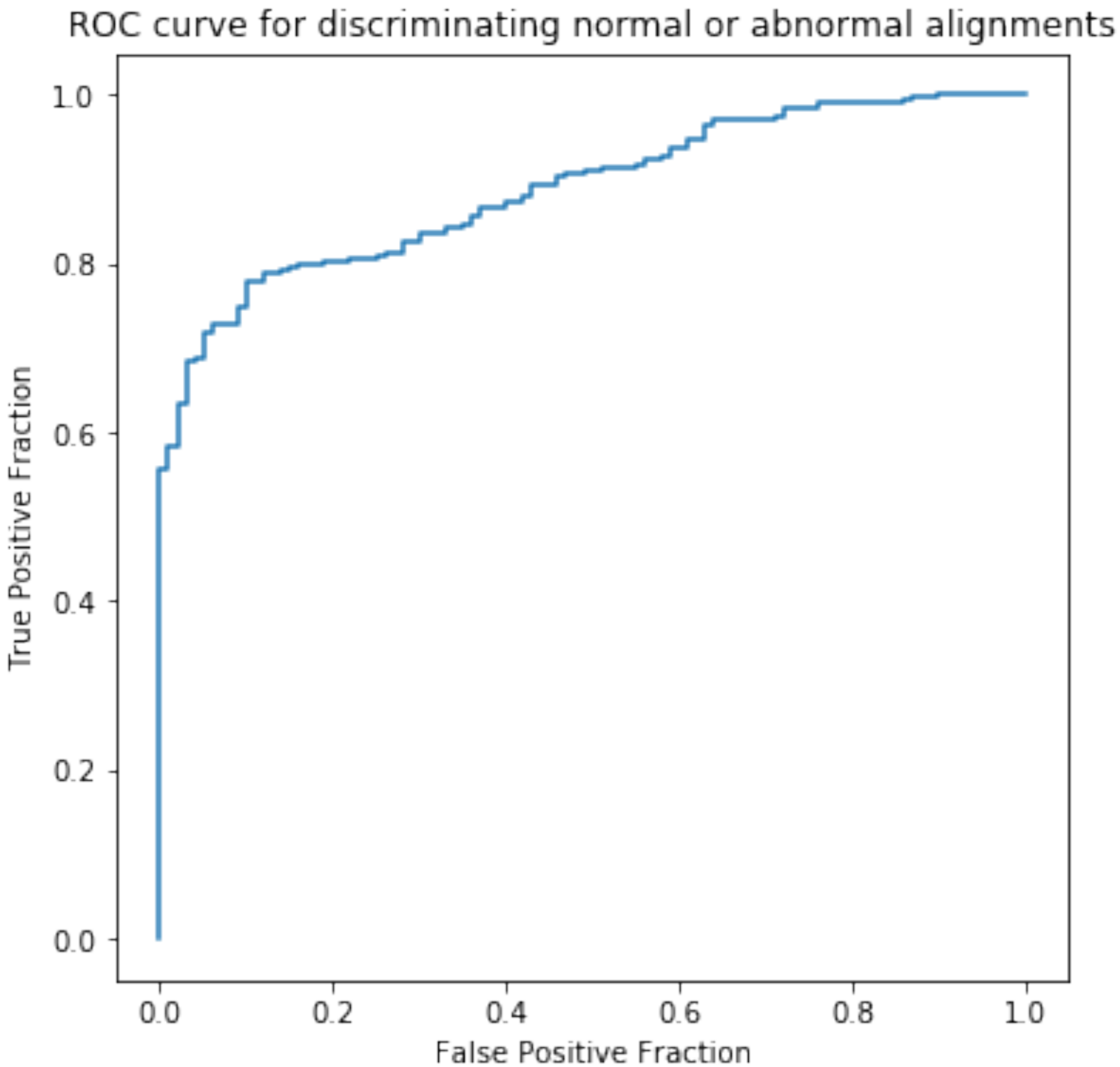

Heat map from abnormal images
Image numbers in abnormal images
Feature numbers from Encoder output


Abnormal
Normal
0
20
40
60
80
100
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200

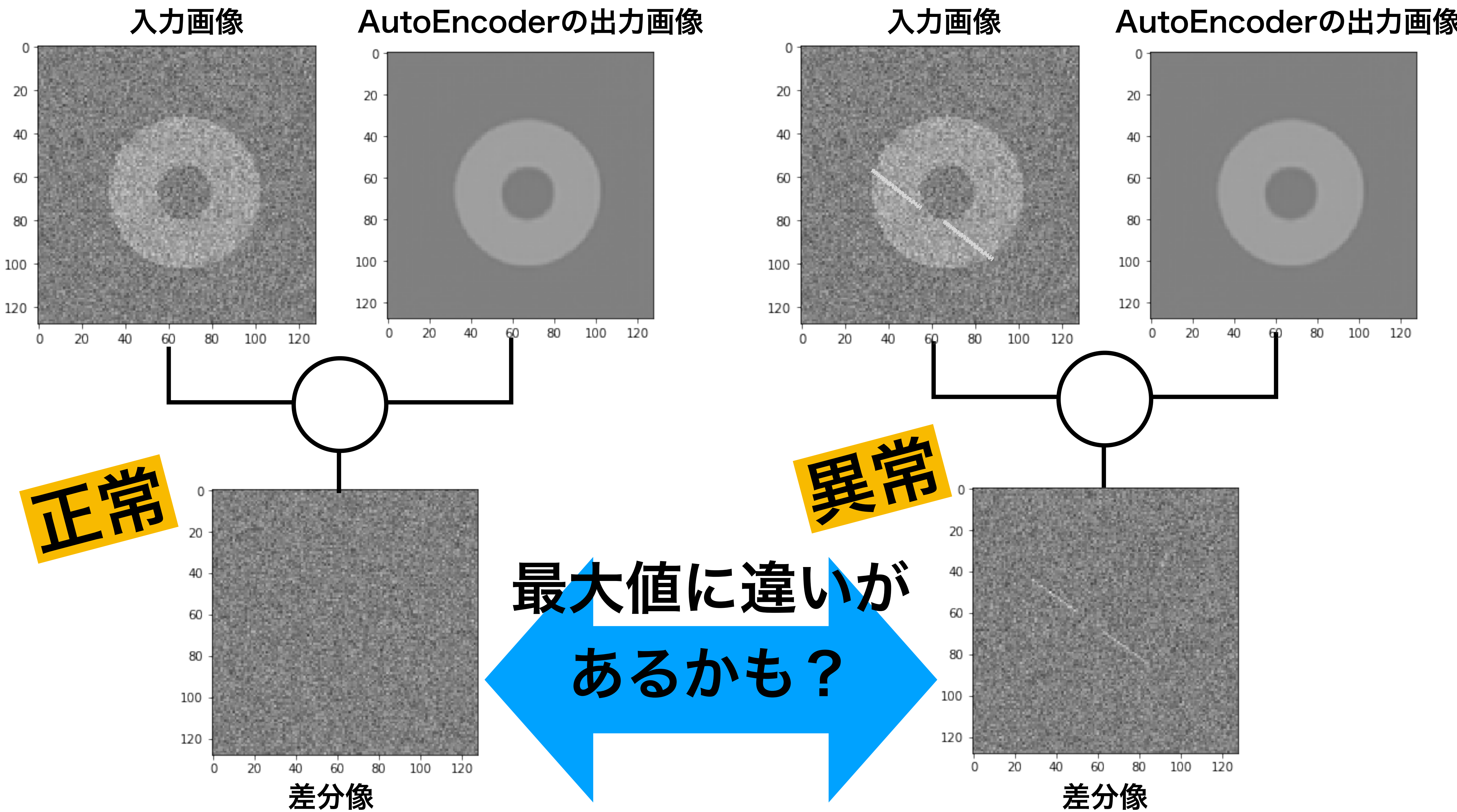

ROC curve for discriminating normal or abnormal alignments
True Positive Fraction
False Positive Fraction
0.0
0.2
0.4
0.6
0.8
1.0

データだけ集めて手法開発





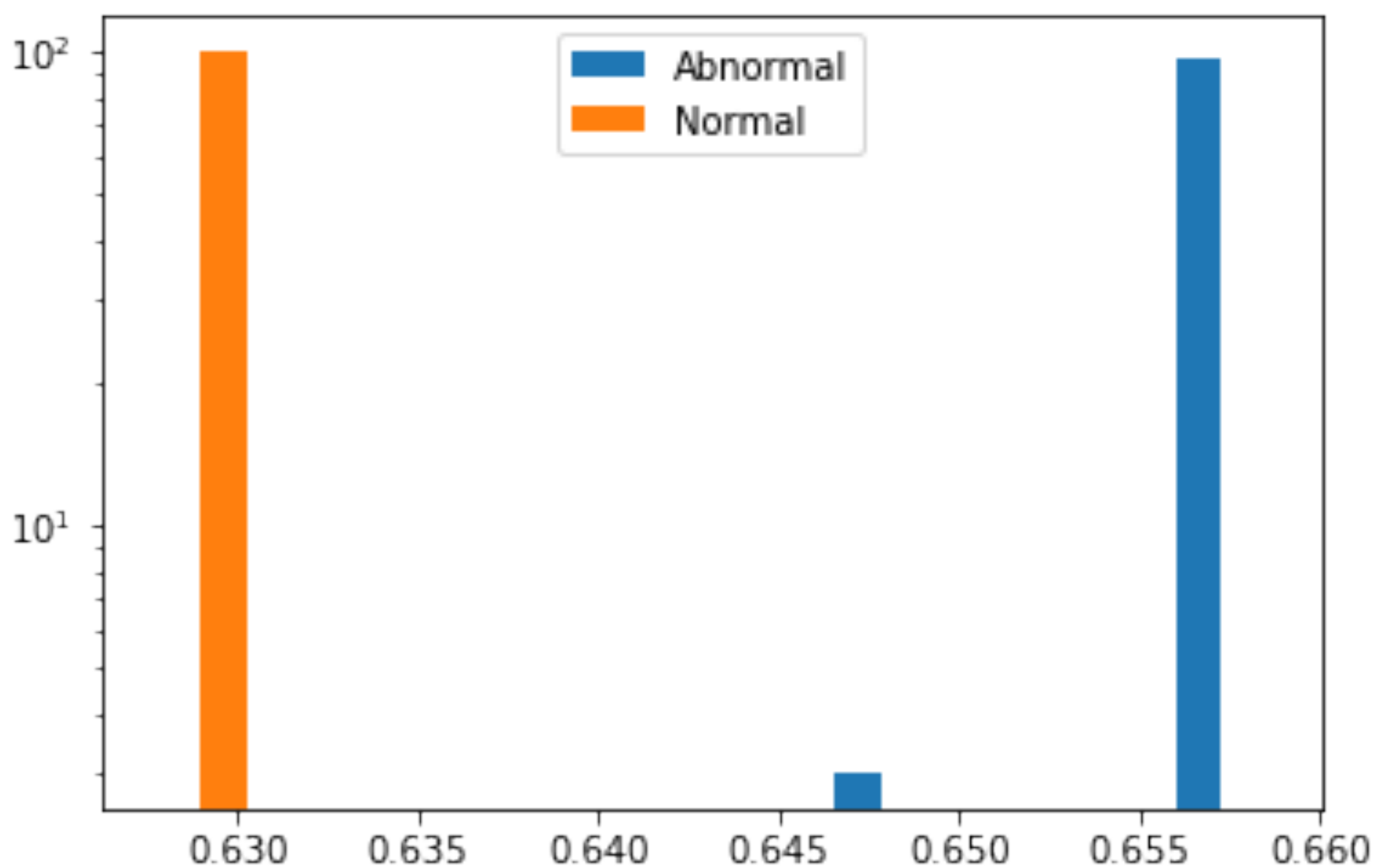
かっこよく言うとend-to-end

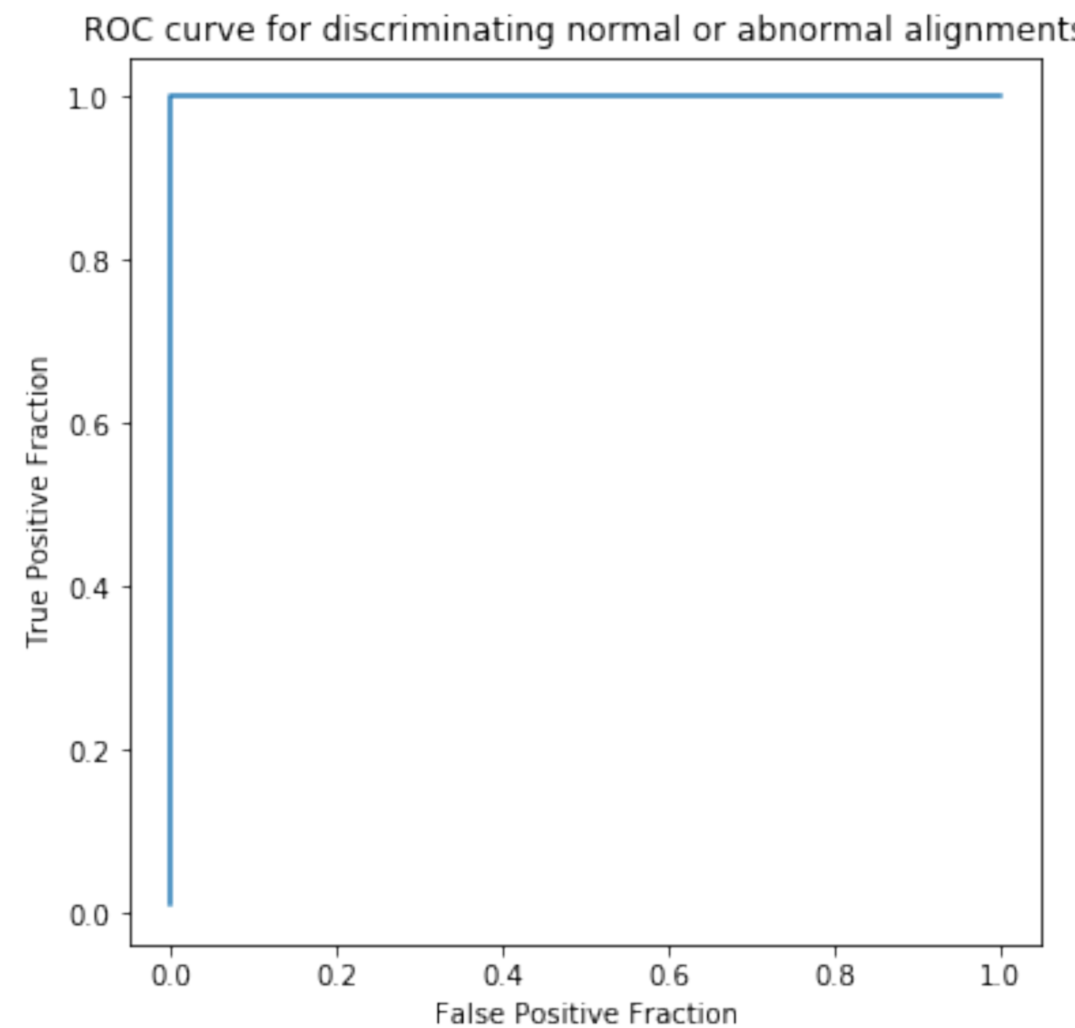


入力画像
AutoEncoderの出力画像
入力画像
AutoEncoderの出力画像
正常
異常
最大値に違いがあるかも？
差分像
差分像

極端な例ですが…

AutoEncoderを前処理としてうまく利用できれば
単純な特徴量（例えば最大値）でも分離可能





といっても
実際には
簡単には
うまくいきません

まとめ

深層学習は，機械学習の方法の一つです．
中でも「教師あり学習」が基本です．

機械学習は，データに基づいて課題を解きます．
30~100例から実験できます．
1000~10000例を目標としましょう．

人が学習するよりもたくさんの美味しい画像を食べさせないと深層学習は十分に成長しません．
Garbage-in, garbage-outです（本当か？）