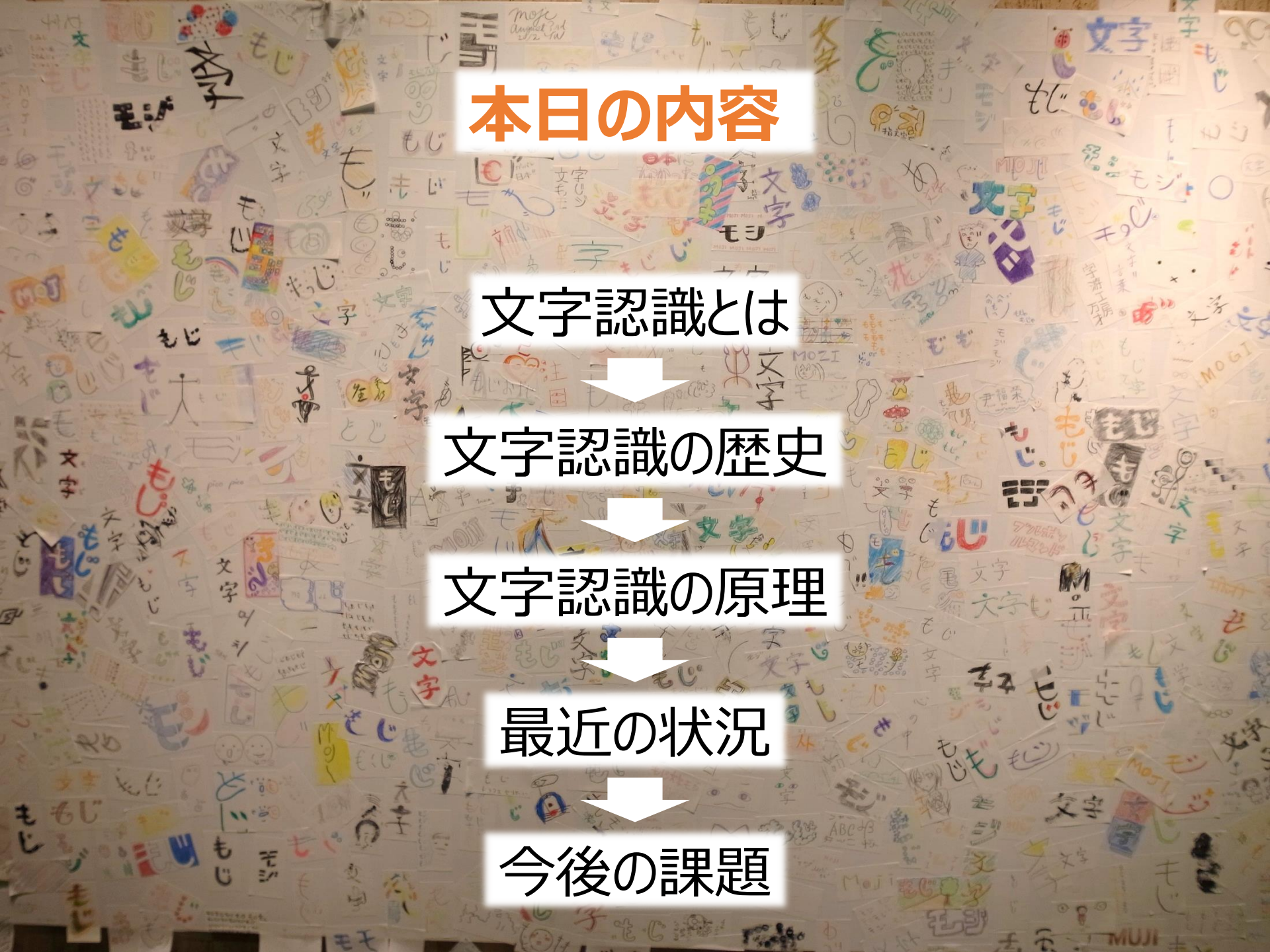


文字認識研究の 過去・現在・未来

日本文化とAIシンポジウム2019
基調講演 AIによるくずし字認識の展望
2019年11月11日13:15-13:50@NII 一橋講堂

九州大学 システム情報科学研究所 内田誠一

The background is a dense, colorful collage of various Japanese characters, including 'も', 'じ', '文字', and 'もじ', along with small drawings and illustrations. The text is overlaid on this background.

本日の内容

文字認識とは



文字認識の歴史



文字認識の原理



最近の状況



今後の課題



文字認識とは

一口に文字認識といっても、いろいろなバリエーションがあります

文字認識とは



入力画像



活字文字画像



認識結果
(クラス)

“春”

文字認識とは



入力画像



手書き文字画像



認識結果
(クラス)

“春”

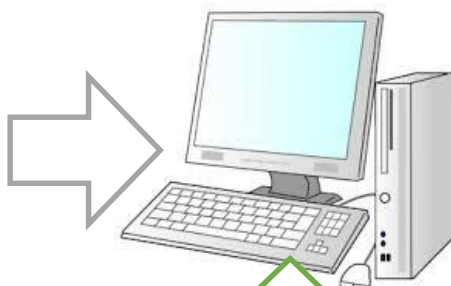
文字認識のバリエーション①

文字列認識

入力画像

昼食後はねむい

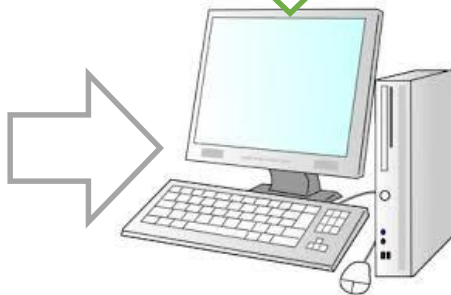
活字文字列画像



認識結果
(クラスの列)

“昼食後はねむい”

技術的には結構違う!



昼食後はねむい

手書き文字列画像

“昼食後はねむい”

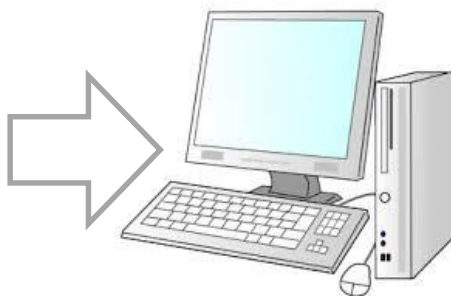
文字認識のバリエーション②

情景内文字認識

入力画像

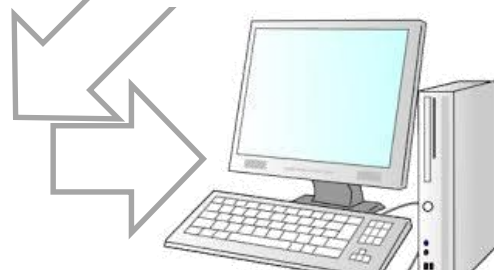


文字検出
(どこに文字列があるか?)



認識結果
(クラス)

文字列認識



“立飲み 入船”

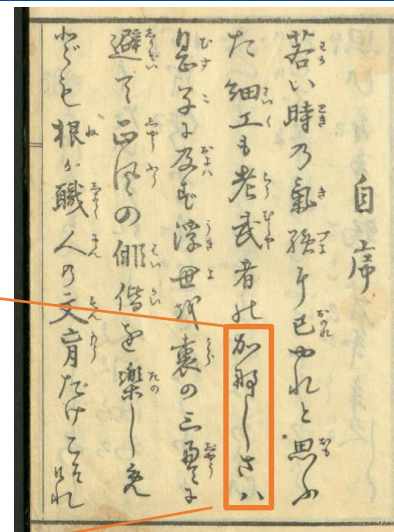
- 文字列が複数あれば、すべてを検出・認識
- 携帯カメラで写したメニューを翻訳するなど最近需要多い

文字認識のバリエーション③

古文文献（くずし字）認識

- 基本は、手書き文字列画像認識
 - ただし、一般的な手書き文字列より遥かに難しい！

“かなしさは” ←



日本古典籍くずし字データセット「虚南留別志」

- 紙面全体が与えられれば、どこに文字列があるかを「検出」する必要もある



日本古典籍くずし字データセット「南総里見八犬伝」

文字認識の歴史もほぼこの順番

活字 ⇨ 手書き ⇨ 情景内 ⇨ 古文文献
(くずし字)

かんたん

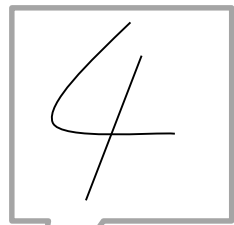
むずかしい

文字認識の歴史はドラゴンボール(?)的
(より強い敵を見つけて戦う)



参考：オンライン文字認識

入力画像？



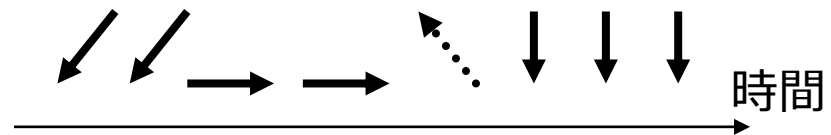
タブレットや
タッチパッド等



認識結果
(クラス)

“4”

その正体，実はペン先座標の
時系列データ！



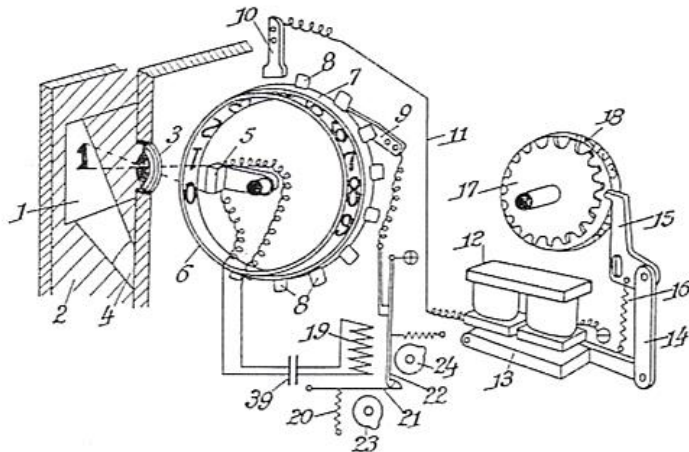
- 画像による文字認識とは全く違った認識手法を利用
- 筆順情報があるので画像よりも高い認識率が得られる

文字認識の歴史

ざくっと、ふりかえってみましょう

文字認識の歴史

1929	画像による文字認識（OCR）の原理が 特許取得（オーストリア）
1955	印刷数字OCR実現（米，ファーリントン社）
1955	より高性能なOCR（英）
	初のオンライン文字認識実験（米，ベル研）
1958	初のOCR実験（日，電気試験所，飯島ら）



Tauschekの特許
（オーストリア、1929）
（橋本、「文字認識概論」より抜粋）

文字認識の歴史

1960年代
前半

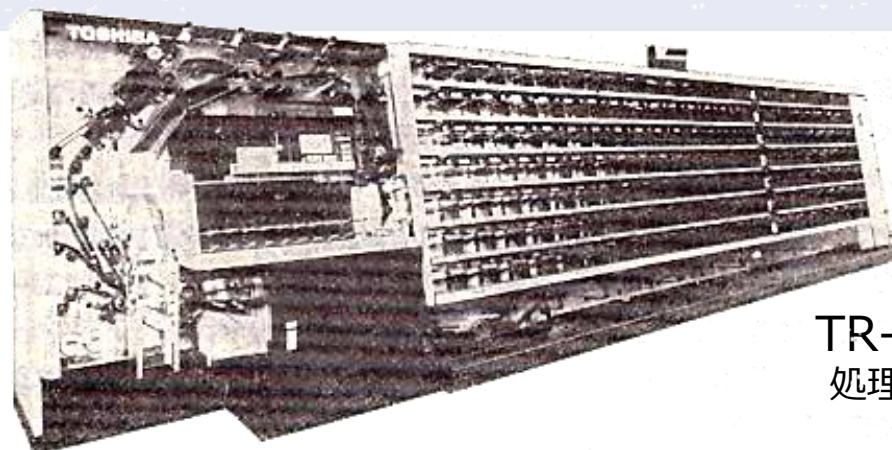
各社手書き・活字OCR発売（米）
REI社の製品は3億円！

1960年代
後半

各社OCR発売（日）
郵便区分機実用化（日, 東芝, 1966）
通産省大型プロジェクト開始（日, 1966）
活字漢字認識についての論文
（米, IBM, Casey-Nagy, 1966）

0 1 2 3 4 5 6
5 6 7 8 9 0 1

www.post.japanpost.jp



郵便区分機
TR-16（東芝、1985）
処理能力 3万～4万通／時
（森、「パターン認識」より抜粋）

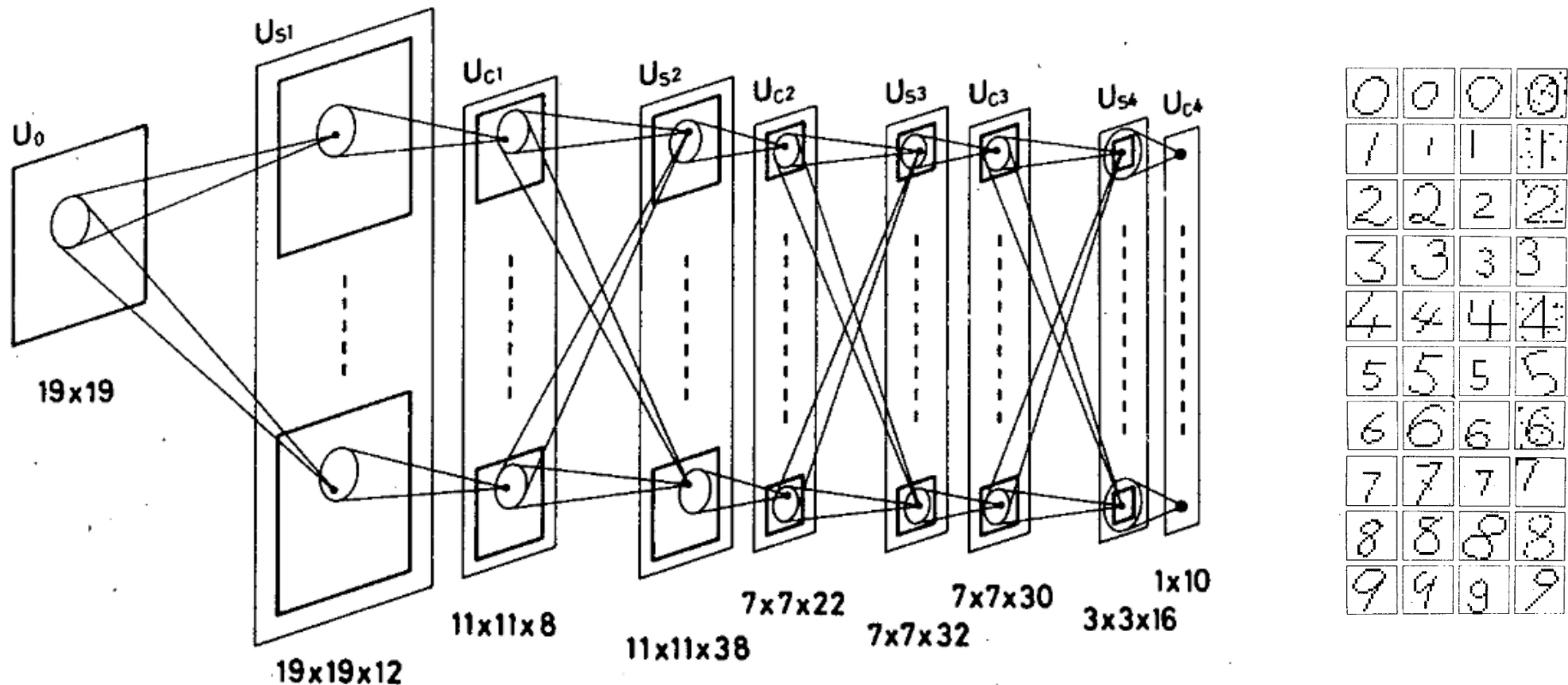
文字認識の歴史

1970 年代	手書き漢字OCR挑戦（日,1970） 印刷漢字認識実験（日,日立,1972） ハントスキャン型OCR（米, 1975） 帳票OCR全盛（日）
1980 年代	手書き漢字OCR発売（日） <u>深層ニューラルネットワーク</u> の元祖,「ネオコグニトロン」の 発明(1979)と文字認識応用（日, 福島@放送技研）
1990 年代	手書きあて名認識実用化（日、米） 手書き小切手認識研究（欧米） 低価格印刷文字OCR普及（世界） 手書き漢字認識率99%超える（日, 三重大, 1996）

最近“AI”と
呼ばれているもの

ネオコグニトロン

(福島, 信学論1979; 福島他, 信学論1983)



- 驚くべきことに，現在隆盛を極めている「畳み込み深層ニューラルネットワーク」とほぼ完全に同一！

文字認識の歴史

2000年代	情景内文字認識の萌芽的研究 タブレットPC, Nintendo DS Google Booksearch 古文獻認識の大型プロジェクト開始（欧州）
2010年代	情景内文字認識全盛 「深層ニューラルネットワーク×大量データ」による 劇的な精度改善（全世界） Evernoteなどのサービス GoogleLensなどのアプリ Robotic Process Automation導入（企業）
2020年代	さて、どうなるでしょう？（後ほど私見を）

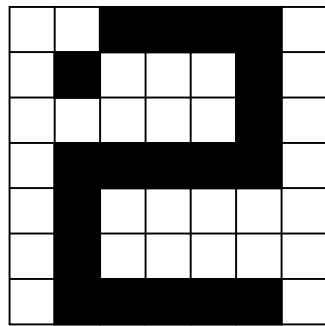


文字認識の原理

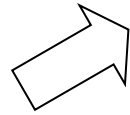
なるべく平易に説明いたします！
(専門家の方、申し訳ございません)

文字認識の基本的流れ①

特徴抽出＝画像を何らかの数字の組で表す

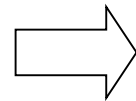


7×7画像



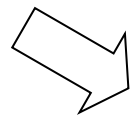
$(3, 2)$

横線の数, 縦線の数



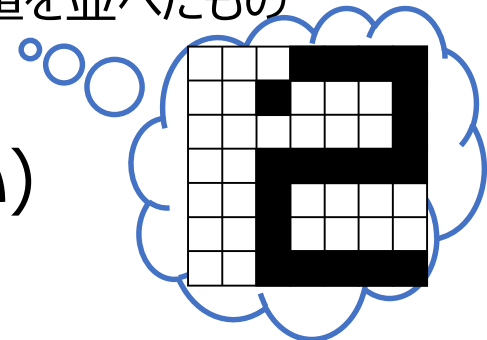
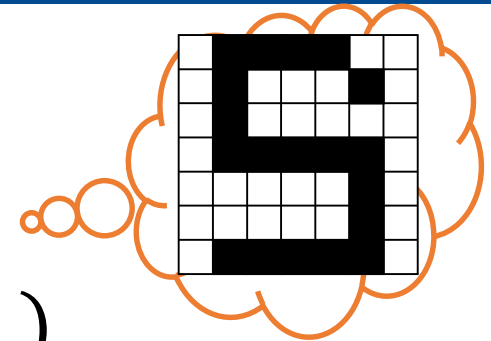
$(4, 2, 1, 5, 1, 1, 5)$

各行の黒画素数



$(0, 0, 1, 1, 1, 1, 0, 0, 1, \dots, 1, 0)$

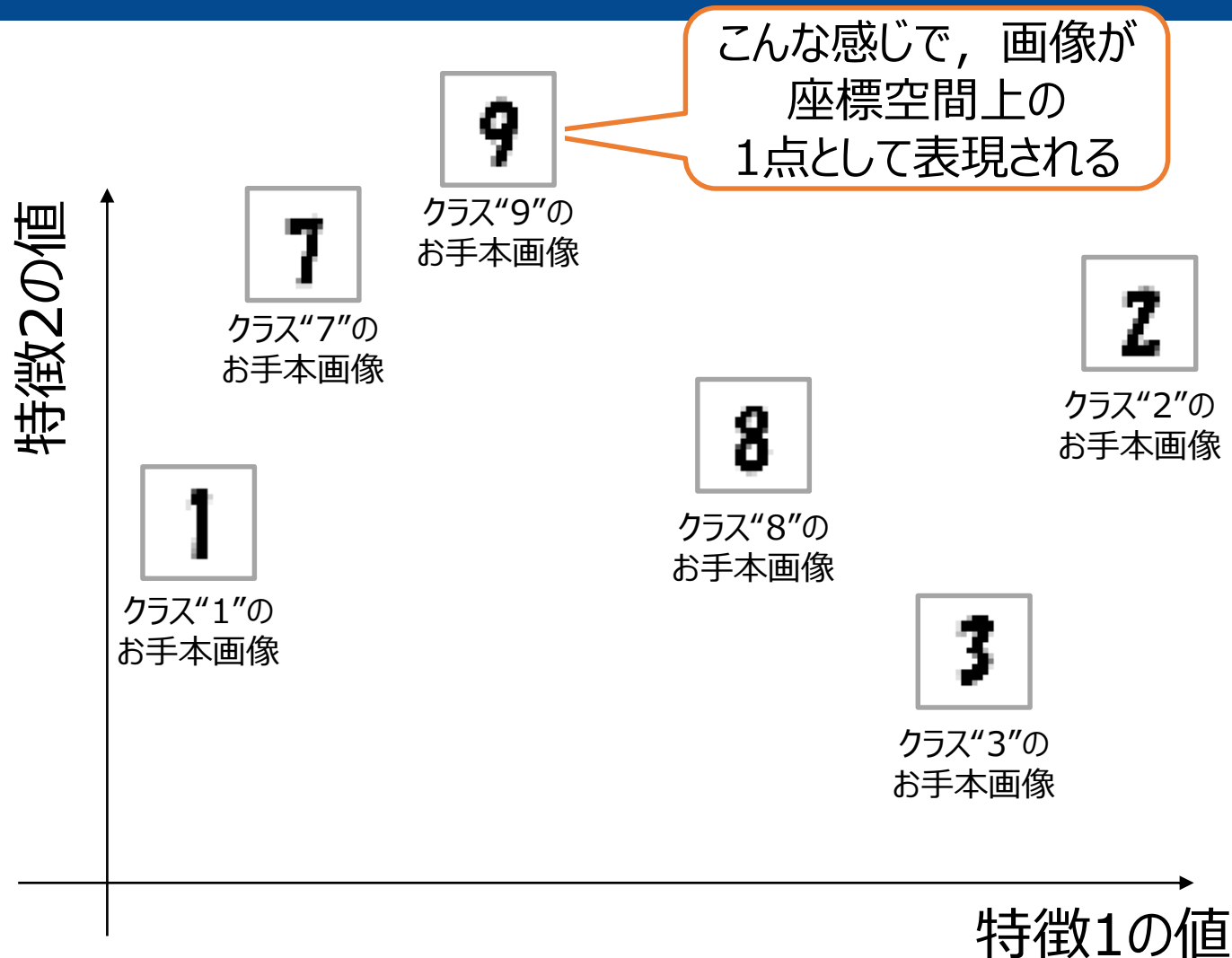
49個の白(0)黒(1)の値を並べたもの



- どれでもOK (絶対これ！という決まりはない)
- ただしうまく選ばないと認識性能に差が…
- 文字認識研究の歴史 = よりよい特徴を探す旅

文字認識の基本的流れ②

「お手本画像」から特徴抽出しておく



特徴2の値

クラス"9"のお手本画像

クラス"7"のお手本画像

クラス"8"のお手本画像

クラス"2"のお手本画像

クラス"3"のお手本画像

クラス"1"のお手本画像

クラス"4"のお手本画像

クラス"5"のお手本画像

クラス"6"のお手本画像

入力

???

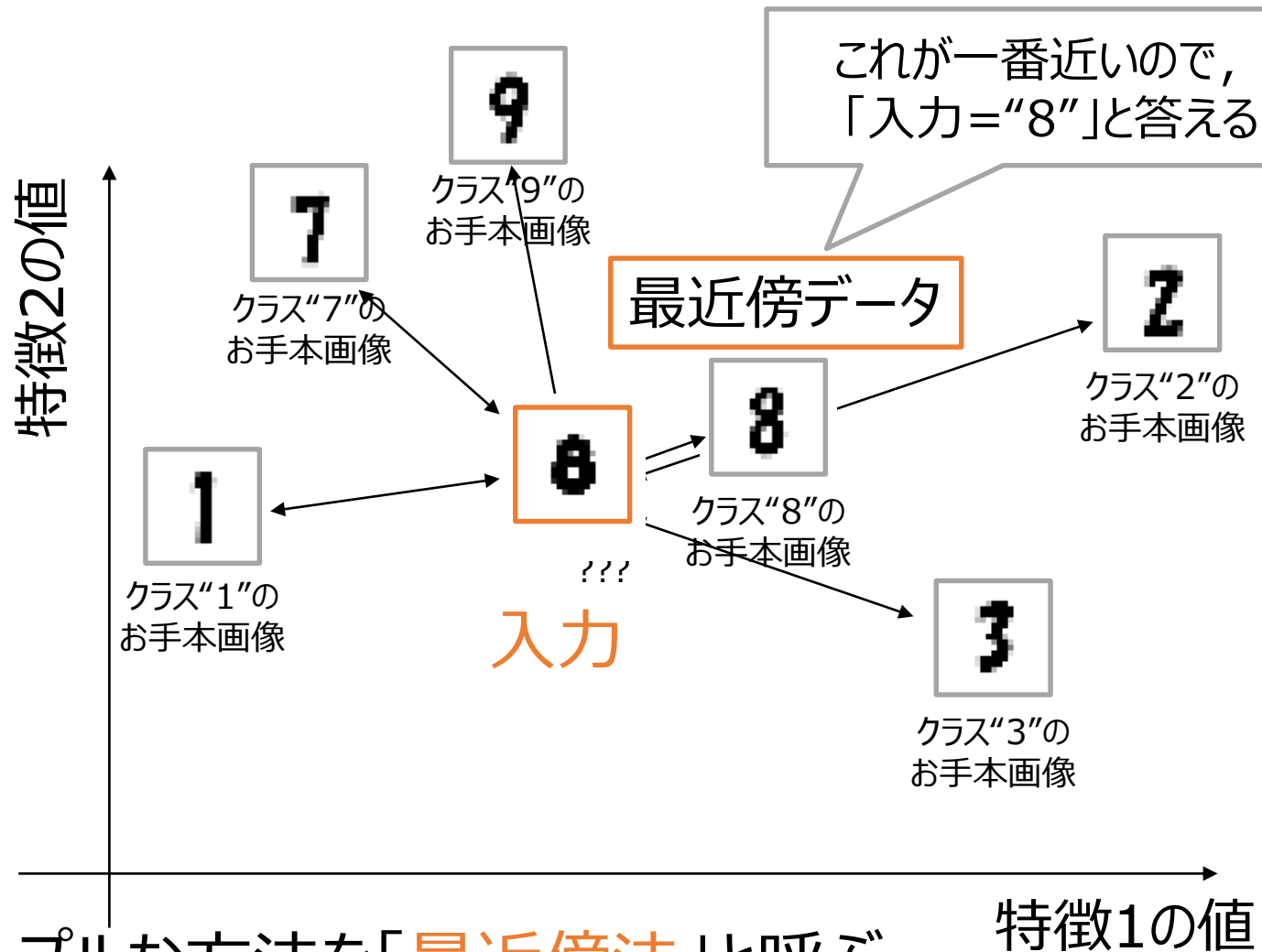
あるかを知りたい！

特徴1の値

これが何であることを知りたい！

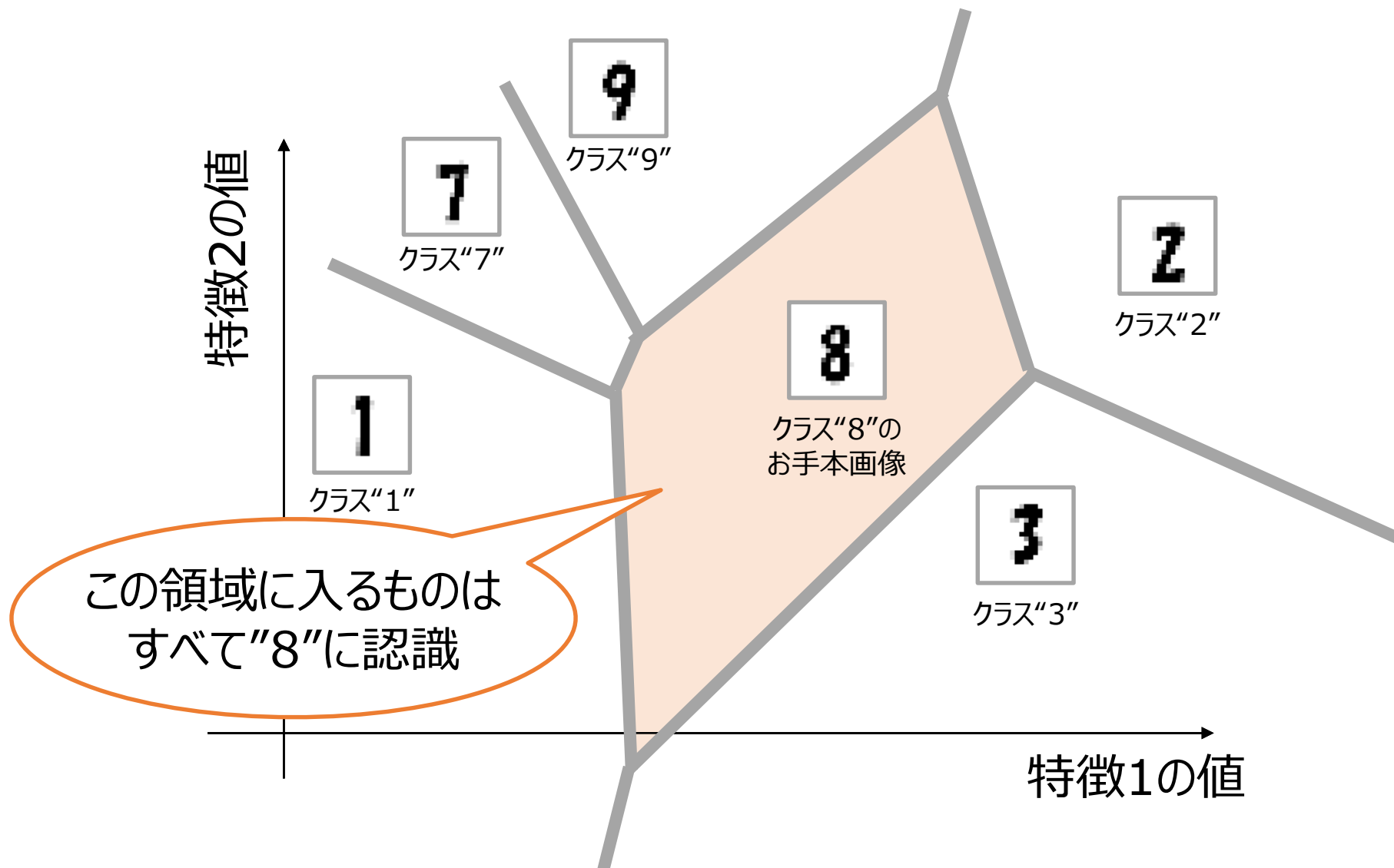
文字認識の基本的流れ④

いよいよ認識！（入力画像のクラス決定）

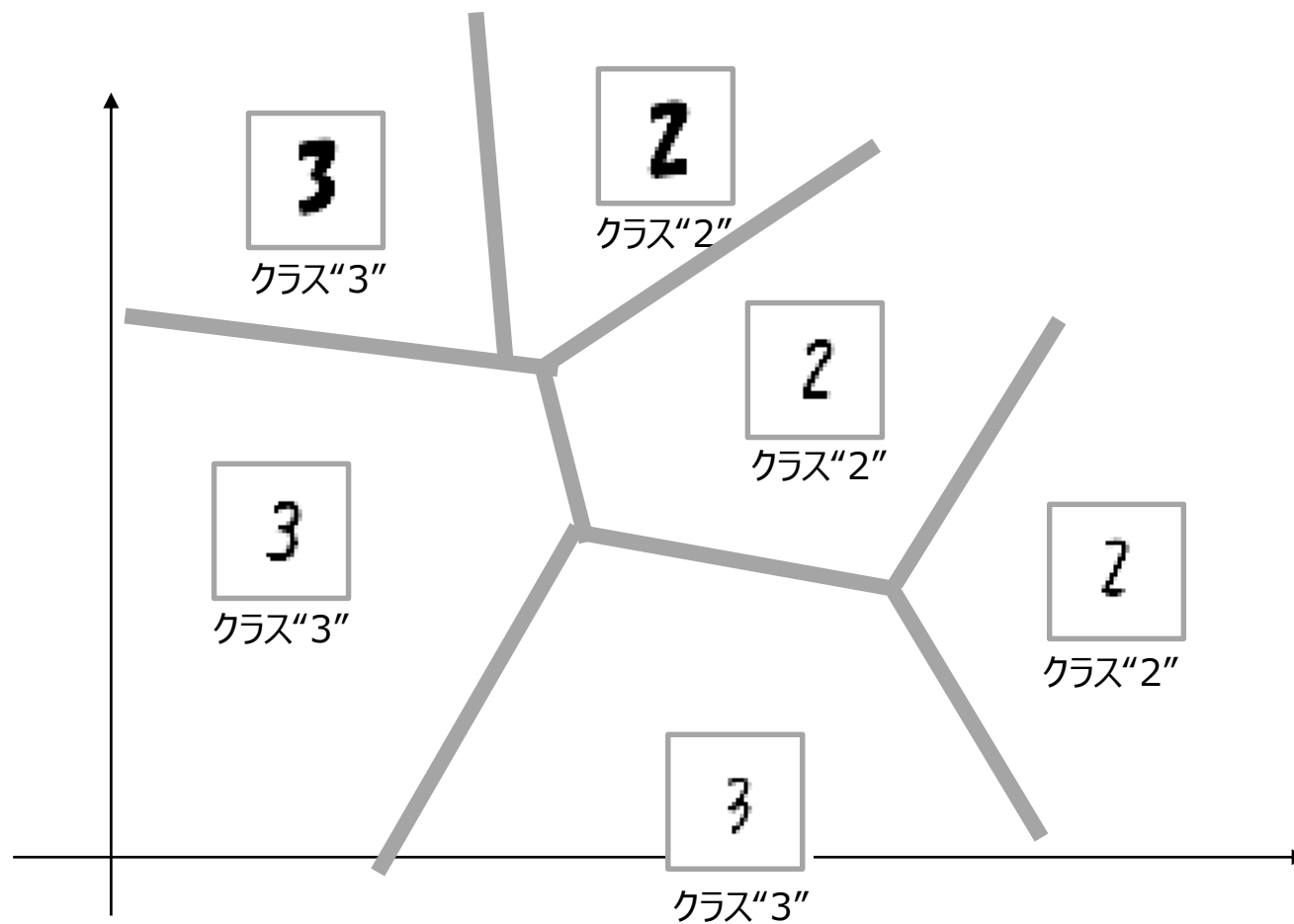


- このシンプルな方法を「最近傍法」と呼ぶ

最近傍法に関する一口メモ： 実は暗に「なわばり」がある



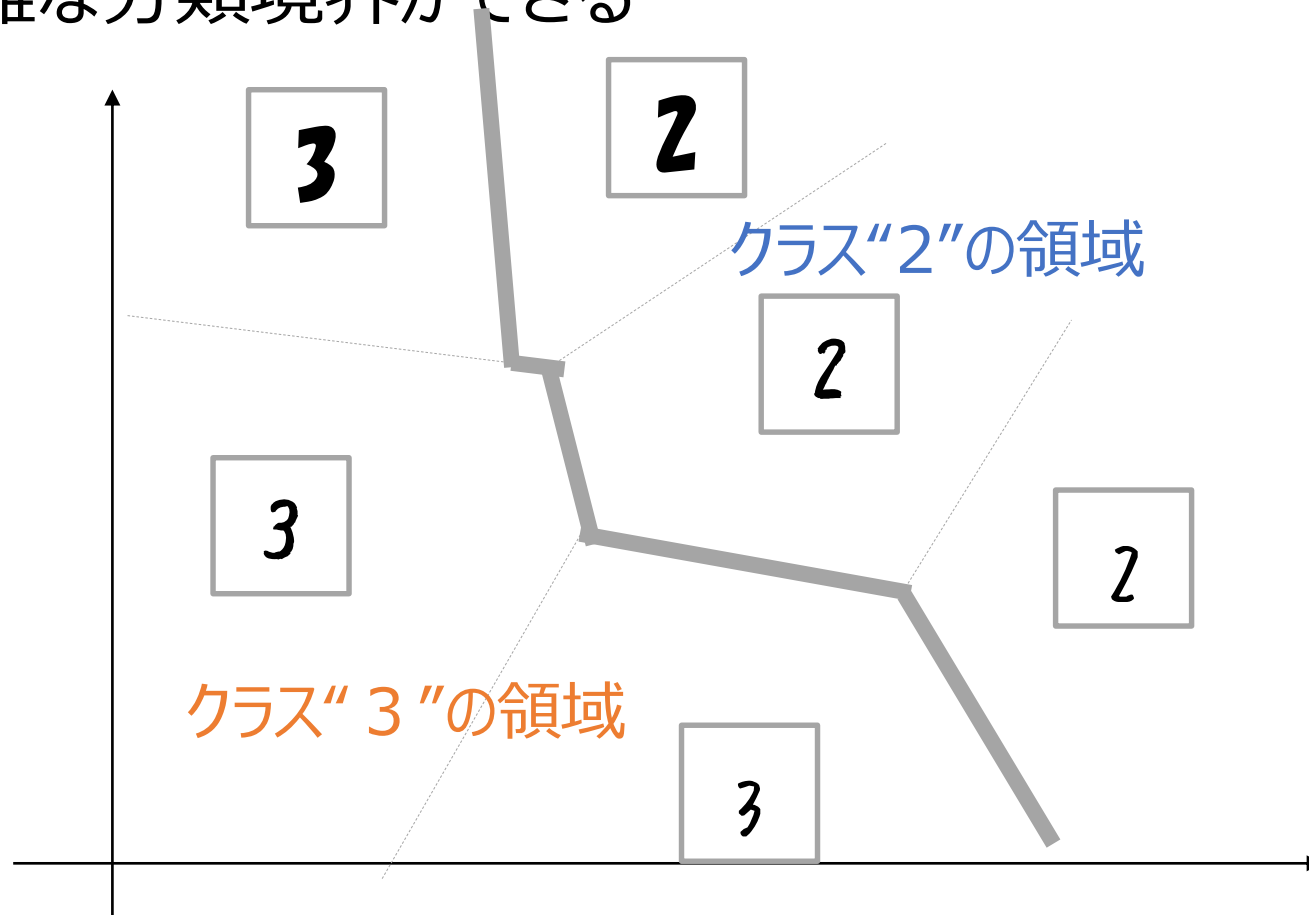
最近傍法に関する一口メモ：
1 クラスあたり複数個を覚えさせてもOK!



最近傍法に関する一口メモ：

1 クラスあたり複数個を覚えさせてもOK!

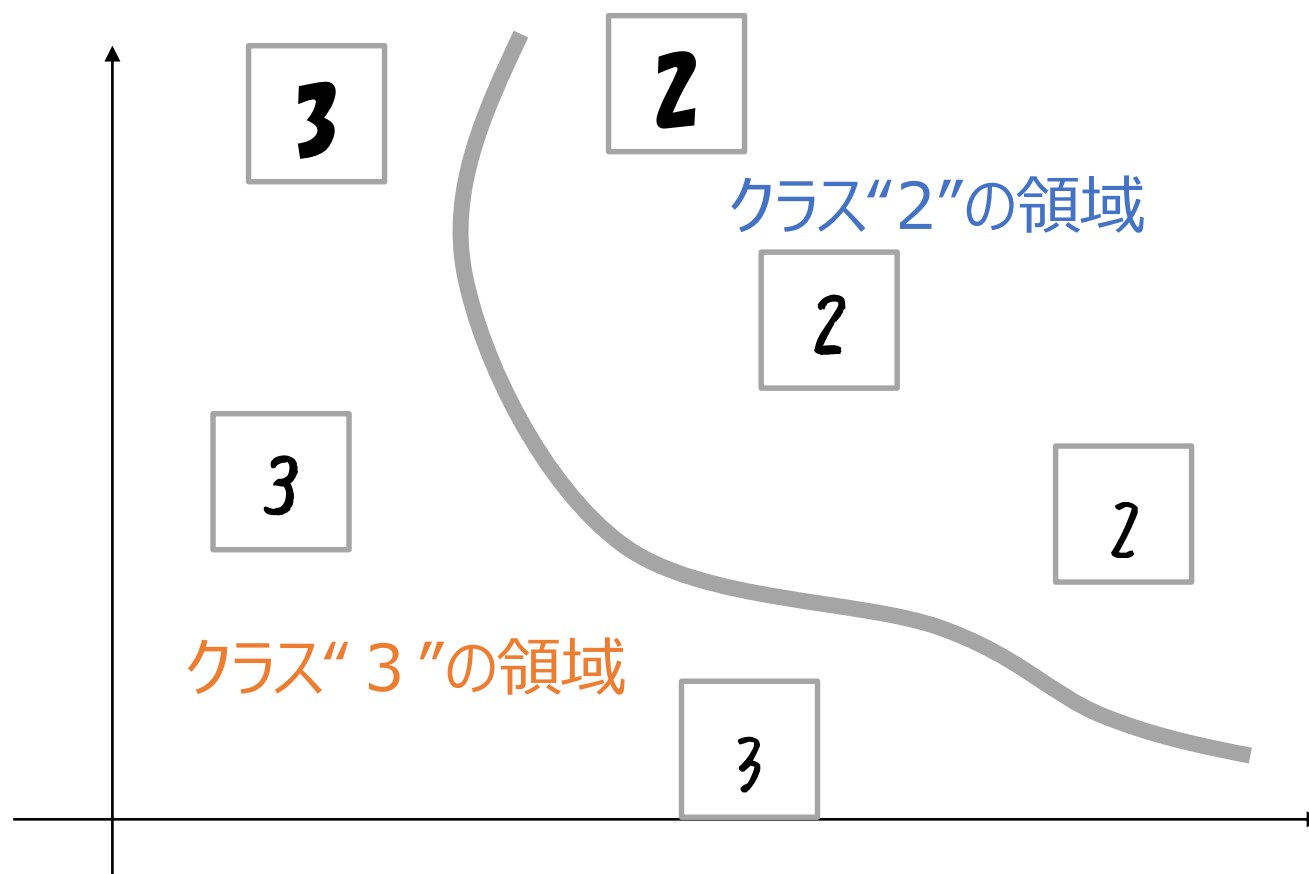
- より複雑な分類境界ができる



- もっとどんどん増やせばもっと複雑にできそう。それだったら…

最近傍法以外の文字認識手法： 適切な境界線を直接作ればいい！

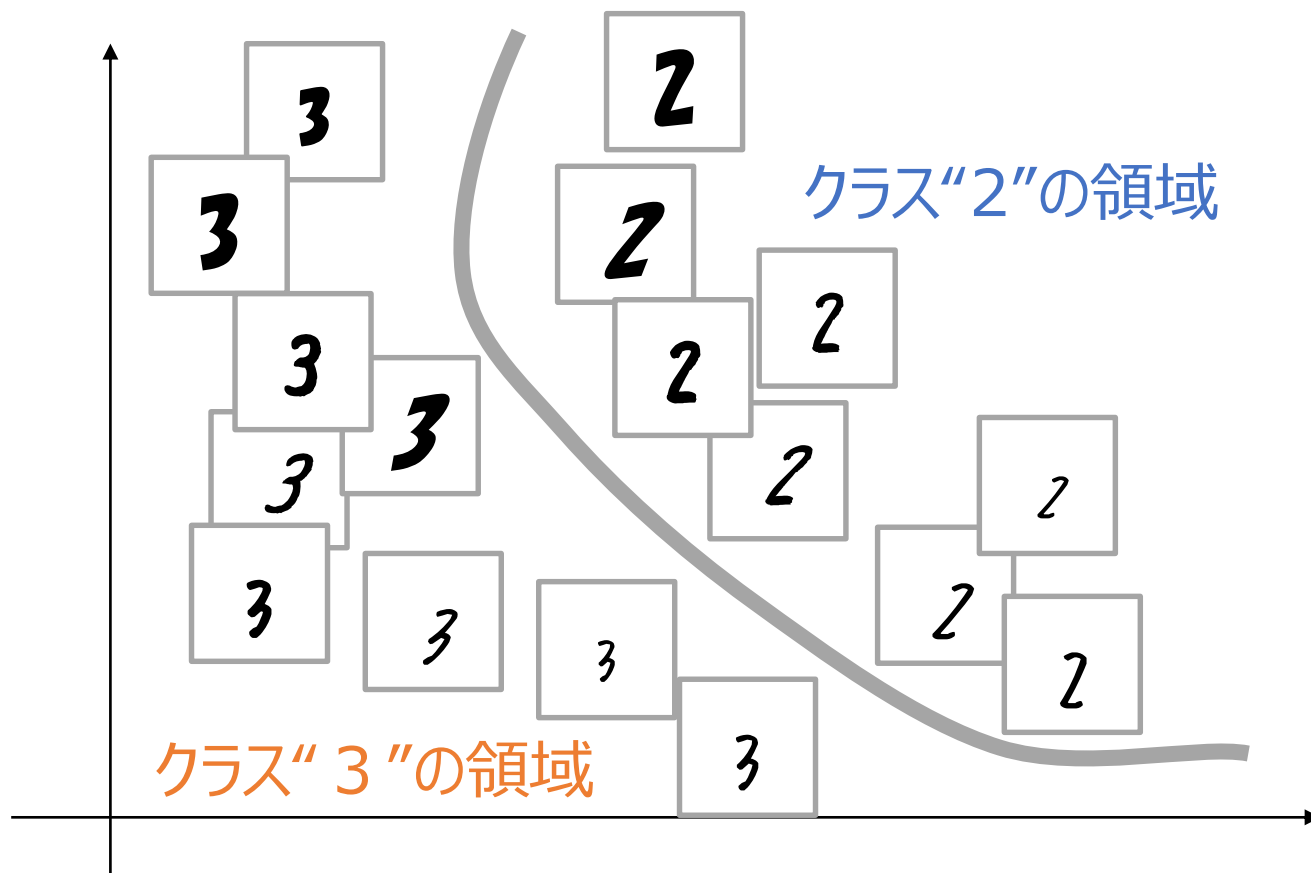
- これを「識別関数法」と呼ぶ



識別関数法の発展版: 深層ニューラルネットワーク

最近“AI”と
呼ばれているもの

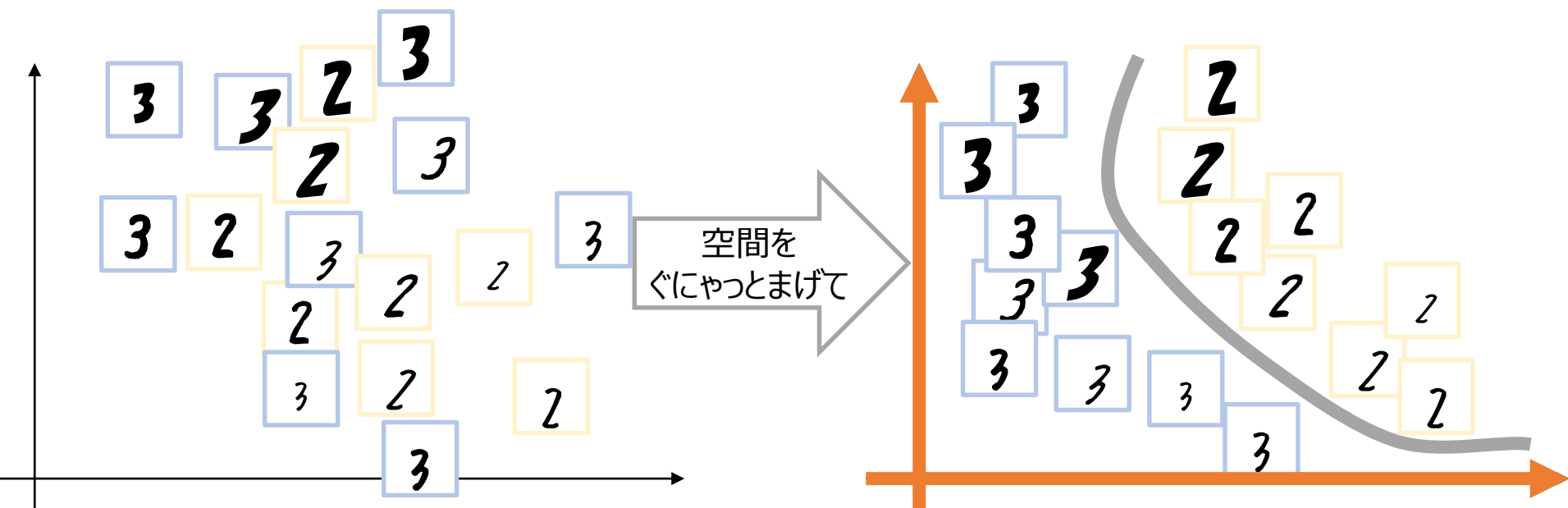
- たくさんサンプルを与えれば極めて適切な境界線を作る



識別関数法の発展版： 深層ニューラルネットワーク

最近“AI”と
呼ばれているもの

- 識別境界を決めるだけでなく，認識しやすいような特徴も自動で決定してくれる点が，ありがたい！



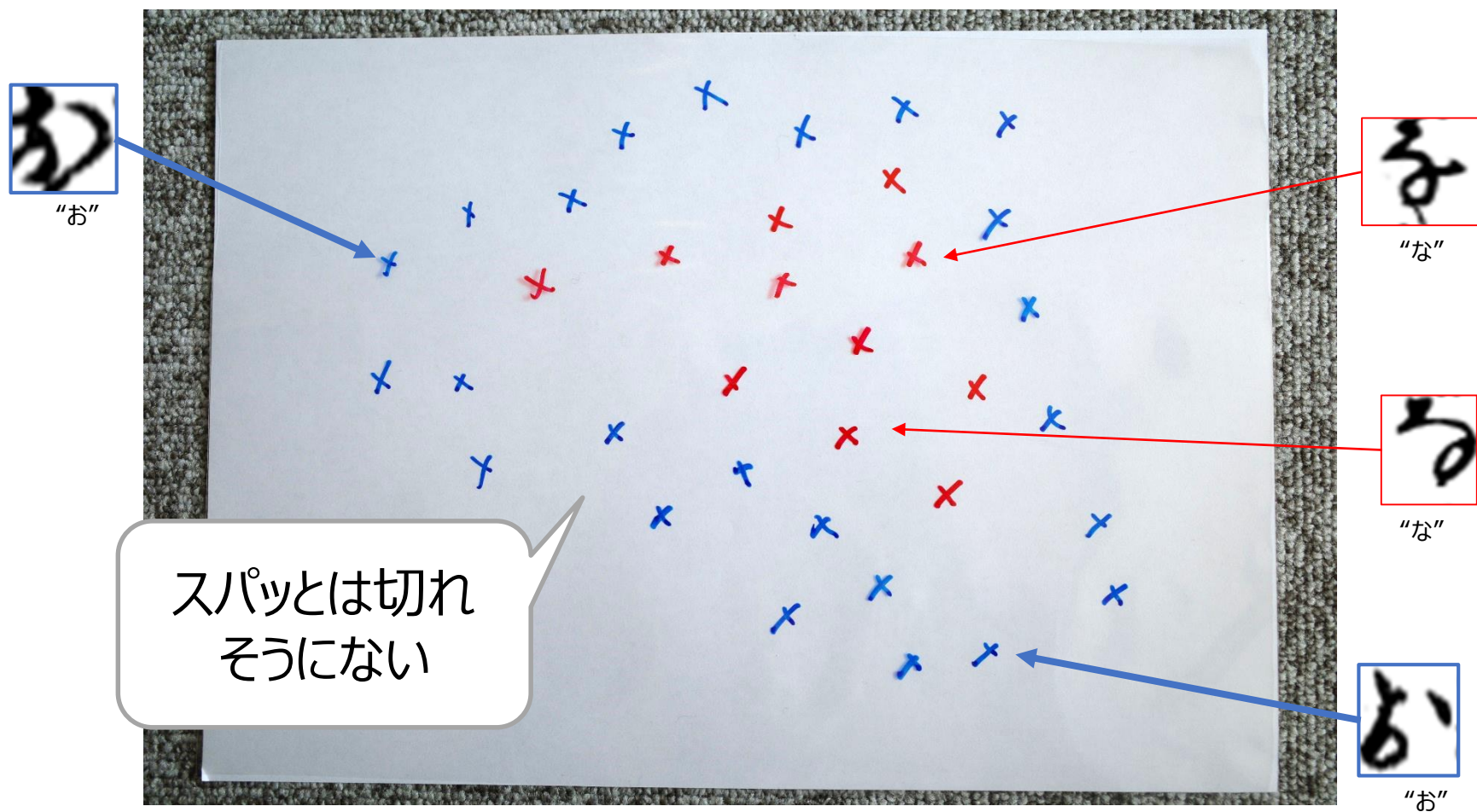
元々は入り混じっているが

スパッと切りやすく！

識別関数法の発展版: 深層ニューラルネットワーク

最近“AI”と
呼ばれているもの

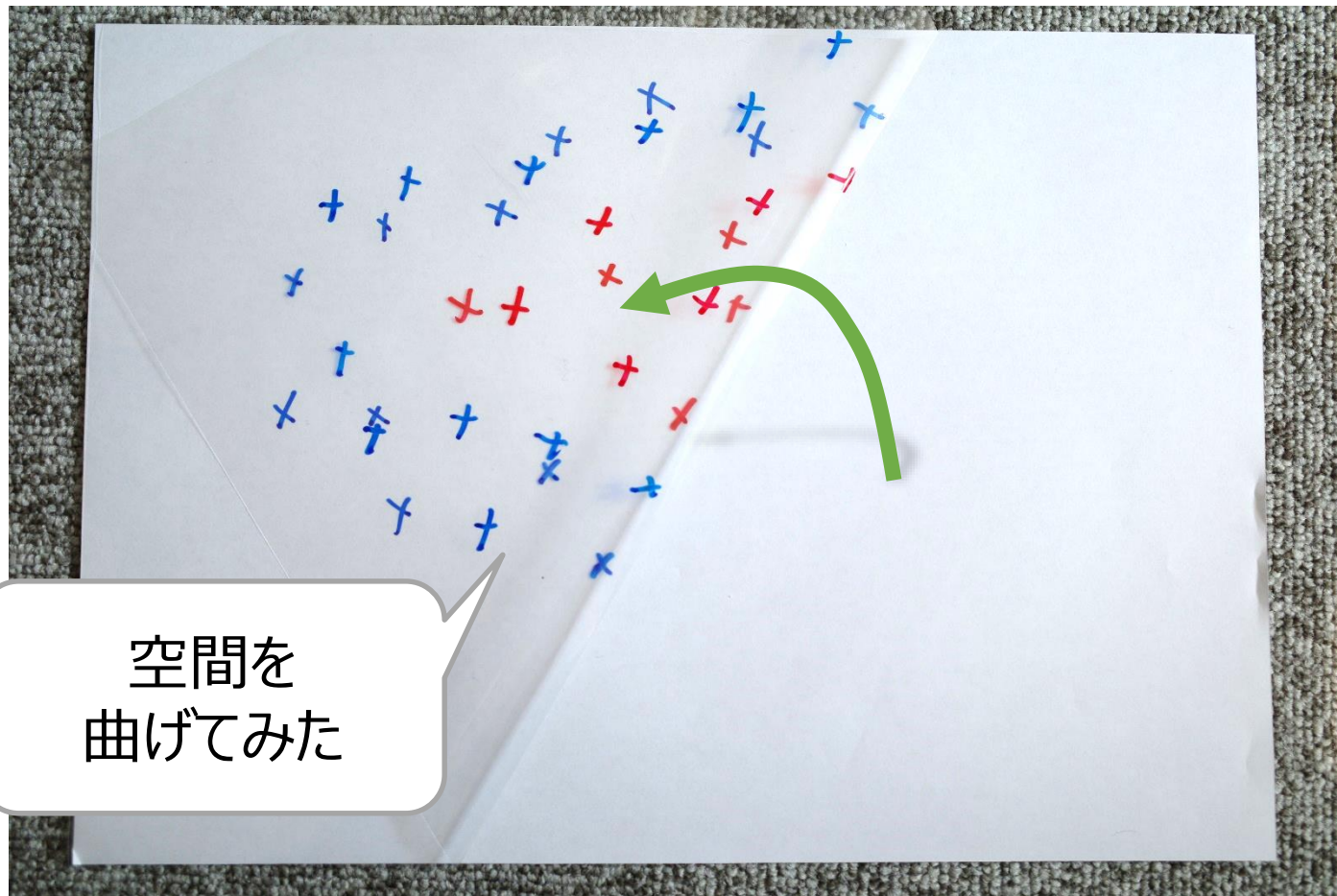
- (誤解を恐れず)もっと簡単に説明すると…



識別関数法の発展版: 深層ニューラルネットワーク

最近“AI”と
呼ばれているもの

- (誤解を恐れず)もっと簡単に説明すると…

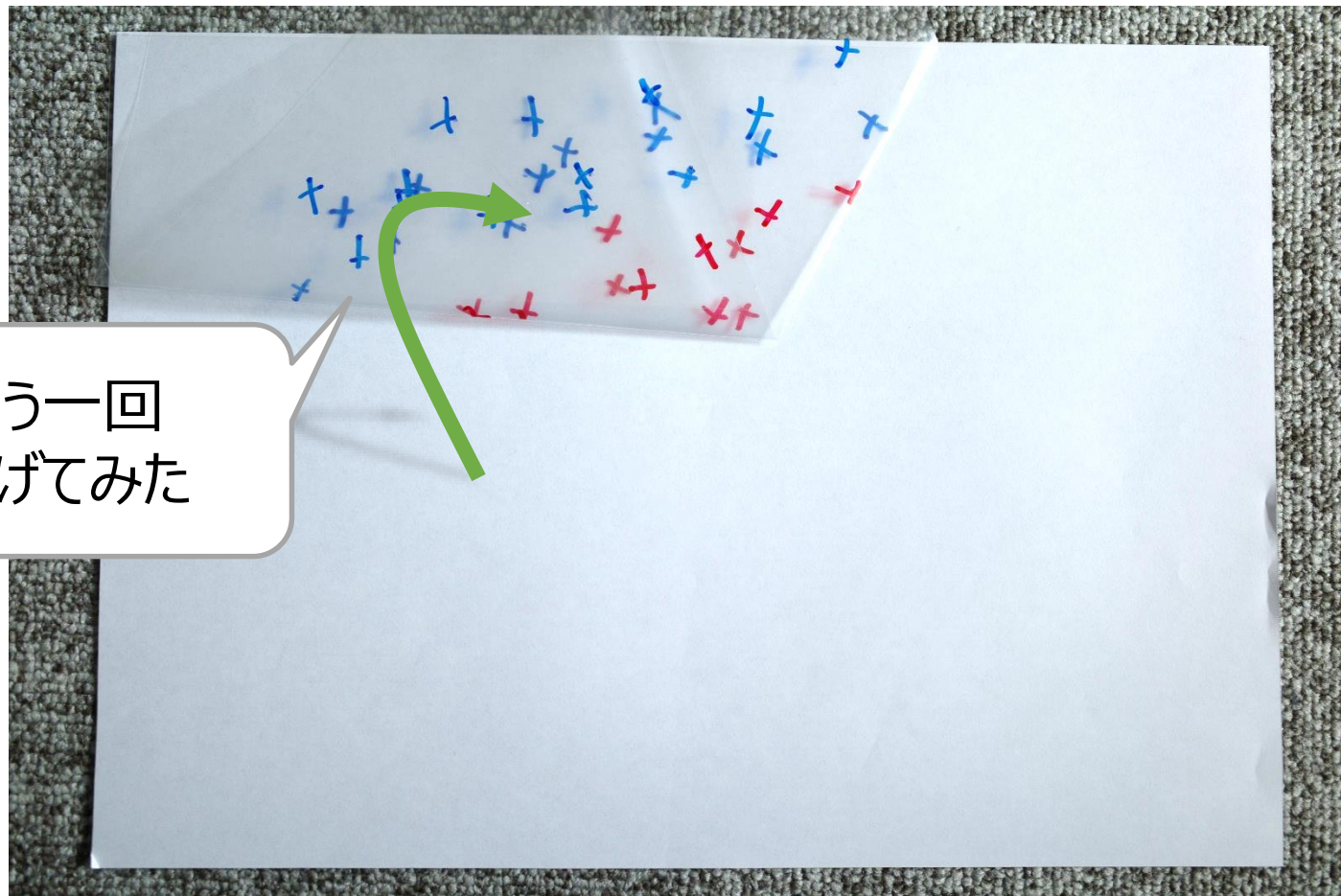


空間を
曲げてみた

識別関数法の発展版: 深層ニューラルネットワーク

最近“AI”と
呼ばれているもの

- (誤解を恐れず)もっと簡単に説明すると…

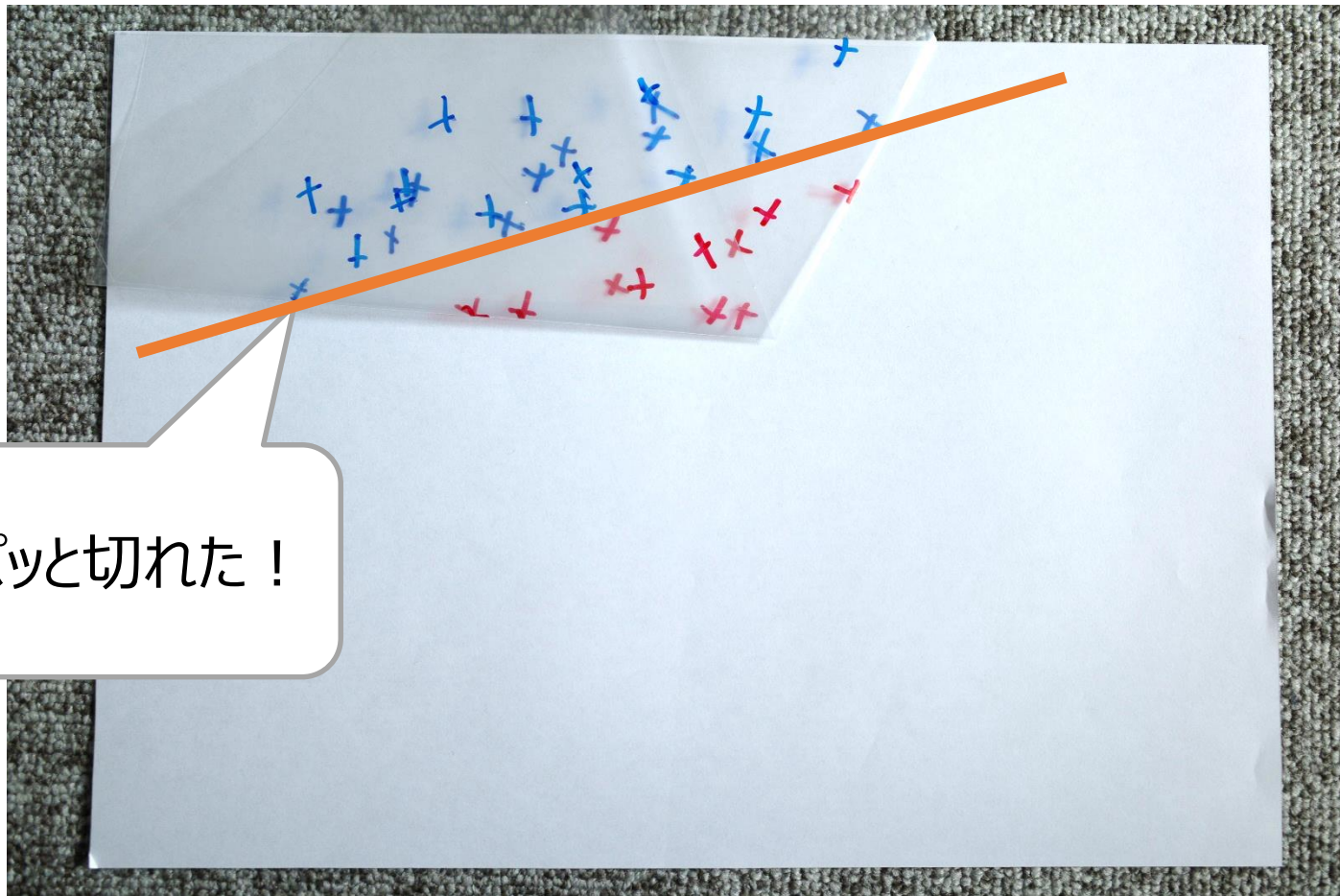


もう一回
曲げてみた

識別関数法の発展版： 深層ニューラルネットワーク

最近“AI”と
呼ばれているもの

- (誤解を恐れず)もっと簡単に説明すると…

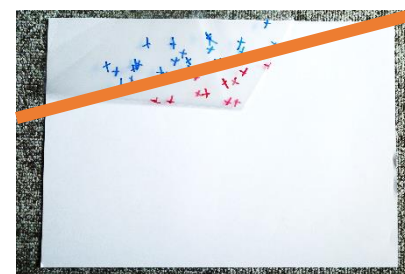
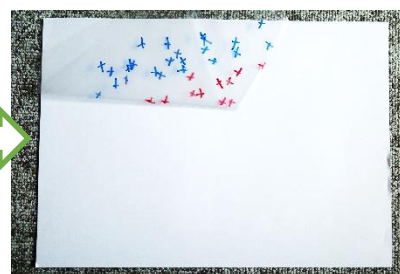
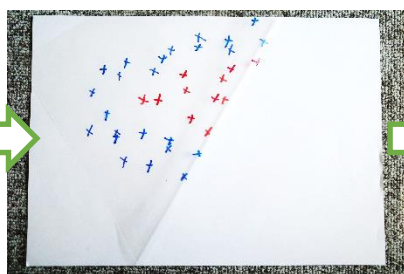
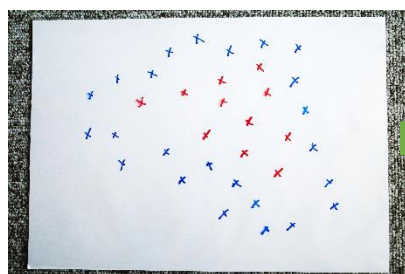
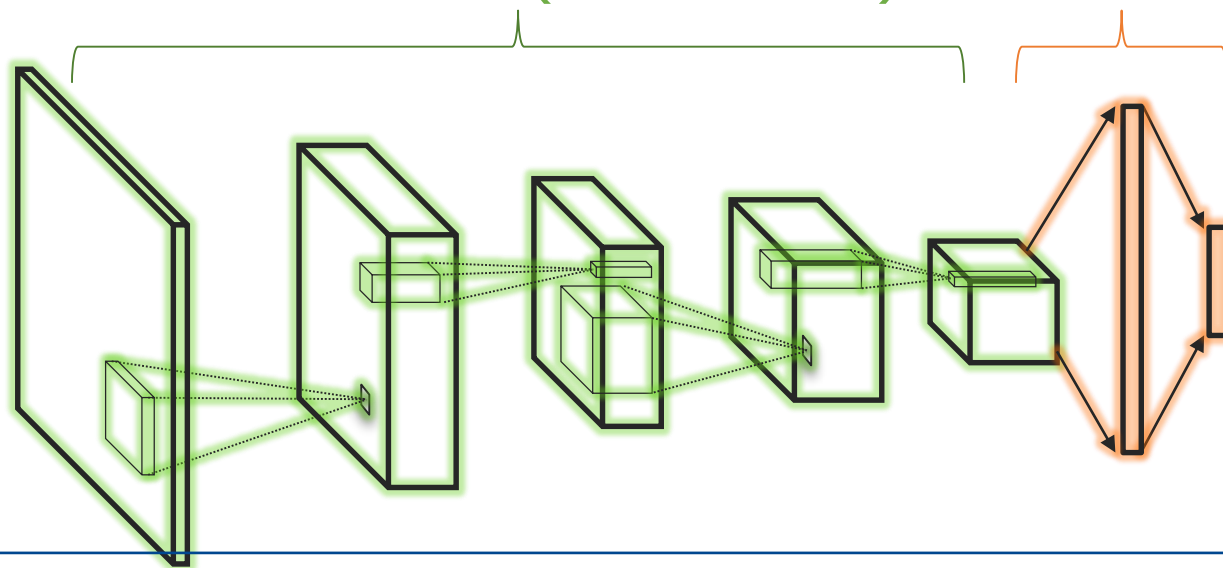


スパッと切れた！

識別関数法の発展版: 深層ニューラルネットワーク

空間を曲げる層(特徴抽出層)

スパッと切る層(識別層)

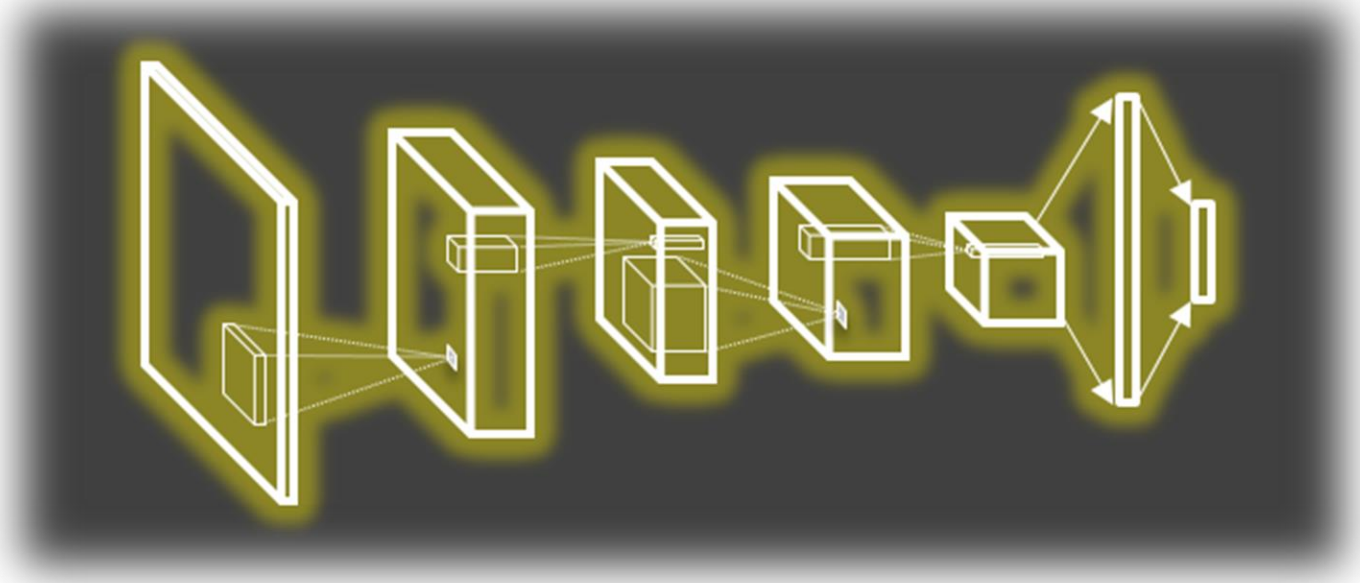


最近の状況

こんなことができるようになっていきます

最近の状況

- 深層ニューラルネットワーク(Deep neural networks, DNN)が全盛！



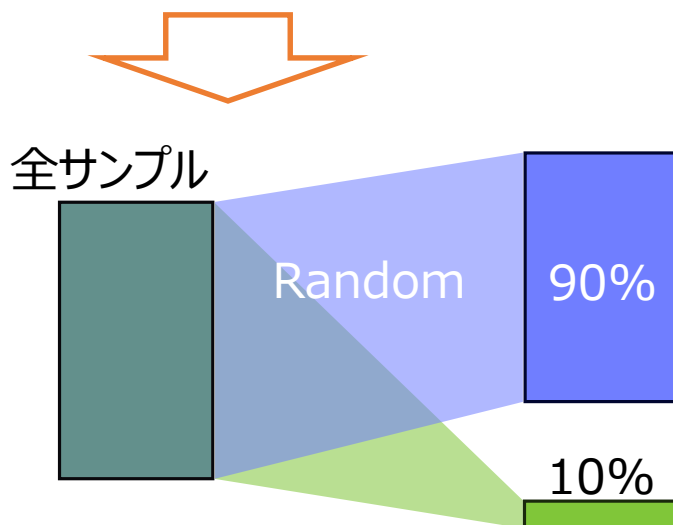
- (ひねくれ者の)私は元々「アンチDNN」だったんですが…
 - 4年前に, DNNに打ちのめされる前までは..

私がDNNに打ちのめされた経緯:

①活字文字認識

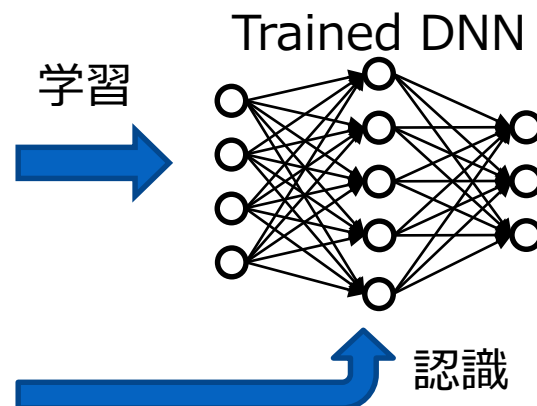
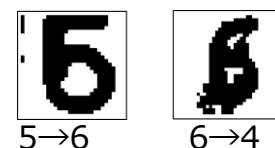
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

スキャンした活字数字
フォントは2種類程度
512,265サンプル



認識率: 99.99 %

(誤認識は2画像のみ!)



私がDNNに打ちのめされた経緯:

②手書き文字認識

20年間の研究で,
99.5%に到達し
ていたのに...

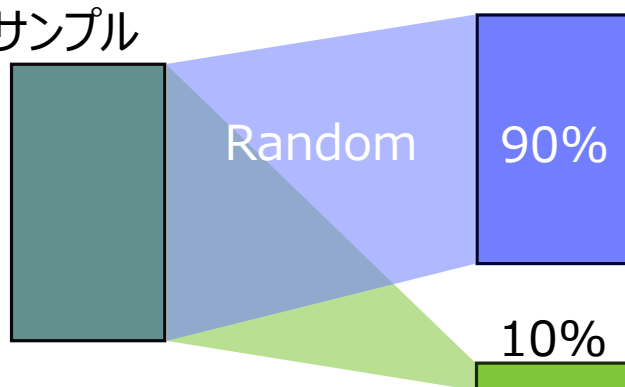


スキャンした手書き数字
不特定多数の筆記者
819,652サンプル

認識率: **99.89 %**
(誤認識は92画像のみ)



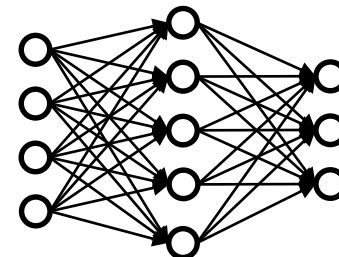
全サンプル



学習



Trained DNN

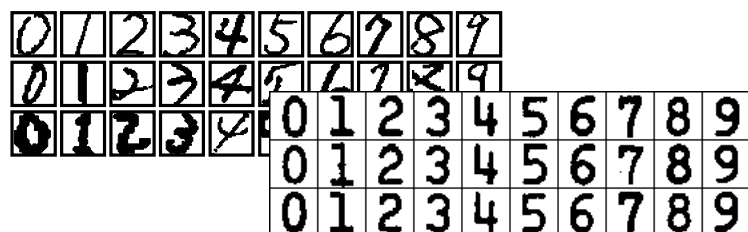


認識



私がDNNに打ちのめされた経緯:

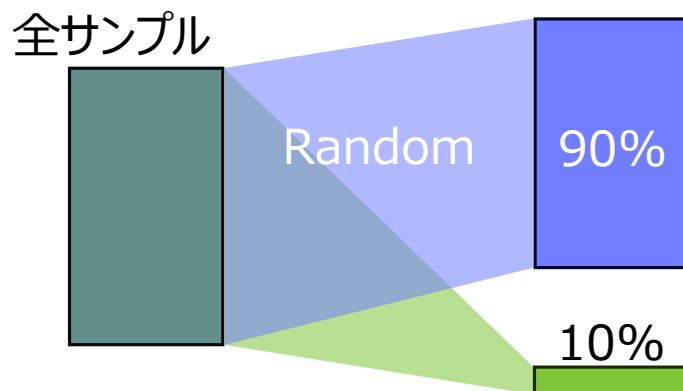
③活字と手書きの混合認識



認識率: 99.92 %

誤認識→115

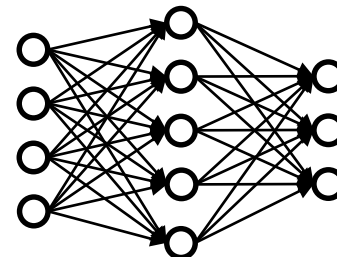
先ほどの2つのデータセットを混合
819,652+52,265サンプル



学習



Trained DNN

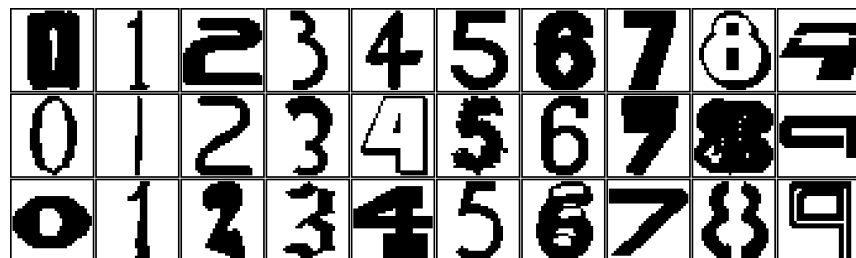


認識

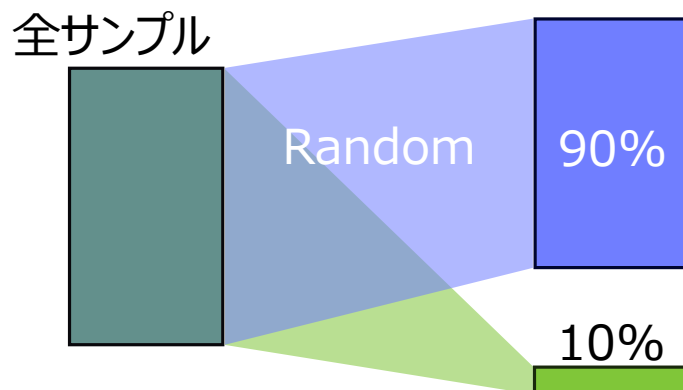


私がDNNに打ちのめされた経緯:

④デザインフォント認識



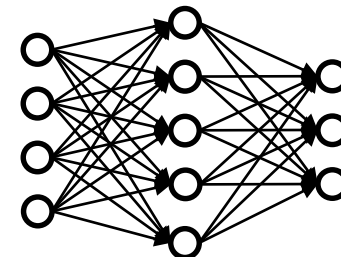
6,721種のフォント
飾り文字的なものも非常に多い



学習



Trained DNN



認識



認識率: 96.4 %





アルファベットも
余裕で認識

私がDNNに打ちのめされた経緯: ⑤難問だった情景内文字検出 & 認識も、脅威の進歩

- 検出 by EAST [Zhou+, CVPR2017]
- 認識 by CRNN [Shi+, TPAMI, 2017]



OPEN

Menu

White

pancake

panopoly

milk skin



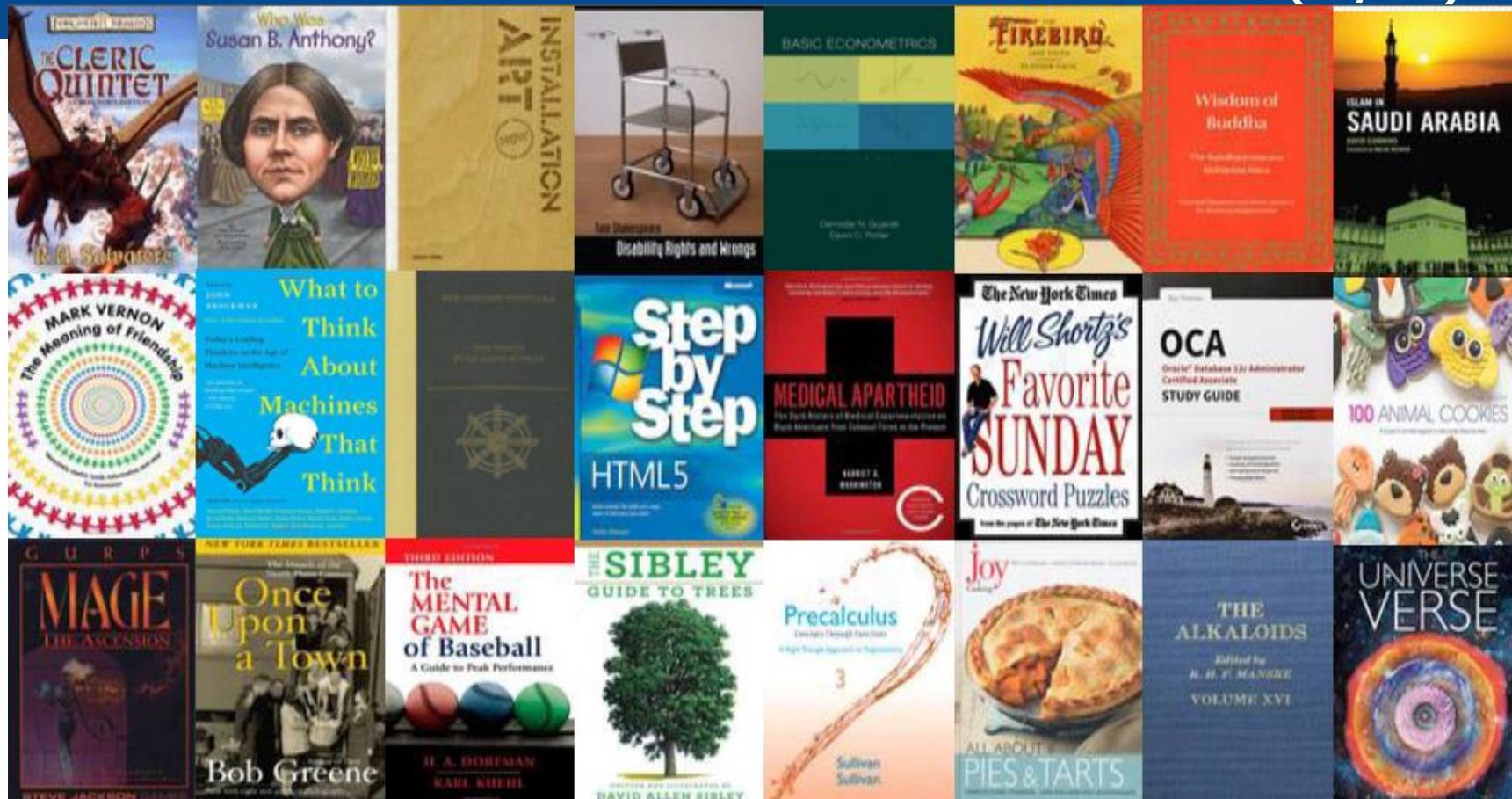
CRNNで読めた文字列

昨日の敵は今日の友！
こうなったらDNNを使い倒そう！

特に、これまで難しく
できなかったことをやろう！



DNNのおかげでできるようになった研究: 書籍ジャンルとタイトルフォントの相関解析(1/2)



- Amazonにある20万冊の書籍表紙画像を対象
- 文字検出・文字認識・フォント分類には，様々な機械学習を多用

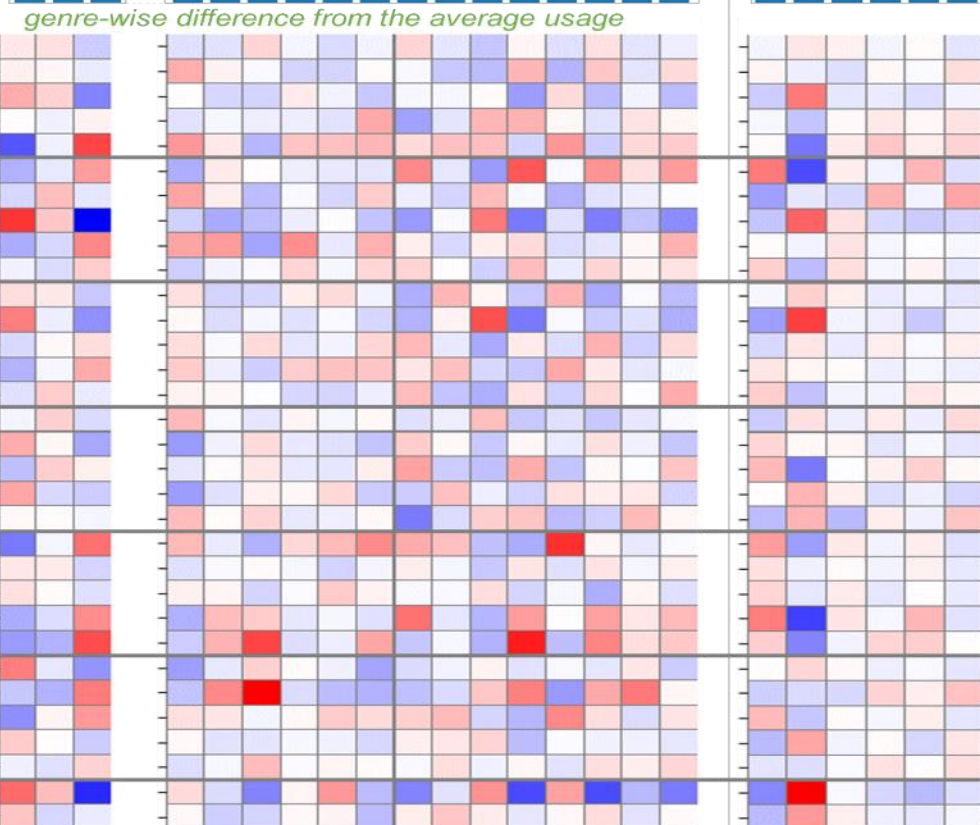
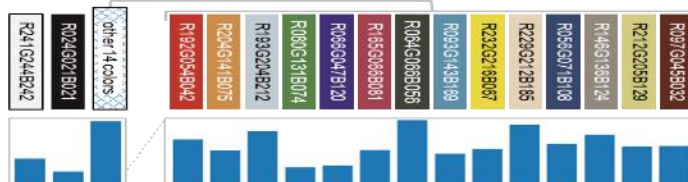
DNNのおかげでできるようになった研究: 書籍ジャンルとタイトルフォントの相関解析(2/2)

タイトル文字の色

フォント

ジャンル

ジャンル	#images
Arts & Photography	6,460
Biographies & Memoirs	4,261
Business & Money	9,965
Calendars	2,636
Children's Books	13,605
Christian Books & Bibles	9,139
Comics & Graphic Novels	3,026
Computers & Technology	7,979
Cookbooks, Food & Wine	8,802
Crafts, Hobbies & Home	9,934
Education & Teaching	1,664
Engineering & Transportation	2,672
Gay & Lesbian	1,339
Health, Fitness & Dieting	11,886
History	6,807
Humor & Entertainment	6,896
Law	7,314
Literature & Fiction	7,580
Medical Books	12,089
Mystery, Thriller & Suspense	1,998
Parenting & Relationships	2,523
Politics & Social Sciences	3,402
Reference	3,268
Religion & Spirituality	7,559
Romance	4,291
Science & Math	9,276
Science Fiction & Fantasy	3,800
Self-Help	2,703
Sports & Outdoors	5,968
Teen & Young Adult	7,489
Test Preparation	2,906
Travel	18,338



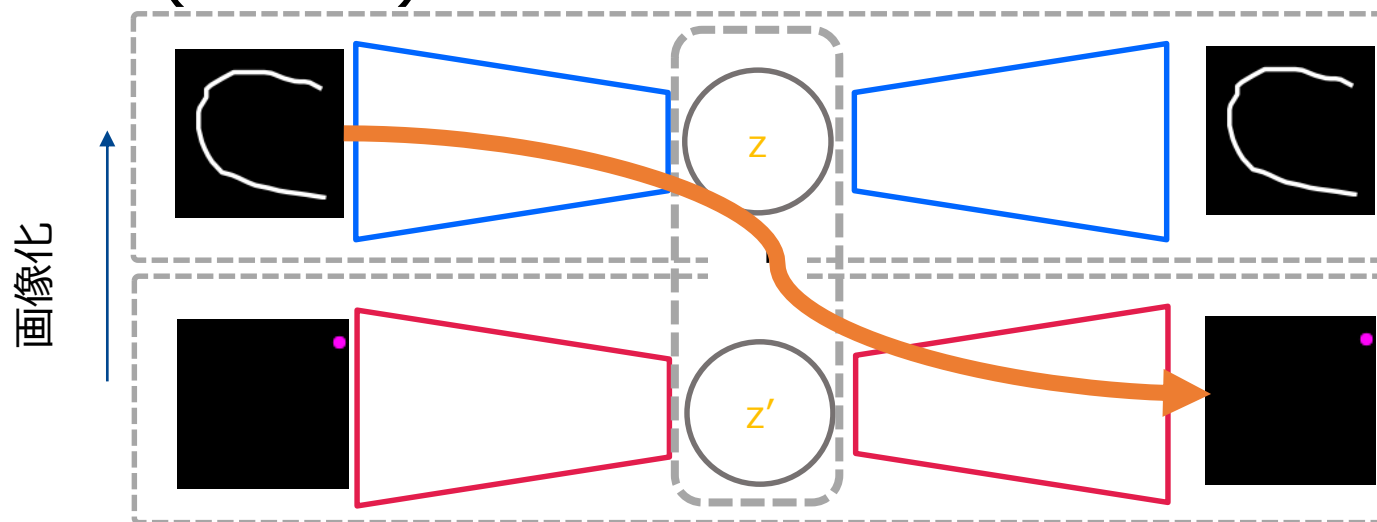
200,000
bookcover images

文字検出 by DNN
文字認識 by DNN
タイトル検出
フォントタイプ分類

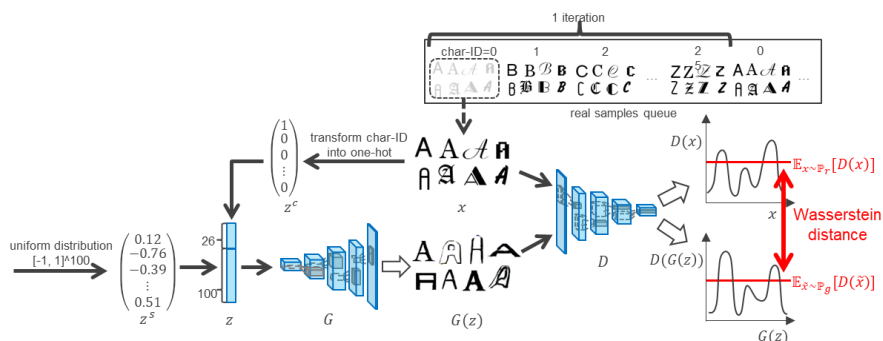
DNNのおかげでできるようになった研究： 筆順の復元(stroke recovery)



- 二つのAI(U-net)を組み合わせて実現



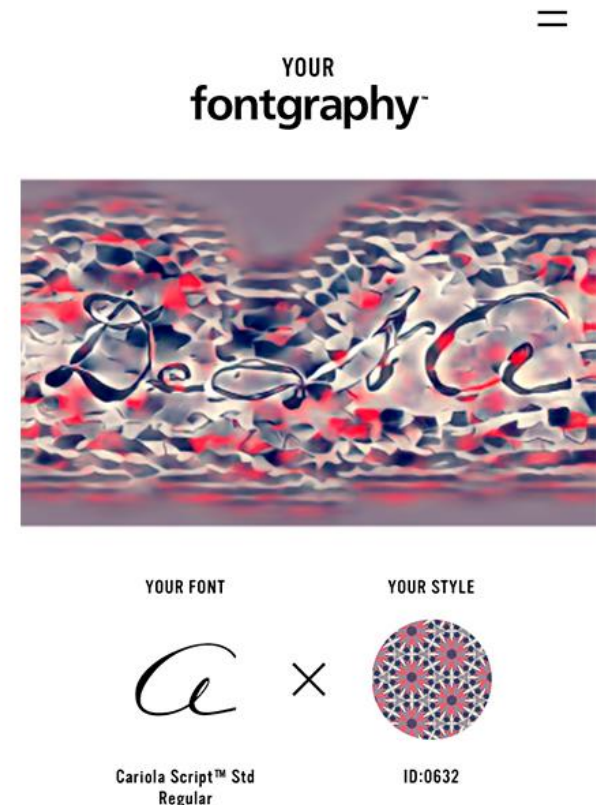
DNNのおかげでできるようになった研究: フォントの自動生成



A B C D E F G
H I J K L M N
O P Q R S T U
V W X Y Z

DNNのおかげでできるようになった研究: Fontgraphy (DeNAさんとのコラボ)

- あなたの声の「質」にあったロゴを生成します



DNNのおかげでできるようになった研究: 文字(単語)と物体の共起関係



OpenImages
174万画像

文字が書かれやすい
物体 top 10

rank	word
1	ambulance
2	calculator
3	scoreboard
4	poster
5	scale
6	ruler
7	envelope
8	fax
9	bus
10	cream

物体と単語の
共起 top 10

rank	object	word
1	car	police
2	bus	bus
3	man	army
4	bus	school
5	book	one
6	book	land
7	book	new
8	tree	park
9	person	army
10	book	book

物体認識DNN
情景内文字検出DNN
情景内文字認識DNN

物体 "person"
情景内単語 "police"

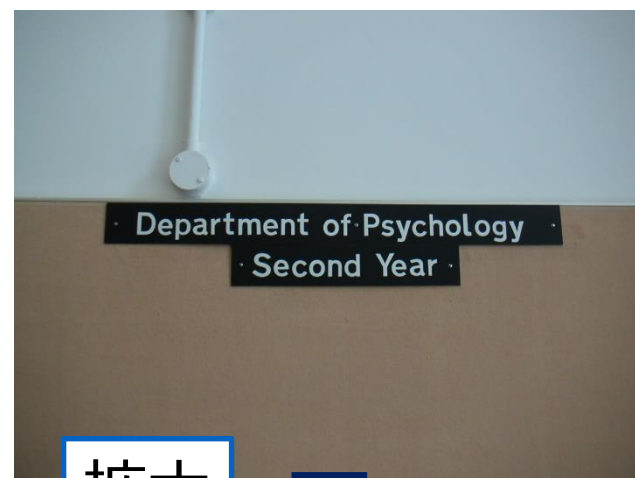
DNNのおかげでできるようになった研究: 文字消しゴム & 文字のみ拡大



消しゴム



[Nakamura+, ICDAR2017]



拡大



[Nakamura+, ICDAR2019]

最近では, こんな国際コンペも: ICDAR 2019 Robust Reading Challenge on Scene Text Visual Question Answering (VQA)

- VQAとは: 画像 + 言葉での質問 → (DNNで) 回答作成



- Scene Text VQA



古文文献認識の国際ワークショップ(HIP2019)でも DNNを利用した研究の発表が多かった

Day 1 - September 20 (Building 10, Aerial Function Centre, Broadway Room)

From To

09:15 09:30 **Welcome**

09:30 10:30 **Session 1 - Datasets** (Chair: Clemens Neudecker)

Dataset of Pages from Early Printed Books with Multiple Font Groups *Mathias Seuret, Saskia Limbach, Nikolaus Weichselbaumer, Andreas Maier and Vincent Christen*

HORAE: an annotated dataset of books of hours *Mélodie Boillet, Marie-Laurence Bonhomme, Dominique Stutzmann and Christopher Kern*

BADAM: A Public Dataset for Baseline Detection in Arabic-script Manuscripts *Benjamin Kiessling, Daniel Stökl Ben Ezra and Matthew*

10:30 11:00 **Break**

11:00 12:20 **Session 2 - Training** (Chair: Christopher Kermorvant)

Generating Cuneiform Signs with Cycle-Consistent Adversarial Networks *Eugen Rusakov, Kai Brandenbusch, Denis Fisseler, Turna So and Gerfrid Müller*

okralact - a multi-engine Open Source OCR training system *Konstantin Baierer, Rui Dong and Clemens Neudecker*

Using Balanced Training to Minimize Biased Classification *Redy Andriyansah, Syed Saqib Bukhari, Martin Jenckel and Andreas Dengel*

Recognition of Japanese historical text lines by an attention-based encoder-decoder and text line generation *Anh Le Duc, Daichi Mima and Ly Tuan Nam*

12:20 13:00 **Discussion**

13:00 14:10 **Lunch**

14:10 15:30 **Session 3 - Forms, Tables, Pattern Spotting** (Chair: Elisa H. Barney Smith)

Crowdsourcing Historical Tabular Data – 1961 Census of England and Wales *Christian Clausner, Justin Hayes and Apostolos Antonacopoulos*

Keeping Informed: Automatic Processing of Residual Functional Capacity Form Images *Julia Porcino and Chunxiao Zhou*

Signature detection as a way to recognise historical parish register structure *Solène Tarride, Aurélie Lemaître, Bertrand Couasnon and Sophie Tardivel*

Pattern Spotting in Historical Documents Using Convolutional Models *Ignacio Ubeda, José M. Saavedra, Stéphane Nicolas, Caroline Petitjean and Laurent Heutte*

15:30 16:00 **Break**

16:00 17:20 **Session 4 - HTR, Matching, Layout Analysis** (Chair: Bertrand Couasnon)

Information Extraction in Handwritten Marriage Licenses Books *Verónica Romero, Emilio Granell, Alicia Fornés, Enrique Vidal and Joan Andreu Sánchez*

Recognition of Japanese Historical Hand-Written Characters Based on Object Detection Methods *Yiping Tang, Kohei Hatano and Eiji Takimoto*

Papy-S-Net : A Siamese Network to match papyrus fragments *Antoine Pirrone, Marie Beurton-Aimar and Nicholas Journet*

Layout Analysis and Text Column Segmentation for Historical Vietnamese Steles *Anna Scius-Bertrand, Lars Voegtlin, Michele Alberti, Andreas Fischer and Marc Bui*

17:20 17:30 **Wrap-up**



盛況でした

バツカスビル

INFORMATION

3 F らん 朱すめ らん 朱すめ II らん 朱すめ III SNACK M Kei シユガ- SUGAR RAY らうんじ 夏目	5 F マロアベル スナック Rロイヤル TIME プレイス.23 AFTER DOLL Bar ゆうこ	7 F KEY-SITE Charlbourg ねん Z いすせ 芳 賀 at Rie Frankly BAR ぷらす+ぷらす	9 F とれぞあ Earth Loisir Lounge ペガサス 心音 COCO-NE Wine Retreat ARCHANGE イタリ-料理 ガルバンゾ MUSIC LOUNGE Brave+
2 F 海 すなっく 寿 SAKI LUNA 居酒屋 てらお そば処 丸山 すし KAN むらかみ 北の味 のつけ 山 吉	4 F ふるか-わ めい 鈴 は な 江に家 ライブ ステイング-レイ Ki Ra Ra 果 琳 旬食酒房 こと Mar maara	6 F 蓮華 マイバフ カザリコ 居酒屋 魚菜亭 ふみ @ chat. lounge Madoka SNACK ギルダ- 蓮華	8 F 彩華 蓮石 Mix Bar Iris Pub salon L.A.La フチキスコチック Bar Reif BAR ひづき Chouette BAR HOKUTO ルームイン 雅やび 明洞 명동 PHANTOM
1 F B1 F 4 F 加藤ビルディング B1F 警備・管理室			

今後の課題

文字研究はこれからがオモシロイ!



さらに**困難な文字認識**課題に挑む！



- もちろん古文獻，特にくずし字
 - 歴史的事実を知るためにも！
- レイアウト自由手書き(≡手書きノート)
 - 数式や手書きの図が混在すると非常に困難！

記述式入試の
自動採点...!?

● 帳票構造認識

Order Form		Order id:		Order date: (dd/mm/yy)	
Customer Name				Gender	
Given	Family				
Address	ZIP			Country (check one)	Pakistan Germany Japan
Order list	Item	Qty	Price (tax)	Note	
			total \$ ____ ()		
Payed? (y/n)		Payment	Cash \$ ____	Credit card \$ ____	Wire trans. \$ ____

● マンガ認識

- オノマトペの
認識は難しそう

ここには
何を書く？

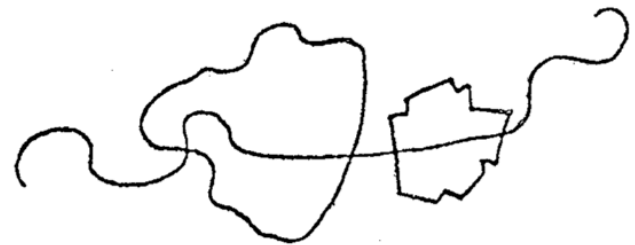
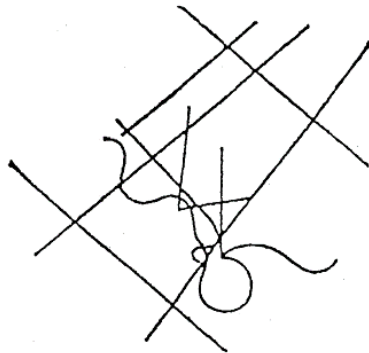
オノマトペ



“Akkerakanjincho” Kobayashi Yuki

「なぜ文字は読めるのか」を考える！

- 文字はノイズが印加されても、下手でも、結構読める！



[小川英光+, IEICE1994]

● 課題

- Q. なぜ下手な字でも読めるのか？
- Q. どれぐらい下手な字でも読めるのか？
- Q. 人間が文字を作った際に、変形耐性を考えたのか？
- Q. どうすれば読めなくなるのか？
- Q. 文字と非文字の厳密な区別は可能か？

手書き文字について考える！

● Q. 筆記を数式で表現できるか(=モデル化できるか)？

- ミクロ = 連続的にペン先を動かす局所的構造
- メゾ = 腕の運動モデル
- マクロ = 全体のバランスを保つための大局的構造

統合すべき？

● Q. 手書きから筆記者の情報を推定できるか？

- Q. 年齢の影響は？
- Q. 性別の影響は？
- Q. 気分・感情・性格の影響は？
- Q. 病気の影響は？



- パーキンソン病, アルツハイマー, 失読症(ディスレクシア)等

Catherine Metzger
13 Octobre 1869

フォントと印象の関係を探る！

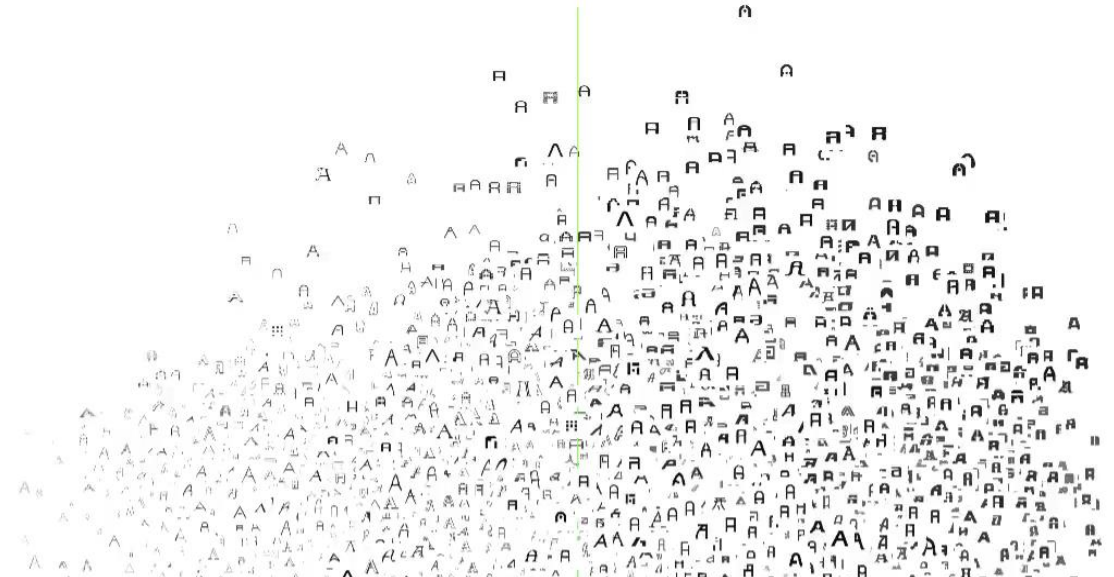
Dior

Dior

Dior

● 課題

- Q. 特定印象に合致したフォントを選ぶか？
- Q. 背景画像/TPOにマッチしたフォントを選ぶか？
- Q. 特定印象を持ったフォントを自動生成できるか？
- Q. フォントのどのような要素がどのように印象を決定？
 - 全体的な形なのか？ 局所的な形なのか？

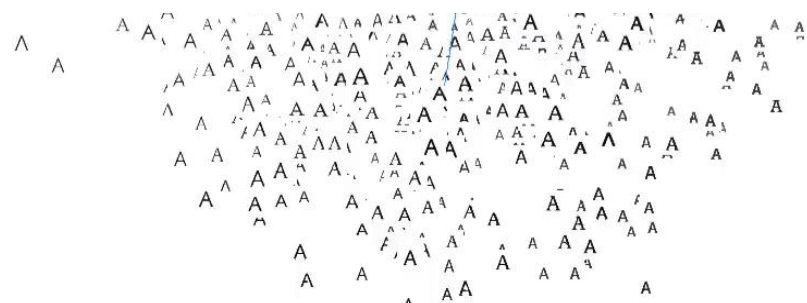


ご清聴ありがとうございました！



Take-home message:

文字研究はこれからがオモシロイ！



(二二から番外編)
kaggleコンペを含み
本日参加させていただいての
メッセージ

九大 情報 内田誠一

メッセージ 1 : 是非新規参入を！

- だれもが輝ける時代！「分野が違う」はそろそろ終わり
 - 大昔に前にできた「学部」「分野」とかは忘れたほうがよい？
 - 豊富なHow-to 情報
 - 世界最先端のソフトウェアが無料で落ちている
 - 1 時間で目の前で世界最先端が動く
- 最初の環境設定さえできれば！
 - その情報もネットにあり．困ったら周りの誰かに頼む？
- AIで職を失うのは、私のような画像認識研究者かも

メッセージ 2 : くずし字に特化したAIができると楽しそう

- 既存手法をそのまま使うだけでなく、くずし字の性質に特化した工夫ができれば、ご検討がさらに深化！
 - 今回も、「注目文字の周囲を見る工夫」などが散見
 - キャンベル先生もいまおっしゃってました
- そういう「対象の性質を見る」スキルは、他の課題を扱う場合でも活きます！
 - 深層ニューラルネットワーク全盛の時代でも、そう思います

メッセージ 3 : 認識結果の解析が、きつともつともつと面白い

- 認識は解析の第一歩目
 - 画像データにアノテーション(テキスト情報)が自動で付いた段階
 - スタート地点に立った？
- テキスト(= 認識結果)の解析にもAIを
 - 自然言語処理もAI全盛の時代
 - 自動現代語訳などできると、さらにオモシロイ？
- 各文字種での字形の解析なんかも、きつとオモシロイ！

文理を問わず



Enjoy!