

Computerの視覚と感性 について

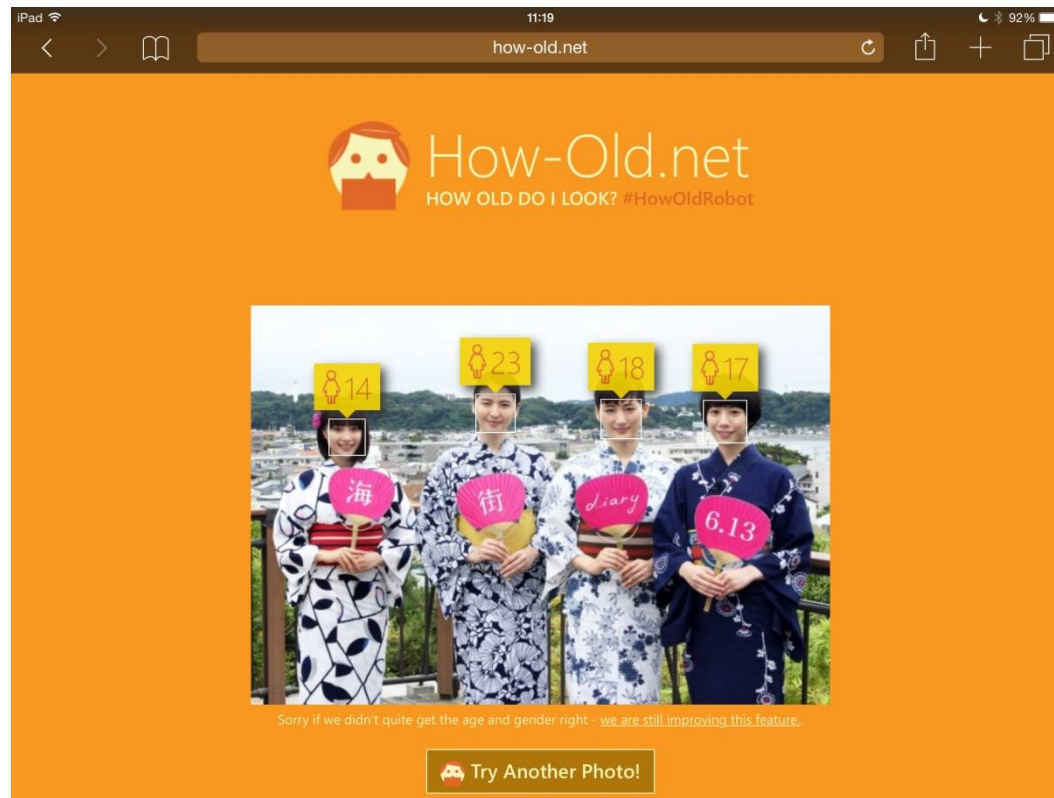
戴 イン

Outline

- ▶ 事例
- ▶ 自動認識システムの仕組み
- ▶ 顔(年齢)認識システム
- ▶ 顔表情から感情を認識するシステム
- ▶ 人間の感性を考慮する画像検索システム
- ▶ Big sensor dataを利用する健康状態推測システム
- ▶ Deep learningを用いた自動認識システム
- ▶ 卒業研究紹介
- ▶ まとめ
- ▶ 参考文献

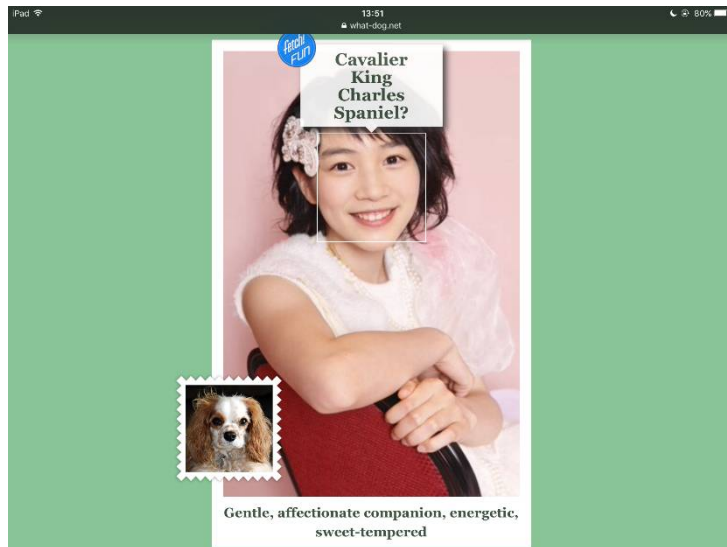
事例

▶ MicrosoftのHow Old Do I Look



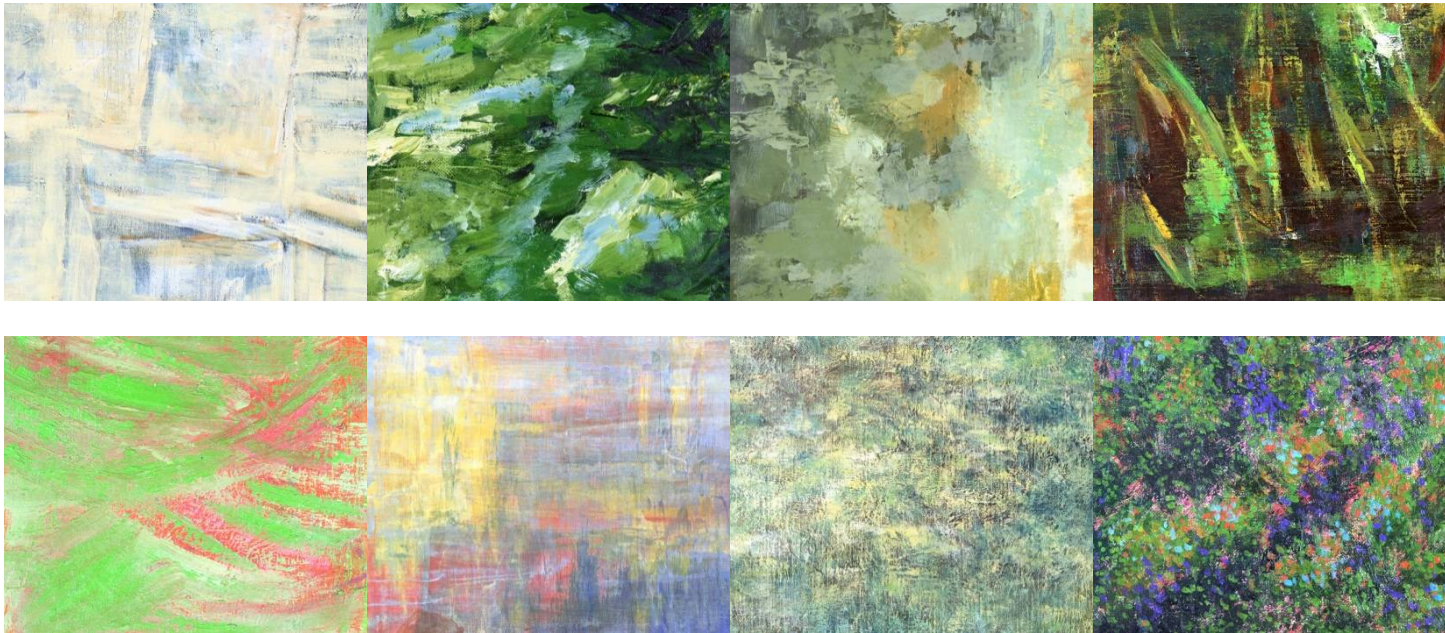
事例

▶ MicrosoftのWhat-Dog.net



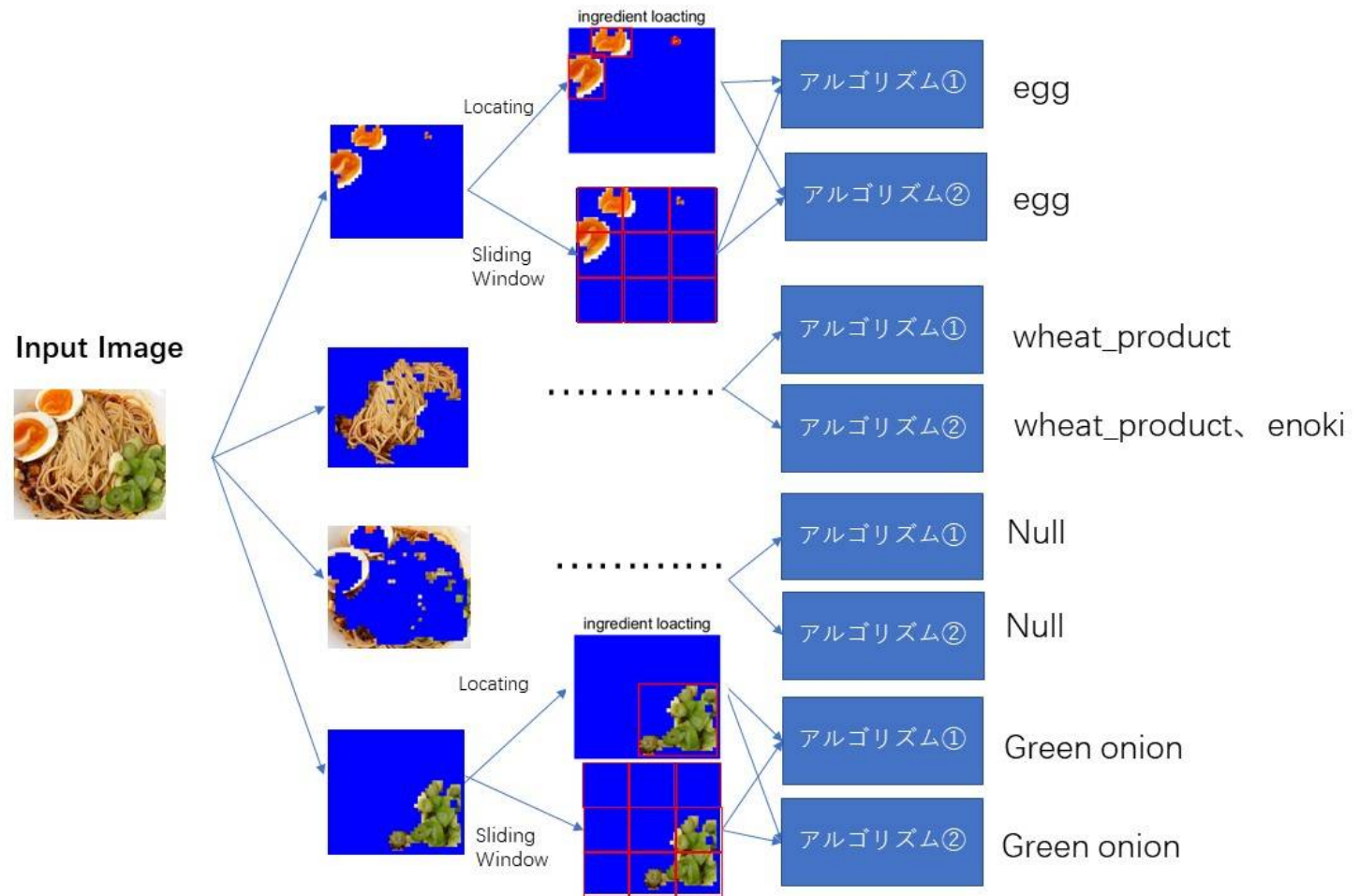
事例: Imageryによる画像検索システム

- ▶ 質問: パステル画の中で、多彩で穏やかに見える画像



Retrieved results

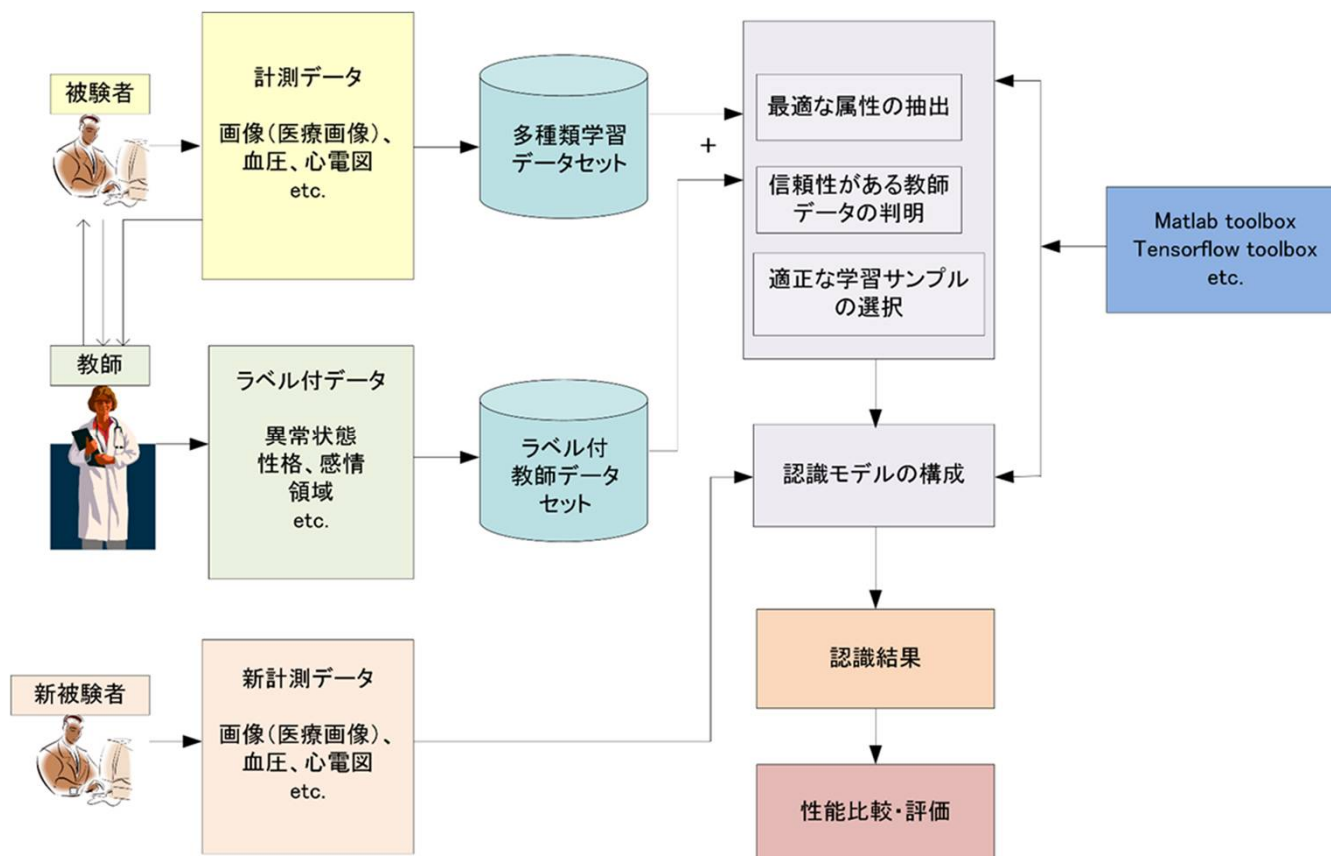
事例：料理画像からの食材識別



課題

- ▶ 機械に人間の視覚や感性を持たせることができる
- ▶ 工夫することにより性能が高める
- ▶ 実現する方法は？

自動認識システムの仕組み



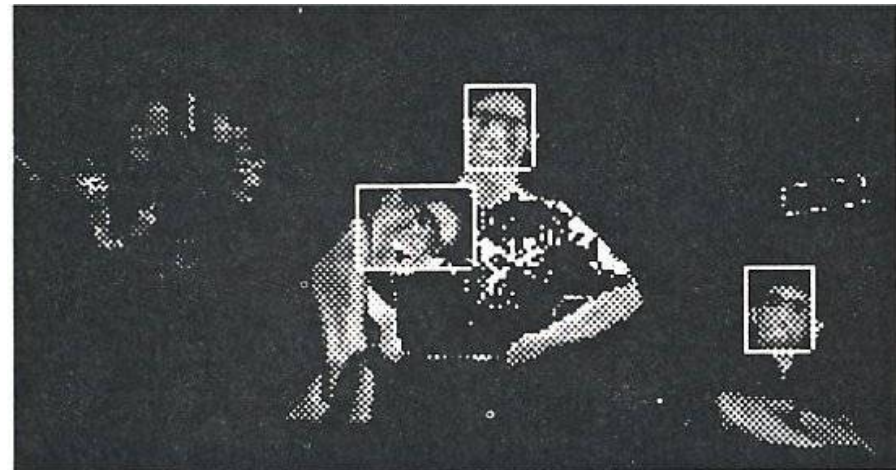
研究キーワード

- ・ 画像認識
 - ・ 画像検索
 - ・ 感性認識
 - ・ deep learning
 - ・ self-revised learning
- etc.

顔(年齢)認識システム

▶ 顔位置の特定

- Ying Dai, et al., “Space Gray-Level Dependence matrix (SGLD) of face pattern for face location”, Pattern Recognition, Vol.29, pp. 1007–1017 (1996)



顔(年齢)認識システム

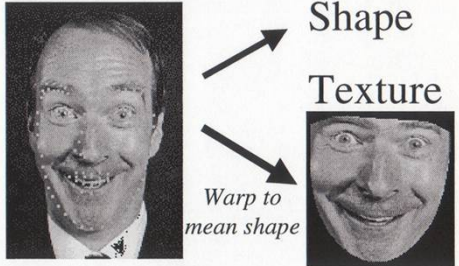
- ▶ 認識ための特徴の抽出
 - Eigen face
 - 顔の各部分の間の距離
 - Appearance models

Cootes *et.al.* University of Manchester, UK


From a training set learn model of shape and texture variation

Shape: $\mathbf{x} = \mathbf{x}_{mean} + \mathbf{Q}_s \mathbf{c}$

Texture: $\mathbf{g} = \mathbf{g}_{mean} + \mathbf{Q}_g \mathbf{c}$



Varying one parameter of model:



顔(年齢)認識システム

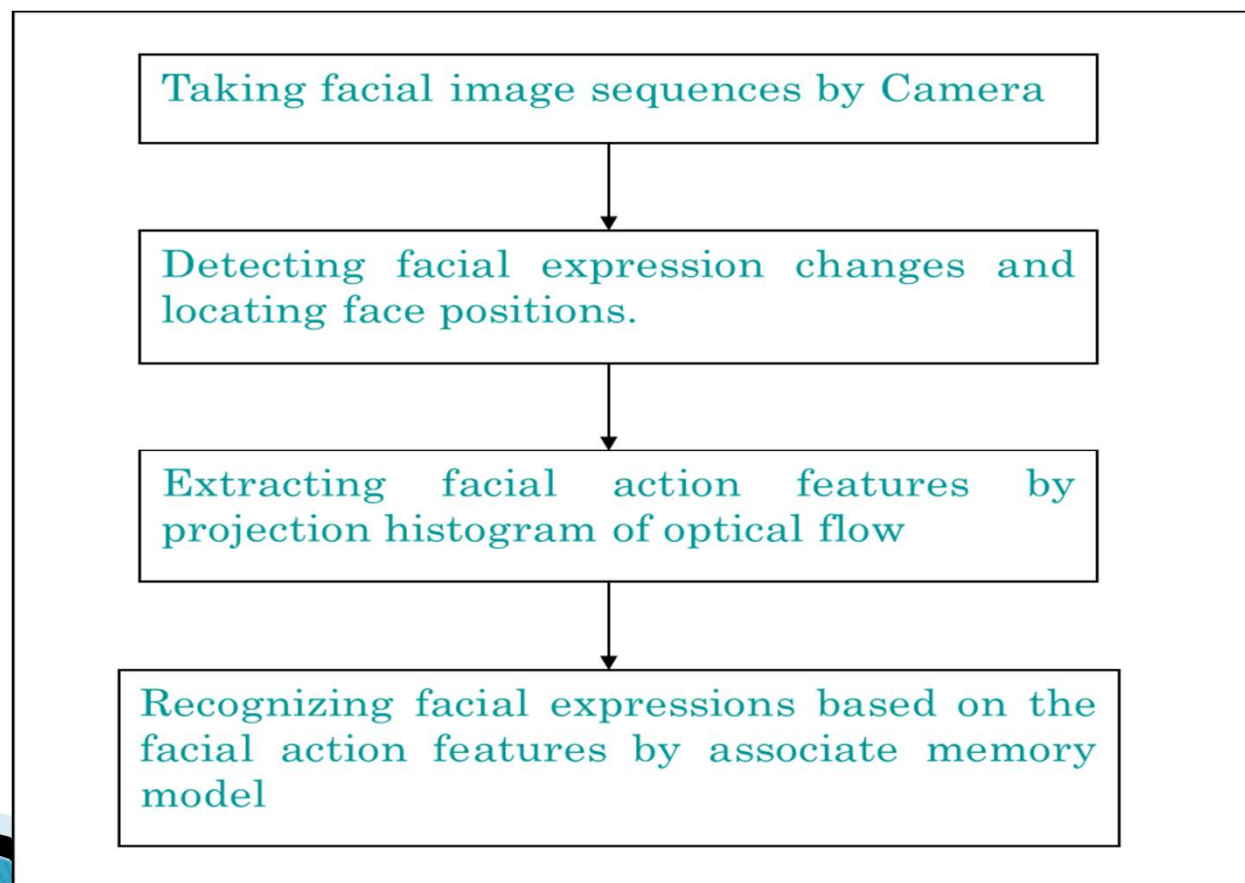
▶ 認識

- 特徴に関わる学習データを用意
- 学習データに訓練させることにより、認識モデルを構築
 - Template matching
 - Neuro network (associated memory model)
 - Deep learning

Y. Dai, et al. Recognition of Facial Images with Low Resolution Using a Hopfield Memory Model, Pattern Recognition, Vol. 31, pp.159-167 (1998)

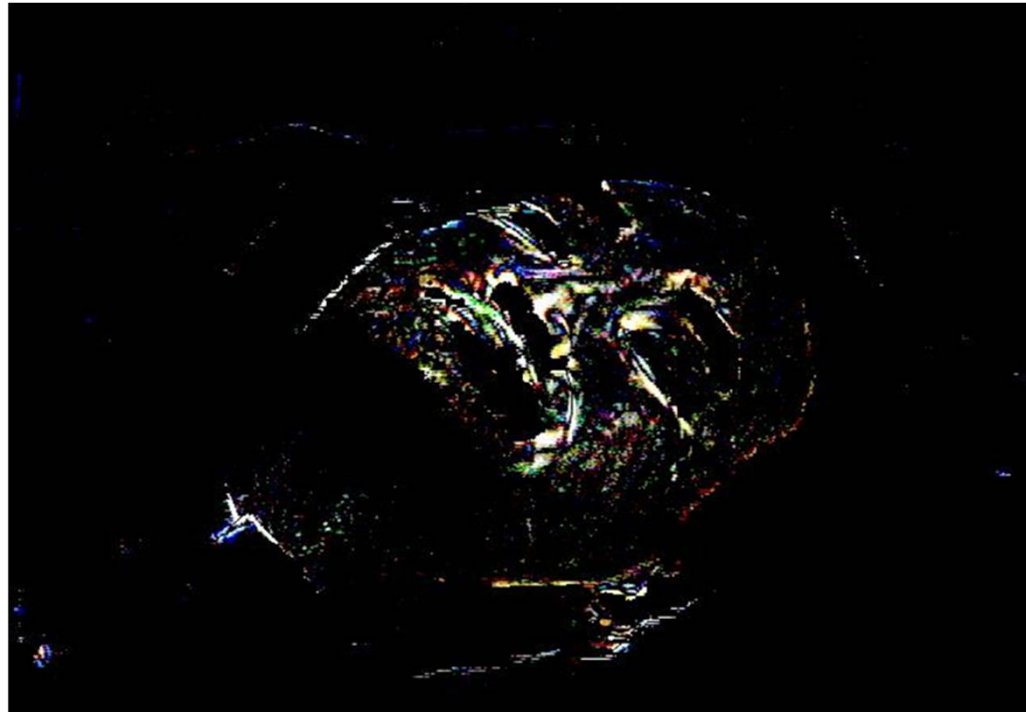
顔表情から感情を認識するシステム

- ▶ Ying Dai, et al. “Understanding of facial expressions by the hierarchical recognition of genuine emotions”, International Journal of Innovative Computing, Information and Control, Vol. 1, Number 2, pp. 203–214, June 2005
- ▶ 流れ図









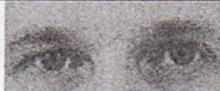








顔表情から感情を認識するシステム

- ▶ 表情変化の抽出



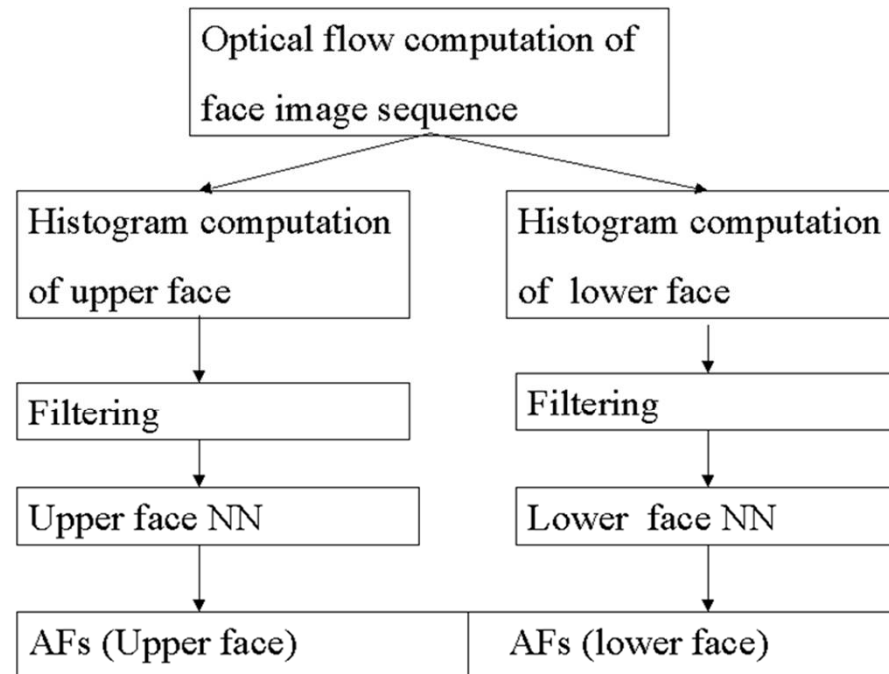
顔表情から感情を認識するシステム

- ▶ Facial action coding system((Ekman and Friesen 1978))

<i>NEUTRAL</i>	AU 1	AU 2	AU 4	AU 5
				
Eyes, brow, and cheek are relaxed.	Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together	Upper eyelids are raised.
AU 6	AU 7	AU 1+2	AU 1+4	AU 4+5
				
Cheeks are raised.	Lower eyelids are raised.	Inner and outer portions of the brows are raised.	Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.
AU 1+2+4	AU 1+2+5	AU 1+6	AU 6+7	AU 1+2+5+6+7
				
Brows are pulled together and upward.	Brows and upper eyelids are raised.	Inner portion of brows and cheeks are raised.	Lower eyelids cheeks are raised.	Brows, eyelids, and cheeks are raised.

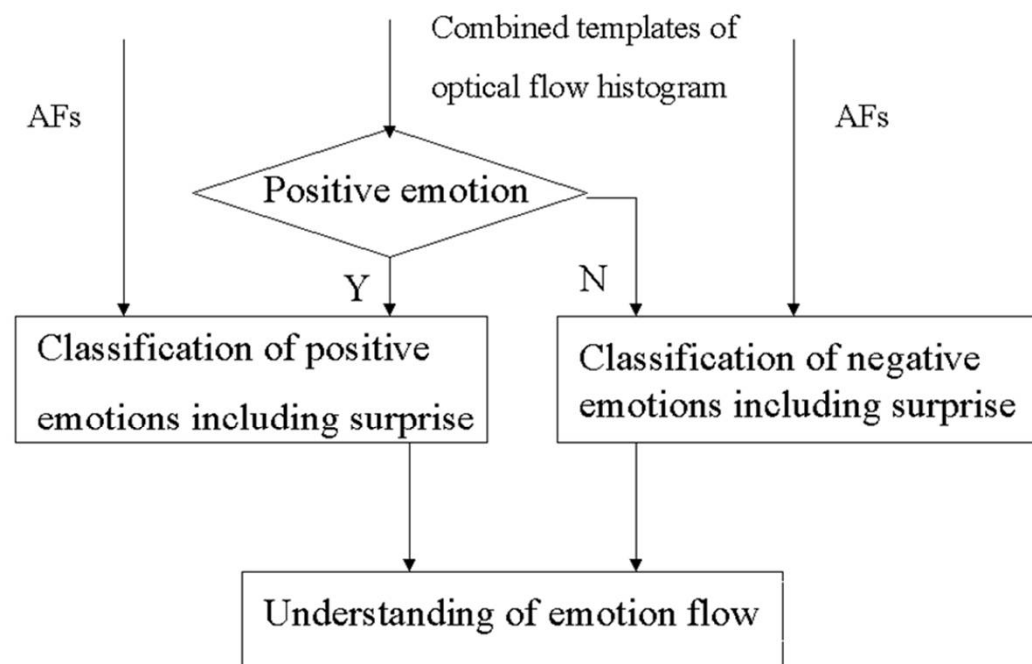
顔表情から感情を認識するシステム

▶ Facial action features (AFs)の認識



顔表情から感情を認識するシステム

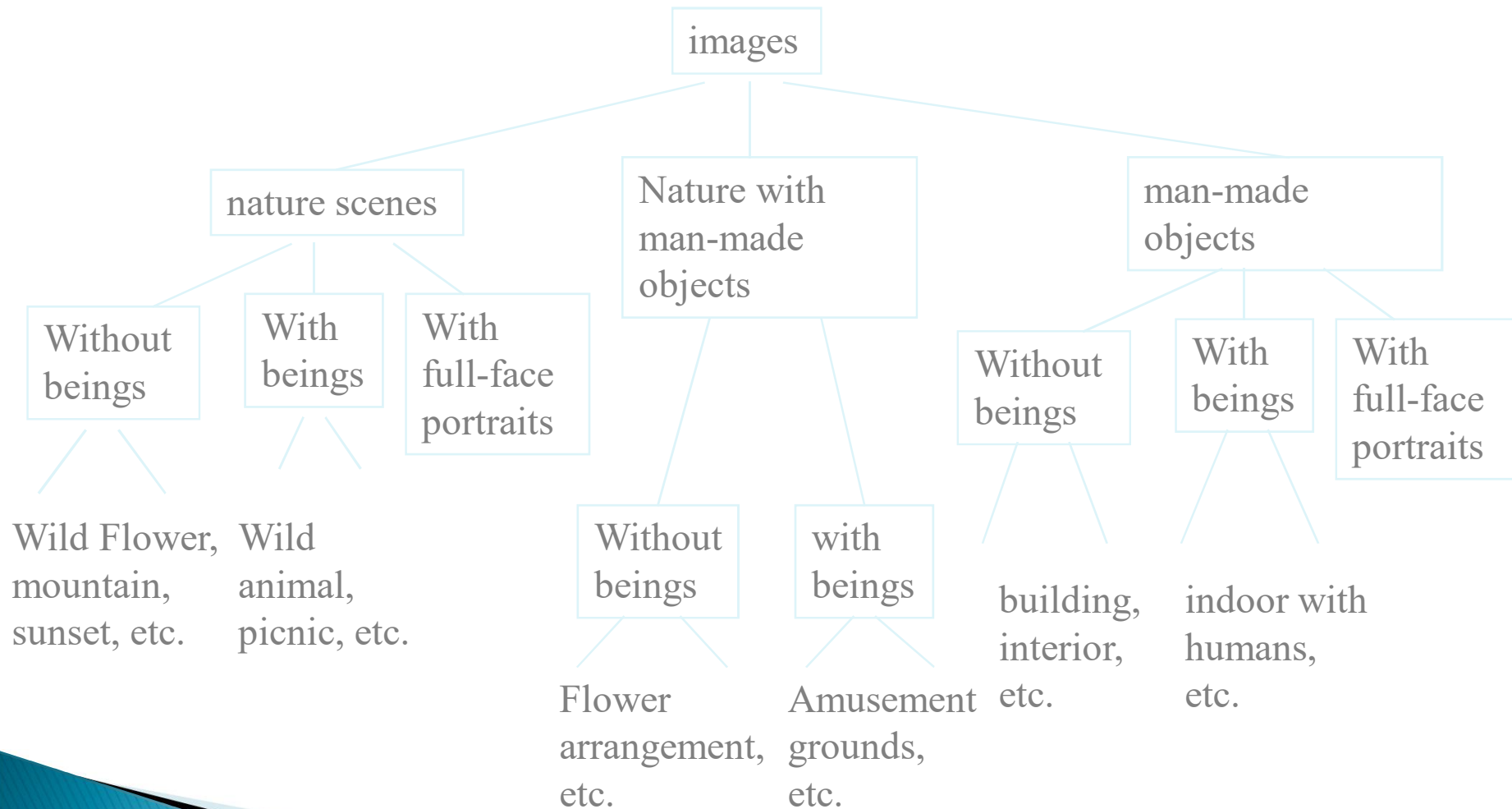
▶ AFsから感情を認識



Positive emotion group: happiness, easiness, surprise

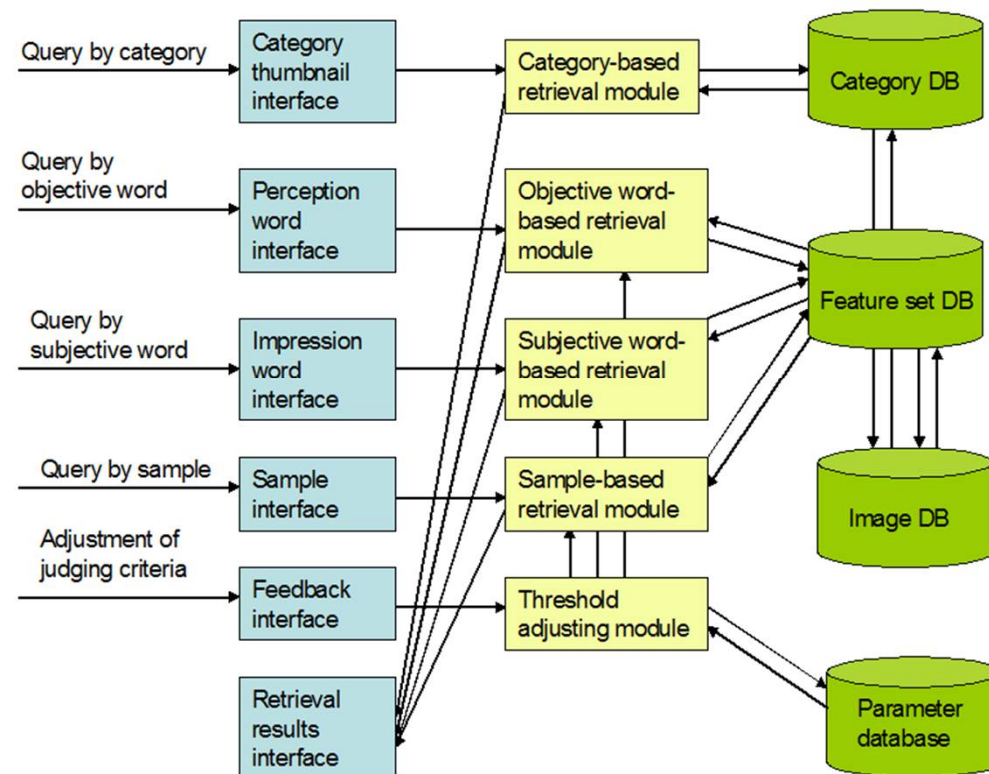
Negative emotion group: uneasiness, disgust, suffering, surprise

画像の意味的なカテゴリー

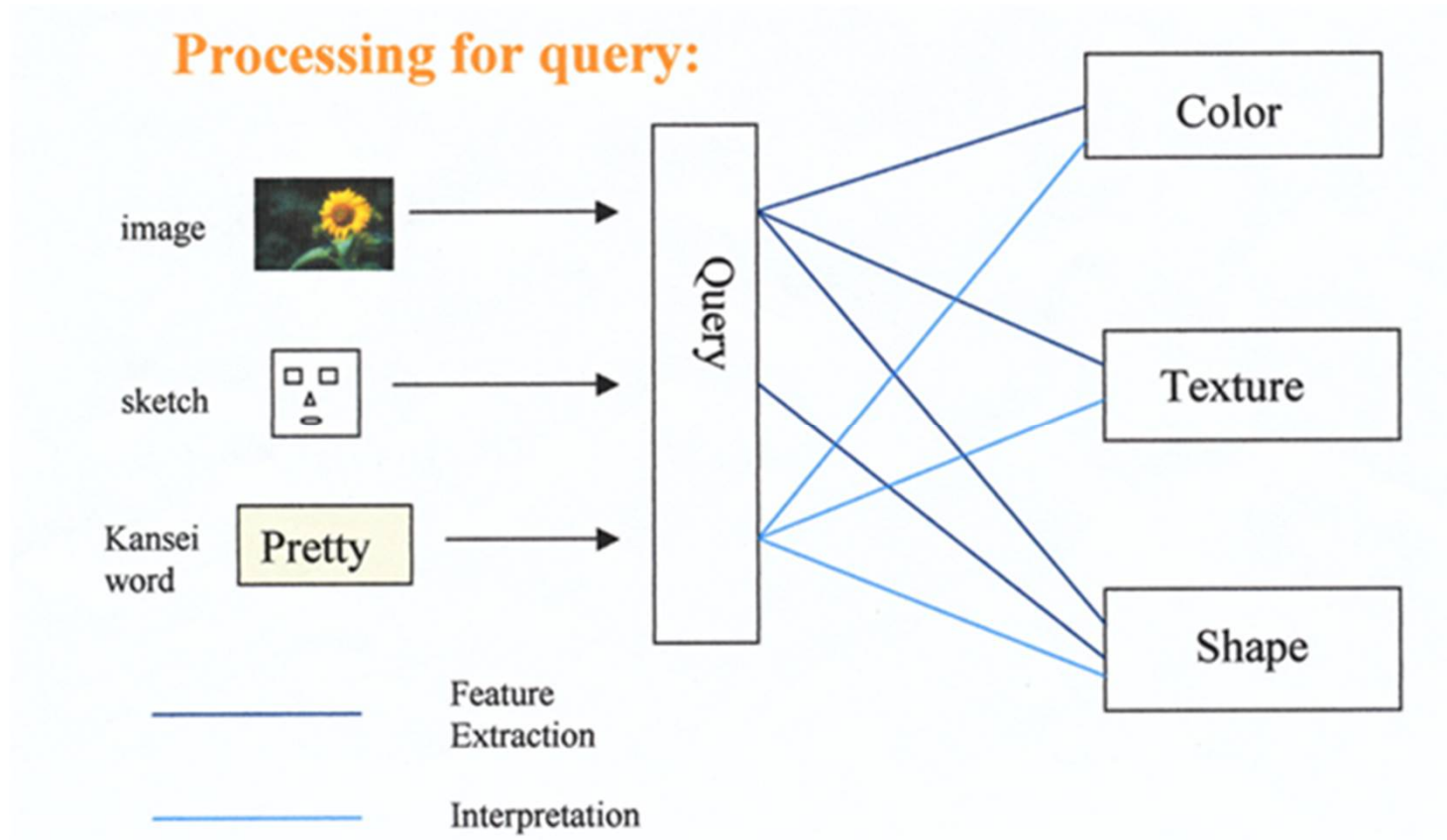


人間の感性を考慮する画像検索システム

- ▶ Ying Dai, “Class-based image representation for Kansei retrieval considering semantic tolerance relation”, Journal of Japan society for fuzzy theory and intelligent informatics, Vol. 21, No. 2, pp. 184-193 (2009).
- ▶ 仕組み



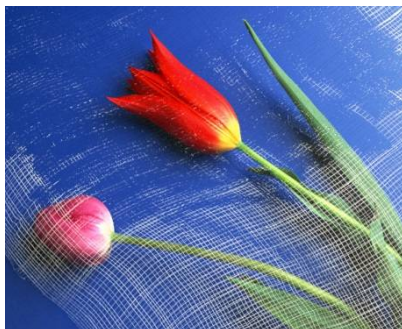
Queryの解析と解釈



画像の物体分け



Original image



Original image

Computerで画像検索の結果

- ▶ 質問: 花の画像の中で、参照画像の色と似ている画像



A reference



Retrieved results (5 best matches)

Some image retrieval results

- ▶ 質問: 花の画像の中で、参照画像の総体的な色と構造と似ている画像



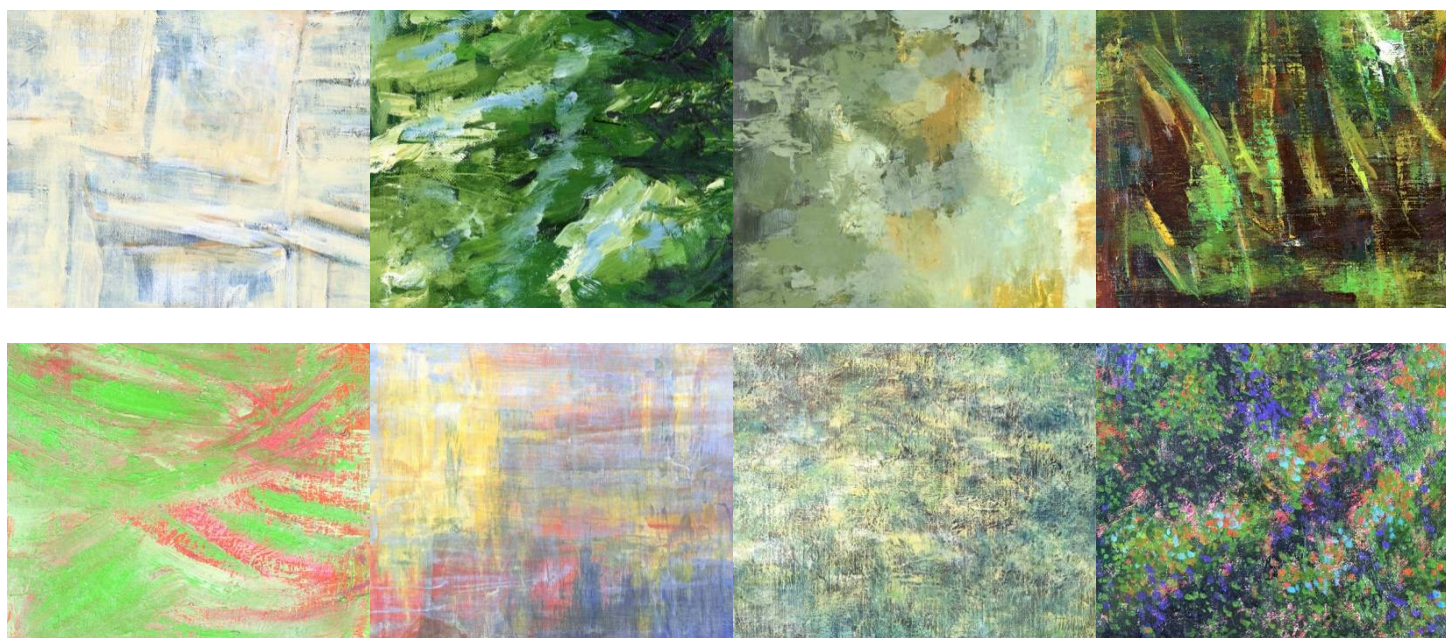
A reference



Retrieved results (6 best matches)

Computerで画像検索の結果

- ▶ 質問: パステル画の中で、多彩で穏やかに見える画像



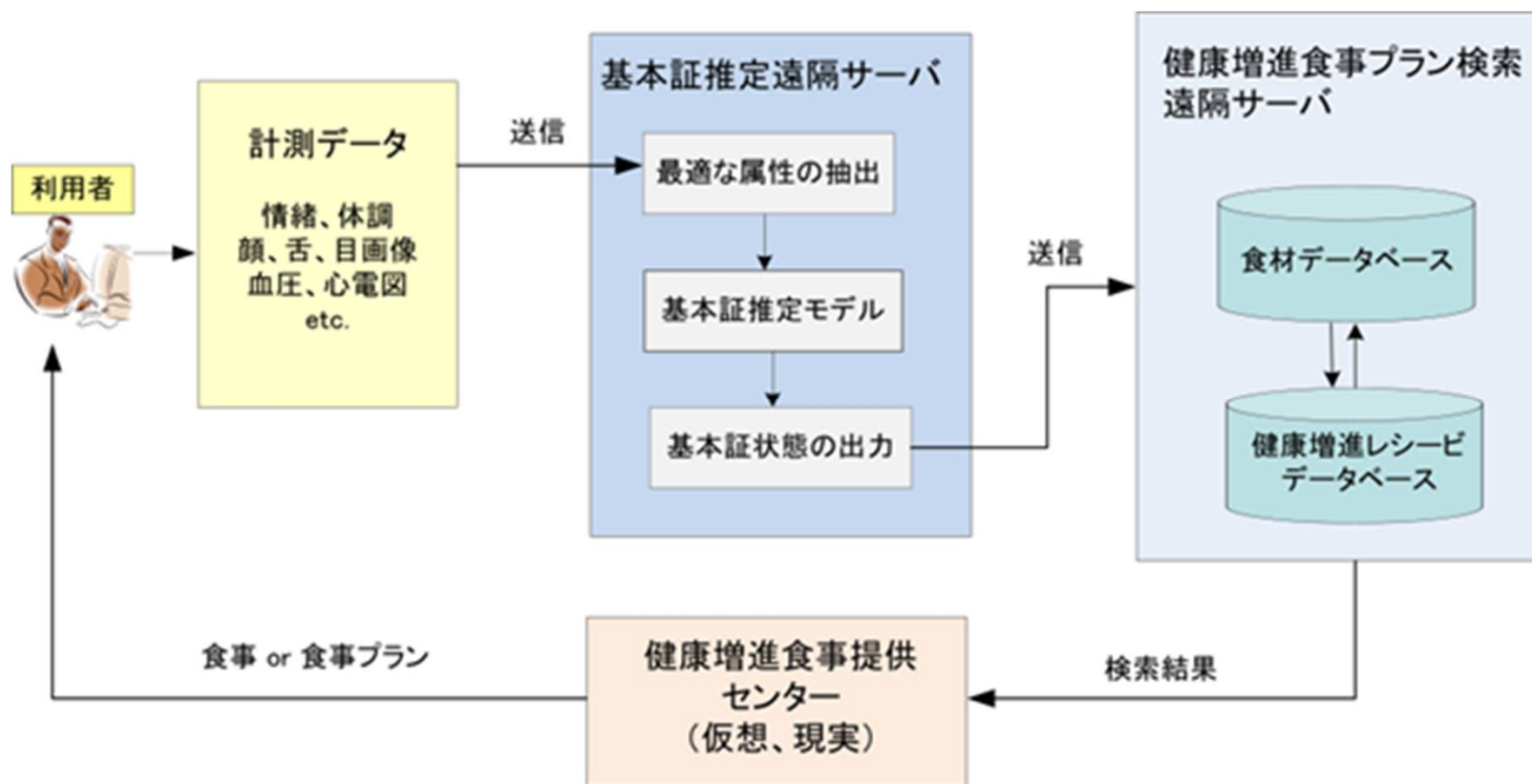
Retrieved results

Big sensor dataを利用する健康状態推測システム

- ▶ 東洋医学の利点を生かし、手間を取らないような顔画像や舌画像などの情報により、利用者の健康状態を予測する。
- ▶ 中医学の基本15証(表寒、表熱、裏寒、裏熱、気虚、気滞、血虚、血瘀、津虚、痰飲、心、肺、脾、肝、腎)による健康状態推測
- ▶ Yi Wang, Ying Dai, et al. “Sensitive-based information selection for predicting individual’s sub-health on TCM doctors’ diagnosis data”, Journal of Japan society for fuzzy theory and intelligent informatics, Vol. 23, No. 5, pp. 749–760 (2011)

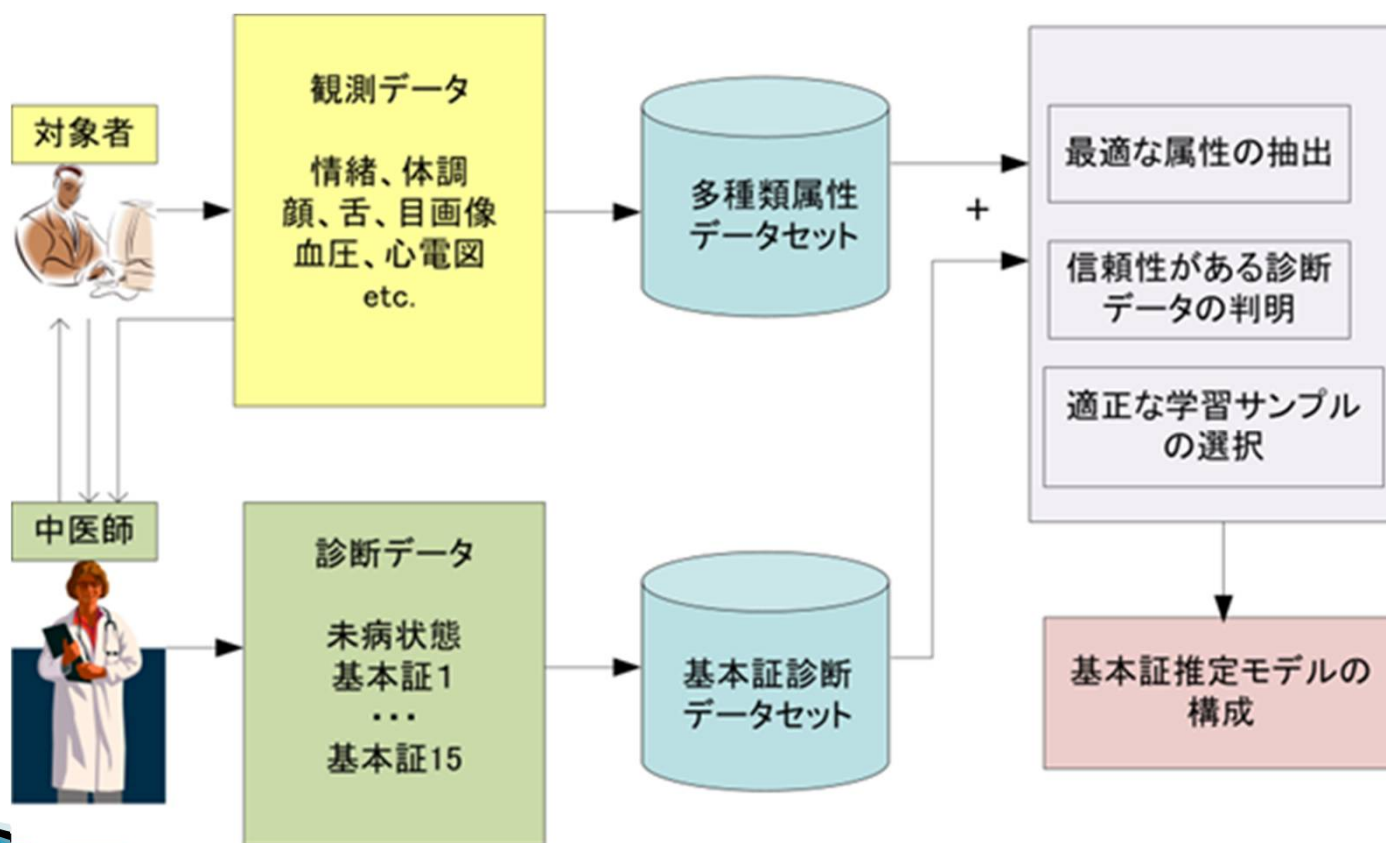
Big sensor dataを利用する健康状態推測システム

システム構成



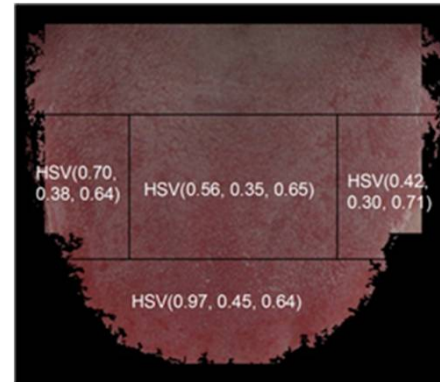
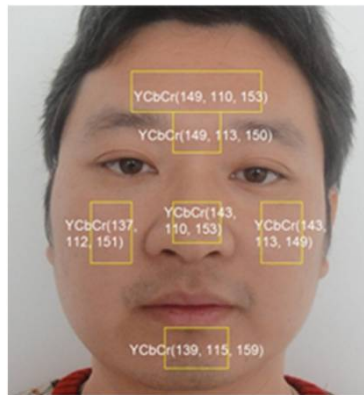
Big sensor dataを利用する健康状態推測システム

▶ モデル構築の流れ



Big sensor dataを利用する健康状態推測システム

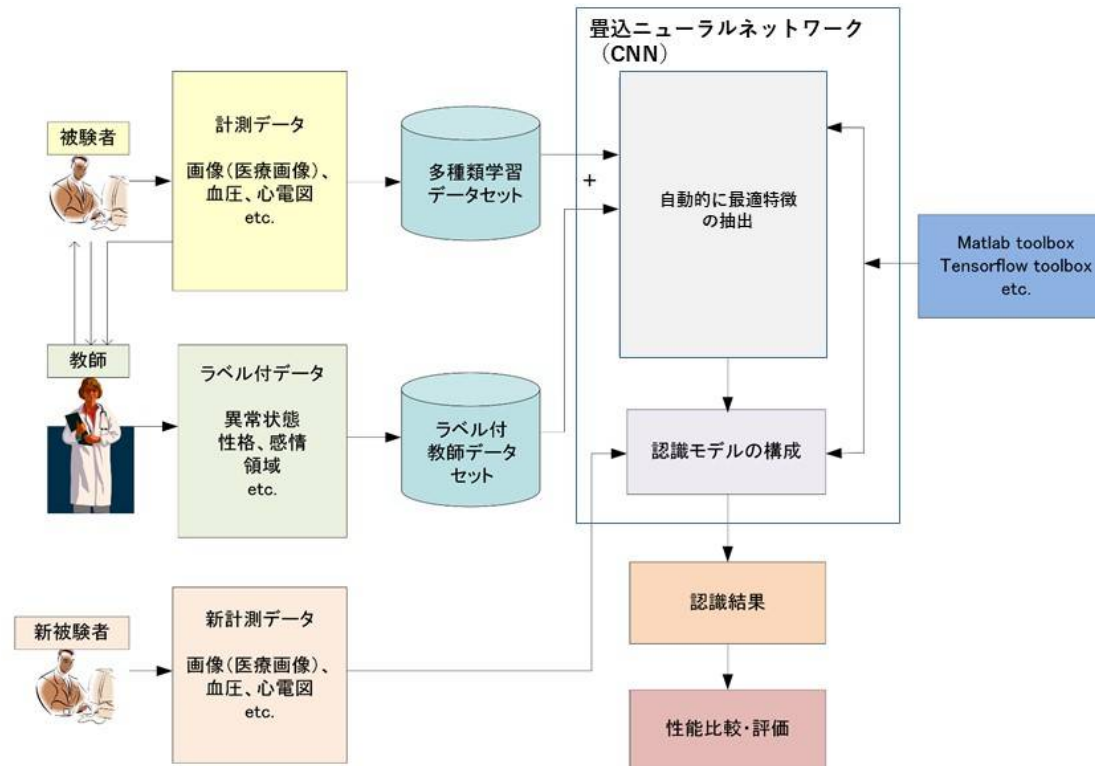
▶ 観測データの例



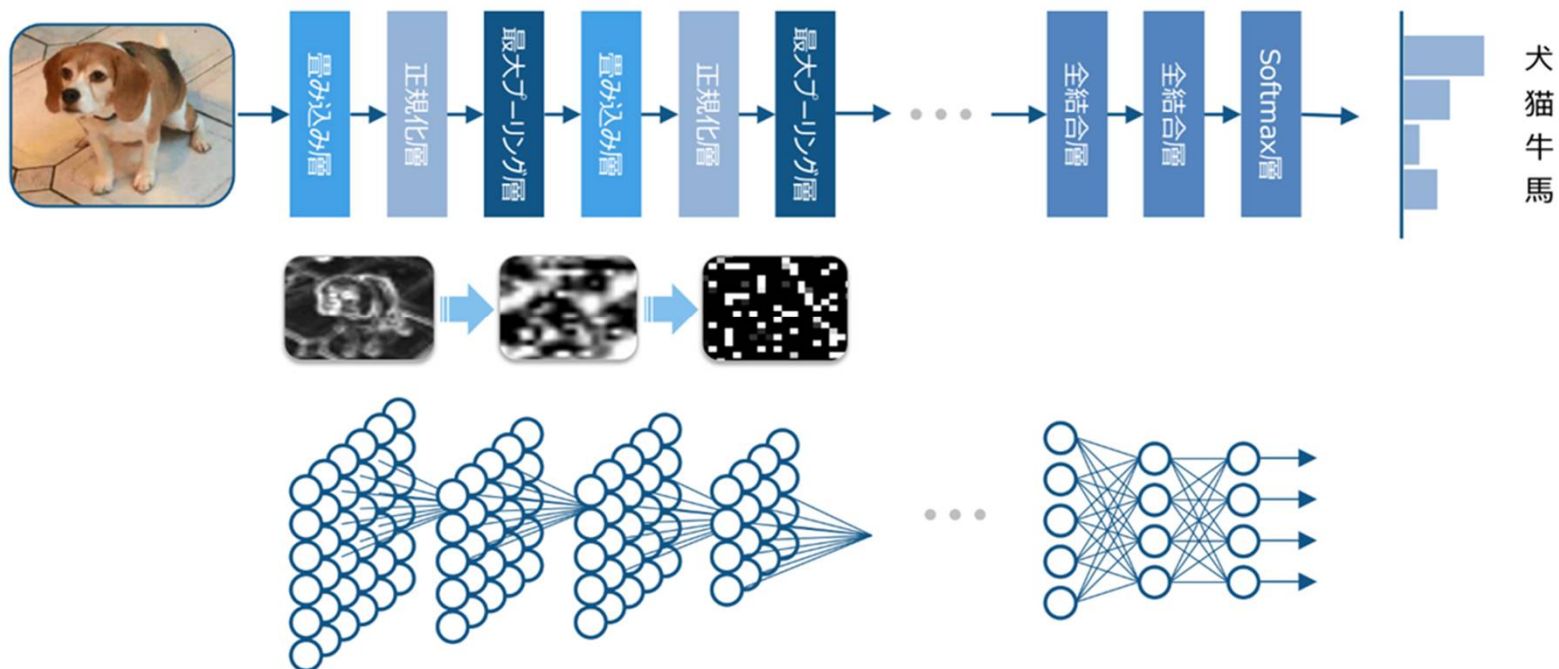
▶ 推測結果の例

寒	熱	気虚	気滞	血虚	血瘀	津虚	痰飲	心	肺	脾	肝	腎	未病
0	1	0.5	0	1	0	0	0	0.5	0	1	0	0	0.5

Deep learning(CNN)を用いた自動認識システムの構築



畳み込みニューラルネットワーク(CNN)



畳み込み層 (convolutional layer):
画像のフィルタ処理に相当する処理、特徴抽出器としての役割
(matlabの深層学習資料により)

料理画像から食材を認識するシステム (Fu, Zhu)

- ▶ 学習と検証データセットの用意
 - ▶ 単一食材分類モデルの構築
 - ▶ 料理画像から複数食材の分割
 - ▶ 分割された食材の識別
-
- ▶ Ziyi Zhu, Ying Dai, “A New CNN-Based Single-Ingredient Classification Model and its Application in Food Image Segmentation”, Journal of Imaging 2023, 9, 205
 - ▶ Kun Fu, Ying Dai, et al. “CNN-based visible ingredients recognition in a food image using decision making schemes”, Proc. of IEEE SMC 2023, pp. 2427–2432



食材の選定

- 食材を下記の標準により選定
 - 新食品成分表 FOODS 〈2021〉
 - 油脂類、菓子類、飲料類を除くカテゴリーをカバーする
- 認識対象の食材が110種

1	abalone	31	crab	61	octopus	91	soybean
2	almond	32	cream	62	okra	92	spinach
3	apple	33	cucumber	63	onion	93	squids
4	asparagus	34	daikon	64	orange	94	strawberry
5	avocado	35	eel	65	other white flesh	95	sunflower seed
6	bamboo shoot	36	egg	66	oyster mushroom	96	sweet potato
7	banana	37	eggplant	67	oysters	97	seaweed
8	bean sprout	38	enoki	68	papaya	98	swine
9	bitter melon	39	fig	69	peach	99	tofu
10	black rice	40	garlic stem	70	peanuts	100	tomato
11	blueberry	41	grape	71	pear	101	tree ears
12	bok choy	42	grape fruits	72	peas	102	tuna
13	bonito	43	green soybean	73	pecan	103	wakame
14	broad beans	44	green onion	74	pepper	104	watermelon
15	broccoli	45	hazel nuts	75	pineapple	105	wax gourd
16	cabbage	46	kidney bean	76	pistachio	106	walnuts
17	carrot	47	kidney beans	77	pitaya	107	wheat product
18	cashews	48	konpu	78	potato	108	yam
19	cattle	49	kiwi	79	poultry	109	yogurt
20	cauliflower	50	lemon	80	pumpkin	110	yellow peach
21	celery stem	51	lettuce	81	pumpkin seeds		
22	celtuce	52	lobster	82	purple laver		
23	cheese	53	lotos	83	raspberries		
24	cherry	54	mackerels	84	rice		
25	chestnuts	55	mango	85	salmon		
26	chickpea	56	mantis shrimp	86	sesame seeds		
27	chinese cabbage	57	melon	87	shiitake		
28	chinese chive	58	millet	88	shimeiji		
29	clam	59	mushroom	89	shrimp		
30	corn	60	meat product	90	snowpea		

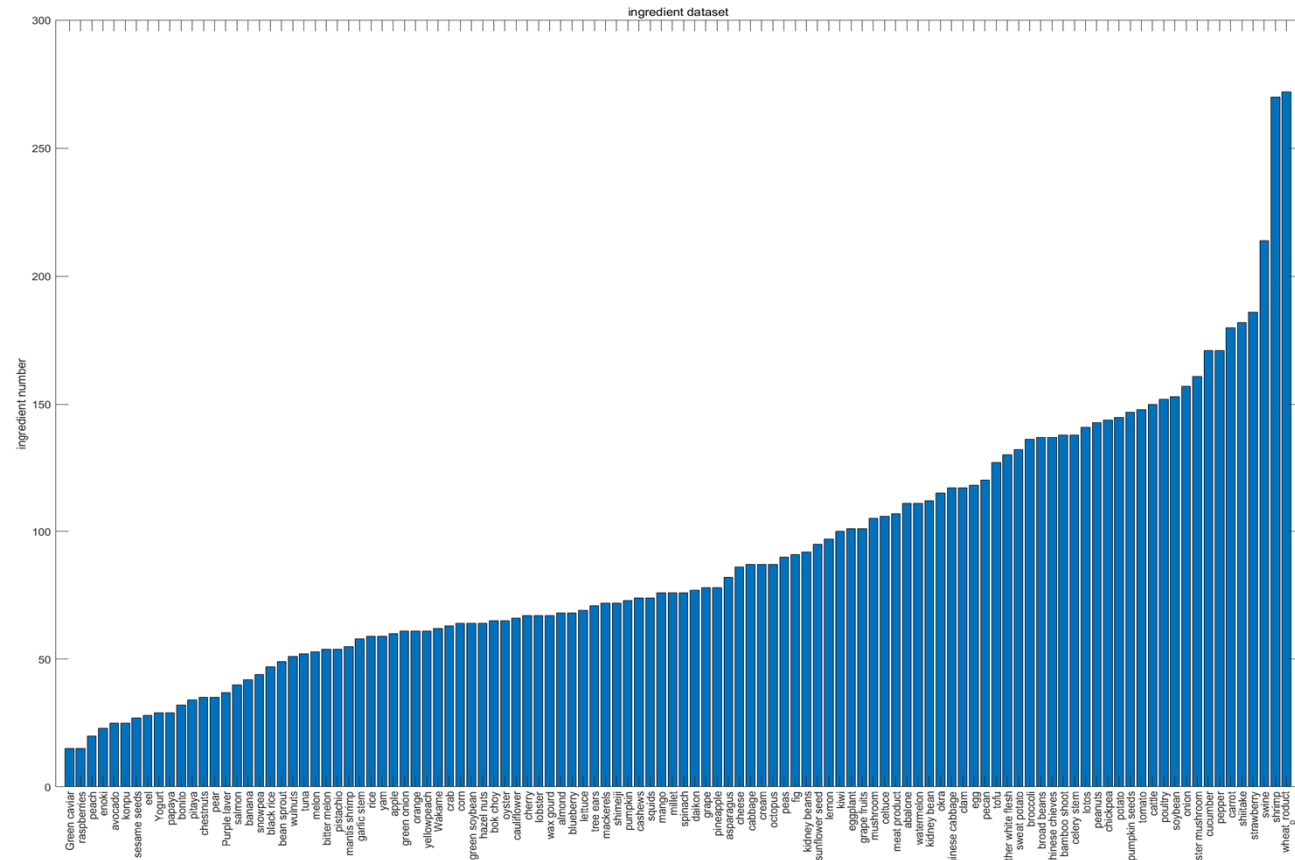
単一食材のデータサンプル収集基準

- 同じ食材に対して
 - 食材の画像は、できるだけ異なる切り方や調理法を反映する画像を含める
 - 似ている画像に対して5~10枚
 - ひとつの画像に食材ができるだけ単一かつ認識できる
- 例



単一食材データセット (SI110)

- 食材種類: 110
- 画像の枚数: 9982



複数食材の料理画像のデータセット (MIF110)

- 複数食材の認識性能評価に利用
- 全部で2099枚の画像がある
- 各画像には、複数の食材がある

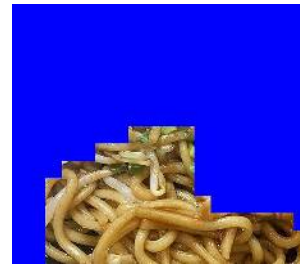
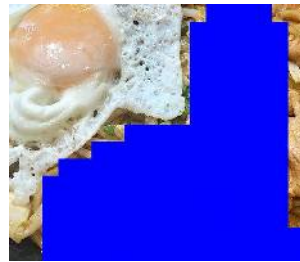


単一食材分類モデルの構築

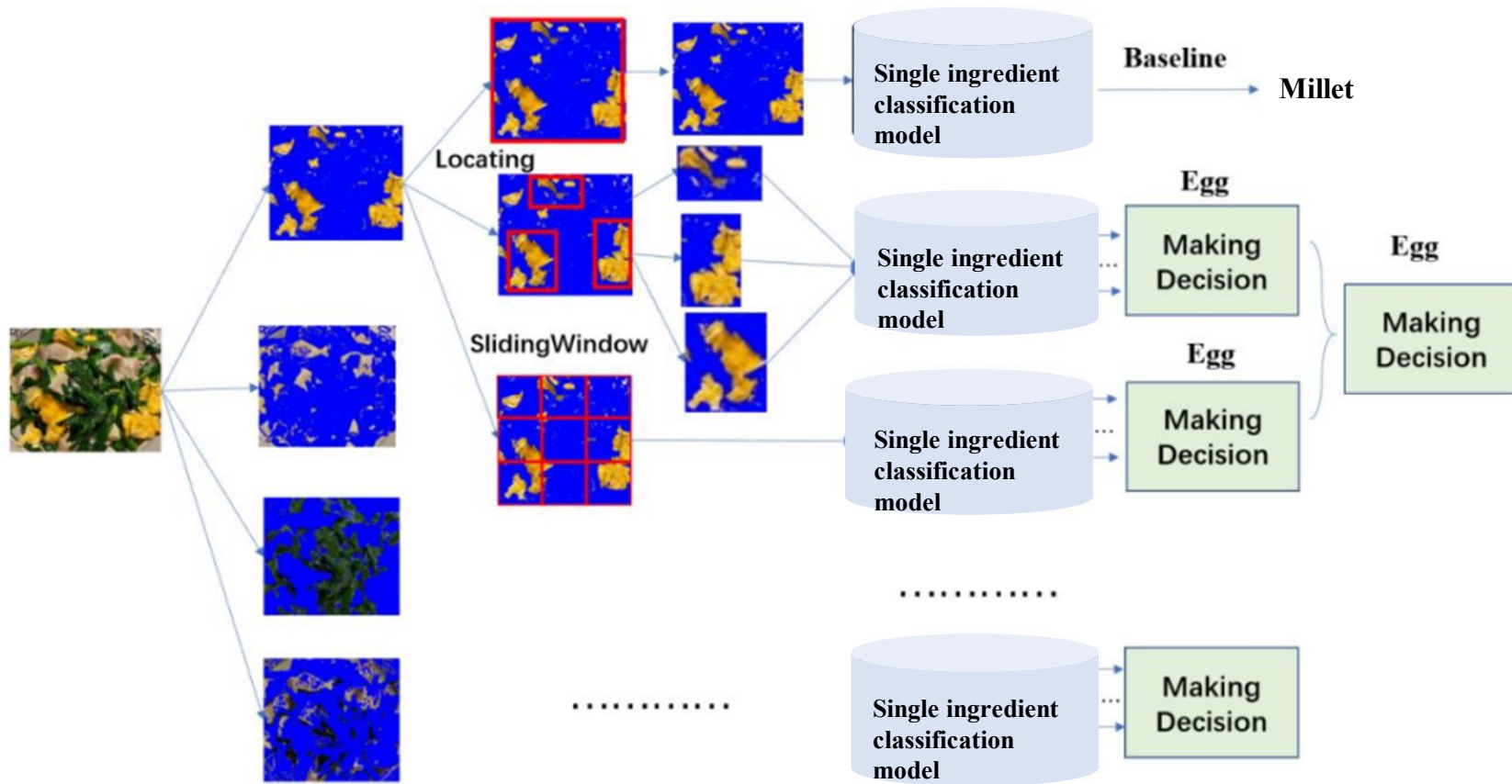
- 学習済みモデルに基づいて転移学習により単一食材分類モデルを訓練
 - 単一食材データセットが学習データセットとする
 - EfficientNet B0が学習済みモデルとする

料理画像から複数食材の分割

K-means clusteringを用い、単一食材分類モデルの特徴マップによって分割
(Kの値が自動予測)



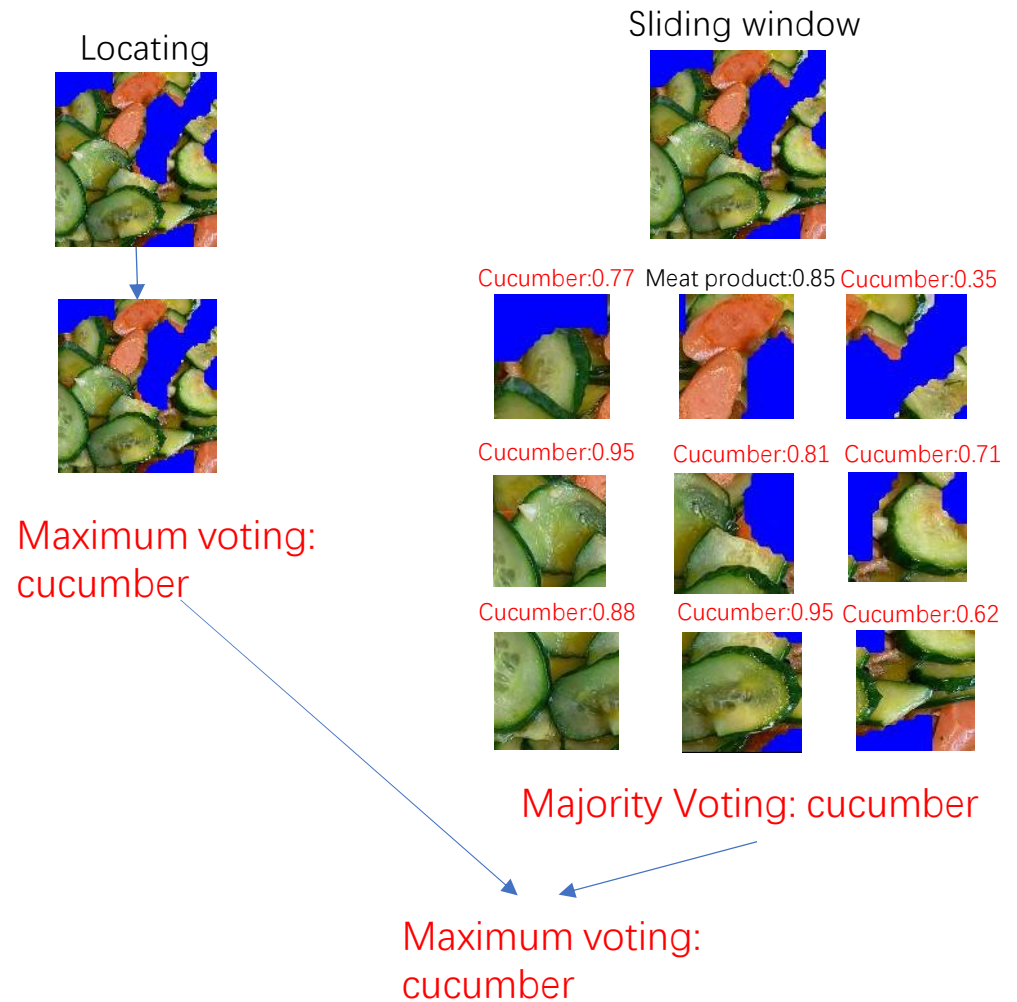
複数食材識別システム



Making decision

アルゴリズム①
出力結果が一つだけ

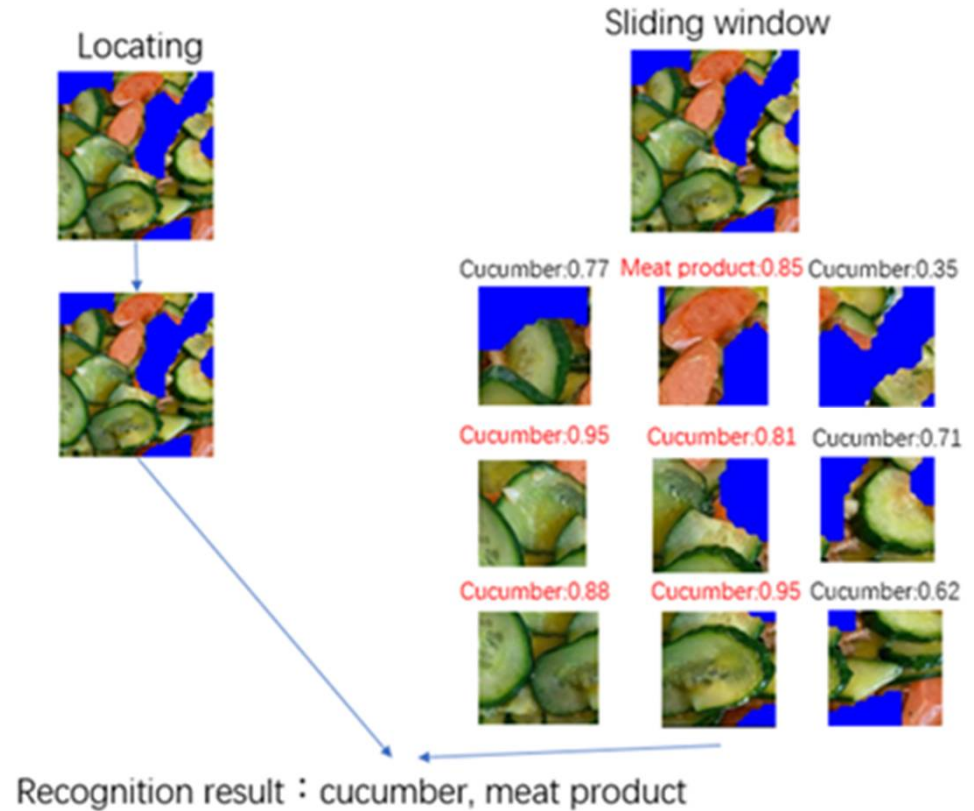
- 各識別結果の個数が違いとき:
Majority voting of averages of class scores
- 各識別結果の個数が同じとき:
Maximum voting of averages of class scores
- Locatingと Sliding Windowの結果を結合
Maximum voting



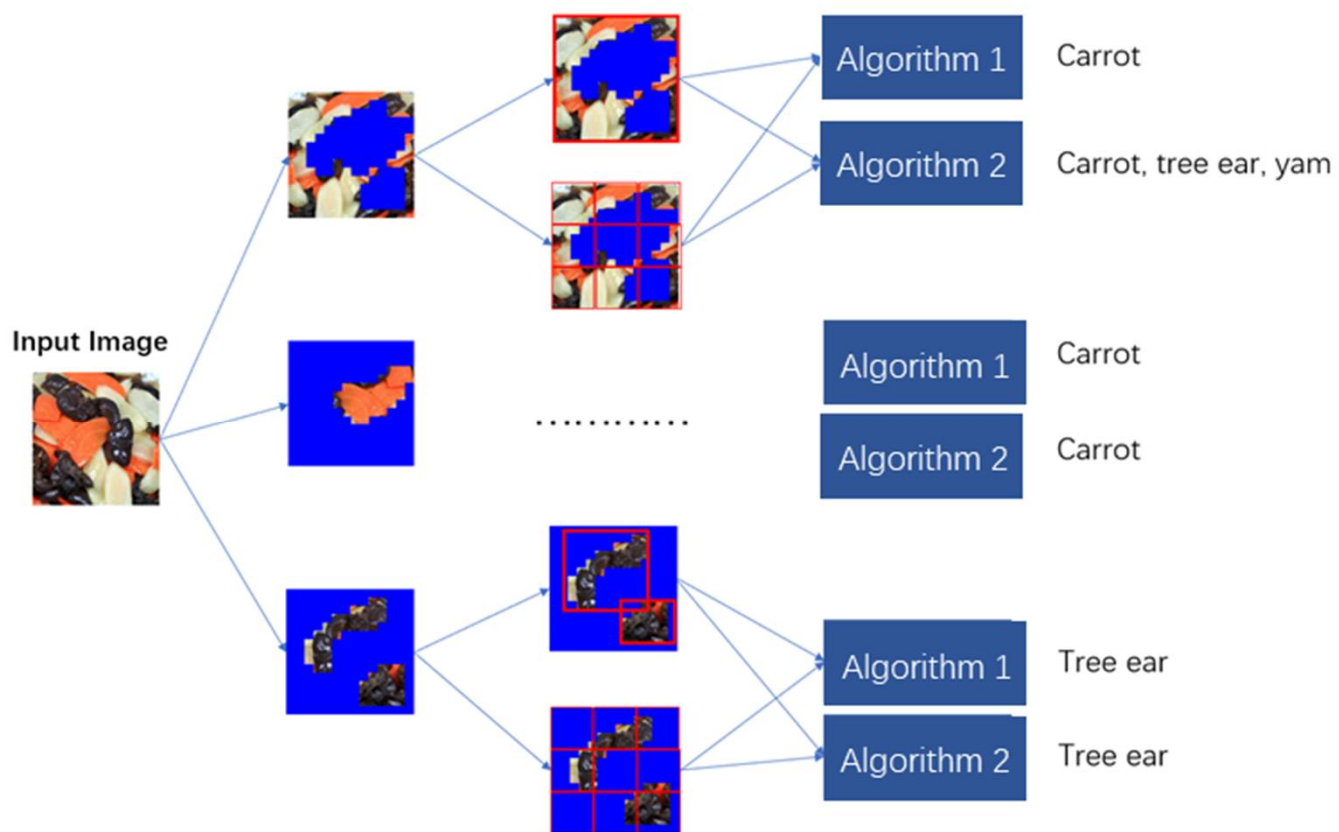
Making decision

アルゴリズム②
出力結果が二つ以上が可能

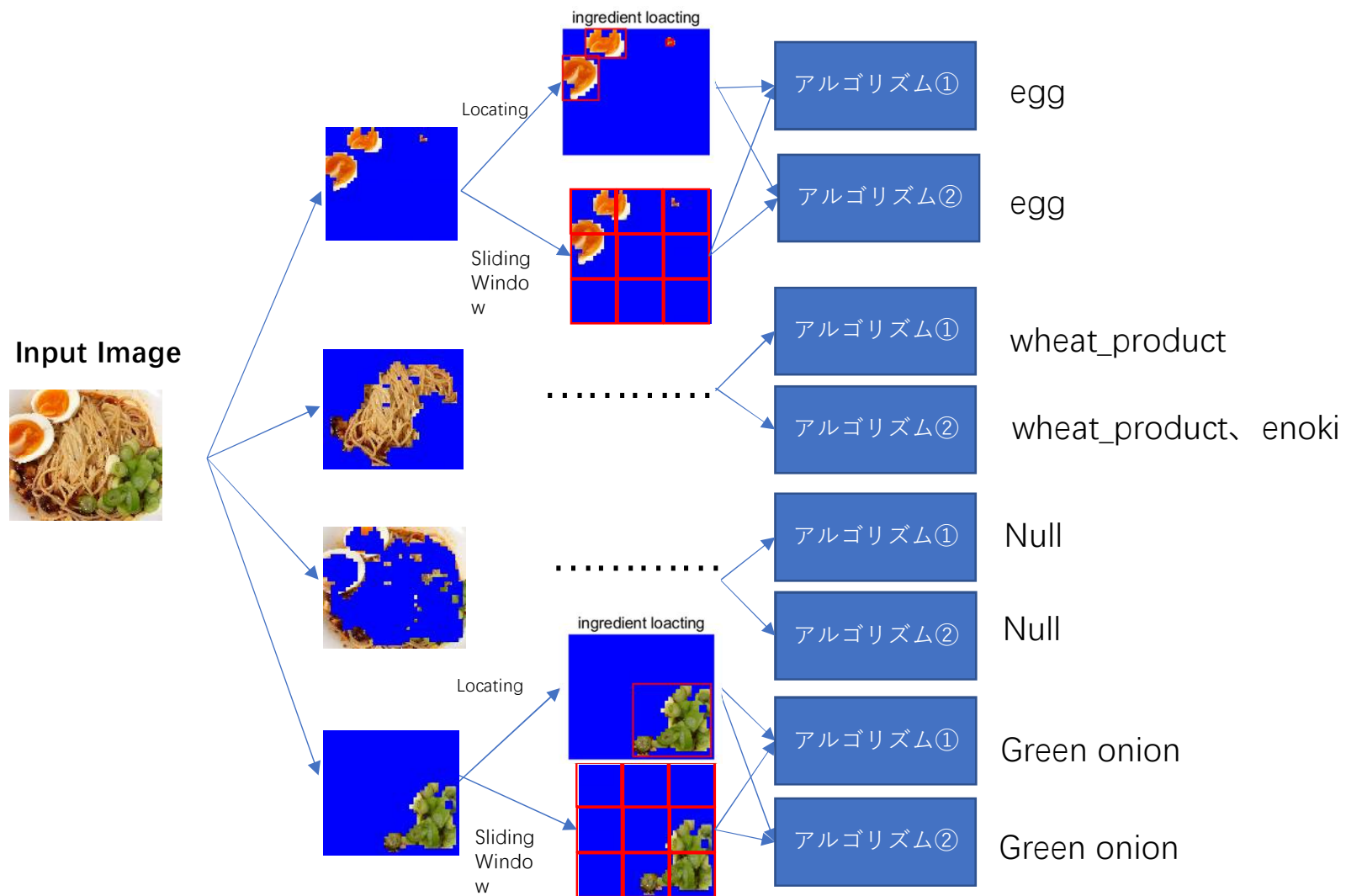
- それぞれの結果を結合
- Scoreがtop nの結果を選択
- 同名の食材を合併



料理画像から複数食材を識別する例

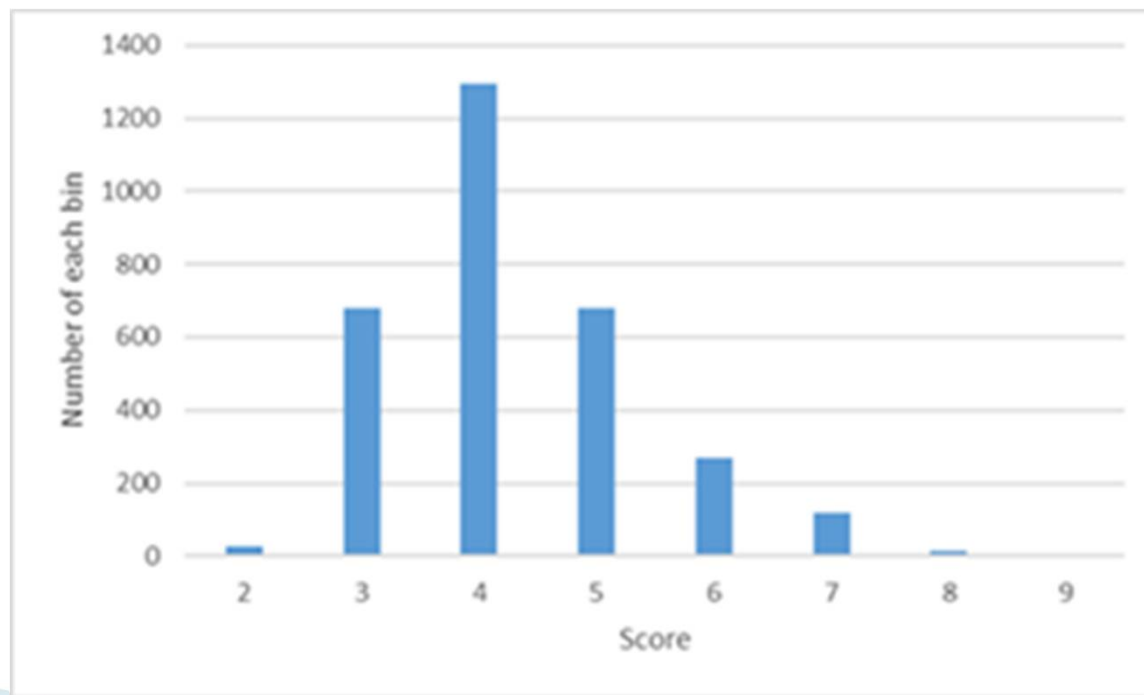


料理画像から複数食材を識別する例



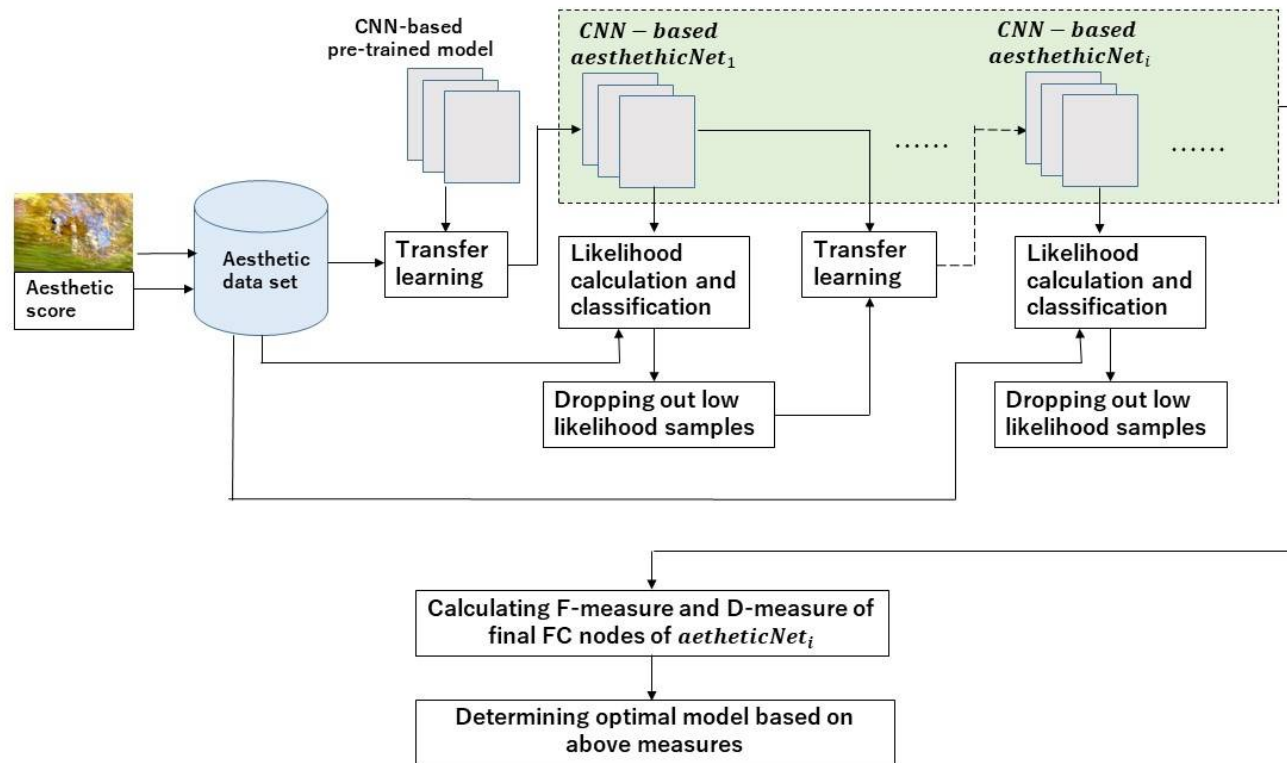
CNNを用いた写真の美的評価モデルの構築

- ▶ Ying Dai, “CNN-based repetitive self-revised learning for photos’ aesthetics imbalanced classification”, Multimedia tools and application, 10.1007/s11042-020-09426-z , 2020
- ▶ 写真の学習データセットの美的評価(10段階)分布



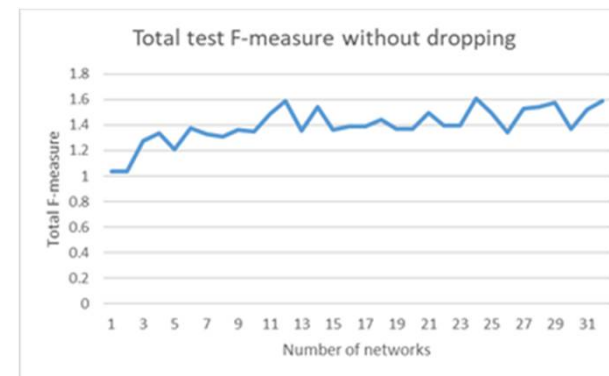
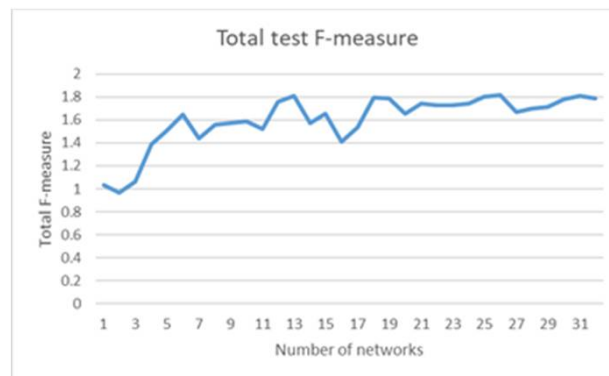
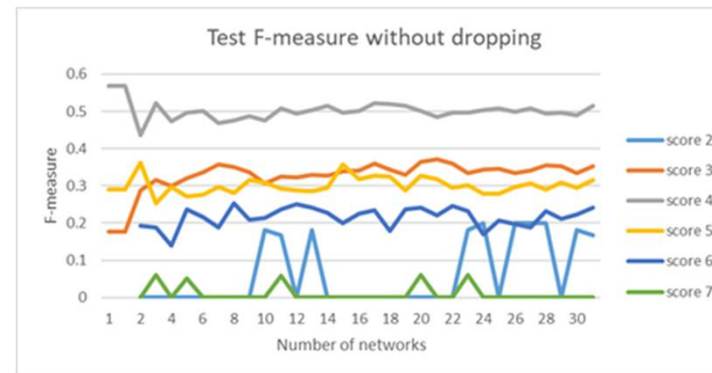
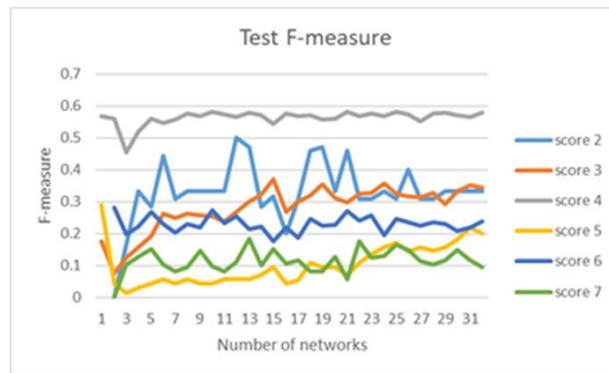
CNNを用いた写真の美的評価モデルの構築

- ▶ CNN-based repetitive self-revised learning (RSRL)



CNNを用いた写真の美的評価モデルの構築

▶ Performance evaluation

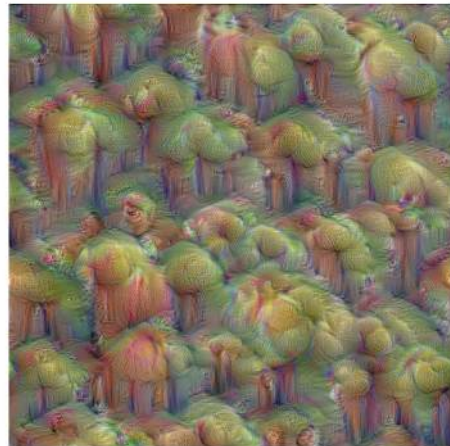
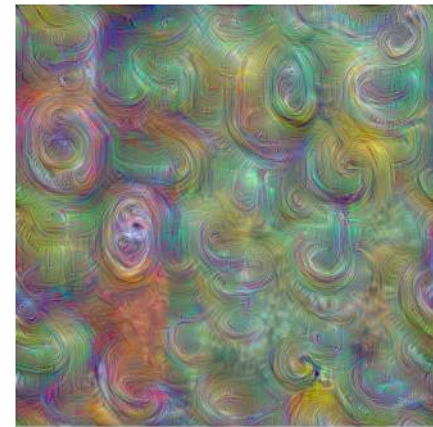
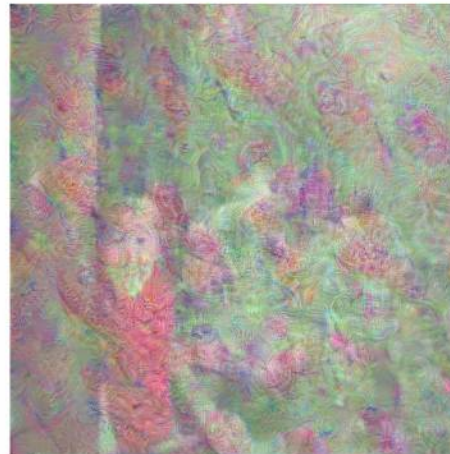


With RSRL

Without RSRL

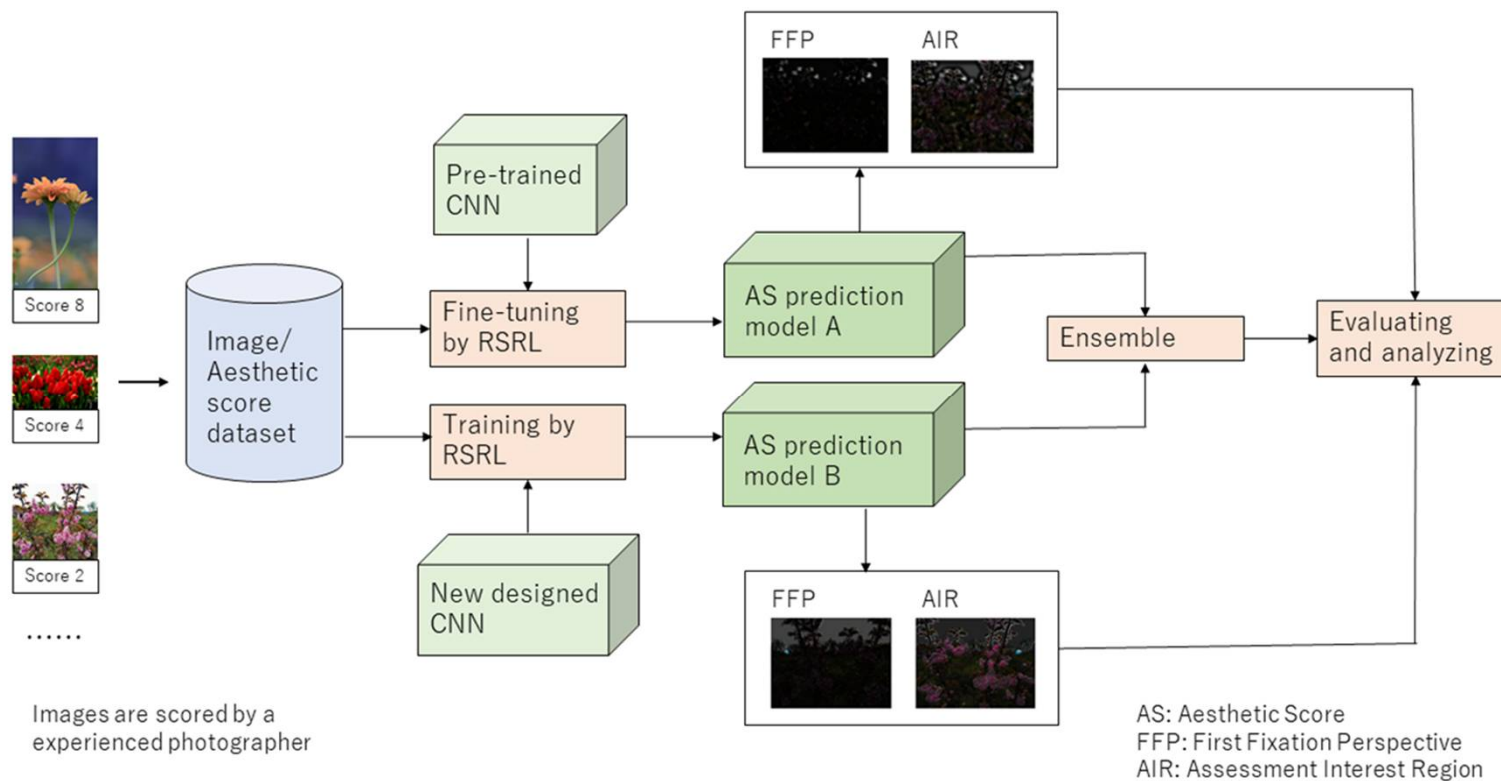
CNNを用いた写真の美的評価モデルの構築

- ▶ deep dream images of different networks



CNNを用いた写真の美的評価モデルの構築

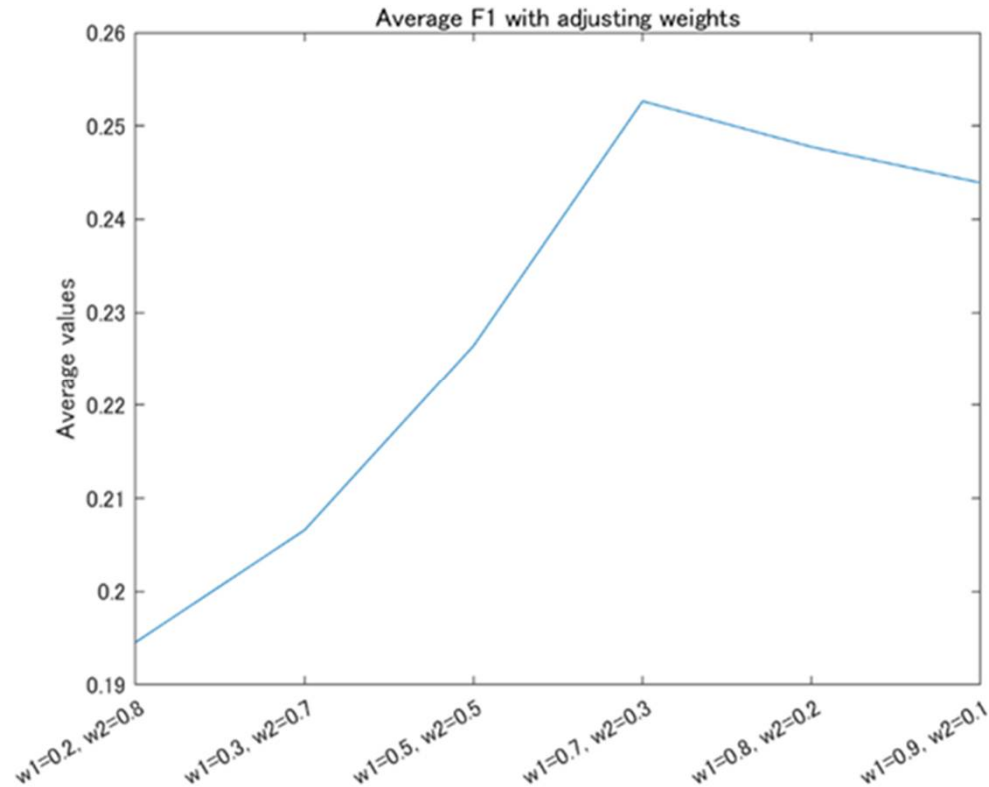
- ▶ Improving the performance by the ensemble



CNNを用いた写真の美的評価モデルの構築

▶ Improving the performance by the ensemble

- $p_S^{ensemble} = w_1 p_S^{model A} + w_2 p_S^{model B}$



CNNを用いた写真の美的評価モデルの構築

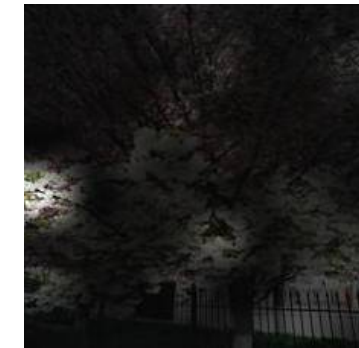
- ▶ Examples of the FFP of a optimal network model



Score 6



Score4



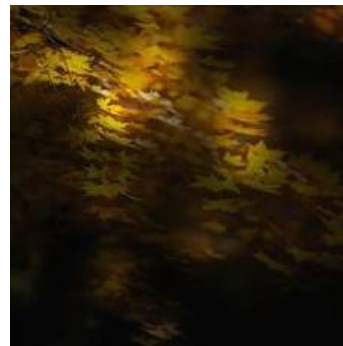
Score2

CNNを用いた写真の美的評価モデルの構築

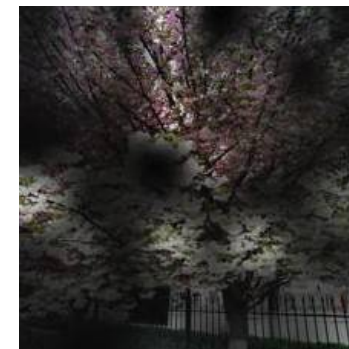
- ▶ Examples of the AIR of a optimal network model



Score 6



Score 4



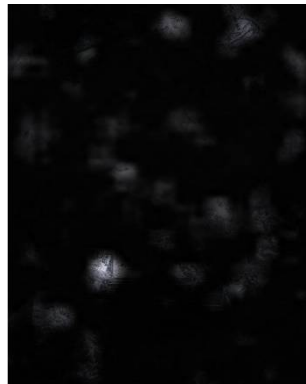
Score 2

CNNを用いた写真の美的評価モデルの構築

- ▶ Examples of the FFP of the optimal network models With different CNN architectures

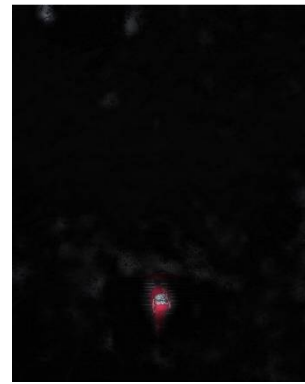


Original image



By resNet

Score 3



By efficientNet

Score 6



By a new designed CNN

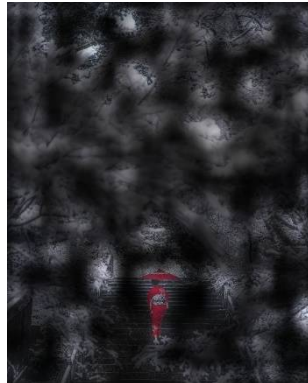
Score 6

CNNを用いた写真の美的評価モデルの構築

- ▶ Examples of the AIR of the optimal network models With different CNN architectures

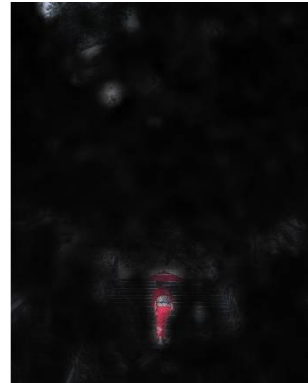


Original image



By resNet

Score 3



By efficientNet

Score 6



By a new designed CNN

Score 6

CNNを用いた写真の美的評価モデルの構築

- ▶ Examples of the aesthetics assessment



Score 2



Score 3



Score 4



Score 5



Score 6



Score 7

卒業研究紹介

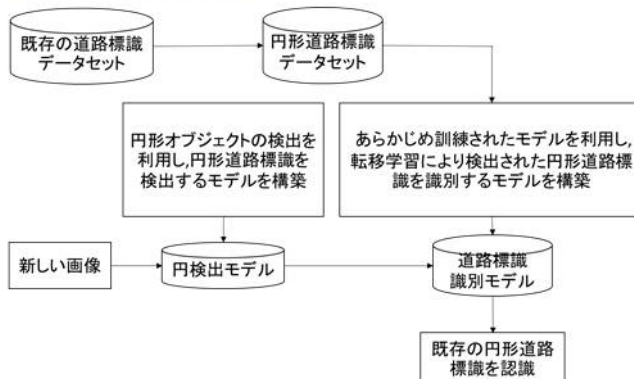
円形道路標識検出と識別システムの構築

岩手県立大学ソフトウェア情報学部
戴研究室 高橋美紗

提案手法

本研究では、既存研究の学習データセットにより、Matlabを使用し、円形道路標識の抽出と識別システムを提案する。既存研究の学習データセットの画像を使用し、抽出モデルにはMatlabの円形オブジェクトの検出関数を使用する。識別モデルは、まず、様々な種類の円形道路標識の位置と種類をラベリングする。それらの情報をあらかじめ訓練されたモデルを転移学習により学習させ、モデルを構築する。

システム概要図



円形道路標識の抽出例



円形道路標識の抽出結果

半径の大きさ		画像一枚当たりの平均	全体
半径20~45	再現率	38.88%	27.66%
	適合率	39.05%	29.21%
半径25~50	再現率	24.09%	13.83%
	適合率	22.18%	27.08%

円形道路標識の識別精度

種類	精度	
AlexNet	Test 89.42%	Train 97.73%
DenseNet-201	Test 66.51%	Train 92.46%
ResNet-50	Test 65.55%	Train 89.47%
GoogleNet	Test 73.68%	Train 97.01%

戴研究室 本田 皓大
岩手県立大学ソフトウェア情報学部

背景

保護性さびが形成されるまで腐食状況を把握するために定期的な点検が必要しかし、目視での判定は判断基準が曖昧で誤判断もしじやすい。

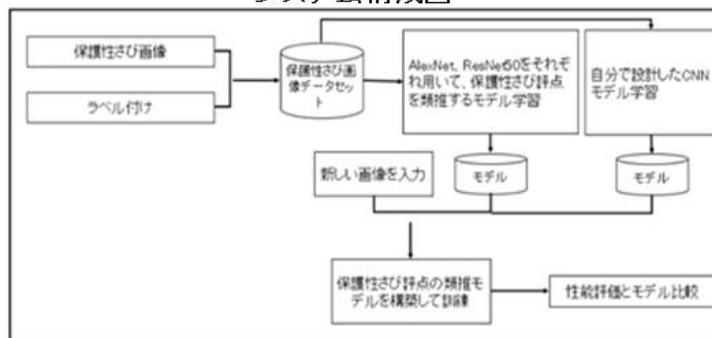
↓
保護性さび評点の判定のバラつきを抑制する判定手法が必要



実現方法

AlexNet、ResNet50、そして自分で設計したCNNにそれぞれ転移学習で訓練させ、保護性さび評点を推測するモデルを構築する。
さらに、ResNet50では関数maxで最も強く活性化するfeature mapを見つけ、それをinput画像として用い、ResNet50-based 転移学習による推測するモデルResNet50FMを訓練する。

システム構成図



まとめ

結果

本研究ではCNN-based転移学習を用いた保護性さび評点を推測するモデルを構築
→結果、ResNet50FMが最高テスト精度62.00%、平均適合率が0.5582、平均再現率が0.5586を示した。
全体を通して適合率、再現率は評点1が最も高かったが、評点3、4は共に低かった。

分析

評点3、4が低かったのは保護性さびの状態が悪くなるまでの途中経過であるため判定がしにくいと考察する。

	評点1	評点2	評点3	評点4	評点5
適合率	0.750	0.667	0.454	0.325	0.595
再現率	0.724	0.694	0.439	0.317	0.619

CNNを用いた手書き文字画像から人の性格を推測するシステム

岩手県立大学 4年
 戴研究室 小川沙己

研究の目的

人の性格判断：手書き文字、写真などから筆跡鑑定士が対人認知により行う

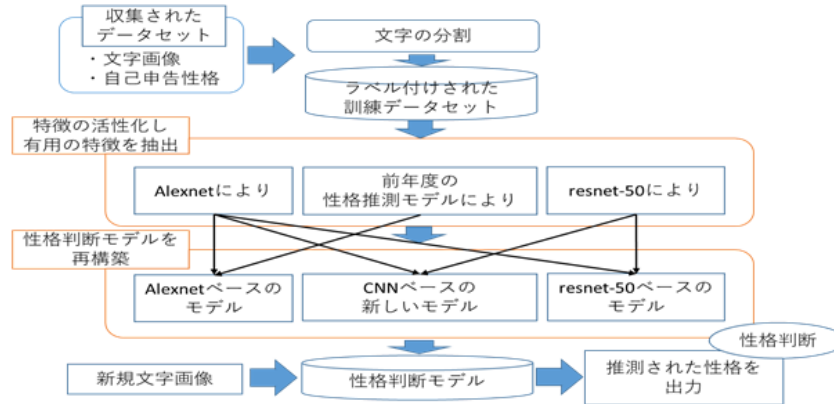
機械により性格判断を行うことで対人認知における曖昧さを排除し自動的に性格判断できるシステムを構築

提案手法

データセット：
 手書き文字画像

- 日本語251枚、アルファベット106枚、数字504枚
- ①CNNを利用し、深い層で最も強く活性化する特徴マップを抽出
- ②①を入力画像として利用して性格判断モデルを構築

システム構成図



検証結果

Alexnetで有用の特徴を抽出し、Alexnetからの転移学習で性格判断モデルを構築するのが最適

5段階の検証精度

性格分類	改革する人	助ける人	達成する人	個性を求める人	謙べる人	信頼を求める人	熱中する人	挑戦する人	平和を求める人
学習精度	37.7	38.4	40.9	38.7	40.8	38.2	39.0	31.8	45.9
テスト精度	36.1	38.2	40.7	37.6	40.1	36.6	39.5	31.4	46.5
前年度卒研のテスト精度	26.7	33.2	29.8	28.5	28.6	30.2	29.4	24.0	35.0
差	+9.4	+5.0	+10.9	+9.1	+11.5	+6.4	+10.1	+7.4	+11.5

単位：%

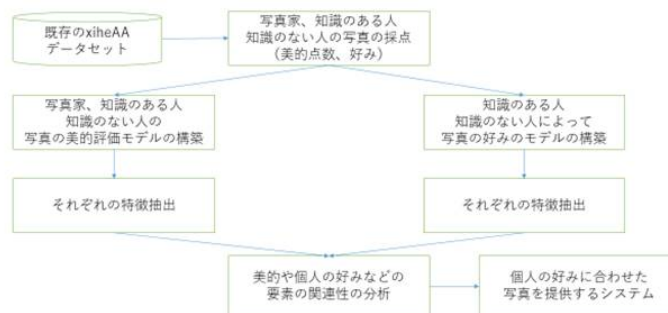
個人の好みに合わせた写真を提供するシステムの構築 Building a system that provides photos that suit individual tastes

岩手県立大学ソフトウェア情報学部
戴研究室 田中敦大

提案手法

本研究では, AlexNet を使用した転移学習等の技術を利用し, 個人の好みに合わせた写真を提供するシステムの構築を目指す. そのために, 本論文では既存のデータセットである xiheAA データセットを用いて複数人の美的評価, 好みの評価を行うことで CNN に基づく複数の美的評価と好みの予測モデルを訓練する. さらに, 構築されたモデルを用い, 写真の美的特徴マップを抽出し分析を行う.

システム構成図



特徴抽出例



性能評価

1	2	3	4	5
56.4%	56.9%	55.3%	54.0%	53.8%

美的評価モデルのテスト精度

1	2	3	4	5
83.3%	70.8%	79.2%	58.3%	70.8%

好みモデルのテスト精度

まとめ

CNN を用いて美的評価, 好みの評価を行うことで評価モデルの構築を行い, その後 CNN モデルを訓練, 構築することにより, 個人の好みの予測を行った. 検証によって美的評価モデルと好みモデルの訓練とテストを行った. 美的評価モデルに対して好みモデルの方がより高い精度を検出した. また, それぞれのモデルから抽出された評価領域などは実際の視点と一致していた.

路面標示の劣化を検出するモデルの構築

岩手県立大学ソフトウェア情報学部
 戴研究室 米倉寿莉

目的

- 路面標示の点検には多大な費用と時間を要する。
- 結果のばらつきや過剰評価の可能性が指摘されている。

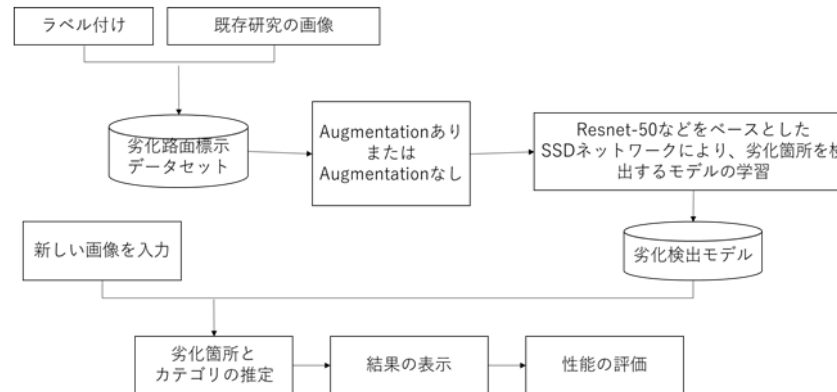


効率的かつ低コストで点検を行う手法の必要性が高まっている。

提案手法

- 既存研究の画像に劣化した路面標示の位置と種類を全国道路標識・標示業協会の基準でラベリング。
- データ拡張
- ResNet-50またはVGG-16をベースネットワークとしたSSD (Single Shot MultiBox Detector) に学習させ、劣化検出モデルを構築する。

システム構成図



実験結果

データ拡張をしていないResNet-50をベースネットワークとしたSSDモデルのmAPが最も高い結果となった。

ResNet-50_SSD (拡張なし) の評価

閾値	mAP	WL	WM	CL	CM	CW
0.5	0.33	0.55	0.30	0.32	0.00	0.48
0.3	0.42	0.62	0.40	0.49	0.00	0.61

VGG-16_SSD (拡張なし) の評価

閾値	mAP	WL	WM	CL	CM	CW
0.5	0.13	0.16	0.12	0.13	0.00	0.23
0.3	0.17	0.19	0.16	0.18	0.01	0.30

ResNet-50_SSD (拡張あり) の評価

閾値	mAP	WL	WM	CL	CM	CW
0.5	0.30	0.29	0.27	0.26	0.00	0.68
0.3	0.35	0.32	0.37	0.34	0.01	0.70

VGG-16_SSD (拡張あり) の評価

閾値	mAP	WL	WM	CL	CM	CW
0.5	0.13	0.14	0.15	0.06	0.00	0.30
0.3	0.16	0.16	0.18	0.08	0.01	0.34

自然体の顔、舌画像からのストレス度推定 Estimating stress level from natural face and tongue images

岩手県立大学4年生
戴研究室 大坪凌

背景

地方の医師不足や心身の健康を意識する傾向にあり、今後、遠隔医療やセルフチェックシステムが拡大していく可能性が高い。その一方で、ストレスは、体と心のさまざまな病気に影響を与える。

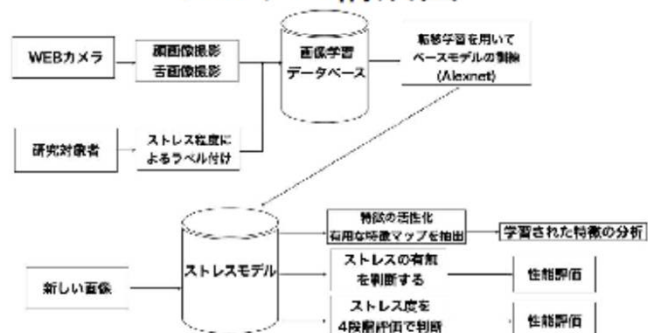


カメラで撮影した顔、舌画像から、ユーザーのストレス推測するシステムを構築

実現方法

Matlab環境で、Alexnetに基づいた転移学習を用いてストレスを推測するモデルを構築する。モデルを訓練するために利用する関数はtrainNetwork、顔と舌画像からストレスを予測するのはclassifyである。次は構築されたモデルにより、activations関数を利用して特徴マップを抽出し、ストレスに関わる特徴の可視化を行う。

システム構成図



結果

ストレスの程度	1H	50	1F	ER
顔画像でのストレスの有無の判断	67%	66%	66%	41%
舌画像でのストレスの有無の判断	55%	69%	62%	52%
顔画像でのストレスのレベルの判断	61%	67%	14%	31%
舌画像でのストレスのレベルの判断	72%	72%	67%	76%

→結果、顔と舌画像を用いたストレス推測モデルでは、全体の精度は約4~6割となり、ストレスの有無の推測では、全体で6~7割程度の精度を得た。

分析

特徴の可視化により、顔画像では目の下の部位が、舌画像では舌を出す動きがストレスと関わりがある可能性があると考えられる。

特徴マップ(顔画像) 特徴マップ(舌画像)



道路交通標識の検出モデルと識別モデルの構築

岩手県立大学4年
戴研究室 赤坂健斗

背景

事故の原因の多くは安全運転義務違反によるもので、速度違反や道路交通標識の見落としなどが挙げられる。

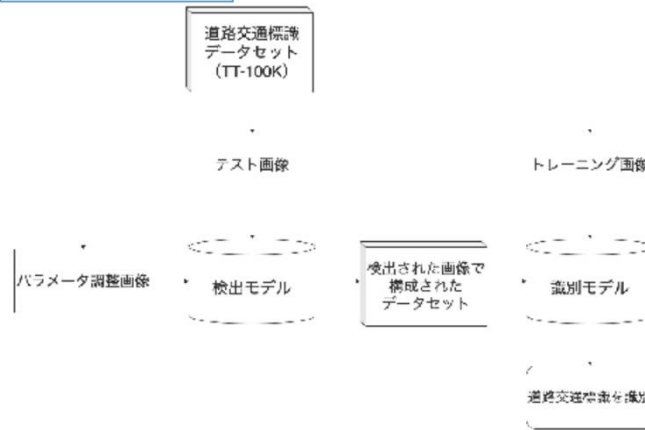


道路交通標識を自動認識しドライバーへの注意喚起や、自動運転にすることにより事故等を減少に繋がる。

提案手法

TT-100Kデータセットを使用し、道路交通標識の検出と識別システムを提案する。検出モデルでは、Pythonを使用しOpenCVの各種関数を利用して、色調と形状特徴ごとにモデルを構築する。識別モデルでは、道路交通標識の座標と種類に関してラベリングを行い、それらの情報を基に予め訓練されたモデルを転移学習により学習させモデルを構築する。

システム構成図



検出例



検証結果

検出結果

標識	適合率	再現率
赤-円形	18.05%	32.69%
青-円形	4.41%	31.47%
黄-三角形	1.33%	39.02%
青-四角形	1.78%	50.00%
赤-その他	1.22%	62.79%
全体	None	34.76%

識別結果

識別モデル	精度		学習に要した時間
	Train	Test	
AlexNet	99.92%	87.46%	0:19:40
DenseNet	98.19%	80.20%	3:37:00
GoogLeNet	97.83%	86.80%	0:35:40
ResNet	97.87%	80.85%	1:48:46

特徴点を用いた手書き平仮名文字上達支援システム

岩手県立大学ソフトウェア情報学部
 戴研究室 森 悠人

目的

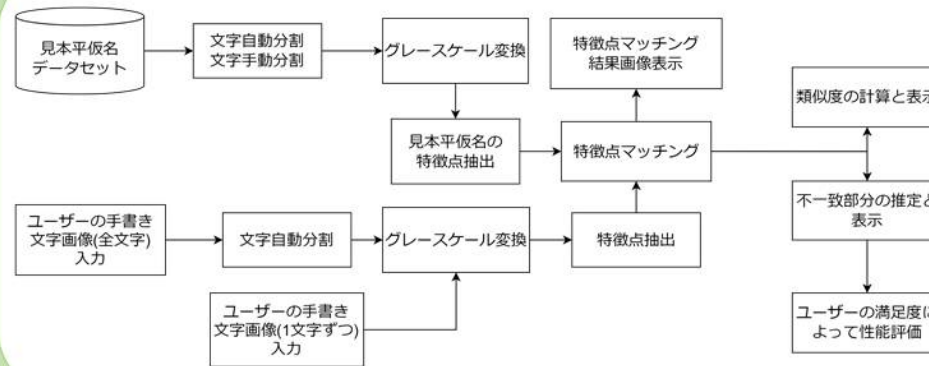
- 自分の書く文字に自信がある人は約2割しかない
- タブレットにペンタブを用いて書いた文字を評価するアプリケーションしかなく手書き文字が上達すると言えない

実際に紙に鉛筆やペンで書いた文字を評価

提案手法

- ユーザー11人分の手書きデータを撮影とスキャンの2種類とサンプルデータ5種類を収集
- AKAZEを用いて特徴点抽出・マッチング・類似度判定
- 特徴点不一致部分に円を描画
- ユーザーの満足度によって性能評価

システム構成図

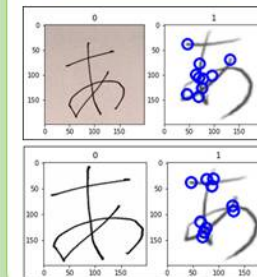


実験結果

性能評価の結果、撮影したものの④以外で50%以上の評価を得られた

アンケート項目	撮影	スキャン
①不一致部分が正しく表示されていると思うか	60.61	63.64
②不一致部分に気を付けることで上達すると思うか	54.55	78.79
③実際に書いて上達したと感じるか	60.61	72.73
④類似度の点数は正しく表示されていると思うか	42.42	60.61
⑤総合的に評価	54.55	100.00

不一致部分の表示例



深層学習を用いた車両台数推定システムの構築 Construction of a vehicle number estimation system using deep learning

岩手県立大学ソフトウェア情報学部
戴研究室 千田 浩正

目的

- 駐車場を探す作業は、時間を浪費するので非効率的である。
- 夜間などの一定の明るさが無い場合、駐車スペース推定の精度が低い。

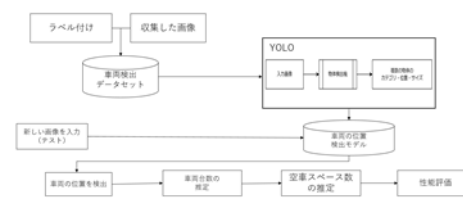


昼夜問わず駐車車両台数を推定することで、空車スペース数を求める。

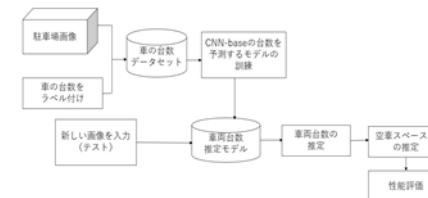
提案手法

- 異なる天候・時間帯に対応するため、7,10,13,16,19時台に画像を収集し、合計3000枚使用。
- YOLOv3-basedとCNN-basedの分類器による車両台数推定モデルを構築。
- 推定した台数とラベルデータの台数を比較し結果が一致した数から精度を求める。

システム構成図



概要図1：YOLOv3-basedの車両位置検出による空車スペース数の推定



概要図2：CNN-basedの車両台数分類器による空車スペース数の推定

実験結果

- 本研究では、YOLOv3をもとに駐車位置から車両台数を推定するモデルと、CNN-basedの車両台数分類器(Resnet50-based,Googlenet-based)をもとに駐車台数を推定するモデルの構築を行い精度から、性能評価・比較する。

$$\text{テスト精度(accuracy)} = (\text{Tr} = p) / T$$

Tr=p:推定した車両台数ラベルとモデルから推測した車両台数が一致した数、T:すべてのテスト画像数

YOLOv3	Resnet50	Googlenet
0.9857	0.8957	0.9567

考察

- 車両台数推定モデルの精度から見ると、どちらのモデルも十分高い精度を示しているが、YOLOv3-basedの精度の方がGooglenet-based分類器の精度よりも約3%高かった。
- しかし、Googlenet-basedの分類器も約95%と高い精度を示しており、学習までのラベル付けにかかった手間を考えると効率よくラベル付けすることができたGooglenet-basedの分類器による車両台数推定の方が実用的だと考える。

自然体の顔、発話時画像からのストレス度推定 Estimating stress level from natural face and speak images

背景

私達は、現代社会においてありとあらゆるストレスに遭遇する。ストレスは、心身に影響を与え、その影響が自殺や交通事故などの重大な事例につながる可能性もある。そのため、各個人のストレスを把握し、適切なストレス対策を施す必要性が求められている。



カメラで撮影した顔画像や発話画像から個人のストレス状態を推測するシステムを構築する

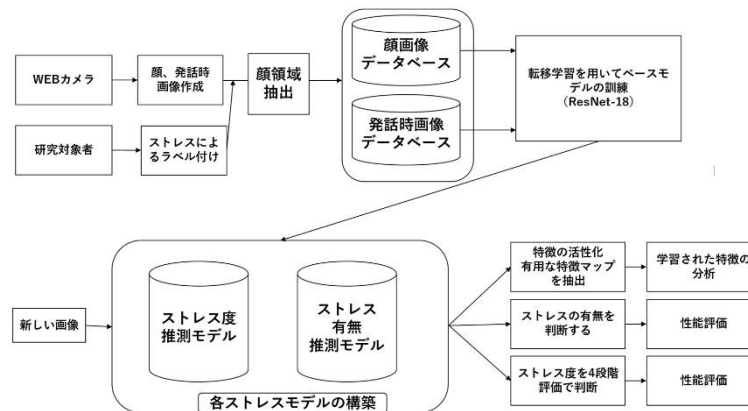
岩手県立大学
戴研究室 菅原陸

実装方法

個人の顔画像、発話時画像データを集め、収集時の感覚によってストレスのラベル付けを行う。このデータセットをMATLAB環境で、ResNet-18による転移学習を行い、ストレス度、ストレスの有無を推測するモデルを構築する。また、モデルの性能を調べるためGrad-CAMやt-SNEを用いた分析も行う。

ストレスの種類：
身体的ストレス (PH)、社会的ストレス (SO)
心理的ストレス (PS)、環境的ストレス (EN)

システム構成図



推測精度

ストレス度推測モデル	PH	SO	PS	EN
被験者A:被験者B	34.67%	36.5%	35.67%	54%
被験者B:被験者A	9%	29.83%	42.83%	56.67%
ストレス有無推測モデル	PH	SO	PS	EN
被験者A:被験者B	60.33%	51.5%	56.83%	51.83%
被験者B:被験者A	71%	59.5%	70%	58.5%

(左の列は学習データの種類と検証データの種類を示す)

結果によりモデルがストレス有無の推測においてはある程度の汎用性を持つことが分かった

分析



この図は、構築したモデルがどこを判断の根拠としているかを表したものである。図は右からストレスの種類がPH、SO、PS、ENとなっている。この図から、主に目や口の周りを根拠としていることとストレスの種類による根拠部分の違いが見て取れる。

The study on visible food ingredient Segmentation in dish images for the identification

戴 研究室
ZHU ZIYI

Background

- Food ingredient classification is an essential task in food-related research
- State-of-the-art research
 - Ingredient-Guided Cascaded Multi-Attention Network for Food Recognition
 - MVNet: Multi-Task Guided Multi-View Attention Network for Chinese Food Recognition
 - Region-Level Attention Network for Food and Ingredient Joint Recognition
 - Limitations
 - can not avoid the interference from other ingredients

Purpose

Segment ingredients from food images for multiple ingredient recognition

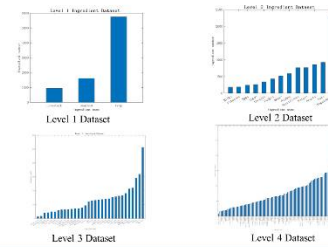
Objective

- Complete the Food ingredient dataset
- Build the Single Ingredient Models base on food ingredient dataset
- train the single-ingredient classification model on the basis of CNN
- Propose a multi-ingredient segmentation framework using the above classification model
- Propose the metrics for evaluating performance of the segmentation

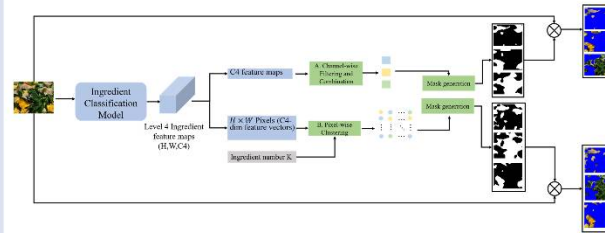
Multi-level Single Ingredient dataset

- Hierarchical Ingredient Structure**
First 3 Level Ingredient Structure
- Image Collection**
 - according to the hierarchical food ingredient structure
 - each image contains only one type of viable ingredient
 - each ingredient should contain as many different visual appearances as possible
- Example of carrot with a different appearance**
Chunk type shredded type slice type

Single ingredient dataset distribution for each level



Framework for multi-ingredient segmentation



Metrics for evaluating ingredient Segmentation

- Instance-wise IOU (intersection over union) : $\frac{|A \cap B|}{|A \cup B|}$
- Unicity of Segment : $\frac{|A \cap B|}{|B|}$
- Entirety of Segment: $\frac{|A \cap B|}{|A|}$
- Loss of GTs: $\frac{|U_i^j A_i - |U_i^j A_i \cap U_i^j B_j|}{|U_i^j A_i|}$
 - A_i : mask of GT
 - B_i : mask of Segment
 - i : the number of GTs for w -th sample
 - j : the number of Segments for w -th sample

Results of multiple ingredient segmentation on FoodSeg103

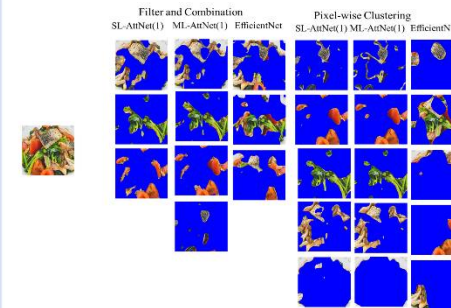
Channel-wise Filtering and Combination

model types		mUnicity	mEntirety	mIoGTs	mIoU
SLM	AttNet(1)	0.6712	0.9177	0.1751	0.6051
	AttNet(1+3)	0.7283	0.7398	0.1693	0.5309
	AttNet(3+1)	0.7221	0.7615	0.1152	0.5368
	AttNet(3)	0.6831	0.9075	0.1525	0.6072
	Resnet18	0.7679	0.4733	0.6592	0.3666
MLM	EfficientNet	0.5987	0.788	0.1925	0.531
	AttNet(1)	0.6102	0.9709	0.1043	0.5874
	AttNet(1+3)	0.616	0.9196	0.0862	0.5559
	AttNet(3+1)	0.6377	0.823	0.1236	0.5552
	AttNet(3)	0.6138	0.9501	0.1174	0.5726

Pixel Clustering

model types		mUnicity	mEntirety	mIoGTs	mIoU
SLM	AttNet(1)	0.8339	0.8003	0.0676	0.6532
	AttNet(1+3)	0.8565	0.7391	0.1276	0.6185
	AttNet(3+1)	0.8255	0.6618	0.1824	0.5548
	AttNet(3)	0.8158	0.6599	0.2072	0.5749
	efficientnet	0.7555	0.5212	0.3606	0.5175
MLM	resnet	0.1079	0.0507	0.9282	0.0317
	AttNet(1)	0.8256	0.7900	0.0837	0.6540
	AttNet(1+3)	0.8373	0.7495	0.1141	0.6199
	AttNet(3+1)	0.7949	0.5001	0.5865	0.2836
	AttNet(3)	0.8026	0.6055	0.2636	0.5090

Visualization of Ingredient Segmentation



- AttNets perform better than the pre-trained EfficientNet and ResNet
- The proposed framework for multi-ingredient segmentation performs well special for the method of pixel-wise clustering

CNN-BASED VISIBLE MULTI-INGREDIENT RECOGNITION IN A FOOD IMAGE USING DECISION MAKING SCHEMES

戴 研究室 FU KUN

Background

- According to the data of the Report on China, more than half of China's adults are overweight. One of the important reasons is that many people have unhealthy lifestyles and bad eating habits, such as increased intake of high energy and high-fat food.
- If foods or ingredients can be identified, various health-related analyses can be performed, such as calorie intake estimation, nutritional analysis and analysis of eating habits.

Purpose

Propose novel methods for multi-ingredient recognition in food images.

Objective

- Construct a standard hierarchical single-ingredient for training ingredient classification model.
- Construct a multi-ingredient food image dataset for validation.
- Train single-ingredient classification models on the above dataset by transfer learning using the pretrained networks of Resnet18 and EfficientNetD0.
- Propose a framework for the multi-ingredient recognition in food images using the above ingredient classification model.
- Propose algorithms for improving the multi-ingredient recognition performance.

Single ingredient dataset (SI dataset)

- 110 kinds of ingredients

1. radish	20. rad	30. radish	40. radish
2. carrot	20. carrot	30. carrot	40. carrot
3. cucumber	20. cucumber	30. cucumber	40. cucumber
4. green pepper	20. green pepper	30. green pepper	40. green pepper
5. egg	20. egg	30. egg	40. egg
6. tomato	20. tomato	30. tomato	40. tomato
7. mushroom	20. mushroom	30. mushroom	40. mushroom
8. onion	20. onion	30. onion	40. onion
9. potato	20. potato	30. potato	40. potato
10. corn	20. corn	30. corn	40. corn
11. carrot	20. carrot	30. carrot	40. carrot
12. cucumber	20. cucumber	30. cucumber	40. cucumber
13. green pepper	20. green pepper	30. green pepper	40. green pepper
14. egg	20. egg	30. egg	40. egg
15. tomato	20. tomato	30. tomato	40. tomato
16. mushroom	20. mushroom	30. mushroom	40. mushroom
17. onion	20. onion	30. onion	40. onion
18. potato	20. potato	30. potato	40. potato
19. corn	20. corn	30. corn	40. corn
20. carrot	20. carrot	30. carrot	40. carrot
21. cucumber	20. cucumber	30. cucumber	40. cucumber
22. green pepper	20. green pepper	30. green pepper	40. green pepper
23. egg	20. egg	30. egg	40. egg
24. tomato	20. tomato	30. tomato	40. tomato
25. mushroom	20. mushroom	30. mushroom	40. mushroom
26. onion	20. onion	30. onion	40. onion
27. potato	20. potato	30. potato	40. potato
28. corn	20. corn	30. corn	40. corn
29. carrot	20. carrot	30. carrot	40. carrot
30. cucumber	20. cucumber	30. cucumber	40. cucumber
31. green pepper	20. green pepper	30. green pepper	40. green pepper
32. egg	20. egg	30. egg	40. egg
33. tomato	20. tomato	30. tomato	40. tomato
34. mushroom	20. mushroom	30. mushroom	40. mushroom
35. onion	20. onion	30. onion	40. onion
36. potato	20. potato	30. potato	40. potato
37. corn	20. corn	30. corn	40. corn
38. carrot	20. carrot	30. carrot	40. carrot
39. cucumber	20. cucumber	30. cucumber	40. cucumber
40. green pepper	20. green pepper	30. green pepper	40. green pepper
41. egg	20. egg	30. egg	40. egg
42. tomato	20. tomato	30. tomato	40. tomato
43. mushroom	20. mushroom	30. mushroom	40. mushroom
44. onion	20. onion	30. onion	40. onion
45. potato	20. potato	30. potato	40. potato
46. corn	20. corn	30. corn	40. corn
47. carrot	20. carrot	30. carrot	40. carrot
48. cucumber	20. cucumber	30. cucumber	40. cucumber
49. green pepper	20. green pepper	30. green pepper	40. green pepper
50. egg	20. egg	30. egg	40. egg
51. tomato	20. tomato	30. tomato	40. tomato
52. mushroom	20. mushroom	30. mushroom	40. mushroom
53. onion	20. onion	30. onion	40. onion
54. potato	20. potato	30. potato	40. potato
55. corn	20. corn	30. corn	40. corn
56. carrot	20. carrot	30. carrot	40. carrot
57. cucumber	20. cucumber	30. cucumber	40. cucumber
58. green pepper	20. green pepper	30. green pepper	40. green pepper
59. egg	20. egg	30. egg	40. egg
60. tomato	20. tomato	30. tomato	40. tomato
61. mushroom	20. mushroom	30. mushroom	40. mushroom
62. onion	20. onion	30. onion	40. onion
63. potato	20. potato	30. potato	40. potato
64. corn	20. corn	30. corn	40. corn
65. carrot	20. carrot	30. carrot	40. carrot
66. cucumber	20. cucumber	30. cucumber	40. cucumber
67. green pepper	20. green pepper	30. green pepper	40. green pepper
68. egg	20. egg	30. egg	40. egg
69. tomato	20. tomato	30. tomato	40. tomato
70. mushroom	20. mushroom	30. mushroom	40. mushroom
71. onion	20. onion	30. onion	40. onion
72. potato	20. potato	30. potato	40. potato
73. corn	20. corn	30. corn	40. corn
74. carrot	20. carrot	30. carrot	40. carrot
75. cucumber	20. cucumber	30. cucumber	40. cucumber
76. green pepper	20. green pepper	30. green pepper	40. green pepper
77. egg	20. egg	30. egg	40. egg
78. tomato	20. tomato	30. tomato	40. tomato
79. mushroom	20. mushroom	30. mushroom	40. mushroom
80. onion	20. onion	30. onion	40. onion
81. potato	20. potato	30. potato	40. potato
82. corn	20. corn	30. corn	40. corn
83. carrot	20. carrot	30. carrot	40. carrot
84. cucumber	20. cucumber	30. cucumber	40. cucumber
85. green pepper	20. green pepper	30. green pepper	40. green pepper
86. egg	20. egg	30. egg	40. egg
87. tomato	20. tomato	30. tomato	40. tomato
88. mushroom	20. mushroom	30. mushroom	40. mushroom
89. onion	20. onion	30. onion	40. onion
90. potato	20. potato	30. potato	40. potato
91. corn	20. corn	30. corn	40. corn
92. carrot	20. carrot	30. carrot	40. carrot
93. cucumber	20. cucumber	30. cucumber	40. cucumber
94. green pepper	20. green pepper	30. green pepper	40. green pepper
95. egg	20. egg	30. egg	40. egg
96. tomato	20. tomato	30. tomato	40. tomato
97. mushroom	20. mushroom	30. mushroom	40. mushroom
98. onion	20. onion	30. onion	40. onion
99. potato	20. potato	30. potato	40. potato
100. corn	20. corn	30. corn	40. corn

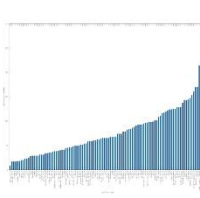
Image Collection

- For the same ingredient, the samples should include as many kinds of ingredient's shapes as possible.
- For the similar shaped ingredient, about 5 to 10 image samples are collected which should be as single as possible and recognizable.
- For example, the egg should include the boiled egg, the scrambled egg, and the fried egg.



Examples of ingredient samples

Data distribution of the SI dataset



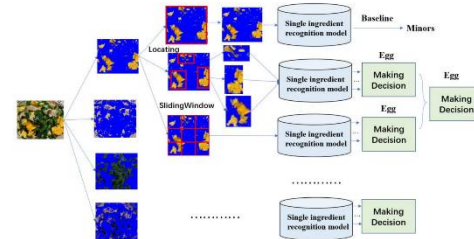
Multi-ingredient food image dataset (MIFI dataset)

- MIFI dataset contains a total of 2166 images.
- Each image contains a multitude of ingredients.
- This dataset is used to evaluate the performance of the ingredient recognition in the food images.

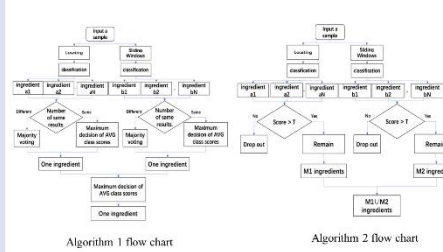


Examples from MIFI dataset

Multi-ingredient identification system



Making decision



Evaluating performance

- Evaluation method
- We use the following metrics for evaluating the performance of the ingredient recognition in the food images

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

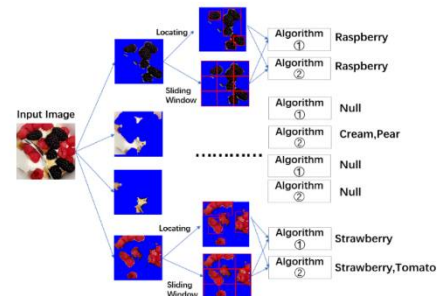
$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

- Experimental results on MIFI dataset

Method	Model	Avg Precision	Avg Recall	Avg F1 score	Method	Model	Precision	Recall	F1 score	
Baseline	Efficientnet	median	33.33	18.74	20.80	Efficientnet	median	33.33	18.74	20.80
		Lower quartile	18.69	6.69	10.53		Lower quartile	18.69	6.69	10.53
		Upper quartile	52.24	28.84	32.54		Upper quartile	52.24	28.84	32.54
Resnet	Resnet	median	39.56	13.81	16.11	Resnet	median	39.56	13.81	16.11
		Lower quartile	4.93	1.82	1.97		Lower quartile	4.93	1.82	1.97
		Upper quartile	62.29	26.57	28.87		Upper quartile	62.29	26.57	28.87
Algorithm1	Efficientnet	median	55.89	32.13	35.80	Efficientnet	median	55.89	32.13	35.80
		Lower quartile	74.14	48.26	53.83		Lower quartile	74.14	48.26	53.83
		Upper quartile	76.18	14.29	21.44		Upper quartile	76.18	14.29	21.44
Resnet	Resnet	median	48.02	38.57	33.03	Resnet	median	48.02	38.57	33.03
		Lower quartile	71.28	13.07	17.24		Lower quartile	71.28	13.07	17.24
		Upper quartile	67.37	43.22	47.20		Upper quartile	67.37	43.22	47.20
Algorithm2	Efficientnet	median	39.84	48.91	37.50	Efficientnet	median	39.84	48.91	37.50
		Lower quartile	53.73	29.48	32.85		Lower quartile	53.73	29.48	32.85
		Upper quartile	25.00	76.81	48.76		Upper quartile	25.00	76.81	48.76
Resnet	Resnet	median	42.73	37.03	39.99	Resnet	median	42.73	37.03	39.99
		Lower quartile	29.42	18.03	18.83		Lower quartile	29.42	18.03	18.83
		Upper quartile	66.87	53.32	49.70		Upper quartile	66.87	53.32	49.70

Visualization of ingredient classification



Stable Diffusionを用いたキャラクターの画風変換に関する研究

岩手県立大学ソフトウェア情報学部
戴研究室 縣 憲世

目的

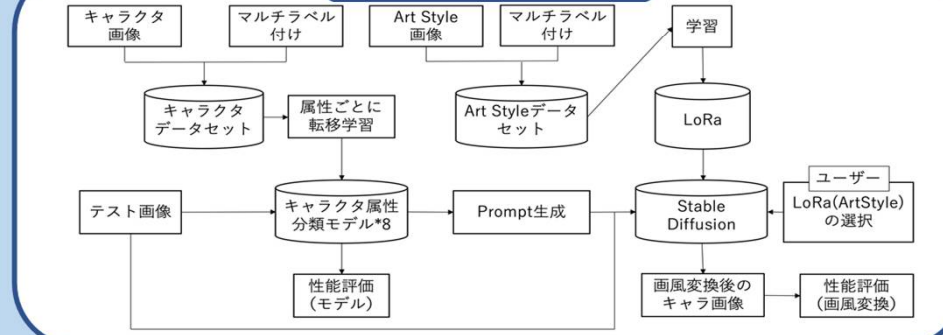
- ・ゲーム制作など多人数で共同開発を行う時、キャラクター原案を作品の画風に直す必要があり、その作業は現在手作業で行われている。
- ・自動で画像変換を行うStableDiffusionのimg2imgはPromptなどの設定を細かく調整する必要がある。

Promptを自動生成し、元画像の属性を保持した画風変換をできるシステムの提案。

提案手法

- ・キャラクター画像の属性を[髪色,髪型,目,性別,年齢,画角,表情,ポーズ,服装]と定義
- ・efficientnetB0で属性ごとにキャラクター属性分類モデルを訓練
- ・分類結果を結合しPromptを生成
- ・画風変換
- ・性能評価

システム構成図



実験結果

キャラクター属性分類モデルと画風変換の属性保持率は高い結果を示した。最終的な画風変換では、属性が足りず20%台と低い結果になった。

model	髪型	髪色	目	性別	年齢	画角	ポーズ	服装	平均
efficientnetB0 (spac100)	83.5%	83.8%	77.7%	79.8%	85.9%	54.0%	72.0%	78.1%	
efficientnetB0 (spac500)	78.8%	88.4%	78.3%	79.8%	70.8%	54.5%	69.8%	67.2%	
efficientnetB0 (unknown (spac100))	84.9%	88.6%	81.0%	-	53.7%	71.4%	70.9%		
efficientnetB0 (unknown (spac500))	75.9%	78.9%	79.7%	-	49.1%	71.5%	74.7%		
efficientnetB0 (DataAug (spac100))	92.0%	97.0%	96.1%	97.5%	96.0%	87.2%	85.6%	94.1%	
efficientnetB0 (DataAug (spac500))	96.0%	97.7%	96.6%	97.0%	87.1%	71.2%	87.8%	93.8%	
efficientnetB0 (unknown (DataAug (spac100))	96.8%	96.6%	96.7%	-	84.5%	86.7%	93.3%		
efficientnetB0 (unknown (DataAug (spac500))	92.0%	97.3%	93.5%	-	83.9%	88.1%	94.1%		

LoRA	髪色	髪型	目	表情	画角	ポーズ	性別	年齢	服装	平均
LoRA無し	100%	93%	97%	96%	100%	100%	93%	97%	97.3%	
Kelion art-style	100%	99%	97%	94%	100%	100%	93%	82%	94.5%	
Naruto art-style	100%	94%	87%	87%	99%	80%	91%	91%	91.9%	
Jiburi art-style	100%	94%	93%	95%	99%	95%	94%	96%	95.8%	
Onepiece art-style	100%	96%	87%	87%	99%	92%	92%	97%	93.8%	
Bleach art-style	99%	93%	87%	93%	99%	96%	92%	96%	94.0%	

属性保持率	LoRA	kelion	naruto	jiburi	onepiece	bleach
本モデル	28%	25%	28%	20%	24%	
BLIP-2	16%	12%	14%	10%	11%	

キャラクター属性変換モデル

画風変換

変換失敗例



元画像

変換結果

道路交通標識の検出と認識モデル構築 Detection and recognition for road traffic signs

背景

2022年度の交通事故件数は300,839件である。事故の原因の多くは安全運転義務違反によるもので、速度違反や道路交通標識の見落としなどが挙げられる。そのため、道路交通標識を自動で認識・分類することで事故の減少に繋がる。



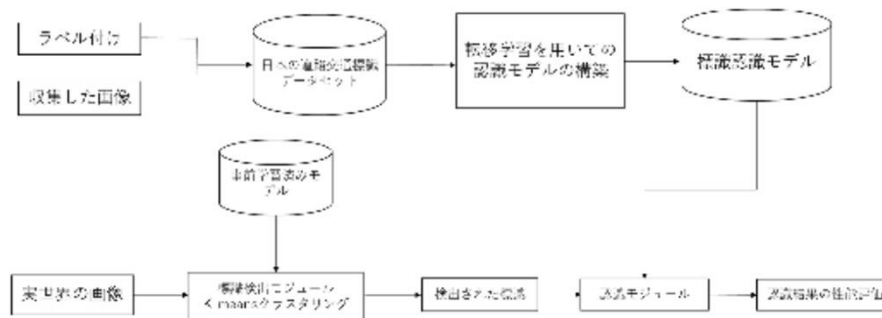
深層学習を用いて、日本の道路交通標識に対応した認識モデルを構築し、それを用いて実世界画像の交通標識を検出し、認識するシステムを構築する。

岩手県立大学
戴研究室 浅沼将人

実装方法

「道路交通標識ハンドブックⅢ」の中の標識を撮影し、日本の道路交通標識のデータセットを作成する。これを用いて転移学習を行い標識分類モデルを構築する。さらに、事前学習済みモデルとK-meansクラスタリングを用いて実世界画像から道路交通標識を検出し、標識分類モデルで標識認識を行う。

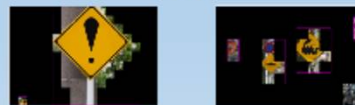
システム構成図



結果

ネットワーク名	EfficientNet-b0	GoogLeNet	ResNet-18	
Accuracy	0.9911	0.9970	0.9985	
	誤一(31)	誤検(148)	正検率(誤一)	正検率(誤検)
EfficientNet-b0	6	4	0.11	0.02
GoogLeNet	18	12	0.35	0.08
ResNet-18	7	8	0.13	0.05

考察



単一画像は、比較的分割が上手くいったが複数画像が上手く分割出来なかった。単一画像の認識はある程度出来たが複数画像の場合は、正しい認識結果が得られたのが1つしかないものが多く、標識以外の部分も標識として認識してしまった。

参考文献

- ▶ Y. Dai, et al. Face Texture Model Based on SGLD and Its Application in Face Detection in a Color Scene, Pattern Recognition, Vol.29, pp. 1007–1017 (1996)
- ▶ Y. Dai, et al. Recognition of Facial Images with Low Resolution Using a Hopfield Memory Model, Pattern Recognition, Vol. 31, pp.159–167 (1998)
- ▶ Ying Dai, et al. “Understanding of facial expressions by the hierarchical recognition of genuine emotions”, International Journal of Innovative Computing, Information and Control, Vol. 1, Number 2, pp. 203–214, June 2005
- ▶ Ying Dai, “Class–based image representation for Kansei retrieval considering semantic tolerance relation”, Journal of Japan society for fuzzy theory and intelligent informatics, Vol. 21, No. 2, pp. 184–193 (2009)
- ▶ Yi Wang, Ying Dai, et al. “Sensitive–based information selection for predicting individual’s sub–health on TCM doctors’ diagnosis data”, Journal of Japan society for fuzzy theory and intelligent informatics, Vol. 23, No. 5, pp. 749–760 (2011)
- ▶ Ying Dai, “Sample–specific repetitive learning for photo aesthetic auto–assessment and highlight elements analysis”, Multimedia Tools and Applications, Vol. 80, pp. 1387–1402, 2021

参考文献

- ▶ Ying Dai, “Exploring Metrics to Establish an Optimal Model for Image Aesthetic Assessment and Analysis”, Journal of Imaging 2022, 8(4), 85, DOI: [10.3390/jimaging8040085](https://doi.org/10.3390/jimaging8040085), 2022
- ▶ Ying Dai, “Building CNN-Based Models for Image Aesthetic Score Prediction Using an Ensemble”, Journal of Imaging 2023, 9(2), 30, DOI: [10.3390/jimaging9020030](https://doi.org/10.3390/jimaging9020030), 2023
- ▶ Ziyi Zhu, Ying Dai, “A New CNN-Based Single-Ingredient Classification Model and its Application in Food Image Segmentation”, Journal of Imaging 2023, 9, 205, DOI: [10.3390/jimaging9100205](https://doi.org/10.3390/jimaging9100205), 2023
- ▶ Kun Fu, Ying Dai, et al. “CNN-based visible ingredients recognition in a food image using decision making schemes”, Proc. of IEEE SMC 2023, pp. 2427–2432, USA, DOI: [10.1109/SMC53992.2023.10394513](https://doi.org/10.1109/SMC53992.2023.10394513), 2023