深層学習による食材認識のための合成画像を用いた学習データ生成

○山辺　貴之（金沢大学），石地　竜也（金沢大学），辻　徳生（金沢大学），関　啓明（金沢大学），
平光　立拓（金沢大学）

Generating Training Data Using Composite Images for Food Recognition by Deep Learning

○ Takayuki YAMABE (Kanazawa University), Tatsuya ISHICHI (Kanazawa University), Tokuo TSUJI (Kanazawa University), Hiroaki SEKI (Kanazawa University), and Tatsuhiro HIRAMITSU (Kanazawa University)

Abstract : Recognition of the quantitative balance of food ingredients in the diet is in increasing demand for health care, product quality control and appearance assessment. In order to achieve highly accurate image recognition by deep learning, it is necessary for humans to assign correct labels to many training images. In this study, we propose a method for efficient training by combining images of only one type of food which can be labeled automatically. In addition, we combine the basic data enhancement methods of cropping, flipping/rotating, and color manipulation to generate images suitable for learning multiple ingredients. In the experiment, we checked the effectiveness of each process and compared the percentage of correct answers for two sets of ingredients.

1. 緒言

深層学習の登場によって一般画像認識の性能が向上し，その応用問題として食事画像の認識も盛んに研究されてきた．現在，一般的な料理カテゴリーの認識に関しては実用的な精度に達しており，実社会での運用が進んでいる [1]．一方，食事に含まれている食材の量的なバランスの認識は未解決の課題であり，健康管理や製造分野における製品の品質管理，見た目の評価のための需要が増大している．食材バランスの認識手法の一つとして，本研究で扱う深層学習によるセマンティックセグメンテーションが挙げられる．セマンティックセグメンテーションは画像に対しピクセル単位で物体の領域を抽出する処理である．食材画像のセマンティックセグメンテーションの一例を Fig. 1 に示す．

画像認識の学習のための学習データの作成は，一般的に人間が画像に対して正解ラベルを付ける作業を行う必要があるため，大きな労力がかかり，多数のデータを準備することが難しい．これに対し，データの作成を効率的に行おうとする研究 [2] や，少数のデータを変形させることでデータの多様性を確保するデータ拡張の研究 [3] が行われている．本研究では，特に食材のセマンティックセグメンテーションに有効な学習データ生成手法を提案する．

ここで，Fig. 2 に示すような背景と複数種類の食材が写っている画像を混合画像 (image of mixed food)，Fig. 3 に示すような背景と 1 種類の食材のみが写っている画像を単種画像 (image of one kind food) と定義する．食材バランスの認識において必要なのは混合画像の識別であり，その学習に必要な画像も混合画像である．混合画像のラベル付けは手動で行う必要があり，食材の種類が増えると作業時間が非常に長くなる．これに対し，単種画像のラベル付けは背景と食材の色の違いから自動で行うことができる．したがって，単種画像同士を合成して混合画像のような合成画像を生成することができればラベル付け作業が容易になる．このような，合成画像を用いて学習データを生成する手法は，農園画像における隠れ果実領域の抽出 [4] という二値化の課題に対して利用されているが，セマンティックセグメンテーションに応用された例は無い．そこで，本研究ではセマンティックセグメンテーションの学習における，画像合成による学習データ生成手法の提案と評価を行う．

本研究では，画像合成に加え，基本的なデータ拡張手法である切り抜き，反転・回転，色操作の処理を組み合わせ，複数食材の学習に適した画像生成を行う．データ生成の過程を Fig. 4 に示す．実験では提案手法の評価のために，データ生成で行う各処理の正解率にもたらす影響の確認と，2 つの食材セットに対する正解率の比較を行った．

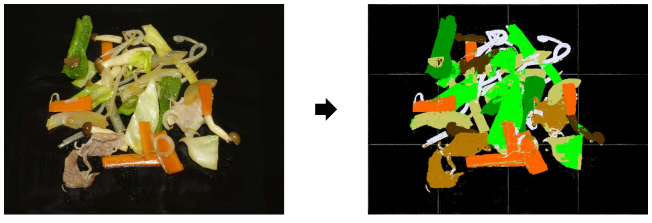

Fig. 1: Example of semantic segmentation

SY0009/21/0000 - 0160 © 2021 SICE

Fig. 2: Images of mixed food

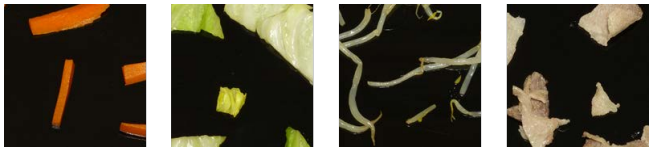

Fig. 3: Images of one kind food

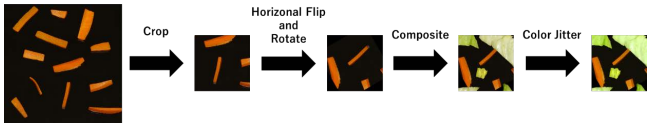

Fig. 4: Process of data generation

2. データ拡張手法

2.1 切り抜き (Crop)

複数回行うことにより，1 枚の画像から様々な食材配置のバリエーションを生成することを目的とした処理である．入力画像からランダムに 1000x1000 ピクセルの領域を切り抜いて出力する．Fig. 5 に一例を示す．

2.2 反転・回転 (Horizonal Flip and Rotate)

食材がどのような角度で置かれていても対応させることを目的とした処理である．反転処理では入力画像を左右反転して出力する．回転処理では入力画像に対して，画像中心を中心として 0〜360 度の範囲でアフィン変換による回転を行って出力する．一連の処理としては，入力画像に対して 50%の確率で反転処理を行い，その後に回転処理を行って出力する．Fig. 6 に一例を示す．

2.3 画像の合成 (Composite)

食材同士の様々な重なり方に対応させることを目的とした処理である．ラベル付けを利用して画像の食材部分のみを切り出し，背景との境界部分のノイズが入らないように領域を収縮した上で合成し出力する．合成後は境界部分に不自然な凹凸が生まれる．そこで，境界部分にのみぼかし処理を行う．Fig. 7 に一例を示す．実際に認識する画像は，一種類の食材しか映っていない場合から，すべての種類の食材が写っている場合まで想定できるため，合成処理は 1 枚の訓練画像に対して 0〜6 回行う．

2.4 色操作 (Color Jitter)

色合いが多少異なっていても同じ特徴を持っていれば同じ食材であると認識させることを目的とした処理である．入力画像の RGB の 3 値のそれぞれを 0.8〜1.2 倍して出力する．Fig. 8 に一例を示す．

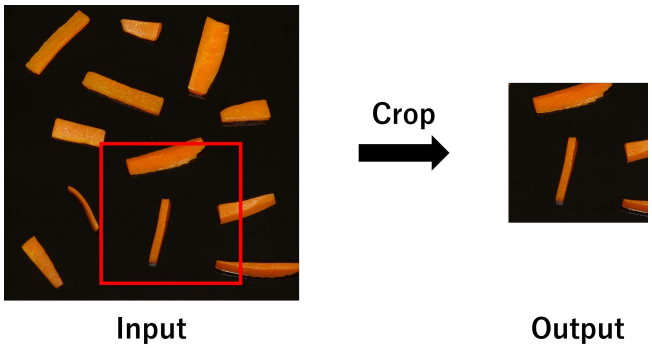

Fig. 5: Example of crop

Fig. 6: Example of horizonal flip and rotate

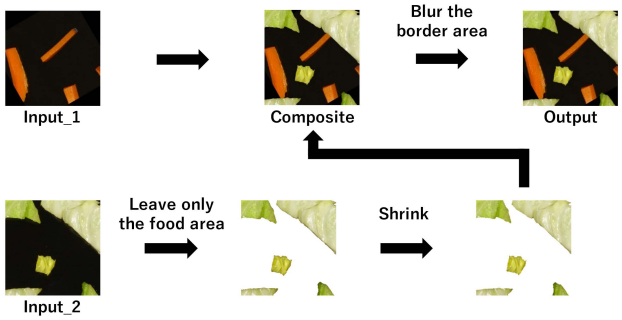

Fig. 7: Example of composite

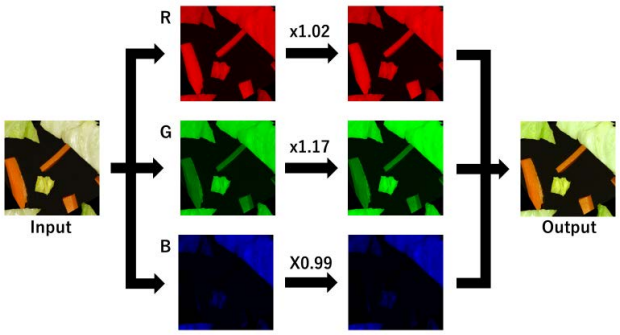

Fig. 8: Example of color jitter

Table 1: Label colors

Color \ Food Set	1	2
	back ground	back ground
	carrot	carrot
	cabbage	cabbage
	sprout	bamboo shoot
	pork	pork
	green pepper	green pepper
	onion	onion
	shimeji	kikurage

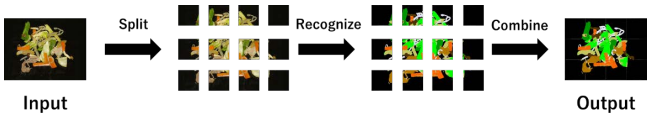

Fig. 9: Process of recognition

3. 食材認識実験

3.1 実験準備

以下に示す 2 種類の食材セットを準備し，黒色の背景上で撮影した．

- 食材セット 1(野菜炒め)

  中火で 5 分間炒めた人参，キャベツ，もやし，豚肉，ピーマン，玉ねぎ，しめじ．

- 食材セット 2(解凍野菜)

  冷凍状態から解凍した人参，キャベツ，たけのこ，豚肉，ピーマン，玉ねぎ，きくらげ．野菜炒めとは調理過程が異なるため同じ種類の食材でも見た目が異なる．

撮影条件はカメラ Panasonic DMC-FZ200，F 値 8.0，SS 1.0，ISO 感度 100，カメラと皿の距離 60cm，ズーム 4 倍，蛍光灯下，順光での撮影とした．

食材の調理後に，単種画像用と混合画像用の食材に分け，単種画像は各種 4 枚ずつの計 28 枚撮影し，混合画像は訓練データ用の 4 枚とテストデータ用の 4 枚の計 8 枚撮影した．撮影においては同じ食材が 2 回以上写らないようにした．単種画像は，HSV 色空間において明度が一定値以下のピクセルを背景，それ以外を食材として自動でラベル付けした．きくらげについては自動でのラベル付けが困難であるため，手動でラベル付けを行った．混合画像は，評価実験のために手動で食材に対応するラベル付けを行い，正解を作成した．各ラベルに対応する色を Table. 1 に示す．

3.2 実験手法

3.2.1 学習と認識

学習では，データ拡張を行った後の 1000x1000 ピクセルの訓練画像をモデルに学習させる．

認識では，学習済みのモデルにテストデータを入力して認識を行う．テストデータとして入力する画像は 4000x3000 ピクセルであり，学習した 1000x1000 ピクセルの訓練画像とは異なる．そこで，入力画像を分割してそれぞれに対するセマンティックセグメンテーションを行った後に元の形に結合して認識結果として出力する (Fig. 9)．

3.2.2 評価関数

本研究のような多クラス分類のモデルの性能を評価するための指標として，機械学習ライブラリの keras [5] が提供している categorical accuracy という評価関数がある．categorical accuracy はピクセル毎にモデルの出力値が最も大きいラベルと正解ラベルが一致しているかを調べ，一致しているピクセルの数をピクセルの総数で割った値である．ここで，本研究で用意した画像は背景の認識が容易であるため，テスト画像の背景の領域が多いほど categorical accuracy が高くなってしまう．このため，食材に対する正解率を適切に評価できない可能性がある．したがって，背景を除いた食材のピクセルのみに対して categorical accuracy と同様の手法で正解率を評価する自作の評価関数 food accuracy を用いる．

3.3 実験条件

共通の学習条件はモデル U-Net [6]，訓練データ数 10000，エポック数 30，バッチサイズ 8，損失関数 Categorical Cross Entropy，最適化手法 Adam [7]，学習率 0.001 とした．

Table. 2 に示すように，訓練データとテストデータ，訓練データを生成する際のデータ拡張手法の条件が異なる 8 つの学習を行った．各条件を以下に示す．

学習 A〜F は食材セット 1 のテストデータで評価する．

- 学習 A

  食材セット 1 の単種画像に対して色操作，反転・回転，切り抜き，合成のデータ拡張を行って生成した訓練データによる学習．

- 学習 B

  学習 A の条件において，色操作のデータ拡張を行わずに生成した訓練データによる学習．

- 学習 C

  学習 A の条件において，反転・回転のデータ拡張を行わずに生成した訓練データによる学習．

- 学習 D

  学習 A の条件において，切り抜きのデータ拡張を行わずに生成した訓練データによる学習．

- 学習 E

  学習 A の条件において，合成のデータ拡張を行わずに生成した訓練データによる学習．

- 学習 F

  学習 A の条件において，単種画像の代わりに混合画像を用いて，合成のデータ拡張を行わずに生成した訓練データによる学習．

学習 G，H では食材セット 2 のテストデータで評価する．

- 学習 G

  学習 A の条件において，食材セット 1 の代わりに食材セット 2 を用いて生成した訓練データによる学習．

- 学習 H

  学習 A の条件において，もやしとしめじの単種画像のみを食材セット 2 のたけのことキくらげの単種画像に置き換えて生成した訓練データによる学習．

学習 A と比較して，学習 B-E は各データ拡張手法の有無の影響，学習 F は混合画像で学習した場合の違い，学習 G，H は異なるデータセットに対する結果を確認した．

3.4 実験結果

それぞれの学習条件について，テストデータに対するエポック毎の food accuracy を Fig. 10-11 に示す．テスト画像とそれに対する学習 A のモデルの出力画像を Fig. 12- 13 に示す．学習後の最終的な food accuracy を Table. 3 に示す．学習 A，F，G，H のモデルのラベル別の認識結果の平均値の混同行列を Table. 4-7 に示す．

Table 2: Study conditions

Study	Food Set for Test	Food Set for Training	Image Type for Training	Color Jitter	Horizonal Flip and Rotate	Crop	Composite
A	1	1	One Kind	○	○	○	○
B	1	1	One Kind		○	○	○
C	1	1	One Kind	○		○	○
D	1	1	One Kind	○	○		○
E	1	1	One Kind	○	○	○	
F	1	1	Mixed	○	○	○	-
G	2	1*	One Kind	○	○	○	○
H	2	2	One Kind	○	○	○	○

*Bamboo shoot and kikurage are from the food set 2

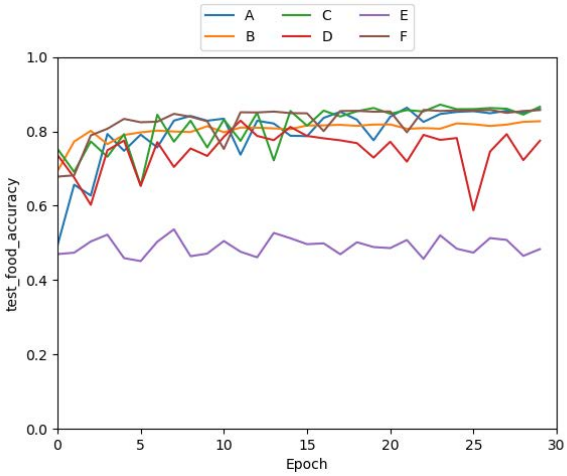

Fig. 10: Food accuracy to test data (A-F)

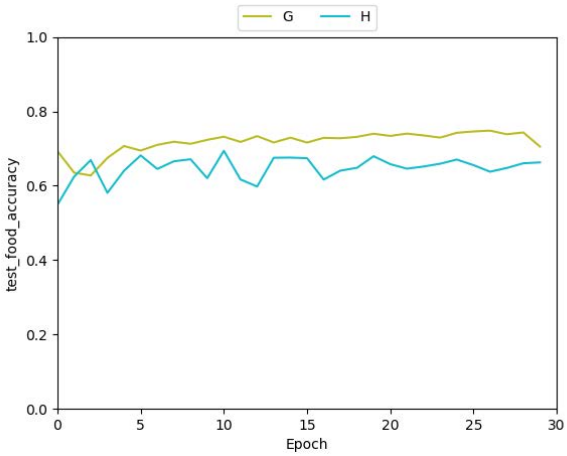

Fig. 11: Food accuracy to test data (G,H)

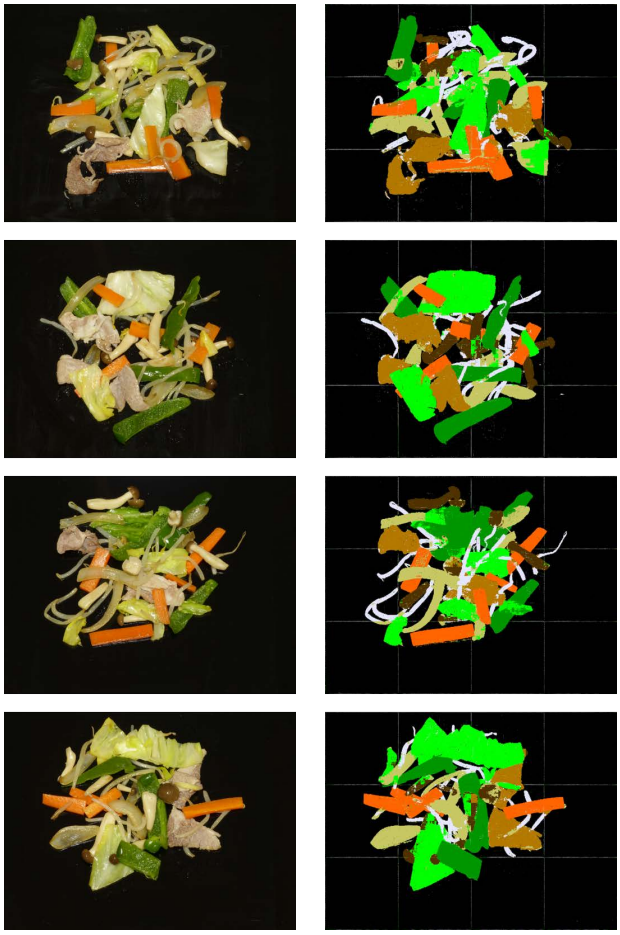

Fig. 12: Input images

Fig. 13: Output images

Table 3: Food accuracy after studying(%)

Study	Food accuracy
A	86.0
B	82.7
C	86.6
D	77.5
E	48.3
F	85.7
G	70.5
H	66.3

Table 4: Confusion-matrix of study condition A (%)

correct \ predict	back ground	carrot	cabbage	sprout	pork	green pepper	onion	shimeji
background	98.63	0.09	0.10	0.75	0.06	0.16	0.14	0.07
carrot	0.63	95.21	0.69	0.80	0.29	0.14	1.97	0.27
cabbage	0.80	0.46	82.38	1.75	0.25	9.25	4.83	0.30
sprout	1.81	3.89	10.38	64.08	4.29	1.07	12.98	1.51
pork	0.83	0.50	0.60	4.11	90.31	0.16	2.58	0.93
greenpepper	0.75	0.08	0.80	0.51	0.05	97.42	0.34	0.08
onion	1.00	0.71	8.21	2.02	0.12	1.62	85.99	0.34
shimeji	2.03	0.86	10.71	3.82	2.59	0.64	9.56	69.79

Table 5: Confusion-matrix of study condition F (%)

correct \ predict	back ground	carrot	cabbage	sprout	pork	green pepper	onion	shimeji
background	99.36	0.06	0.07	0.22	0.08	0.07	0.09	0.07
carrot	1.59	90.08	1.03	0.67	0.46	0.23	5.60	0.36
cabbage	1.09	0.21	77.78	2.39	1.24	8.39	6.38	2.53
sprout	2.91	1.02	7.67	71.25	5.57	1.19	8.89	1.52
pork	1.28	0.27	1.73	4.09	89.68	0.18	1.22	1.58
greenpepper	1.64	0.12	1.22	0.41	0.11	95.65	0.46	0.39
onion	1.72	0.65	1.47	7.81	2.06	1.07	82.65	2.58
shimeji	1.75	0.64	3.82	6.10	4.80	0.50	2.74	79.67

Table 6: Confusion-matrix of study condition G (%)

correct \ predict	back ground	carrot	cabbage	bamboo shoot	pork	green pepper	onion	kikurage
background	98.78	0.05	0.18	0.02	0.07	0.14	0.03	0.74
carrot	0.38	92.88	3.54	0.09	2.54	0.13	0.07	0.39
cabbage	0.27	0.61	82.53	3.56	11.93	0.19	0.81	0.10
bambooshoot	0.29	0.62	33.93	48.70	16.16	0.06	0.13	0.11
pork	0.24	0.84	33.46	12.55	52.57	0.11	0.08	0.16
greenpepper	0.42	1.84	54.11	0.30	0.28	42.80	0.15	0.12
onion	0.50	0.58	39.99	19.85	22.29	0.12	16.41	0.27
kikurage	15.08	5.54	2.18	0.13	1.37	0.94	0.05	74.73

Table 7: Confusion-matrix of study condition H (%)

correct \ predict	back ground	carrot	cabbage	bamboo shoot	pork	green pepper	onion	kikurage
background	96.84	0.03	0.05	0.10	0.03	0.05	0.05	2.86
carrot	0.38	87.12	2.52	0.08	1.91	0.10	7.17	0.73
cabbage	0.29	0.52	64.43	15.04	8.01	0.67	10.91	0.14
bambooshoot	0.32	0.62	18.80	61.24	14.58	0.09	4.17	0.19
pork	0.28	0.76	16.16	10.68	66.13	0.13	5.65	0.22
greenpepper	0.40	0.85	18.08	0.35	0.10	79.50	0.64	0.10
onion	0.49	0.73	16.83	52.88	20.75	0.16	7.73	0.43
kikurage	2.77	0.89	1.02	0.21	0.96	0.39	5.65	88.11

4. 実験結果の考察

Fig. 10-11 において，学習後期の正解率が安定していることから，最終結果を用いて比較を行う．

4.1 データ拡張手法に関する比較

Table. 3 において，学習 A に比べ学習 B，D，E の食材の正解率が低いことから，色操作，切り抜き，合成のデータ拡張による正解率の増加を確認できた．しかし，学習 A と学習 C の正解率にはほとんど違いが無かったことから反転・回転のデータ拡張による正解率の増加は確認できなかった．これは，訓練データに食材の向きという点において十分なバリエーションがあったことで反転・回転が無意味なデータ拡張になったためと考えられる．また，学習 A と学習 F の正解率にほとんど違いが無いことから，本研究の条件においては単種画像を合成した画像は混合画像と同等の訓練データとして活用できることが確認できた．

Table. 4- 5 から，合成画像で利用した学習 A では混合画像を利用した学習 F より，もやしの正解率が低いことを確認できた．この原因として，Fig. 12- 13 から，半透明なもやしが背後にある別の食材の色を透過している部分の誤認識が挙げられる．この誤認識は，本研究の単種画像の合成手法では背後の色を透過するという処理を行っていないために，そのような状況を学習できずに発生したものと考えられる．

4.2 食材セットに関する比較

Table. 3 において，学習 G の正解率は学習 A の正解率と比較して 15.5%低いことが確認できた．また，Table. 6 において，学習 G では他の食材と大きく異なる色をしている人参については学習 A と同様に高精度の認識を行うことができているが，たけのこ，豚肉，ピーマン，玉ねぎがキャベツと誤認識されている割合が大きい．これは，食材セット 2 は人間でも判別が困難な食材が多いということや，前述のもやしのように半透明な食材に対応できなかったことが原因である．Table. 7 において，学習 H では，たけのこときくらげ以外は別の調理過程を経た食材の画像を用いたにもかかわらず，玉ねぎ以外の食材については 60%以上の正解率を示している．したがって，調理過程が異なる入力画像に対しても対応できる汎化性能を確認できた．

5. 結言

食材画像のセマンティックセグメンテーションを実装し，提案手法によって，ラベル付けが容易な単種画像を合成することで効率的な学習を行えることを実験から確認した．データ拡張手法については，色操作，切り抜き，合成のデータ拡張による正解率の増加を確認した．提案手法によって学習したモデルの正解率は野菜炒めに対しては 86.0%，解凍野菜に対しては 70.5%であった．共通する食材を野菜炒めの食材で学習したモデルを，解凍野菜で評価すると玉ねぎ以外の食材の正解率は 60%以上であり，汎化性能を確認できた．提案手法では半透明な食材に対して，背後に別の食材があるときに誤認識を起こしやすいことが分かった．

参考文献

[1] 柳井啓司． 食事画像認識の現状と今後． 人工知能, Vol. 34, No. 1, pp. 41–49, 2019.

[2] 田中裕隆, 新納浩幸. 物体検出における教師データの効率的な作成. 人工知能学会全国大会論文集 第 34 回全国大会 (2020), pp. 3Rin455–3Rin455. 一般社団法人人工知能学会, 2020.

[3] Agnieszka Mikołajczyk and Michał Grochowski. Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)*, pp. 117–122. IEEE, 2018.

[4] 高井亮磨, 小林一樹． 農園画像における深層学習を用いた隠れ果実領域の抽出． 人工知能学会全国大会論文集 第 33 回全国大会 (2019), pp. 1F3OS17a03–1F3OS17a03. 一般社団法人 人工知能学会, 2019.

[5] François Chollet, et al. Keras. `https://keras.io`, 2015.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

[7] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.