

## 腹部超音波画像からの腫瘍検出

B4 原 英吾

### 1 研究背景および目的

- 背景
  - － 器具の操作と診断を同時に行わなければならない高難易度
  - － 肝臓は沈黙の臓器と呼ばれ初期には自覚症状がほとんどない
    - \* 症状を自覚している時には重症化している場合が多い
  - － 機械学習による診断のサポート
    - \* 良性・悪性を見分けることが重要視される
    - \* 図 1 の様に明らかなラベル不足<sup>1</sup>のある画像が存在する
- 目的
  - － 既存の研究を踏まえたモデルの精度向上
    - \* 良性・悪性判別の高精度化
  - － 超音波支援システムの開発
    - \* 早期発見につながると良い

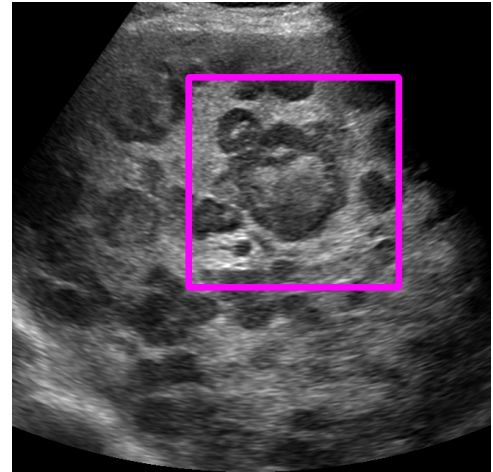


図 1: ラベル不足のある診断画像例

### 2 これまでの研究のまとめ

- データセット
  - － 国立研究開発法人日本医療研究開発機構 (AMED)<sup>2</sup>が提供している延べ 8 万枚に及ぶ以下のデータが付随
    - \* 腹部超音波画像, ROI
    - \* 年齢, 性別

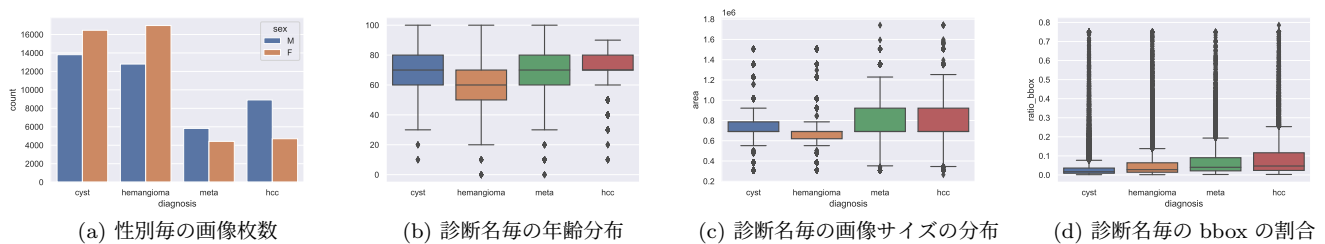


図 2: データセットにおけるデータの分布

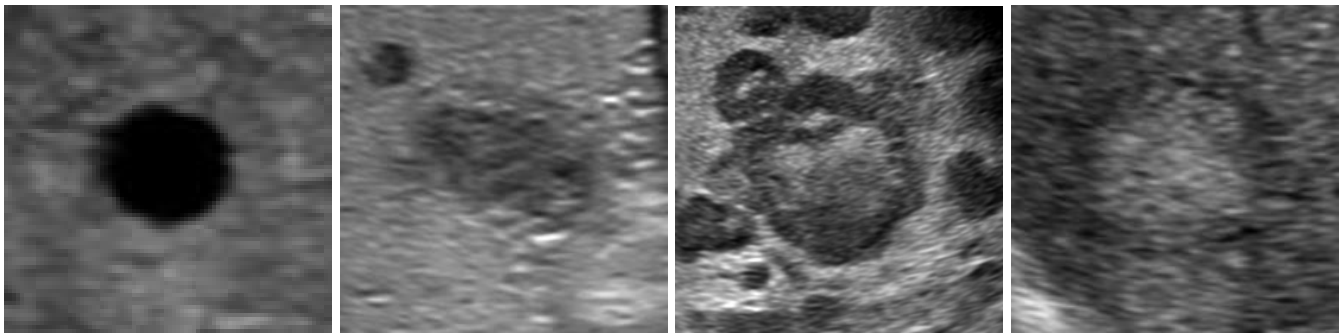
- － 性別 (図 2a)
  - \* HCC(肝細胞癌) は男性が罹患しやすい
    - ・ 昔は男性の方が飲酒・タバコが多く癌に罹りやすかったという時代背景があるかもしれない
  - \* hemangioma(血管腫) は女性が罹患しやすい
  - \* Meta(転移性肝癌) は他の症状よりも少ない
- － 年齢 (図 2b)
  - \* cyst(単純嚢胞), hemangioma(血管腫) の分布にはあまり特徴がない
  - \* hemangioma(血管腫) は比較的若年層でも罹患する

<sup>1</sup>今回は図 1 の様なアノテーションが不足しているものを指す

<sup>2</sup><https://www.amed.go.jp/>

- \* Meta(転移性肝癌) における 0 歳はラベルミスである可能性が高い
- \* HCC(肝細胞癌) は比較的高齢者が罹患しやすい
- 画像サイズ (図 2c)
  - \* hemangioma(血管腫) は比較的画像サイズが統一されている
    - ・ 腫瘍の大きさが血管に依存するためあまり偏りが生じていない?
- bbox の画像に占める割合 (図 2d)
  - \* cyst(単純嚢胞) は他の診断と比べて bbox の割合が低い ( $\frac{1}{2}$  程度) である
  - \* HCC(肝細胞癌) は画像に占める bbox の割合が高い

● 症状毎の特徴を調査



(a) cyst(単純嚢胞)

(b) hemangioma(血管腫)

(c) Meta(転移性肝癌)

(d) HCC(肝細胞癌)

図 3: 症状毎における腫瘍の超音波画像

- cyst(単純嚢胞) (図 3a)
  - \* 液体が貯留されている状態
  - \* 症状がでないことが多いため大きな腫瘍になって発見されることが多い
  - \* 嚢胞の内腔に向けて増殖するため転移することは少ない
- hemangioma(血管腫) (図 3b)
  - \* 肝臓にできる良性腫瘍の中で最も多い
  - \* 女性ホルモンが原因で女性が罹患しやすいと言われているが詳しくは解明されていない
  - \* 血管が無数に絡み合うことによって出来た血管の塊であることから血流が遅いという特徴がある
  - \* 他の臓器に浸潤したり転移することは無いと言われている
- Meta(転移性肝癌) (図 3c)
  - \* 門脈を介して大腸癌などの消化器癌から転移する割合が多い
  - \* 類似したエコーパターンをもつ腫瘍が多発してみられることが多い
- HCC(肝細胞癌) (図 3d)
  - \* 肝臓にできる悪性腫瘍の中で最も多いと言われている
  - \* 約 90 %がウイルス感染症が原因
    - ・ B 型肝炎ウイルス (HBV) が約 20%
    - ・ C 型肝炎ウイルス (HCV) が約 70%

● データクレンジング

1. 400 × 400 以下の画像の除外
2. Perceptual Hash を利用した類似画像の除外
3. 青色や黄色のスケールの除去

● 提供されているデータを COCODataset の形式に変換

- train data : test data : val data = 67122 : 8390 : 8391
- 見やすいようにインデントしたファイル<sup>3</sup>も作成

<sup>3</sup>[//aka/work/hara.e/AMED/lib/dataset/annotations/train\\_large.json](https://aka/work/hara.e/AMED/lib/dataset/annotations/train_large.json) など

- 1 クラス (腫瘍) での学習結果

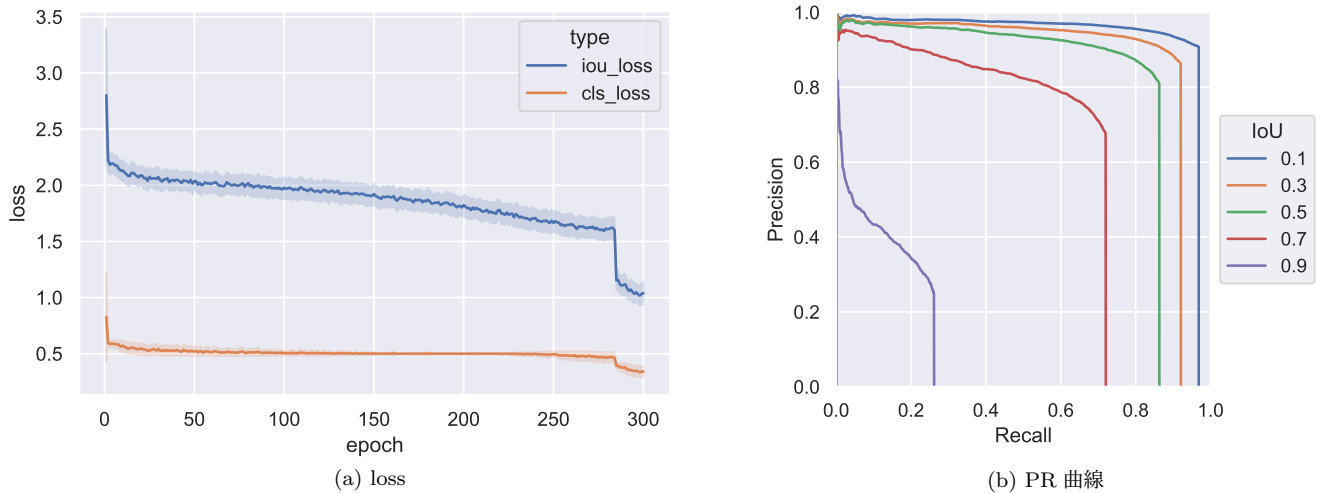


図 4: YOLOX[2] で 1 クラスの検出を行った時の loss と PR 曲線

表 1: 1 クラスの学習で得られた精度

model	backbone	classes	epoch	size	batch_size	IoU			area <sup>4</sup>		
						mAP	AP <sub>50</sub>	AP <sub>75</sub>	AP <sub>S</sub>	AP <sub>M</sub>	AP <sub>L</sub>
YOLOX[2]	DarkNet	1	300	512	64	0.519	0.839	0.558	-	0.639	0.631

- 考察

- 図 4a から

- \* IoU の loss が class 分類の loss に比べて大きい
  - ・ 1 クラス分類のため cls\_loss は小さくなる傾向がある
  - ・ Noisy Label の影響?
- \* IoU の loss でより大きな Double Descent[1] が起きている
  - ・ Noisy Label が最大の要因
  - ・ iou\_loss と cls\_loss は backbone として DarkNet を共有していることから cls\_loss にも影響が出ている?

- 図 4b から

- \* Recall が Precision に比べて低い
  - ・ Recall が低いということは腫瘍を検出できていないということ
  - ・ 医用で Recall が低いのは問題

<sup>4</sup>Small < 32<sup>2</sup> < Medium < 96<sup>2</sup> < Large

### 3 前回の GM からの進捗

- 4 クラス分類での学習結果

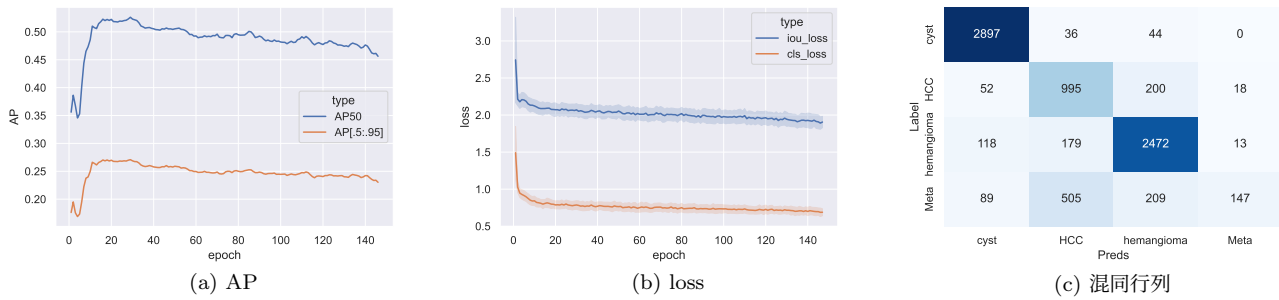


図 5: YOLOX[2] で 4 クラス分類を行った時の AP と loss および混同行列

表 2: 4 クラス分類を行った時の未検出数

Diagnosis	cyst	HCC	hemangioma	Meta
未検出数	60	106	180	71
データ総数	3037	1371	2962	1021
未検出割合	0.0198	0.0773	0.0608	0.0695

- 図 5a から
  - \* 30epoch 辺りから過学習が起きている
- 図 5b から
  - \* IoU が精度を下けている大きな要因
- 表 2 から
  - \* cyst (単純嚢胞) は比較的未検出数が少ない
- 図 5c と表 2 から表 3 の様な値を得ることができる

表 3: 評価指標毎の値

Diagnosis	accuracy	precision	recall	f1-score
cyst	0.8789	0.9179	0.9539	0.9356
HCC	0.4758	0.5801	0.7257	0.6448
hemangioma	0.7238	0.8451	0.8346	0.8398
Meta	0.1397	0.8258	0.1440	0.2452

- 表 3 から
  - \* 精度が要求される診断名の精度が低い
    - ・ 実用上最も問題がある診断名は Meta (転移性肝癌)
    - ・ 予測が最も信用出来ない診断名は HCC (肝細胞癌)
- まとめ
  - \* cyst (単純嚢胞)
    - ・ ほぼ問題ない
  - \* hemangioma (血管腫)
    - ・ cyst (単純嚢胞) や HCC (肝細胞癌) と誤分類してしまうことが多い
    - ・ 未検出の多さが目立つ
  - \* HCC (肝細胞癌)
    - ・ hemangioma (血管腫) と誤分類してしまうことが多い
    - ・ 未検出が多い
  - \* Meta (転移性肝癌)
    - ・ HCC (肝細胞癌) や hemangioma (血管腫) と誤分類してしまうことが多い
    - ・ 正確な予測ができていない (予測がバラバラ) ことが多い

- 1 クラス分類と 4 クラス分類の比較

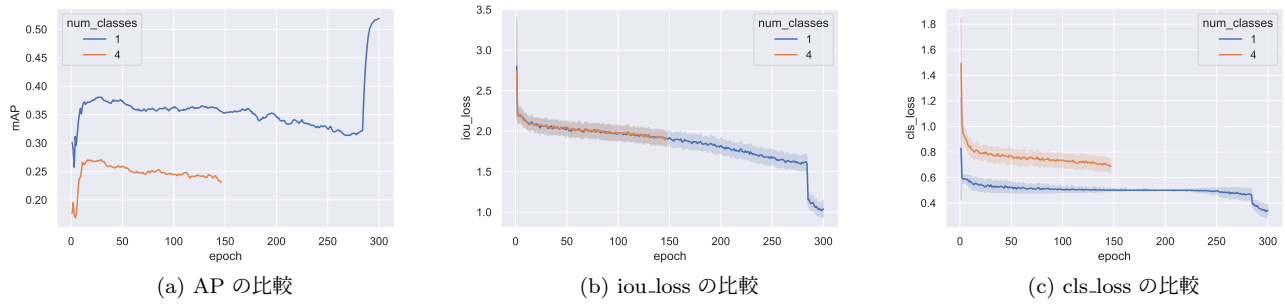


図 6: 1 クラス分類と 4 クラス分類の比較

- 図 6a から
  - \* 4 クラス分類の方が AP が低い
  - \* クラス数を増やしているから当たり前
- 図 6b から
  - \* 1 クラスと 4 クラスは同じくらいの学習速度
  - \* 1 クラスと 4 クラスでは iou\_loss の差異が小さい
  - \* クラス数を増やしても IoU の精度はあまり変化が無い
- 図 6c から
  - \* 1 クラスと 4 クラスでは cls\_loss の差異が大きい
  - \* クラス数を増やしているから当たり前

表 4: 1 クラスと 4 クラスの学習で得られた精度

model	backbone	classes	epoch	size	batch_size	IoU			area		
						mAP	AP <sub>50</sub>	AP <sub>75</sub>	AP <sub>S</sub>	AP <sub>M</sub>	AP <sub>L</sub>
YOLOX[2]	DarkNet	1	300	512	64	0.519	0.839	0.558	-	0.639	0.631
		4				0.279	0.526	0.248	-	0.221	0.288

- 表 4 から
  - \* AP が半分程度になっている
  - \* IoU が 0.5 の時は AP が比較的高い
  - \* 4 クラス分類では腫瘍が大きいものは分類しやすい傾向がある？

- これから用いるモデルについて考えた

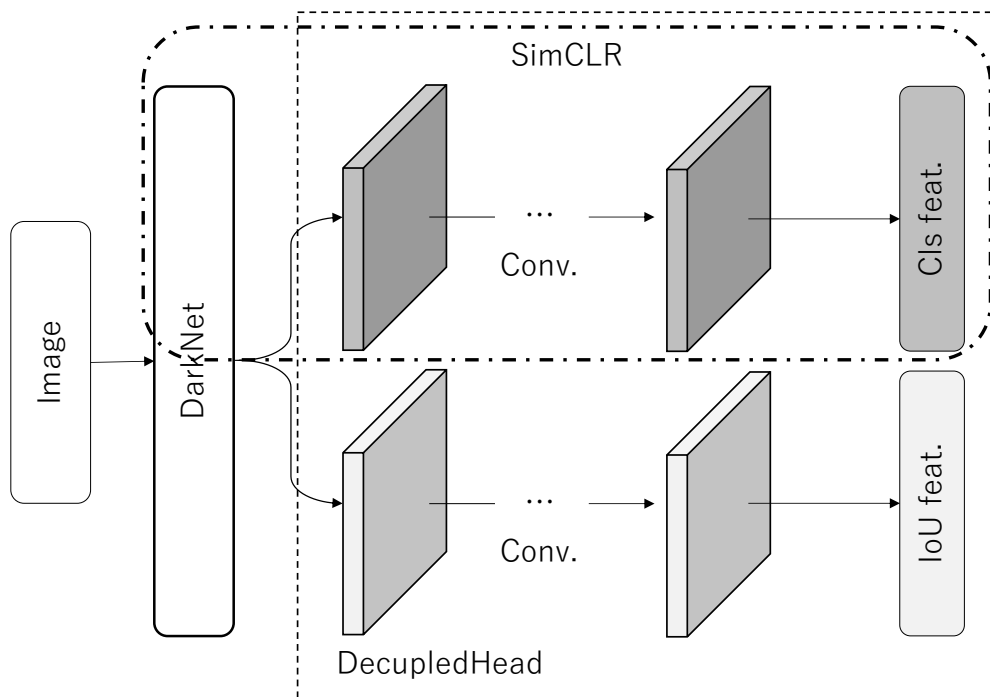
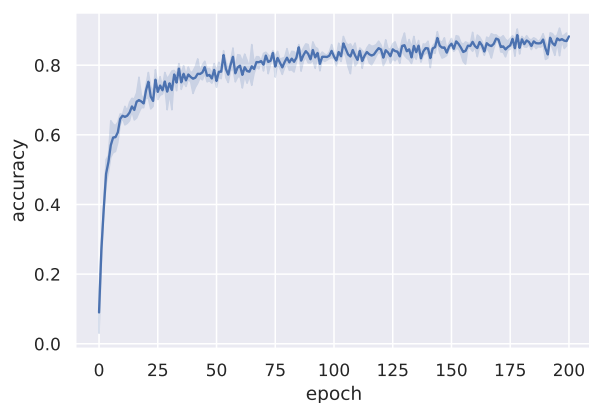
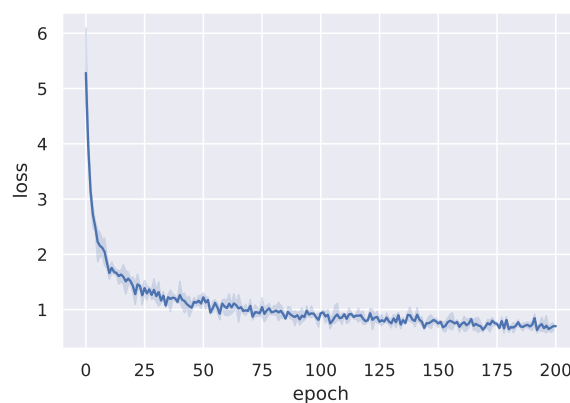


図 7: YOLOX[2] と SimCLR[3] の複合モデル

- クラス分類の部分の model を SimCLR[3] で事前学習
- 一部 (クラスの特徴量を抽出する層) だけ freeze する
- SimCLR[3] を使ってみた
  - ハイパーパラメーター
    - \* backbone: ResNet18
    - \* batch\_size: 256
    - \* epoch 数: 200
    - \* size: (96, 96)



(a) accuracy



(b) loss

図 8: SimCLR[3] で対照学習を行なった結果 (train データ)

## 4 今後の課題&スケジュール

- 5/31 までに
  - 対照学習の結果を (遅くとも 6/14 までに) 出すことを目標にする
  - 結果と言っても (できれば同じ backbone を使って) test データで Precision や Recall を出すところまで

## 参考文献

- [1] Anonymous authors. [DEEP DOUBLE DESCENT: WHERE BIGGER MODELS AND MORE DATA HURT](#) ICLR 2020, 2020.
- [2] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. [YOLOX: Exceeding YOLO Series in 2021](#), 2021.
- [3] Ting Chen, Simon Kornblith, Mohammad Norouzi, Geoffrey Hinton. [A Simple Framework for Contrastive Learning of Visual Representations](#), 2020.