

## 物体検出モデルによる造影・非造影CT画像からの腎腫瘍の検出

深沢 貴希<sup>†</sup> 小川 悠太<sup>††</sup> 亀谷 由隆<sup>†</sup> 山田 啓一<sup>†</sup> 堀田 一弘<sup>†</sup>  
 高橋 友一<sup>‡</sup> 佐々 直人<sup>‡‡</sup> 松川 宜久<sup>§</sup> 岩野 信吾<sup>§</sup> 山本 徳則<sup>§</sup>

<sup>†</sup>名城大学大学院理工学研究科 <sup>††</sup>名城大学工学部 <sup>‡</sup>Meis Technology Inc.  
<sup>‡‡</sup>愛知医科大学医学部 <sup>§</sup>名古屋大学大学院医学系研究科

## 1 はじめに

近年、深層学習技術を腹部CT画像の分析に応用する試みがみられる。例えば腎臓の造影CT (contrast-enhanced CT, CECT) 画像のデータセット KiTS [1] や、画像認識モデルを用いた腎腫瘍の検出 [2], 腎腫瘍の良性・悪性の識別 [3], KiTSと物体検出モデルを用いた位置特定と腎腫瘍の検出 [4] がある。

田中らが非造影CT (unenhanced CT, UCT) 画像とCECT画像からの腎腫瘍の早期発見を目指し、畳み込みニューラルネットワーク (convolutional neural network, CNN) による画像分類に基づいて腹部CT画像からの腎腫瘍の検出を行い、学習において腎臓領域を切り出したマスク画像の使用が効果的であることを示した [2]。更に、CNNよりも、Transformerベースの画像分類モデルの方が検出精度の面で優れていることを示した。しかし、マスク画像を使う場合は新規患者に対する診断においてもCT画像に対してマスク処理をする必要があり、人的作業もしくは追加のセグメンテーションが必要になる。加えてCECT画像では組織間のコントラストが明確化されることで診断が容易になるが、CECT撮影前に投与する造影剤へのアレルギーや腎機能の低下が懸念される。

物体検出をCECT画像に用いた研究に用いた研究は存在する [4]。しかし、UCT画像に用いた研究は少ない。前述のKiTSでもCECT画像しか提供されていない。そこで本研究では、物体検出モデルであるSSD (Single Shot MultiBox Detector) [5], RetinaNet [6] やTransformerベースの物体検出モデルであるDETR (Detection Transformer) [7] を用いて腹部CT画像からの腎腫瘍の検出を試みる。また、どの物体検出モデルが腎腫瘍検出に有効なのかを評価する。マスク画像を用いた画像分類と異なり、物体検出ではアノテーションが必要なのは学習時のみであるため、予測時のコストを削減できる。また、物体検出モデルは予測ボックス内でクラス分類を行うため、マスク画像を用いた画像分類に近い予測精度が期待できる。

本論文の構成は以下の通りである。はじめに2節で本研究で対象とするデータについて記述する。そして3節で実

Kidney Cancer Detection from Contrast-Enhanced and Unenhanced CT Images Using Object Detectors

<sup>†</sup> Takaki Fukazawa, Yoshitaka Kameya, Keichi Yamada, Kazuhiro Hotta

<sup>††</sup> Yuta Ogawa

<sup>‡</sup> Tomoichi Takahashi

<sup>‡‡</sup> Naoto Sassa

<sup>§</sup> Yoshihisa Matsukawa, Shingo Iwano, Tokunori Yamamoto

Graduate School of Science and Technology, Meijo University (<sup>†</sup>)

Faculty of Science and Technology, Meijo University (<sup>††</sup>)

Meis Technology Inc. (<sup>‡</sup>)

Department of Medicine, Aichi Medical University, (<sup>‡‡</sup>)

Graduate School of Medicine, Nagoya University (<sup>§</sup>)

表1: データセットの構成

データセット	画像単位				患者単位			
	UCT		CECT		UCT		CECT	
腫瘍の有無	有	無	有	無	有	無	有	無
学習用	702	4979	719	5106	147	147	148	146
検証用	136	963	139	1005	28	32	30	30
テスト用	173	1323	192	1368	39	40	39	40



(a) UCT 画像



(b) CECT 画像

図1: DETRでの検出例

験方法、その後4節にて実験結果を示す。5節で論文のまとめを行い、今後の課題を述べる。

## 2 準備

本研究では名古屋大学医学部附属病院で収集されたUCT画像と、造影剤注入後30秒経過した後に撮影されたCECT画像に前処理を施して使用した。本研究で使用したグレースケール画像はCT値を-110から190に制限して作成した。そして、医療専門家が提案した腎臓と腎腫瘍(存在する場合)の上端と下端の軸方向の位置を基に、腎臓と腎腫瘍の有無を示すクラスラベルを紐づけた。また、腎臓周辺の他の臓器からの不要な影響を軽減するために、左右それぞれの腎臓が写るように、サイズ(512, 512)のCT画像からサイズ(256, 256)の範囲の領域を切り取った。更に、本研究の実験では右の腎臓に焦点を当てる。左右の腎臓は概ね対称なことから、専門医の助言の下、左の腎臓のCT画像を水平反転させて、別患者の右腎臓と見なしている。

物体検出モデルのデータセットには、画像ファイルに加えて検出対象の物体を囲うボックスの位置情報とクラス情報が必要である。田中らが使用した人手によるマスク画像 [2] に対してテンプレートマッチングを行い、各画像における腎臓のボックスの位置情報を取得した。マスク画像作成時に画素値がわずかに変わってマッチングミスした画像では、手動で正しい腎臓領域の座標を与えた。また、腎臓領域が非常に小さい画像等を田中らのデータセットから除外した。その結果、腫瘍ありの患者数も一部変化した。

最終的に、腫瘍の有無の割合を維持したまま、データセットを患者単位でランダムに学習用データセット、検証用データセット、評価用データセットに分割した。表1に、分割したデータセットの構成を示す。

### 3 実験方法

#### 3.1 物体検出モデルの学習

物体検出器の学習では、クラス数を3（腫瘍あり、腫瘍なし、背景）に、エポック数はSSDとRetinaNetが200、DETRが1000に設定した。RetinaNetはハイパーパラメータチューニングの結果、UCTでは $\alpha = 0.7$ 、CECTでは $\alpha = 0.4$ とし、他の設定はデフォルトにした。RetinaNetのハイパーパラメータ $\alpha$ は1未満のとき、背景の損失を小さく、物体（腫瘍あり、腫瘍なし）の損失を大きくする。また、SSDは事前学習済みのVGG-16をファインチューニングした。RetinaNetとDETRではResNet50を用いた。過学習を防ぐために、学習曲線を観察してSSDとRetinaNetは100エポック目のモデルを、DETRは500エポック目のモデルを使用した。比較のため、画像分類モデルであるSwin Transformerの事前学習済みモデルを100エポックでファインチューニングした。

#### 3.2 画像単位と患者単位での検出

本研究では、2段階で患者の腎腫瘍を検出した。まず最初の段階では、画像単位での検出を行い、患者の腹部CT画像を、腫瘍を含むもの（陽性クラス）と含まないもの（陰性クラス）に判定する。今回の実験で用いる画像には必ず1つの腎臓が写っているため、それを前提として必ず1つの腎臓を検出し、陽性クラスと陰性クラスに分類する。物体検出器は、背景、陽性クラス、陰性クラスそれぞれを検出したボックスとその確信度を同時に出力する。確信度は0から1の範囲をとる。そのため、陽性クラスを検出したボックスにおける最大の確信度を最大正確度、陰性クラスを検出したボックスにおける最大の確信度を最大負確信度と定義して、各画像における最大正確度と(1-最大負確信度)との調和平均を、画像単位での陽性クラスの確信度とする。なお、最大正確度が高く、かつ最大負確信度が低い場合のみ陽性と判定するために調和平均を用いている。

次の段階では、患者単位での検出を行うために、患者の腹部CT画像に対する画像単位での結果を組み合わせる。具体的には、画像単位での陽性クラスの確信度を各患者内で移動平均をとった時の最大値を患者単位での陽性クラスの確信度とした。

本研究での決定閾値は、画像単位・患者単位での検出ともに、検証用データセットにおいてF値が最も高くなる閾値とした。画像もしくは患者の確信度が各々の決定閾値以上なら陽性と判定する。

#### 4 実験結果

学習を行った物体検出モデルの画像・患者単位の検出精度を陽性（腫瘍を含む）クラスに対する適合率、再現率、F値、AUROCで評価する。臨床的な意味では患者単位の検出だけが意味をもつが、物体検出モデルの予測性能そのものは画像単位で評価の方が直接的である。まず、物体検出の出力例を図1に示す。これらの画像では腎臓領域を捉えることができおり、クラス分類も正しく行えている。予測結果のボックスと正解のボックスとのIoUの平均は0.8を超えており、IoUが0.5未満の割合も1%程度であったため、十分に腎臓領域を捉えることができていると言える。

次に、画像単位での検出結果を表2に、患者単位での検出結果を表3に示す。画像単位では、物体検出モデルにお

表 2: 画像単位でのモデル間比較結果

CT	Model	適合率	再現率	F 値	AUROC
UCT	SSD	0.567	0.462	0.510	0.819
	RetinaNet	0.505	0.572	0.537	0.801
	DETR	0.506	0.468	0.486	0.817
	Swin (参考)	0.474	0.642	<b>0.545</b>	<b>0.863</b>
CECT	SSD	0.794	0.682	0.734	0.920
	RetinaNet	0.714	0.677	0.695	0.878
	DETR	0.679	0.828	<b>0.746</b>	<b>0.947</b>
	Swin (参考)	0.805	0.688	0.742	0.935

表 3: 患者単位でのモデル間比較結果

CT	Model	適合率	再現率	F 値	AUROC
UCT	SSD	0.591	0.667	0.627	0.762
	RetinaNet	0.750	0.846	<b>0.795</b>	<b>0.810</b>
	DETR	0.815	0.564	0.667	0.756
	Swin (参考)	0.771	0.692	0.730	0.795
CECT	SSD	0.854	0.897	0.875	<b>0.951</b>
	RetinaNet	0.729	0.897	0.805	0.913
	DETR	0.841	0.949	<b>0.891</b>	0.933
	Swin (参考)	0.740	0.949	0.831	0.919

いてUCTではRetinaNetが、CECTではDETRがF値で最も優れていた。また、CECTではDETRが人手によるマスク画像を用いたSwin TransformerをF値とAUROCで上回った。患者単位では、UCTでRetinaNetが、CECTでSSDとDETRがSwin Transformerを上回る結果となった。全体としては、アノテーションコストの低い物体検出モデルの検出精度が人手によるマスク画像を用いた画像分類モデルの検出精度と同等であることが示された。

#### 5 おわりに

物体検出モデルを用いてUCT画像とCECT画像から腎腫瘍の検出を行い、3種類の物体検出モデルと画像認識モデルで検出精度を比較した。その結果、アノテーションコストを含めて考えたときの物体検出モデルの有効性が示された。今後は患者単位の検出方法の改善を行いたい。

#### 参考文献

- [1] N. Heller et al.: The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes, arXiv:1904.00445 (2019)
- [2] T. Tanaka et al.: Kidney Cancer Detection from CT Images by Transformer-Based Classifiers, Proc. of the 13th Int'l Conf. on Smart Computing and Artificial Intelligence (2023)
- [3] T. Tanaka et al.: Differentiation of Small ( $\leq 4$  cm) Renal masses on Multiphase Contrast-Enhanced CT by Deep Learning, American J. of Roentgenology, Vol. 214 (2020)
- [4] S. Mahmud et al.: Kidney Cancer Diagnosis and Surgery Selection by Machine Learning from CT Scans Combined with Clinical Metadata, Cancers, Vol. 15 (2023)
- [5] W. Liu et al.: SSD: Single Shot MultiBox Detector, Proc. of ECCV 2016 (2016)
- [6] T. Lin et al.: Focal Loss for Dense Object Detection, Proc. of ICCV 2017 (2017)
- [7] N. Carion et al.: End-to-End Object Detection with Transformers, Proc. of ECCV 2020 (2020)