

博士論文

深層学習を用いた
肝疾患の超音波検査映像に対する
診断支援システムの構築

唐子 顕児

目次

1. 序論	5
1.1 肝臓の腫瘍性疾患	5
1.2 超音波検査	6
1.3 画像診断.....	7
1.4 肝疾患の超音波検査映像に対する診断支援システム.....	8
1.5 本論文の目的と構成.....	11
2. 関連先行研究	13
2.1 基礎的な深層学習技術	13
2.1.1 単純パーセプトロン	13
2.1.2 多層パーセプトロン	14
2.1.3 誤差逆伝播法	15
2.1.4 畳み込みニューラルネットワーク	17
2.2 画像向け分類・物体検出モデルと評価指標	21
2.2.1 画像分類モデル - DPRN -	22
2.2.2 画像分類モデルの評価指標	26
2.2.3 物体検出モデル - Faster R-CNN -	28
2.2.4 物体検出モデル - Mask R-CNN -	33
2.2.5 物体検出モデルの評価指標	34
2.3 画像診断・検査への深層学習技術の応用と課題	36
2.3.1 医療向け画像診断・検査への応用研究と超音波検査の複雑性	36
2.3.2 データ多様性増加の必要性和応用研究	37
2.3.3 診断支援システムとしての評価	39
3. 腫瘍・血管検出モデルの実現.....	41
3.1 背景・目的	41
3.2 データセット	41
3.3 Faster R-CNN による腫瘍検知タスクの構築.....	43
3.3.1 Faster R-CNN 導入目的	43
3.3.2 Faster R-CNN による腫瘍・血管検出モデル構成	43
3.3.3 Faster R-CNN による腫瘍・血管検出性能の評価	45
3.4 Optical Flow Network 導入による時間変化情報補完モデルの構築.....	47
3.4.1 Optical Flow Network 導入目的	47
3.4.2 学習・評価データセット	48
3.4.3 Optical Flow Network 導入による腫瘍・血管検出モデル構成.....	48

3.4.4 Optical Flow Network 導入による腫瘍・血管検出性能の評価.....	51
3.5 まとめ.....	54
4. 肝臓内物体の多種分類タスクの実現.....	57
4.1 背景・目的	57
4.2 Power-Law Shot-Noise Model.....	58
4.3 データセット	61
4.4 Power-Law Shot-Noise Model の肝腫瘍超音波検査画像への適応可能性の検証	62
4.5 肝臓内物体の多種分類タスクへの性能評価	67
4.6 まとめ.....	71
5. 肝臓超音波検査向け扇形回転トリミングによるデータ多様性の確保.....	73
5.1 背景・目的	73
5.2 データセット	73
5.3. 肝臓超音波検査向け扇型回転トリミングの提案.....	74
5.3.1 提案背景	74
5.3.2 扇形回転トリミング	74
5.3.3 Mask R-CNN の構成・学習設定.....	76
5.3.4 テストデータに対する検出性能.....	79
5.3.5 評価用データに対する検出性能.....	79
5.3.6 未学習データに対する検出性能.....	81
5.4 比較対象として既存画像データ水増し手法の導入	82
5.4.1 データ水増し手法導入背景	82
5.4.2 各種データ水増し手法	82
5.4.3 Mask R-CNN の構成・学習設定.....	86
5.4.4 テストデータに対する検出性能.....	87
5.4.5 評価用データに対する検出性能.....	90
5.4.6 未学習データに対する検出性能.....	92
5.5 扇形回転トリミングと既存データ水増し手法の比較.....	95
5.6 まとめ.....	96
6. 肝疾患診断支援システムが実診断に与える影響.....	99
6.1 背景と目的	99
6.2 肝腫瘍診断支援システムの有無による影響の検証方法	99
6.3 肝腫瘍診断支援システムの有無による影響の検証結果	103
6.4 まとめ.....	112
7. 結論と今後の展望	115
7.1 研究内容のまとめ	115
7.2 結論	117

7.3 今後の展望	118
参考文献	119
謝辞.....	125
補論 研究内で使用される超音波検査データについて	126

1. 序論

1.1 肝臓の腫瘍性疾患

日本における肝がんによる死亡者数は 2019 年では 25,264 人で、部位別のがんによる死亡者数で見ると男性では第 5 位、女性では第 7 位となっている[国立がんセンター 21]。肝臓は C 型慢性肝疾患患者・B 型慢性肝疾患患者で発生が多く見られる。採血によるスクリーニング検査でこれらのウイルス性疾患の罹患評価は容易であることから、検診などでウイルス感染が指摘された症例は肝臓内科専門医師や肝臓外科専門医師により肝臓診療ガイドラインに沿った定期的なスクリーニングが行われることが多い。肝臓診療ガイドラインでは肝臓癌の早期発見を目的として、高リスク患者に対する定期的な検査が推奨されている[日本肝臓学会 17]。高リスク患者に対しては、3~6 ヶ月間隔での定期的な腹部超音波検査や腫瘍マーカー検査測定が推奨される。特に肝硬変の患者などの超高リスクの患者は Computed Tomography (CT) 検査や Magnetic Resonance Imaging (MRI) 検査の併用もおこなわれているが被曝リスクや医療コストの面からこれらの検査は経過やリスクを勘案しながら施行の是非を判断する必要がある。一方で、C 型肝炎・B 型肝炎に罹患していないグループでもアルコール性肝炎や非アルコール性脂肪肝を背景とした肝臓癌の発生が見られる。第 22 回全国原発性肝臓癌追跡調査報告[工藤 21]によると肝臓癌症例での HCV 抗体陽性率は肝細胞癌：49.2%、肝内胆管癌：12.8%、混合型肝臓癌 29.0%であり、HBs 抗原陽性率は肝細胞癌：22.7%、肝内胆管癌：23.1%、混合型肝臓癌：21.5%であった。これら以外の症例については C 型慢性肝疾患や B 型慢性疾患を背景としない肝臓に発生した肝臓癌であると考えられる。生活習慣病などで一般診療所に通院している患者にも潜在的な肝臓癌リスク症例は存在しており、ウイルス感染既往のない慢性疾患症例については無症状や軽度の肝機能障害にとどまるケースもみられる。このような患者は肝臓専門医の診察を受けることなく、一般診療所のかかりつけ医師の診察で診察が終了する可能性がありうる。肝臓のスクリーニング検査では、後述するように腹部超音波検査が極めて重要であるが肝臓に対する超音波画像診断には修練を要し、画像診断は検査者の技量により得られる結果が大きく左右される。このことから、そもそも一般診療所では多少の肝機能障害がみられていたとしても腹部超音波検査は行わずフォローが行われず採血検査のみで経過をみるケースもある。かかりつけ医の段階で技量によらないスクリーニング検査を行うことができるようになれば肝炎ウイルスに罹患していない症例であっても早期の肝臓癌診断が可能となりえる。肝臓超音波検査は肝臓腫瘍診断においては核心的な診断であり、術者の技量によらない診断システムを開発することにより一般診療所で肝臓専門医以外による超音波診断の質を向上させることが可能ではないかと考え今回の研究の着想に至った。

1.2 超音波検査

超音波検査は超音波を患部に照射した際の反射波を映像化することで患部の状態をリアルタイムに観察すること可能な検査方法である。超音波が探触子から体深部に伝達していく過程で特性が異なる体組織の境界にて超音波が反射する。この反射波が探触子までに到達する時間で反射箇所の深さ情報を求め、反射波の強弱によって画像化を行っている。音波を使用するため CT と異なり被爆なしに検査が可能である。特別な準備なし探触子を監査する部位に当てることによって観察が可能である。超音波は液体や個体を良く伝導するため、肝臓や腎臓、胆嚢などの臓器の観察によく利用される。一方で空気や骨が経路上にある場合、超音波が全反射、散乱され観察が難しくなり、肺や胃腸の臓器の観察が困難である。

肝臓の定期的な超音波検査には、腹部の上から探触子を当て肝臓内部を観察する腹部超音波検査が実施される。超音波検査によって図 1.1 に示すようなモノクロ画像がリアルタイムに観察できる。図 1.1 にあるように超音波検査によって肝細胞癌が観察可能である。また、画像のように血管や肝細胞癌、肝実質細胞それぞれ特性が異なるため、各部位の輪郭を観察することが可能である。医師は映像に含まれる肝臓の断面や血管の位置・種類、腫瘍のパターンを観察することで、腫瘍の鑑別を行う。正確な診断を行うには超音波検査だけでは不十分で結節等が確認できた場合は追加で CT 検査、MRI 検査を実施し詳細な鑑別が行われる。

超音波検査は肝臓の定期検査だけでなく手術中の検査にも利用される。通常の腹部超音波検査では皮膚の上から超音波を当てるため、脂肪等により深部まで超音波が伝わらず、深部に存在する肝がんの観察が困難な場合がある。そのため、手術中に肝臓の表面から直接超音波を当てることで、細部までの観察が可能となる。術中の超音波検査では腫瘍の正確な所在地の把握だけでなく、肝がんと血管の位置関係を観察することで、血管を傷つけずに安全に肝切除を行うことが可能となる。また、手術前に実施された腹部超音波検査、CT 検査、MRI 検査にて発見することが難しいような小さい腫瘍を発見する場合もある。

このように超音波検査は肝臓の定期検査による早期発見及び、手術での肝臓の観察する上で、有効な手段となる。しかしながら、超音波検査を利用して、正確な診断を実現するには専門的な医学知識だけでなく、熟練の経験が要求されるという問題点が存在する。具体的には肝臓内には肝静脈やグリソン鞘が多数絡み合っており、肝臓の断面から現在観察している肝領域を把握するには肝臓の医学的な構造の知識に加えて、リアルタイム映像から 3 次元的な構造を観察者の脳内にて再構築し把握する必要があり、経験が求められる。

また、正常肝・慢性肝炎・肝硬変といった背景肝の状態によっても得られる画像の質が大きく異なることから、解剖学的な位置把握や三次元的なイメージのみならず得られた画像

の評価には高度な修練を要する。

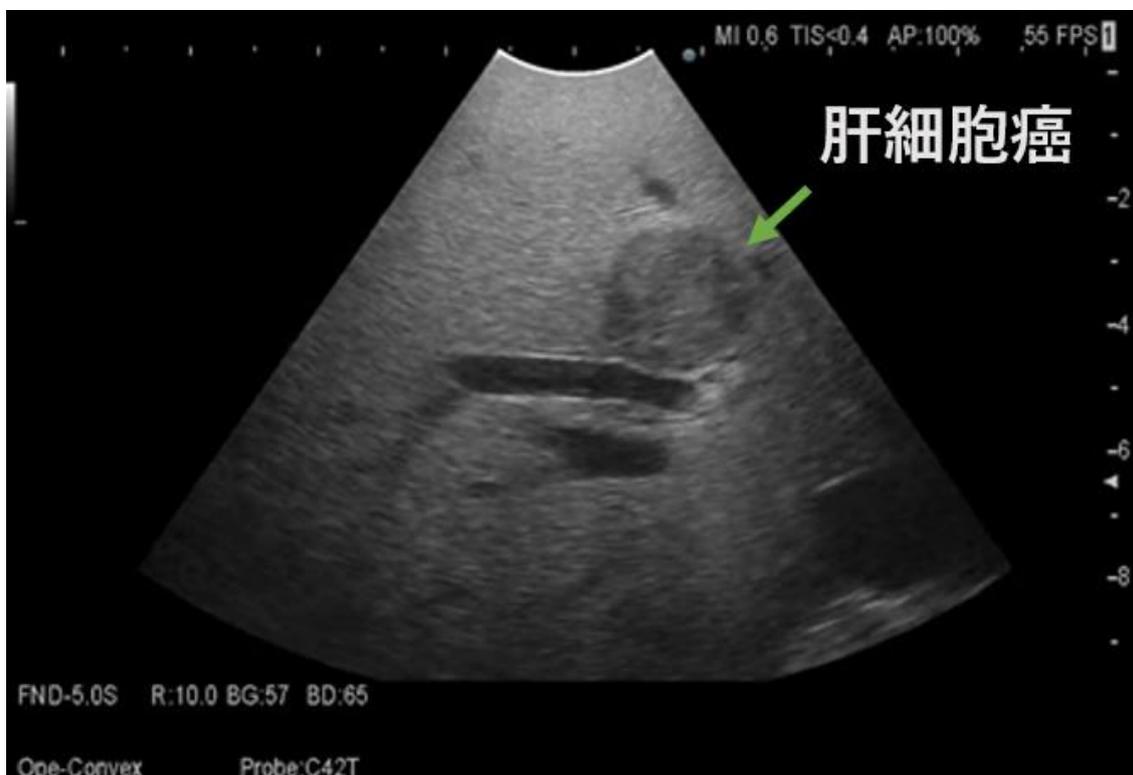


図 1.1 肝臓の超音波検査のサンプル

1.3 画像診断

近年、画像認識分野において、深層学習と呼ばれる、多層のニューラルネットワークを連結させたモデルが注目を浴びている。このニューラルネットワークのベースとなっているモデルは McCulloch が 1943 年に提案した神経細胞を数理的に模した人工ニューロンである [McCulloch 43]。この人工ニューロンを多数連結させることでモデルの表現能力を向上させることを期待して 1958 年にパーセプトロンが提案された [Rosenblatt 58]。パーセプトロンによりニューラルネットワーク自身が学習可能なタスクが広がったが、層が増えるほど各ニューラルネットワークのパラメータの学習がうまく行かない問題が存在した。これを解決するため、2006 年に各パラメータをランダムな初期値でなく、オートエンコーダと呼ばれるモデルを使用し予め学習済みのパラメータを利用することで多数のニューラルネットワークにおいても学習がスムーズに行われるモデルが提案された [Hinton 06]。2012 年には開催された画像認識コンテストである ImageNet Large Scale Visual Recognition Challenge (ILSVRC) において深層学習を用いたモデルが従来の認識精度を 10% 以上優れた性能を示したことで、ディープラーニングは大きな注目を浴びた

[ImageNet 12]。加えて、2015年に開催された ILSVRC[ImageNet 15]では、人間の認識エラー率が 5.1%であった画像認識タスクにおいて Microsoft Research が提案した深層学習モデルが 4.9%のエラー率を示した[He 15]。人間でも間違えてしまう画像認識タスクで人間より優れた性能を深層学習が示したことでさらなる注目を浴び、研究が盛ん行われている。

上記の背景により、深層学習の医療分野への多数の応用研究が実施されており、CT 画像[Huang 17][Zhou 17] [Gerard 18][Wang 21]や MRI 画像[Pereira 16][Suk 16] [Moeskops 16][Ismael 20]、X 線画像[Wang 17][Rajpurkar 17] [Rajkomar 17] [Hemdan 20]から疾患の診断や腫瘍の分類では医師に匹敵するほどの精度を示した研究も報告されている。他には皮膚の画像から皮膚がんの診断[Esteva 17]、内視鏡映像から癌の検出[Sakai 18]、顕微鏡映像から病理診断[Rezaeilouyeh 16][Liu 17]を行う深層学習モデルが提案されており、映像データをベースにした自動診断を実現するために様々な疾患に対して研究が実施されている。その中で、超音波検査映像に対する応用研究としては非アルコール性脂肪肝炎[Reddy 18]、肝硬変[Xu 17]のステージ診断、手術中の左心室の自動検出[Carneiro 13]、3次元超音波検査映像から乳がんの検出[Xu 19]といった研究が報告されている。超音波検査映像への応用研究は他の CT 検査や MRI 検査と比較して少ない。その要因として超音波検査は CT 検査や MRI 検査と異なり任意の方向から対象を観察することができるが、深層学習を用いてモデルを使用する場合に学習すべき状況が多く、高い認識精度を示すには多様な学習データが必要になりそのデータの確保が難しいと考えられる。そういった状況の中 2019 年には物体検出モデルである YOLO を使用して、甲状腺結節の超音波検査映像からリアルタイムに結節を識別する手法が提案されており、熟練の放射線医師の Receiver Operating Characteristic (ROC) curve が 0.859 であったが、提案モデルは 0.902 を示し、熟練医師よりも優れた結果を示す研究も存在する[Wang 19]。深層学習はここ数十年で飛躍的発展した分野であるが、その画像認識性能は医療分野でも優れていることを示しており、今後もさらなる発展が期待されている。一方、肝臓に関してはリアルタイムで腫瘍等を検出する研究は存在しない。

1.4 肝疾患の超音波検査映像に対する診断支援システム

1.1 節、1.2 節で述べたように近年ではウイルス性ではなく肥満やアルコールといった生活習慣が起因となる肝細胞癌が増えている。ウイルス性であれば定期検査として超音波検査によって早期の発見が可能であるが、そうでない場合には専門的な医療機関に掛からない場合もある。この状況を解決する手段の一つとして肝疾患の超音波検査映像に対する診断支援システムが考えられる。これまで、熟練の専門医師でなければ正確な診断が難しい肝臓の超音波検査に対して診断支援システムを導入し一般診療所でも超音波検査が簡便に

利用できる検査手法となれば肝臓の超音波検査のハードルが下がり、初期の一般診療所に患者が訪れたタイミングで肝細胞癌の危険性について検査が可能となる。

しかしながら、この診断支援システムを実現する上で問題点も存在する。その一つが肝臓の超音波検査映像に対する検出タスクの複雑性である。1.3節で述べたように近年深層学習技術が大きく注目され医療分野、CT 画像や MRI 画像に対して広く応用研究が行われており、医師による診断精度を超える結果を示した例も存在するが、これらと肝臓の超音波検査映像に対する検出タスクの大きな違いが撮影される映像の自由度の高さである。CT や MRI ではある決まった角度でデータが取得され診断に利用されるが、肝臓の超音波検査では肝臓内の状態を多角的な方向から物体の連続性や 3 次元的構造を観察し診断が行われる。また、診断にあたっては肝臓内で複雑に絡み合っている肝静脈やグリソン鞘を把握し、そこから腫瘍を検出することが必要であり、これらの肝臓内物体は撮影角度によって断面形状が大きく変化する点に対応する必要がある。また医師は物体の連続性や 3 次元的構造を観察し診断が行われることから単純に既存深層学習技術による静止画ベースによる検出がどれほど有効であるかといった懸念も存在する。

この複雑な分類・検出タスクの実現に有効であると期待される手法として冪乗則による特徴量抽出が存在する。これはもともと超音波検査画像の特徴抽出手法として冪乗則を利用したモデルが研究されていたものである [Kutay 01]。これらの研究では検査で取得されるデータを物体内にある散乱体による反射波の集合と仮定し、その反射波を Power-Law Shot-Noise Model を利用して表現することで超音波に対する物体の散乱特性・減衰特性より、反射波のパワースペクトラムが冪乗則に従うことからスケーリング係数を特徴量として利用している。このスケーリング係数は物体の特性によって変化するため、スケーリング係数を算出することで領域内の組織を区別することが可能となることが期待される。実際に乳がんの超音波検査画像からスケーリング係数を算出することで腫瘍とそれ以外の組織の区別を実現している [Kutay 03]。

タスク自体の複雑性の他に学習データの多様性の確保といった問題も存在する。これは深層学習技術を用いた一般的な分類・物体検出タスクを実現する上で常に存在する問題である。教師あり学習を行う深層学習モデルではタスクの実現には学習データが必要不可欠であり、学習データに含まれるデータパターンが実際に現れうるデータパターンを網羅することは高性能なモデル構築にとって必要不可欠である。特に肝臓の超音波検査映像では撮影視点の自由度の高さからさらなるデータ多様性が求められる。しかしながら他の医学的な応用研究と同様に肝疾患を有する患者からデータを収集が必要でありデータ多様性を確保するだけのデータ収集が困難であり、またデータ作成には熟練医師の協力が必要でデータの大量収集・作成には大きなコストが発生する。このような問題点にたいして、一般的にデータ水増し手法を利用したデータ多様性の確保が行われる。

一般的なデータ水増し手法ではデータに対して、現実により得るような微量な変化を発生させそれを新しい学習データとして利用することで、本来学習データに存在しなかったデータパターンを拡充させる手法である。基本的に光学的に撮影された静止画に対しては画像処理を行うことで画像中の物体に対する変化させて、実在により得るデータパターンを合成するものである。よく使用されるデータ水増し手法[Shorten 19]としてはノイズ付加、コントラストの調整、反転、回転が存在する。ノイズ付加はデータ全体に対して微量なノイズを追加することで、画像の微量変化を再現するもので、これは画像に限定されない様々なデータ水増しに利用される。一方でコントラストの調整、反転、回転はカメラによって撮影された画像を想定し、その撮影時のコンディションによって変化する光量やカメラの向き、物体の立ち位置の変化を再現することでデータ多様性を増やしている。これを超音波検査映像に適用させる場合、超音波検査は撮影対象を撮影方向にスライスして画像を得るものであり、カメラでの撮影画像と大きく変わる要因となりそのままでは適用できない可能性も存在する。また、光学的に撮影された画像に対しては様々なデータ水増し手法として Random Erasing[Zhong 20]、Random Resize Crop[Pytorch 17]、Random Image Cropping and Patching[Takahashi 19]といった様々な手法が提案されている。Random Erasing では画像中に含まれる物体の一部をランダムに削除することで擬似的に物体とカメラとの間に障害物を発生させる。これによってニューラルネットワークが物体の全体像だけでなく物体の各種パーツによりフォーカスさせて学習を促進することが可能である。このような物体の一部が隠れるような状態は超音波検査にも存在し、撮影方向等によって今まで見えていた腫瘍、血管の一部が消えるといったことが発生する。これは肝臓の超音波検査映像に対する診断支援システムのデータ多様性を確保するにあたって有効的な手法と期待できる。Random Erasing と逆に Random Resize Crop では画像の一部をランダムに切り取る学習データとして利用する手法である。Random Erasing と近い効果があり、切り取った部位によりフォーカスされ、更に切り取りによって撮影時に発生する障害物等に遮りを再現することができる。これに関しても超音波検査でも発生しうる状況であり効果が期待される。一方で Random Image Cropping and Patching は異なる種類の4枚の画像を各画像でランダムな位置で4分割し、各種別から分割された画像1枚を抽出して4枚の画像を1つの画像として再構築したものを利用する手法である。複数の物体をまとめた画像を予測分類させることで、モデルの性能を大きく向上させた手法となっている。この手法は超音波検査に適用する上で血管や腫瘍の絡み合っている状況を再現するために有効であると見られるが、肝臓内の本来存在する血管の位置的情報を破壊するような手法となり、適さない手法と思われる。このような画像操作ベースのデータ水増し手法が存在する中で、Generative Adversarial Networks [Goodfellow 20] (GAN)といった深層学習技術をベースとした合成画像生成手法もデータ水増し手法の利用として期待されている。これは学習したデータに類似した合成画像を生成する手法である。これはCT画像やMRI画像といった様々な医療データの合成[Han 18] [Sun 20] [Islam 20]を成功している先行研

究も存在する。応用例として肝臓病変の CT 画像に対する適用例[Frid-Adar 18]も存在する。ただ、これを超音波検査映像に対して適応する場合 GAN も深層学習のためデータ多様性の確保の困難といった問題が同様に存在してしまう。

最後に肝疾患の超音波検査映像に対する診断支援システム自体の有効性について懸念が存在する。本来、長年の経験と医学的知識から肝臓の診断を行っているため、一般診療所の専門的経験が無い医師であっても単純な超音波検査映像に対するアノテーションだけで肝疾患に対する診断の補助が実現できるのかというのは大きな問題点となる。

1.5 本論文の目的と構成

以上に述べたように、肝臓の検査として超音波検査は簡便に実施でき有能であるが、検査者の知識及び経験が肝腫瘍の鑑別に大きな影響をもたらす検査法である。また、肝臓の超音波検査を行う医師には肝臓専門医としての長年の経験が要求される。本研究では一般診療所にて肝臓の専門医師でなくとも肝臓の診断として超音波検査による診断が期待できる肝臓の超音波検査映像に対して肝腫瘍の疑いのある箇所を自動検出しアノテーションにて表示することで肝腫瘍の発見の支援を行う検出システムの構築を行う。本研究ではこのシステム実現にあたって、近年大きく発展している画像診断技術である深層学習技術を利用する。しかしながら、実際にシステムを実現する上で3つの課題が存在する。1つ目は超音波検査映像に対する物体検出タスクの複雑性である。超音波検査によって取得される映像は物体の断面映像であり、また肝静脈やグリソン鞘が複雑に絡みまた、撮影方向によってその形状が大きく変形するこのような一般的な物体を光学的に撮影した画像から物体検出を行うことと比べ複雑なものとなる。また、肝臓の超音波検査映像中に存在する物体としては腫瘍だけでなく肝静脈、下大静脈、下行大動脈、グリソン鞘といった管状の類似した物体が複数存在しこれらの単純な検出だけでなく分類もまた困難な複雑性の要因となる。2つ目にデータの多様性確保の困難さである。深層学習を用いる場合、汎化性能を実現するには出現しうる対象の状況を十分に表現されたデータが必要不可欠となる。超音波検査に適用する場合、超音波検査中に出現する物体は撮影方向による変化だけでなく、肝臓の状態や腫瘍の形状やパターンも異なるため、その多様性を学習できるようなデータが必要となる。また、単純に肝細胞癌を有する患者からデータを収集し、熟練の専門医師によってデータを作成する必要がある、多数のデータ確保にはコストが要求される。3つ目は診断支援システム自体の有効性である。一般診療所で専門的経験がない医師の利用を想定しているため、物体の検出・アノテーションのみでは腫瘍の診断に効果がない懸念がある。

本研究では肝臓超音波検査映像に対する診断支援システムを構築することを通じて上記の3つ課題について明らかにする。まず、「肝臓超音波検査映像に対する物体検出タスクの複雑性」の課題解決のため、既存物体検出モデル及び新規に超音波検査向けに静止画デ

ータに欠落している時間経過による変化情報を補うモデルを提案し、肝臓超音波検査映像に対する腫瘍・血管検出タスクの実現性の検証及びベースモデルの構築を行う。また、肝臓超音波検査映像中に存在する腫瘍と多様な血管に関して分類タスクの実現性について既存深層学習技術及び物体の超音波に対する散乱特性から抽出した特徴量を組み合わせることで明らかにする。「データの多様性確保の困難さ」に対しては、光学的に撮影された画像向けデータ水増し手法と新規に超音波検査映像の変化に着目したデータ水増し手法を提案し、各データ水増し手法の肝臓超音波検査映像に対する物体検出タスクに対する有効性の検証を通じて汎化性能を持つ腫瘍・血管検出モデルを構築する。最後に構築した検出モデル自体の有効性を明らかにするため性能評価を行う。評価にあたっては日本医療研究開発が発行する医用画像診断支援システム（人工知能技術を利用するものを含む）開発ガイドライン2019（手引き）[日本医療研究開発 19]において本研究にて構築するシステムは Computer-Aided Detection (CADe) と呼ばれる画像上に存在する病変の疑いのある場所を自動検出し、マーキング等によって検出の支援を行うものに当たるため、このガイドライン及びコンピュータ診断支援装置に関する評価指標 [厚生労働省医薬食品局 11] に従って、システム自体の腫瘍検出性能を評価するのではなく、利用対象となる医師として想定した肝胆膵外科の研修医が本システムを使用した際の腫瘍検出性能評価を通じて、システム自体の有効性について明らかにする。

本論文の構成は下記のようになる。第2章では既存の深層学習による画像認識・物体検出モデルを外観する。第3章では深層学習を用いた肝臓の超音波検査映像に対する腫瘍・血管検出タスクを実現する基礎となるモデルを構築する。第4章では複雑からみ存在する多様な血管及び腫瘍の分類タスクを実現する基礎モデルを構築する。第5章では限られたデータ量の中でデータ多様性確保を実現するために既存のデータ水増し手法及び新規に扇形回転トリミングを提案し、有効性の検証を通じて、未学習データに対して汎化性能の存在する検出モデルを構築する。第6章ではこれまで構築した診断支援システムの有効性検証のため研修医に対する腫瘍の認知へ寄与を評価する。第7章は結論である。

2. 関連先行研究

本章では肝疾患の超音波検査映像に対する診断支援システムを実現にあたって重要となる深層学習技術とそれに関連する先行研究及び課題について記述する。2.1節では肝臓内の物体検出を実現するモデルの基礎となる深層学習技術について記述する。2.2節では画像分類、物体検出に特化した深層学習技術とそれに用いられる評価指標を記述する。最後に2.3節ではこれまで医療分野において画像診断・検出タスクを実現した先行研究を上げ、肝疾患の超音波検査映像に対する診断支援システムの実現における課題についても明らかにする。

2.1 基礎的な深層学習技術

本研究では肝疾患の超音波検査映像に対する診断支援システムを実現するにあたって、近年画像認識分野で高い性能を示した深層学習技術をベースに利用する。本節ではその基礎となる技術について説明する。1.3節で述べた医学分野に様々な画像認識タスクにはこの深層学習技術が利用されており、基本となるものが人工ニューロンと呼ばれるモデルを多数連結したニューラルネットワークと呼ばれるものである。このニューラルネットワークはニューロン（神経細胞）を数学モデル化した人工ニューロン[McCulloch 43]を多層に連結させた構造をしており、学習を通じて与えられたタスクを模倣する事が可能である。

2.1.1 単純パーセプトロン

はじめにニューラルネットワークではなく単一の人工ニューロンについて記述する。この単一の人工ニューロンは単純パーセプトロンとも呼ばれ、神経細胞であるニューロンをモデル化したものである。図 2.1 に示すような複数の入力 \mathbf{x} と単一の出力 $f(u)$ を持ち、次のように定式化される。

$$f(u) = \begin{cases} 1 & (u \geq 0) \\ 0 & (u < 0) \end{cases} \quad (2.1)$$

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \quad (2.2)$$

$$\mathbf{w} = (w_1, w_2, \dots, w_n) \quad (2.3)$$

$$u = \sum_{i=1}^n w_i x_i + b \quad (2.4)$$

ここで \mathbf{w} は各入力に対する重み成分を表し、 b は単純パーセプトロンが有するバイアス成分を示す。 $f(u)$ は活性化関数と呼ばれ、ここでは単位ステップ関数を使用されている。単純パーセプトロンの動作としては n 個の入力 \mathbf{x} を受け取り、 u が一定値以上に達すると信号を出力

する。これを「発火する」と表現する。この重み成分 w とバイアス成分 b をタスクに合わせて調整することにより、二値分類問題を解くことが可能である。

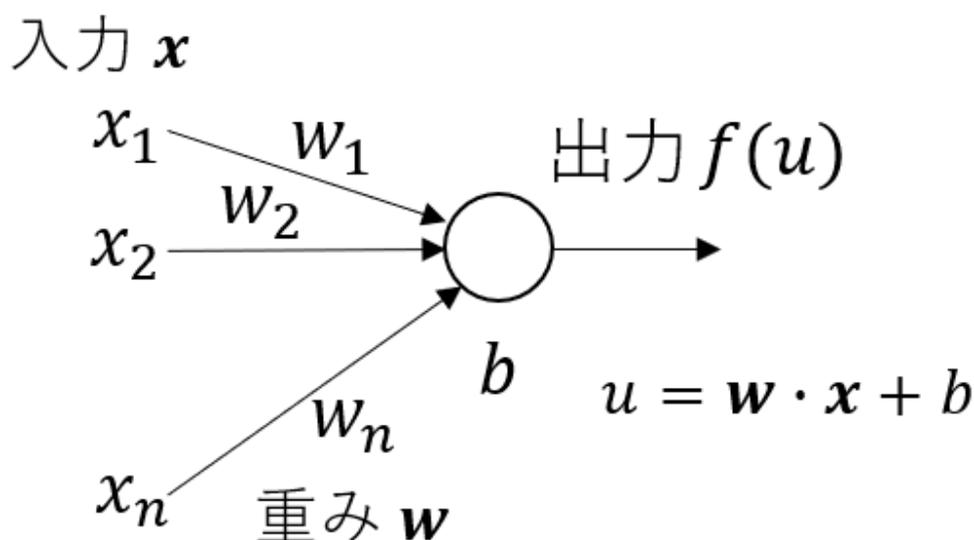


図 2.1 単純パーセプトロンのモデル図

2.1.2 多層パーセプトロン

多層パーセプトロン[Rosenblatt 58]は深層学習で一般的に使われるニューラルネットワーク構造の前身である。また、全結合ニューラルネットワークや Linear Neural Networks と呼ばれる場合もある。図 2.2 に 3 層パーセプトロン構造を示す。この図のように多層パーセプトロンは単純パーセプトロンを多重に連結したモデルである。入力を受け取るパーセプトロン層を入力層、最終的な出力を行うパーセプトロン層を出力層、それ以外のパーセプトロン層をまとめて隠れ層（中間層）と呼ぶ。各パーセプトロンは前層の各パーセプトロンの出力を受け取り。個別に重み成分とバイアス成分を持つ。タスクの学習を通して各パラメータが決定される。2.1.1 節の単純パーセプトロンとは異なり、活性化関数には下記式にて表現されるシグモイド関数がよく使用され、出力を連続値によって表現することが可能となる。

$$f(u) = \frac{1}{1 + e^{-u}} \quad (2.5)$$

全 K 層で k 番目の層の人工ニューロンが N_k 個存在する K 層パーセプトロンを仮定すると、 $k + 1$ 番目の層の出力 l^{k+1} は k 番目の層の出力 l^k を使用して次のように表現できる。

$$l^{k+1} = f(\mathbf{w}^{k+1}l^k + \mathbf{b}^{k+1}) \quad (2.6)$$

ここで、 \mathbf{w}^{k+1} と \mathbf{b}^{k+1} は $k+1$ 番目の層の各人工ニューロンの重みとバイアスを示す。この多層パーセプトロンを使用し与えられたタスクを実現するには各パラメータをタスクに合わせて調整する学習工程が必要となる。この多層パーセプトロンの各重み成分とバイアス成分の学習には誤差逆伝播法[Rumelhart 86]がよく使用される。

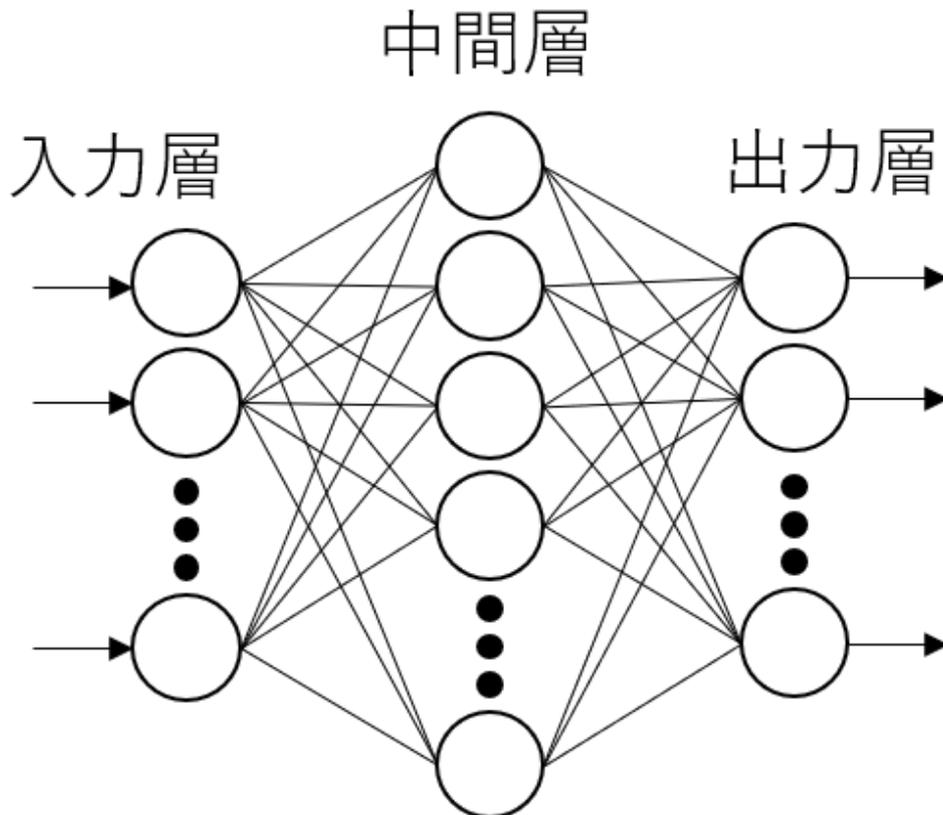


図 2.2 3層パーセプトロンのモデル構造

2.1.3 誤差逆伝播法

ニューラルネットワークの学習は見本となる教師データとなる入力データと出力データのセットをニューラルネットワークが再現するように各パラメータを調整する工程となる。その工程として誤差逆伝播法[Rumelhart 86]が使用される。誤差逆伝播法では教師データをニューラルネットワークに入力した際のニューラルネットワークの出力と正答データの誤差を算出し、その誤差をもとに出力層から順々に前層へ向かって各層の誤差を伝搬しながら算出することで、各ニューロンの重み成分とバイアス成分の変化によって誤差が減少していく方向を求め、各パラメータを更新し、より正答に近い出力ができるようにニューラルネットワークを更新していく手法である。

簡略化のため全 K 層で k 番目の層の人工ニューロンが N_k 個存在するバイアス成分 $b = 0$ の K 層パーセプトロンを仮定する場合、誤差逆伝播法を使用し下記ステップによって学習が実現される。

1. 各人工ニューロンの重み成分 \mathbf{w} を一様乱数で初期化する。
2. 学習用データを入力し、出力を求める。
3. 出力と正答データの誤差（損失関数）を計算する。
4. 損失関数に対する各層の重みの偏微分を計算する。
5. 各人工ニューロンの重み成分 \mathbf{w} を誤差が小さくなるよう更新する。
6. 2~5 を学習回数 E だけ繰り返す。

ステップ 2 における K 層パーセプトロンの出力を $\mathbf{o}^K = (o_1^K, o_2^K, \dots, o_{N_K}^K)$ とすると、損失関数に残差平方和を使用すると誤差 L は次のように表現できる。

$$L = \sum_{n=1}^{N_K} (o_n^K - T_n)^2 \quad (2.7)$$

$$L = \|\mathbf{o}^K - \mathbf{T}\|^2 \quad (2.8)$$

ここで $\mathbf{T} = (T_1, T_2, \dots, T_{N_K})$ は正答データを表す。

誤差 L の第 k 層の重み $w_{i,j}^{k-1,k}$ による偏微分は次の式で与えられる。

$$\frac{\partial L}{\partial w_{i,j}^{k-1,k}} = \frac{\partial L}{\partial i_i^k} \cdot \frac{\partial i_i^k}{\partial w_{i,j}^{k-1,k}} \quad (2.9)$$

ここで i_i^k は第 k 層 i 番目の人工ニューロンの入力値である。

式(2.9)の $\frac{\partial i_i^k}{\partial w_{i,j}^{k-1,k}}$ は入力 i_i^k を前層の出力 o_j^{k-1} を使用して表現することで下記のように簡略化できる。

$$\frac{\partial i_i^k}{\partial w_{i,j}^{k-1,k}} = \frac{\partial \sum_{l=1}^{n_{k-1}} w_{i,j}^{k-1,k} o_l^{k-1}}{\partial w_{i,j}^{k-1,k}} = o_j^{k-1} \quad (2.10)$$

次に式(2.9)の $\frac{\partial L}{\partial i_i^k}$ に関して連鎖率を使用し分解すると

$$\frac{\partial L}{\partial i_i^k} = \frac{\partial L}{\partial o_i^k} \frac{\partial o_i^k}{\partial i_i^k} = \frac{\partial L}{\partial o_i^k} \frac{\partial f(i_i^k)}{\partial i_i^k} = \frac{\partial L}{\partial o_i^k} f'(i_i^k) \quad (2.11)$$

と表現できる。ここで $f(i_i^k)$ は活性化関数である。

$\frac{\partial L}{\partial o_i^k}$ は第 $k+1$ 層の入力を使用して次のように表現できる。

$$\begin{aligned} \frac{\partial L}{\partial o_i^k} &= \sum_{l=1}^{n_{k+1}} \left(\frac{\partial L}{\partial i_l^{k+1}} \frac{\partial i_l^{k+1}}{\partial o_i^k} \right) \\ &= \sum_{l=1}^{n_{k+1}} \left(\frac{\partial L}{\partial i_l^{k+1}} \frac{\partial \sum_{h=1}^{n_{k+1}} w_{l,h}^{k,k+1} o_i^k}{\partial o_i^k} \right) \\ &= \sum_{l=1}^{n_{k+1}} \left(\frac{\partial L}{\partial i_l^{k+1}} w_{l,h}^{k,k+1} \right) \end{aligned} \quad (2.12)$$

式(2.10)、(2.11)、(2.12)を使用して式(2.9)は

$$\frac{\partial L}{\partial w_{i,j}^{k-1,k}} = o_j^{k-1} f'(i_i^k) \sum_{l=1}^{n_{k+1}} \left(\frac{\partial L}{\partial i_l^{k+1}} w_{l,h}^{k,k+1} \right) \quad (2.13)$$

と表現できる。第 k 層重み成分で誤差を偏微分の値は実際に教師データを入力した際の第 $k-1$ 層の出力、第 k 層の活性化関数の微分、誤差を第 $k+1$ 層の入力で偏微分した値を算出することで止めることができる。このときに誤差に対する偏微分の値を出力層から入力層側へ順々に渡しながら各層での誤差に対する重み成分の偏微分を算出するために誤差逆伝播と呼ばれる。最終的に式(2.13)によって各ニューラルネットワークの重みの変化に対する誤差の勾配が求まるため、重みとバイアスを次のよう更新する。

$$w_{i,j}^{k-1,k} = w_{i,j}^{k-1,k} - \eta \frac{\partial L}{\partial w_{i,j}^{k-1,k}} \quad (2.14)$$

η は学習率を示すパラメータである。この教師データを入力し、正答データとの誤差を小さくする方向へのパラメータ更新を行うことでタスクの学習が可能となる。

2.1.4 畳み込みニューラルネットワーク

畳み込みニューラルネットワーク (Convolutional Neural Networks) [LeCun 89] は CNN と略されるが、近年画像分類タスクで大きな成功を収めたニューラルネットワークである。この節では CNN の代表的な構造として 2 次元畳み込みニューラルネットワークについて

記述する。従来の多層パーセプトロンでは全結合と呼ばれる前層のすべてのニューロンの出力を各ニューロンは受け取っていたが、2次元CNNでは各ニューロンは2次元空間的に近傍にあるニューロンの出力のみを受け取る構造をしている。従来のパーセプトロンにおいて各ニューロン間に空間的な関連性がしなかったが、2次元CNNでは各ニューロンは2次元空間内に存在し、それぞれ近傍に隣接する前層の出力を受け取る構造となっている。そのため、画像のような空間情報を含むデータに対して特徴量の抽出性能が優れている。具体的な2次元CNNの構造を図2.3に示す。2次元CNNでは入力に対して、畳み込み操作を行う畳み込み層と出力をまとめて、出力の次元数を削減するプーリング層によって構成される。画像認識に使用する際はこのCNNを多重に連結して使用する。

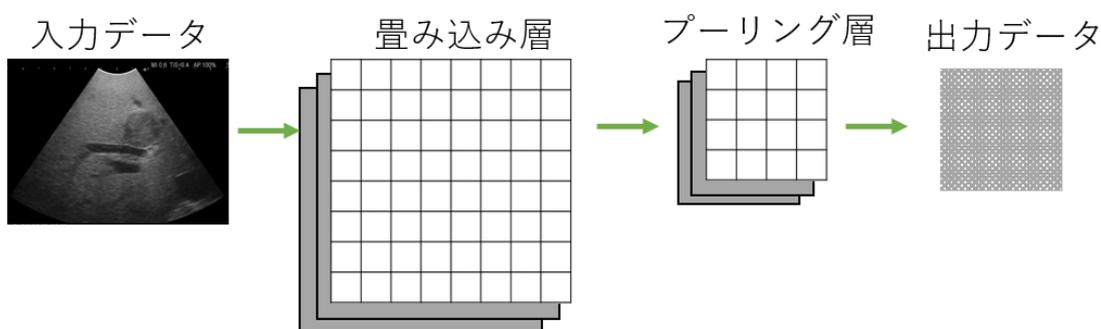


図 2.3 畳み込みニューラルネットワークのモデル図

・畳み込み層

畳み込み層では入力されたデータに対して畳み込み処理が行われる。図2.4にカーネルサイズ 3×3 の畳み込み層による出力イメージを示す。図2.4のように座標 j, k に存在する畳み込み層のニューロンはそのニューロンを中心とした縦 $2M + 1$ 、横 $2L + 1$ 個の範囲にある前層の出力を受け取り、出力 $o_{j,k}$ は次式となる。

$$o_{j,k} = f \left(\sum_{l=-L}^L \sum_{m=-M}^M w_{l,m} x_{j+l, k+m} + b \right) \quad (2.15)$$

$$f(u) = \begin{cases} 0 & (u \leq 0) \\ u & (u > 0) \end{cases} \quad (2.16)$$

ここで f は活性化関数であり、CNNの活性化関数には Rectified linear unit (ReLU 関数) [Glorot 11]がよく使用される。出力 $o_{j,k}$ が示すように畳み込み層では前層の出力値を2次元空間として捉え、その局所データの畳み込み演算を行う操作が行われる。畳み込み操作によって、局所空間に存在するデータのパターンを抽出することが可能となる。一般的に縦 $2M + 1$ 、横 $2L + 1$ 個の畳み込みを行う領域をカーネル、その範囲をカーネルサイズと呼ぶ。またカーネルは並列して複数存在可能でカーネル数が N 個の場合、 N をカーネル数またはチャネ

ル数と呼び、畳み込み層の構成を $2M + 1 \times 2L + 1 \times N$ と表現する。例えば、高さ 32、幅 32 の RGB で構成された画像データ（ $32 \times 32 \times 3$ で構成された入力値）を $3 \times 3 \times 16$ で構成された畳み込み層に入力する構成の場合、出力される値は $32 \times 32 \times 16$ のデータ形式となる。

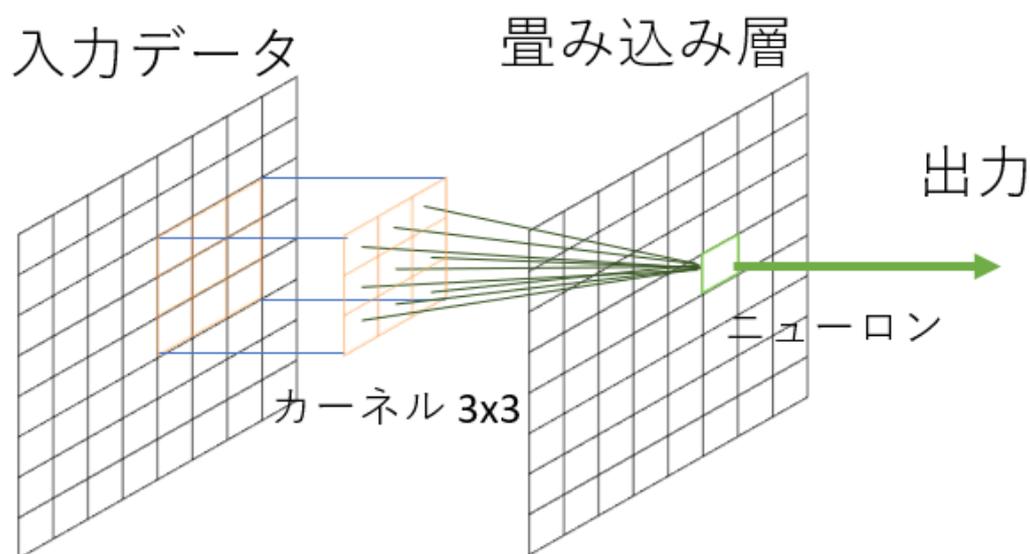


図 2.4 畳み込み層の出力イメージ

・プーリング層

プーリング層では畳み込み層が出力した値を抽出し、次元圧縮を行う。よく使用される処理として Max pooling と Average pooling が存在する。Max pooling によるプーリング操作を図 2.5 に示す。図 2.5 のように縦 2 横 2 の前層の出力から最大値または平均値を次の出力として利用する。この際、入力サイズが $N \times N$ のとき、プーリング層の出力サイズは $\frac{N}{2} \times \frac{N}{2}$ となり次元が削減される。またプーリング層のカーネルサイズを $\frac{N}{2} \times \frac{N}{2}$ と表現する。

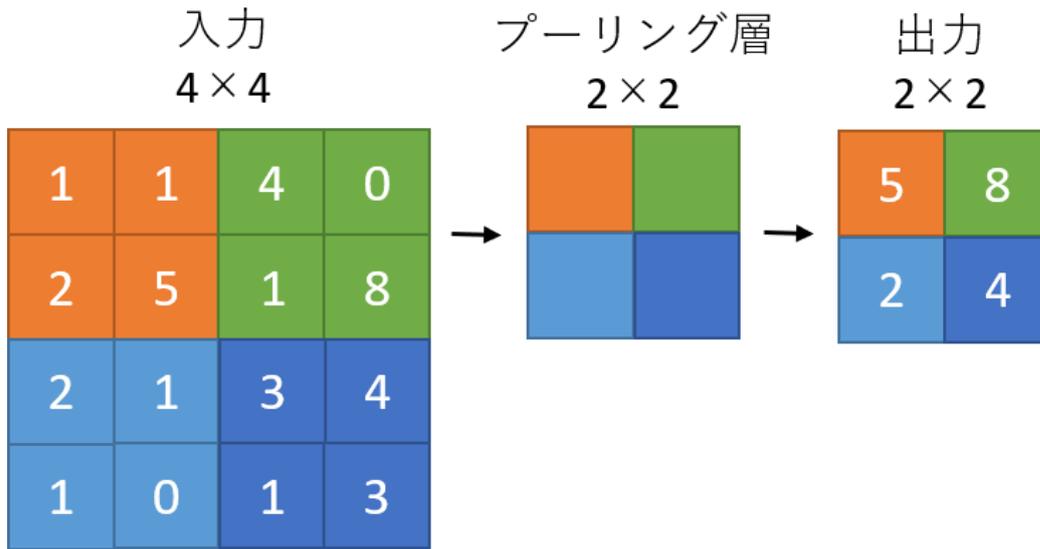


図 2.5 プーリング層の出力イメージ

実際に畳み込みニューラルネットワークを使用する際はこの畳み込み層とプーリング層を図 2.6 に示すように連続的に連結させて、学習を行う。このとき最終の層には全結合ニューラルネットワークを使用する。畳み込み層とプーリング層を繰り返し連結することで、画像から特徴量の抽出を行う。全結合ニューラルネットワークでは抽出された特徴量を受け取り、画像に含まれる物体の種類を確立にて推測する。 C 種類の物体の中から画像に含まれる物体を予測する際、全結合ニューラルネットワークは C 個のニューロンを持つ。このとき、畳み込み層とプーリング層によって抽出された特徴量マップが $N \times N \times F$ のサイズであるとき各ニューロンの出力 $o_{j,c}$ は次の式となる。

$$o_{j,k} = f \left(\sum_{j=1}^{N \times N \times F} w_{j,c} x_{j,c} + b_c \right) \quad (2.17)$$

$$f(u_c) = \frac{e^{u_c}}{\sum_{c=1}^C e^{u_c}} \quad (2.18)$$

ここで $w_{j,c}$ は c 番目の全結合ニューラルネットワーク中のニューロンと $N \times N \times F$ サイズの特徴量マップが入力されたときの j 番目の入力間の重み成分である。 b_c は c 番目の全結合ニューラルネットワーク中のニューロンの重み成分である。式(2.18)では活性化関数としてソフトマックス関数を利用する。このソフトマックス関数により C 個の出力の合計が1となり、各出力層が確率で出力すること可能とする。 c 番目の出力が入力された画像に含まれる物体が c の種別である確率を示している。このような CNN を用いた画像認識モデルを構築した際、CNN 内による物体識別過程を可視化した研究[Zeiler 14]によれば多層 CNN の第 1, 2

層目の CNN レイヤーでは画像に含まれる微小な模様 of 抽出が行われ、さらに第 3~5 層目では物体を特徴づけるようなパターン（動物の目の形や耳、口の模様）を抽出することができる。このパターンの抽出を連続的に行うことで、単純な模様からより複雑な模様を抽出することができる。最終的には各模様の含まれる割合によって種別の認識を行う。

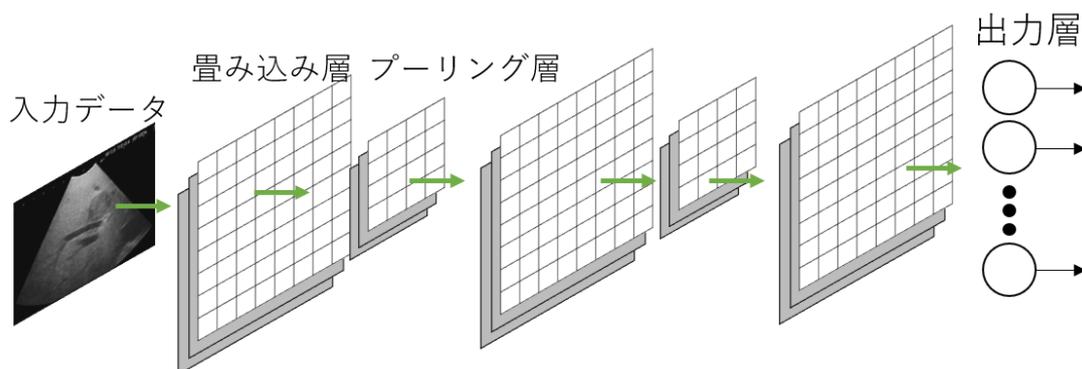


図 2.6 畳み込みニューラルネットワークによる画像認識の流れ

・ He の初期値

一般的に畳み込み層の活性化関数として ReLU 関数が使用されるが、この際に各ニューロンのパラメータの初期値には He の初期値[He 15]と呼ばれる値が使用される。具体的には各ニューロンの重みは層内のニューロン数 N に対して平均 0、標準偏差 $\sqrt{\frac{2}{N}}$ の正規分布によって初期値が設定される。

2.2 画像向け分類・物体検出モデルと評価指標

2.1 節では基礎的な深層学習技術を紹介した。本節では診断支援システムの実現に重要となる画像向けの分類及び物体検出タスクに使用される深層学習モデルを紹介する。近年深層学習技術大きく発展し、画像認識分野で人間に匹敵する性能を示した例[ImageNet 12]も存在し、医療分野でも多く利用されている。これまでの基礎的な深層学習技術との大きな違いはモデルに使用されるニューラルネットワークの数である。これまでニューラルネットワークを多数（多くても 10 程度）重ねてモデルが構築されていたが、近年ではそれを 50 や 100 まで拡張し、モデルに含まれるパラメータ数を増加させモデル自身の表現能力を拡充させた。更にモデルごとに役割をもたせそれらの組み合わせによって物体検出といった複雑なタスクに対しても高性能なモデルの実現もされている。

2.2.1 画像分類モデル - DPRN -

2012 年に ILSVRC[ImageNet 12]にて深層学習を利用したモデルが画像分類タスクにおいて優れた性能を示してから、深層学習が大きく注目され著しく技術が発展してきた。特に本節では近年、画像分類タスクで高性能を示したニューラルネットワークである Deep Pyramidal Residual Networks[Han 17]について記述する。Deep Pyramidal Residual Networks (DPRN)は画像から 10 種類物体識別するタスクの CIFAR-10 と 100 種類物体識別するタスクの CIFAR-100 において[Krizhevsky 09]、従来好成績を示していたニューラルネットワーク構造である Residual Network[He 16]よりも高精度を示したニューラルネットワークモデルである。Deep Pyramidal Residual Networks は図 2.7 に示すように Residual unit 及び Bottleneck の連結で構成される。

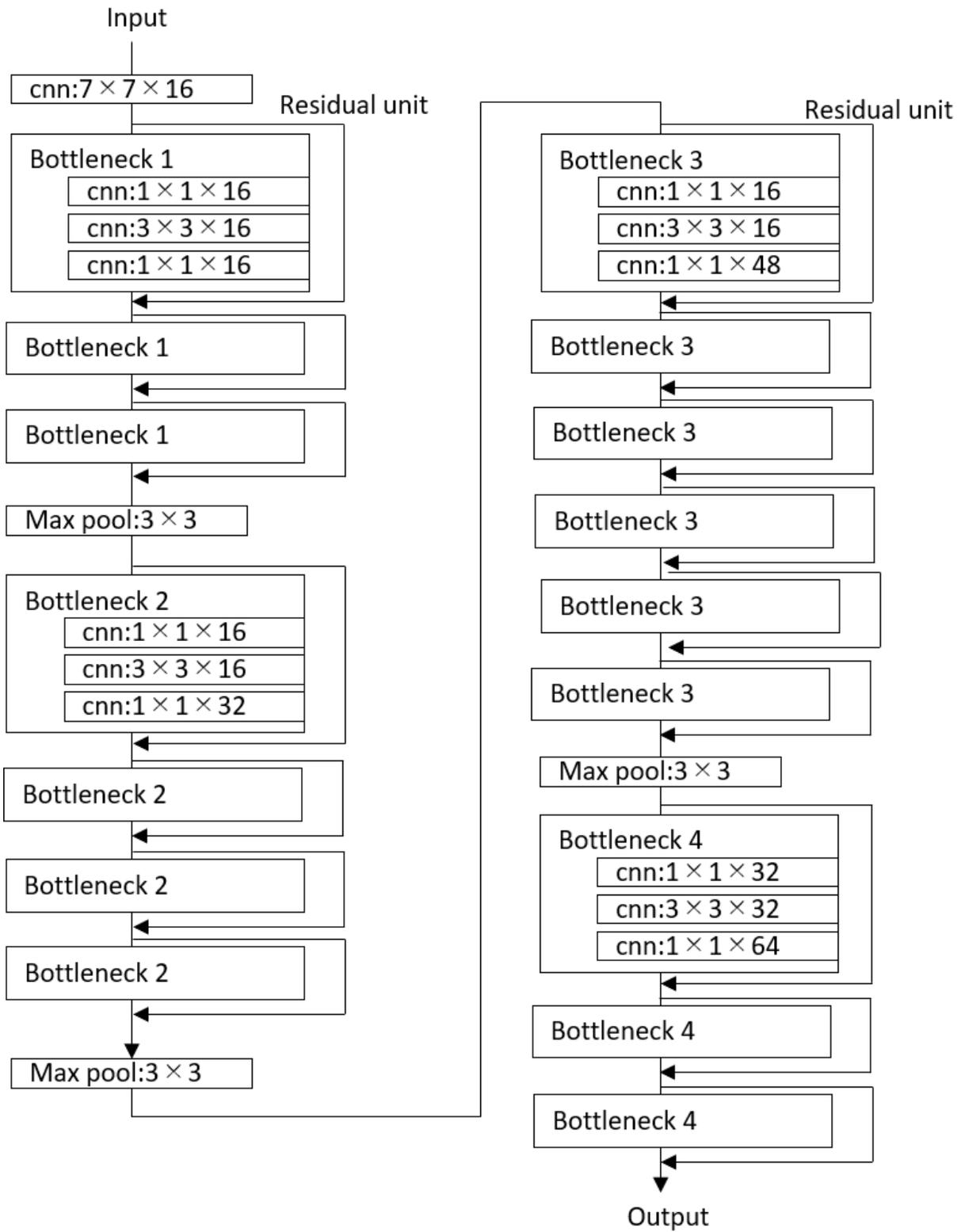


図 2.7 Deep Pyramidal Residual Networks のモデル構造

・ Residual unit

通常の深層ニューラルネットワークでは多数 (二桁以上) のニューラルネットワークが連結されているが、このネットワークを深くすることでネットワーク自身の出力の表現能力が向上されて来た。一方で深く連結することで学習自体が非常に難しくなる傾向がある。そこで学習が効率的に実現できるように Residual unit は提案された[He 16]。Residual unit は図 2.8 に示すように入力 x をそのまま伝達するルートと処理 F を行うルートを持ち、最終的な出力 o は次式によって表される。

$$o = F(x) + x \quad (2.18)$$

一般的に処理 F には図 2.7 の Bottleneck の内部のように CNN 層が多数連結される。この構造の特徴としては、Residual unit は学習時に入力値との残差の処理 F のみを学習すればタスクが実現できることである。実際に誤差逆伝播する際に、通常のニューラルネットワークは深く連結されていると誤差が途中の層で 0 となってしまう、うまく学習が進まないことがあるが、Residual unit では入力との差を学習する構造となっているため、誤差が次の層へと伝達しやすくなり、従来のニューラルネットワークと比較して学習が効率的に行われるようになる特徴がある。Deep Pyramidal Residual Networks のベースとなった Residual Network では Residual unit を利用し、処理 F にはフィルタ数 16 の CNN を入力層として、層が深くなるごとにフィルタ数を 2 倍ずつ増加させた構造を利用した。 k 番目の Residual unit の出力次元数 D_k は Residual unit が所属しているグループ番号 n として次のように表現できる。

$$D_k = \begin{cases} 16, & \text{if } n(k) = 1 \\ 16 \cdot 2^{n(k)-2} & \text{if } n(k) \geq 2 \end{cases} \quad (2.19)$$
$$n(k) \in \{1, 2, 3, 4\}$$

ここで $n(k)$ は k 番目の Residual unit が所属しているグループ番号を示す。処理の流れとしては RGB の 3 次元数で構成された画像データを入力すると第 1 層において畳み込み処理が行われて、16 次元数の特徴量マップが出力される。第 2 層では 16 次元数の特徴量マップが 32 次元数の特徴量マップに変換されこれが繰り返し行われる。最終的に 64 次元の特徴量マップが生成される。

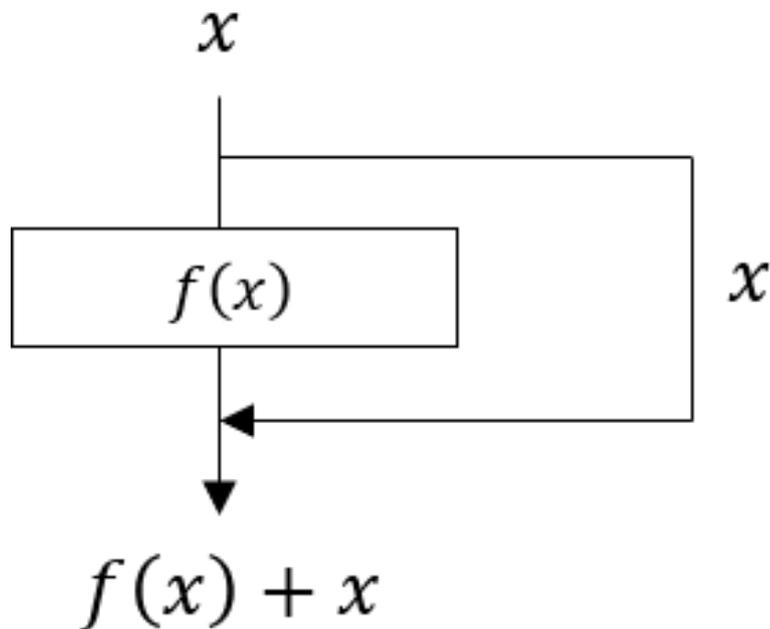


図 2.8 Residual unit の構造

- Bottleneck

Residual Network による高精度での物体認識には Residual unit だけではなく Bottleneck と呼ばれる構造が採用されている [He 16]。一般的に CNN のカーネルサイズはモデルの中では不変で同一の値を採用するが、Residual Network では図 2.7 に示すように Residual unit の処理 F にはカーネルサイズ 1×1 、 3×3 、 1×1 の 3 つの CNN 層が採用されている。Bottleneck 構造では入力された次元数よりも小さい次元マップに一旦変換してから再度もとの入力された次元数に戻すことにより、各 CNN のパラメータ数削減することで計算コストを下げ、学習にかかる時間を下げる効果がある。

- Deep Pyramidal Residual Network

Residual Network による画像認識精度について [Veit 16] が検証したところ、各層をランダムに削除した際に認識精度は大きく変化しないものの、フィルタ数が 2 倍に変化する層を削除すると精度が低下し、構造的にフィルタ数が 2 倍に変化する層に依存していることが明らかとなった。これを解決するため Deep Pyramidal Residual Network [Han 17] は図 2.7 に示すようにこの 2 倍に増加している層への依存度を減らすようにフィルタ数を層が深くなるごとに単調増加させるようなモデルを提案した。 k 番目の Residual unit の出力次元数 D_k は次のように表現できる。

$$D_k = \begin{cases} 16, & \text{if } k = 1 \\ \lfloor D_{k-1} + \frac{\alpha}{N} \rfloor, & \text{if } 2 \leq k \leq N + 1 \end{cases} \quad (2.20)$$

$$N = \sum_{n=2}^4 N_n \quad (2.21)$$

ここで n は各 Residual unit が所属するグループ番号で、 N は Residual unit の総数、 N_n はグループ n に所属する Residual unit の総数を示す。 α は次元数の変化パラメータを示し、各グループの最後の Residual unit が出力する特徴量マップの次元数は $16 + (n - 1)\alpha/3$ となる。 $\alpha = 200$ として CIFAR-10 と CIFAR-100 のデータセットで Residual Network と比較したところ、誤差率が 11.47%だけ Deep Pyramidal Residual Network は優れた性能を示した。

2.2.2 画像分類モデルの評価指標

Deep Pyramidal Residual Network の主な応用先は画像の分類タスクとなる。例えば画像が腫瘍であるかどうかを分類する 2 値分類タスクにおいては、画像を受け取った Deep Pyramidal Residual Network は多層の CNN を通じて画像に含まれる模様パターンを抽出し $16 + (n - 1)\alpha/3$ の次元数を持つ特徴量マップが生成される。その後、マップを更に全結合のニューラルネットワークに入力され最終的に画像が腫瘍であるかどうかを確立にて出力する。この Deep Pyramidal Residual Network を腫瘍の分類器として評価する際、表に示すように予測された結果を実際の正答データと比較しそれをもとに評価を行う。ある画像に対して分類器の予測が Positive (腫瘍である) のとき、実際に腫瘍の画像である場合は True Positive(TP)、腫瘍でない場合は False Positive(FP)となる。逆に分類器の予測が Negative (腫瘍でない) のとき、実際に腫瘍の画像である場合は False Negative(FN)、腫瘍でない場合は True Negative(TN)となる。これを評価用データに対して分類器による予測結果を評価し、TP、FP、FN、TN をカウントする。ここで TP は腫瘍であると予測したうち正解だった数、FP は腫瘍と予測し間違った数、FN は腫瘍でないと予測し間違った数、TN は腫瘍でないと予測し正解だった数を表す。

表 2.1 混同行列

		分類器の予測	
		Positive	Negative
正答データ	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

これらの予測結果をもとに Accuracy、Precision、True Positive Rate、True Negative Rate、

False Negative Rate、False Positive Rate、Receiver Operating Characteristic curve 等によって分類器が評価される。以下では各評価指標について記述する。

Accuracy

Accuracy は評価データ全体の中で正しく分類ができた例 (TP と TN) がどれだけ存在するかを示す指標であり次の式で表される。

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.22)$$

Precision

Precision は精度や適合率とも呼ばれ、Positive (腫瘍である) と分類器によって予測された例 (TP と FP) のうち正解だった数 (TP) の割合であり、つまりどれだけ Positive に関する予測が正しいかを示す。

$$Precision = \frac{TP}{TP + FP} \quad (2.23)$$

True Positive Rate (TPR)

TPR は再現率や真陽性率、Recall とも呼ばれ、評価データ中に存在する Positive (腫瘍である) データの内どれだけ Positive (腫瘍である) として予測できたかを表す。

$$TPR = \frac{TP}{TP + FN} \quad (2.24)$$

True Negative Rate (TNR)

TNR は真陰性率とも呼ばれ TPR の Negative (腫瘍でない) 版であり、評価データ中の Negative (腫瘍でない) データの内どれだけ TN (正しく腫瘍でない) として予測できたかを表す。

$$TNR = \frac{TN}{FP + TN} \quad (2.25)$$

False Negative Rate (FNR)

FNR は偽陰性率とも呼ばれ、全 Positive (腫瘍である) データの内、Negative (腫瘍でない) と分類器に判定された割合を示す。割合が高いほど腫瘍を見逃していることを示す。

$$FNR = \frac{FN}{TP + FN} \quad (2.26)$$

False Positive Rate (FPR)

FPR は偽陽性率と呼ばれ、Negative (腫瘍でない) データを間違っ て Positive (腫瘍である) と予測された割合を示す。割合が高いほど間違っ て腫瘍であると予測していることとなり、分類器を使用するユーザーに不信感を与える。

$$FPR = \frac{FP}{FP + TN} \quad (2.27)$$

Receiver Operating Characteristic curve (ROC 曲線)

これまでの評価指標の計算において TP、FP、FN、TN をカウントしていたが、これは分類器が Positive か Negative のどちらかを出力する場合に成り立つものである。一般的な分類器では Positive、Negative の予測を確率にて出力するものが存在する。この場合、TP、FP、FN、TN は入力データが Positive である予測確率が一定のしきい値以上のときに分類器が Positive と予測したと判定し、そうではない場合は Negative と予測したと判定するものである。しきい値の設定によって分類器の評価は大きく変化する。そこで ROC 曲線が分類器の評価指標としてよく使用される。ROC 曲線ではしきい値を変化させ、しきい値ごとにおける TPR を縦軸、FPR を横軸として図示したものである。ROC 曲線の面積を Area Under Curve (AUC) と呼ぶ。この面積は 0 から 1 となり、ランダムな分類器の AUC 値は 0.5 となる。

・医用画像診断支援システム向け評価指標

人工知能技術が発展し医療分野へと応用され始めていることで、その技術を使ったシステムの評価、開発についてガイドライン[日本医療研究開発 19][厚生労働省医薬食品局 11]が経済産業省/国立研究開発法人日本医療研究開発から発行されている。それによれば医用画像診断支援システムとは医師が様々な画像撮影装置を用いて画像診断を行う際に参考となる情報を提供するシステムである。基本的には Computer-Aided Detection (CADe) または Computer-Aided Diagnosis (CADx) と呼ばれる機能を持つ。CADe の機能とは画像上に存在する病変の疑いのある場所を自動検出し、マーキング等によって検出の支援を行う。CADx は病変の疑いのある場所の検出に加えて、病変の良悪性鑑別や進行度の分析結果を出力し、診断支援を行う。本研究では腫瘍の検出し医師の診断を支援するため CADe にあたる。CADe の性能評価にはシステム本来の診断支援が実施できることを示す必要がある。つまり、単純に分類器自身の Accuracy や ROC 曲線の評価するのではなく、CADe を用いて行った医師の診断と通常時の診断で Accuracy や ROC 曲線を比較し観察者特性を評価することが求められる。分類器単独の性能に対して、診断医がその出力を使用した場合の性能が下回る場合もあれば、上回る場合もある。

2.2.3 物体検出モデル - Faster R-CNN -

CNN や Deep Pyramidal Residual Network は画像に含まる物体の種類を識別するために

提案されたモデルであるが、画像に含まれる特徴を取り出すことに優れた性能を示し、類似したタスクである物体検出タスクにおいてもその技術は利用されている。2015年には図2.9のように画像中に複数存在する物体をBOX状の領域として検出し、更にその物体の種類を識別するような物体検出タスクをリアルタイムに実現するアルゴリズムとして Faster Regions with Convolutional Neural Networks(Faster R-CNN)が提案された[Ren 15]。Faster R-CNN はすべての処理がニューラルネットワークによって行われており、1秒あたり4枚分の画像を処理することが可能で、リアルタイムでの物体検出も実現すること可能である。Faster R-CNN の基本構造は図2.10に示すように Feature Extractor (FE)、Region Proposal Network (RPN)、Classification Network (CN)の3つのニューラルネットワークから構成されている。物体検出処理の流れについて記述する。はじめに入力画像を受け取ったFEは画像に含まれる物体のパターンを抽出し、1ピクセルに多次元情報を含む特徴量マップを生成する。次にRPNは特徴量マップから物体が存在する長方形領域を複数予測する。最後にRPNに予測された領域情報(4点の座標情報)とその領域内に存在する特徴量マップから通常の画像識別タスクと同様にCNがその領域に含まれる物体の種類を予測することで実現する。Faster R-CNNの学習には画像とその画像内に含まれる物体の領域情報と種別情報が必要となる。

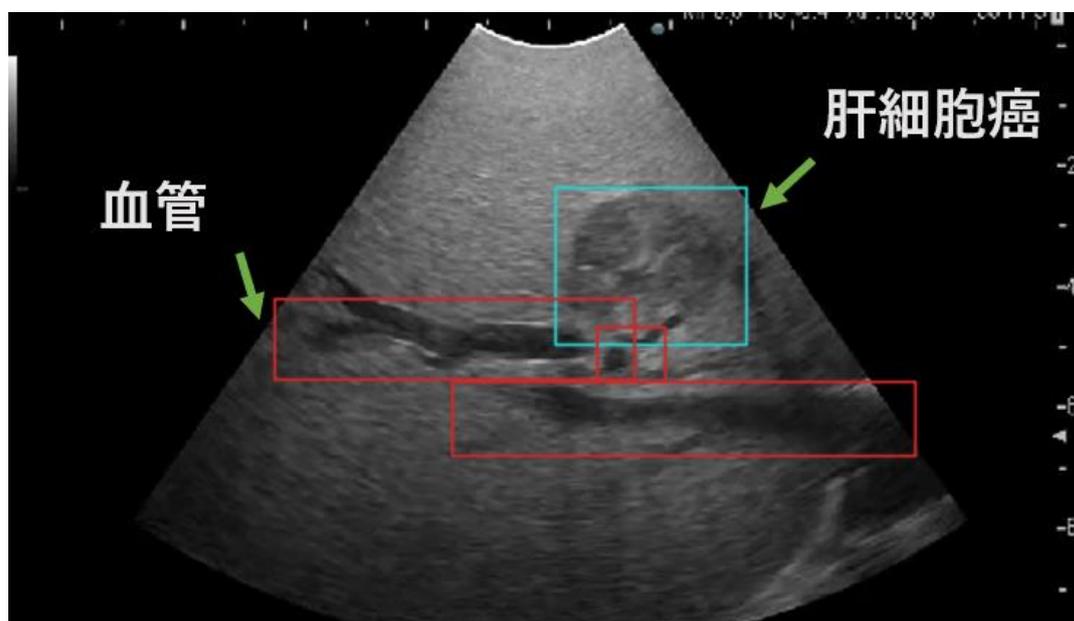


図 2.9 Faster R-CNN による物体検出

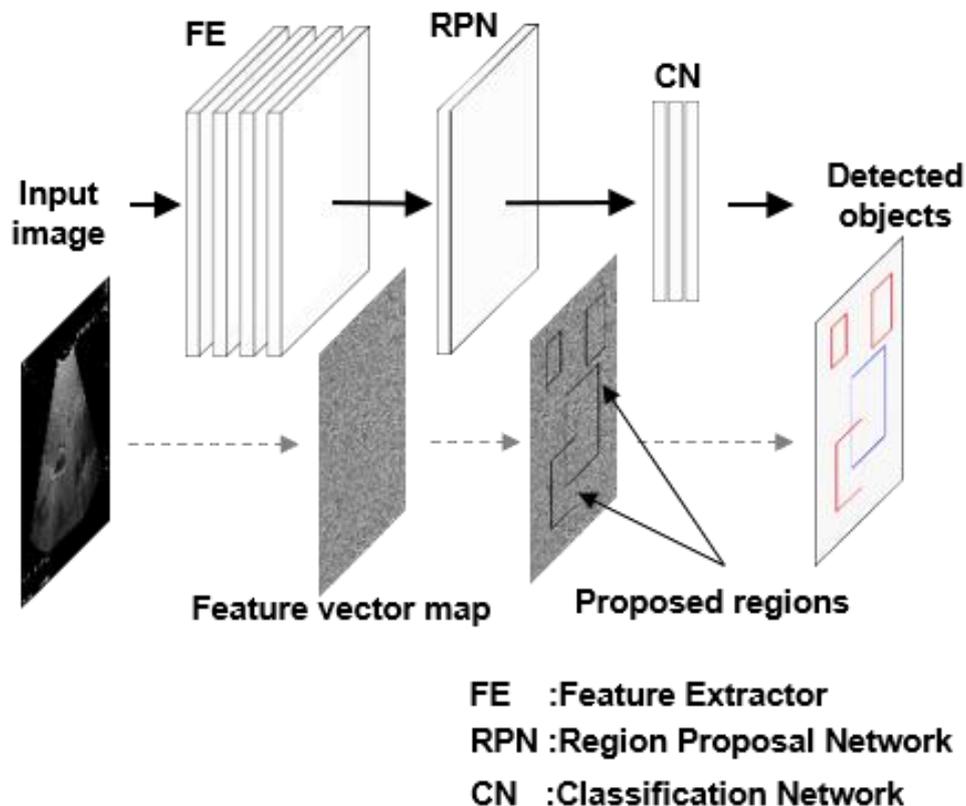


図 2.10 Faster R-CNN の基礎構造

・ Feature Extractor (FE)

FE は入力された画像から特徴量を抽出するためのニューラルネットワークである。一般的には多層 CNN や Deep Pyramidal Residual Network といった画像認識タスクにて高精度を示したモデルが採用される。また、FE に使用されるニューラルネットワークの各重み成分とバイアス成分の初期値としては事前に CIFAR-10 と CIFAR-100 といった大量の教師データを使って学習させ、物体の識別を高精度に実現できるモデルのパラメータを使用することで物体検出タスクを学習する際の学習時間の短縮及び精度向上が期待できる。この別タスクにて学習済みのモデルを利用することを転移学習と呼ぶ。

・ Region Proposal Network (RPN) と Classification Network (CN)

RPN は FE が生成した特徴量マップから物体が存在しうる領域の候補を予測するニューラルネットワークである。RPN では特徴量マップのピクセルごとに固定の物体の領域の候補を生成し、その領域にある特徴量マップから CNN を用いてその領域に物体があるかないかを予測する。CN は全結合のニューラルネットワークによって構成されており、RPN によって予測された領域とその領域に存在する特徴量マップを受け取り、物体の種類を予測する。図 2.10 のように RPN によって物体が存在すると予測された領域情報（特徴量マッ

プにおける矩形の2次元座標)をもとにFEによって生成された特徴量マップが切り取られCNに入力される。

RPNでは図2.11に示すようにFEによって抽出された特徴量マップをもとに $N \times N$ 個のAnchorと呼ばれる座標を設定する。RPNでは各Anchorを中心として9個の異なるBox状のAnchor Boxと呼ばれる領域を設定する。各Anchor Box内の特徴量マップを入力としてCNNを使用して各Anchorに対して9個のBoxの領域に物体が含まれるか、または背景かを予測する。つまり 9×2 の分類問題を行っている。更に、物体かどうかの予測と並行して9個のBox毎に正答データとのズレを予測される。

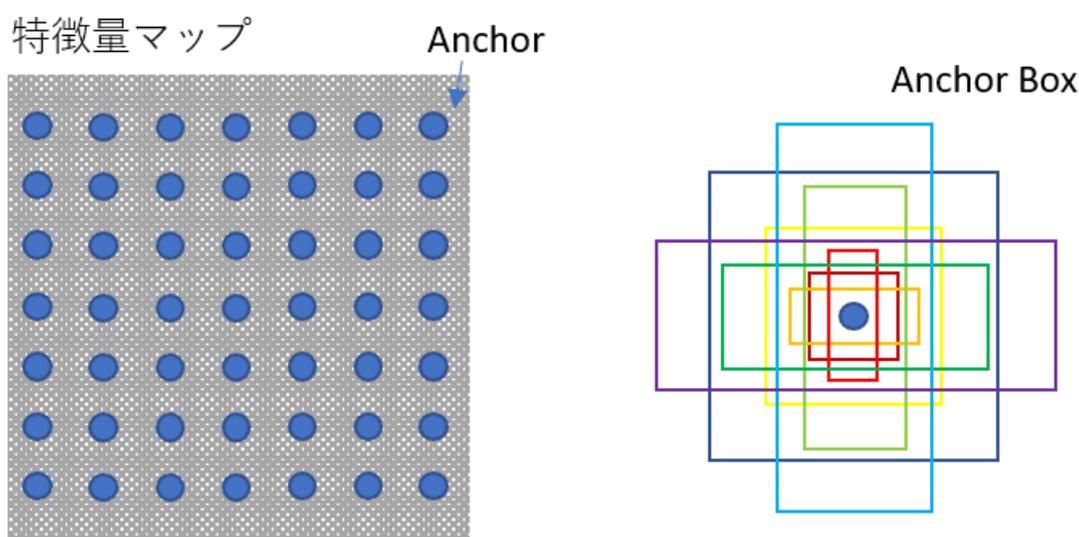


図 2.11 Region Proposal Network によって生成される Anchor

具体的には Box の領域を (S_a, S_b, S_c) のように3種類、縦横比 $(1:1, 1:2, 2:1)$ の3種類として組み合わせた $k = 9$ 種類の形状のAnchor Boxを設定する。このとき図2.12に示すように $2k$ 個の分類問題と $4k$ 個のBoxのズレの予測を行う。 $2k$ はAnchor Boxごとに領域内に物体が含まれるかそれとも背景かを2値分類する。 $4k$ ではAnchor Boxごとに正答データのBoxの中心座標 (x, y) の差、と高さ h と幅 w の長さの差の4つを予測する。このときRPNの誤差関数には次の式が使用される。

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i L_{reg}(t_i, t_i^*) \quad (2.28)$$

$$L_{cls}(p_i, p_i^*) = -(p_i^* \log p_i + (1 - p_i^*) \log(1 - p_i)) \quad (2.29)$$

$$L_{reg}(t_i, t_i^*) = S(t_x - t_x^*) + S(t_y - t_y^*) + S(t_w - t_w^*) + S(t_h - t_h^*) \quad (2.30)$$

$$S(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases}$$

$$\mathbf{t}_i = \{t_x, t_y, t_w, t_h\}$$

ここで i は i 番目の Anchor Box であり、 p_i は i 番目の Anchor Box が物体であるか予測した確率、 p_i^* が正答データ Anchor Box 内に物体が含まれていれば1そうでなければ0となる。 t_i は Anchor Box のズレを示す4次元ベクトルとなる。 L_{reg} は実際に Box 内に物体が存在する倍のみに算出される。 λ は重みパラメータであり $\lambda = 10$ が使用される。また、 $\mathbf{t}_i, \mathbf{t}_i^*$ は CN によって予測された Box の中心点 (x, y) 、幅 w 、高さ h を使用して次のように表現される。

$$\begin{aligned} t_x &= (x - x_a)/w_a, t_y = (y - y_a)/w_a, \\ t_w &= \log(w/w_a), t_h = \log(h/h_a), \\ t_x^* &= (x^* - x_a)/w_a, t_y^* = (y^* - y_a)/w_a, \\ t_w^* &= \log(w^*/w_a), t_h^* = \log(h^*/h_a), \end{aligned} \tag{2.31}$$

ここで x_a, y_a, w_a, h_a は Anchor Box の中心点座標、幅、高さを示す。 x^*, y^*, w^*, h^* は正答データの中心点座標、幅、高さを示す。

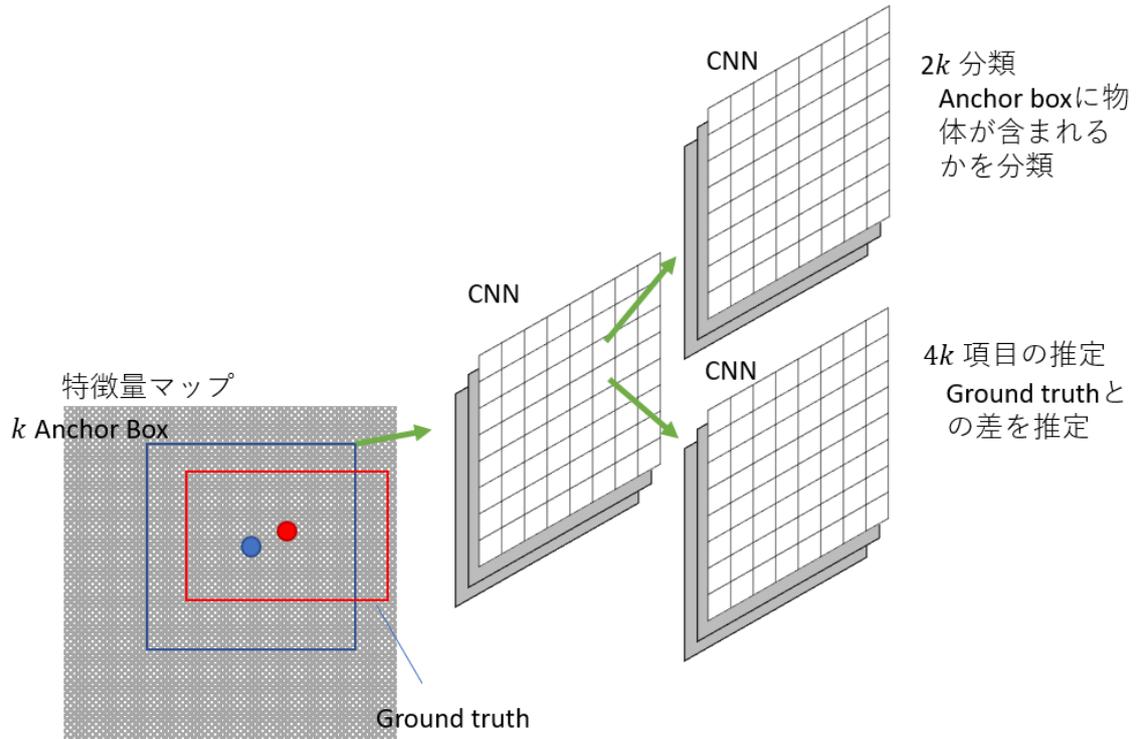


図 2.12 Region Proposal Network による Anchor Box の推定

・ Classification Network (CN)

CN は全結合のニューラルネットワークによって構成されており、RPN によって予測された領域とその領域に存在する特徴量マップを受け取り、物体の種類と Box 上左と下右の頂点の座標を予測する。図 2.10 のように RPN によって物体が存在すると予測された領域情報（特徴量マップにおける矩形の 2 次元座標）をもとに FE によって生成された特徴量マップが切り取られ CN に入力される。

2.2.4 物体検出モデル - Mask R-CNN -

Mask Regions with Convolutional Neural Networks (Mask R-CNN) [He 17]は Faster R-CNN の改良モデルであり、従来は物体の存在する領域を矩形で検出していたが Mask R-CNN ではより詳細な物体の領域自体も検出することが可能となる。Mask R-CNN では図 2.13 に示すように Faster R-CNN の構造をベースに物体の詳細領域（Mask 領域）を予測するためのニューラルネットワーク Mask Predictor (MP)が追加されている。

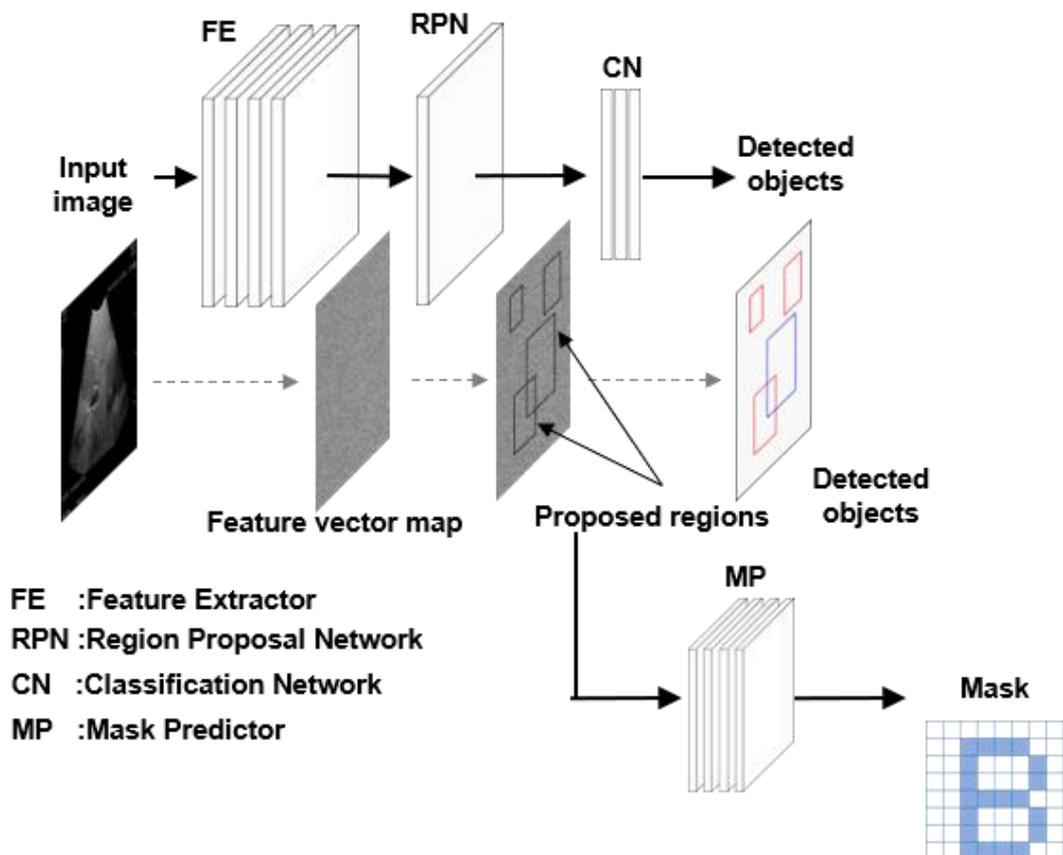


図 2.13 Mask R-CNN の基礎構造

・ Mask Predictor (MP)

MP では Faster R-CNN 内の RPN によって予測された矩形の領域から物体の存在する領域をピクセルごとに予測するため、多層 CNN で構成されたニューラルネットワークである。図 2.14 のように CN と並列して MP は存在し MP に含まれる CNN はカーネルサイズ 1×1 の出力層とそれ以外のカーネルサイズ 3×3 の中間層で構成されている。最終的にはサイズ $28 \times 28 \times 1$ の領域情報として出力され、各ピクセルの値がしきい値 0.7 以上の場合に物体が存在する画素として決まる。

2.2.5 物体検出モデルの評価指標

2 値分類タスクは表 2 に示したように TP、FP、FN、TN として予測結果が評価されていたが物体検出タスクではそのまま適用できない。一般的な物体検出タスクの評価指標として mean Average Precision(meanAP) 及び Average Recall[COCO 17a]が存在する。この指標は一般的な物体のデータを多数含む COCO データセット[COCO 17b]の学習時に使用される評価指標であり、よく他データセットに対してもこの評価指標が使用される。この指標では Intersection over Union(IoU)と呼ばれる予測領域と正答領域の重なっている割合を示す値が一定以上のときにその予測結果が正解としてその適合率を計算するものである。

IoU は図 2.14 に示すような予測領域と正答領域がある場合、次の式で表現される。

$$\text{IoU} = \frac{\text{Overlapping area}}{\text{Overlapping area} + \text{Other area}} \quad (2.32)$$

検出された領域の IoU がしきい値以上の場合正答な検出結果として TP となる。逆にしきい値未満の場合 FP となる。また FN は検出されなかった正答領域である。一方で、TN の算出ができない。これは物体検出タスクでは物体のある場所を予測するものであり、検出されなかった領域すべてが TN となり、TN として多重に領域をカウントできるため評価指標として使用できない。そのため、Precision と Recall (TPR)が物体検出タスクの性能を示す指標となる。また計算式は 2 値分類タスクと変わらず下記式で表される。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.33)$$

$$\text{Recall (TPR)} = \frac{TP}{TP + FN} \quad (2.34)$$

また、ROC 曲線と類似して縦軸を Precision、横軸を Recall としてモデルの評価が行われる。しかしながら、2 値分類タスクとことなり物体検出タスクでは 1 つの入力データに対して複数の物体が存在する可能性がある。データごとに Precision と Recall が変動するため、

IoU=50%のときの平均的な精度を求める場合 Average Precision を次のように計算される。

予測領域と正答領域の IoU が 50%を超えるとときにその予測結果が正しいとして画像中に含まれる物体がすべて検出されるまで物体の予測を行う。その際、下記式のように適合率 p と再現率 r を一回の予測ごとに計算する。

$$p = \frac{\text{計算時点での予測結果が正しかった数}}{\text{計算時点で検出した領域総数}} \quad (2.35)$$

$$r = \frac{\text{計算時点での予測結果が正しかった数}}{\text{画像中に含まれる物体の領域の総数}} \quad (2.36)$$

予測の施行ごとに求めた適合率 p と再現率 r をそれぞれ横軸に r 、縦軸に p として図示する。この際、AP は次のように計算される。

$$AP = \int_0^1 p(r) dr \quad (2.37)$$

より精度が高ければ AP は 1 に近づく。meanAP では IoU=50%から 95%まで 5%ごとに AP を算出した値の平均となる。他に IoU=50%,75%の時の AP を AP_{50} , AP_{75} と表現し、meanAP より緩い条件の指標として利用される。

同様に Average Recall は IoU=50%から 95%まで 5%ごとに算出した画像に対する式 (2.36)の平均となる。Average Recall は画像中に含まれる物体をどれだけの割合で検出するかを表している。

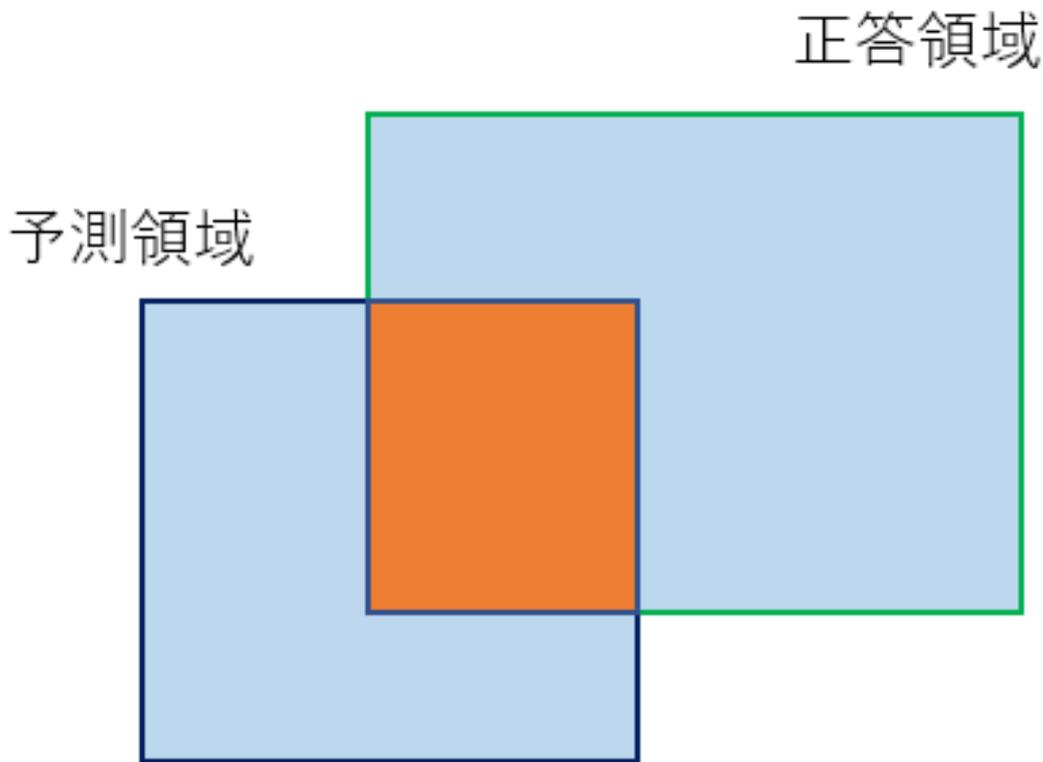


図 2.14 Intersection over Union

2.3 画像診断・検査への深層学習技術の応用と課題

2.3.1 医療向け画像診断・検査への応用研究と超音波検査の複雑性

2.2 節で紹介した深層学習技術は医療分野において特に画像ベースで行われる診断・検査と親和性が高く、様々な応用研究が行われている。医療分野における画像ベース診断は CT 検査、MRI 検査、X 線検査、超音波検査、内視鏡検査、顕微鏡による病理診断、皮膚の目視観察といった多岐にわたって存在する。これらの検査方法の中で内視鏡、顕微鏡、皮膚の目視によって得られる画像はこれまで紹介した分類・物体検出モデルが有効性を示した光学的に撮影された画像と同様で、医師による診断とモデルによる診断のプロセスに違いがない。実際に GoogleNet Inception v3 CNN [Szegedy 16] を用いた皮膚がんの診断 [Esteva 17] では 129,450 枚の大量のデータを確保し皮膚科医と同等の分類性能を示した。また、内視鏡映像から癌の検出 [Sakai 18] では単純な CNN ベースの自動検出モデルによって Accuracy : 82.8% の性能を示した。また、この内 X 線検査で得られる画像は基本的に 1 枚でその画像のみを観察し診断されるため医師の診断と深層学習による診断で情報的に齟齬が存在せず、先行研究 [Wang 17] [Rajpurkar 17] [Hemdan 20] [Rajkomar 17] でも大きな深層学習モデルの改造無し画像分類に使用される AlexNet [Krizhevsky 12], GoogleNet

[Szegedy 15], VGGNet-16 [Simonyan 14], ResNet-50 [He 16]の利用によって放射線科医を超える分類・検出性能を示している。一方でCT検査、MRI検査、超音波検査は臓器の断面を3次元空間に渡ってデータを取得する検査となる。そのため、静止面に優れた性能を示している深層学習技術が空間的に広がりを持つ断面データの集まりに対しても有効であるかといった懸念が存在する。この懸念に対してCT検査、MRI検査への応用例として単純に1枚の断面画像として扱いGoogleNet Inception v3 CNN、AlexNet、ResNet-50による分類・検出を行っている例[Wang 21] [Moeskops 16] [Ismael 20]と連続した断面画像を3次元空間に広がっている情報として抽出することを目的とした通常の畳み込みニューラルネットワークを3次元畳み込みニューラルネットワークやSeg3DNetと呼ばれる3次元物体のセグメンテーションを実現するニューラルネットワークの提案を通じて分類・検出研究[Huang 17][Zhou 17] [Gerard 18] [Pereira 16]が存在する。どちらも高精度での分類・検出を実現しているが、単純な画像ベースでは特定の断面でしか検出されないといった恐れが存在する。また、超音波検査は臓器の空間的な断面データを取得できるがCT検査、MRI検査とは異なりその連続的な断面の広がりが直線ではなく可変となる。これはCT検査、MRI検査では身体を固定し、特定の直線方向に微小にずれながら断面データを取得するのに対し、超音波検査では操作者がプローブを観察対象の表面に当てながら断面データを取得するためである。これによって撮影される断面が物体に対して多角的な角度の断面となり、他の検査によって取得されるデータと比較し大きく自由度が増加し、タスクとしての複雑性が増している。先行研究ではこの可変方向に対する断面データの広がりといった問題点に対しては取り組まれておらずあくまで肝臓の超音波検査に対しては疾患部分の画像にフォーカスし単純なCNNベースで構築されたモデルを使用して分類、ステージ診断[Reddy 18] [Xu 17]をおこなう研究があるのが現状である。肝臓以外の超音波検査に対しては甲状腺結節[Wang 19]に関してYOLOv2[Redmon 17]を使用して検出研究例が存在するがこれらは撮影対象が小さく、撮影される角度の自由度も肝臓と比べて小さいものとなる。肝臓の超音波検査によるスクリーニングでは肝臓全体を観察する必要があり、この多角的に撮影される断面映像とまた付随して血管が絡み合っている肝臓内部が要因となり分類、検出タスクが複雑なものとなっている。

2.3.2 データ多様性増加の必要性和応用研究

肝臓の超音波検査に対する診断支援システムの実現にあたってタスクの複雑性が一つの要因となるが、それに付随してデータ多様性を確保することが重要な要素となる。2.1、2.2節で述べたように深層学習は基本的に教師データから物体の特徴を学習することで未学習データに対しても分類、検出を実現する。肝疾患の超音波検査映像に対する診断支援システムを実現にあたっては典型的な腫瘍等の検出だけでなく多角的に撮影される状況でも検出の実現が求められる。この問題に対して一般的にデータ増強手法が利用されるが、関連先行研究である甲状腺結節の超音波検査画像に対する物体検出を実現した[Wang

19]では使われておらず、超音波検査画像から非アルコール性脂肪肝炎のステージ分類 [Reddy 18]では一般的な光学的に撮影された画像向けの画像反転、回転、拡大等が使われているのみである。一般的によく使用されるデータ水増し手法のノイズ付加、回転、左右反転、上下反転、Random Resize Crop についてサンプル例を図 2.15 に示す。このように光学的に撮影された画像ではこのようなデータ水増し手法によってカメラの角度の変化や物体の移動した場合のデータを生成できる。一方でこれをそのまま超音波検査データに利用する場合学習に悪影響をもたらす可能性もある。現状では超音波検査向けのデータ水増し手法が存在せず、データ多様性の向上に有効なデータ水増し手法の探索が必要となる。また、医学データ向けのデータ水増し手法として Generative Adversarial Networks [Goodfellow 20] (GAN)が期待されており、実際に CT 画像、MRI 画像の合成 [Han 18] [Sun 20] [Islam 20]を実現しているものも存在するがあくまで1枚の静止画を生成する手法であり空間的に連続した断面画像を生成するものではない。これを超音波検査に適用する場合、本来存在しない位置に血管が出現するといった問題が発生してしまう。他に利用可能性のあるデータ水増し手法としては Random Erasing [Zhong 20]、Random Resize Crop [Pytorch 17]がある。Random Erasing は画像の一部をランダムに削除することで擬似的に物体とカメラとの間に障害物を発生させる。このような物体の一部が隠れるような状態は超音波検査にも存在しうるもので、撮影方向変化によって今まで見えていた腫瘍、血管の一部が消える状態を再現することが期待できる。Random Erasing と逆に Random Resize Crop では画像の一部をランダムに切り取る教師データとして利用する手法である。Random Erasing と近い効果があり、切り取った部位によりフォーカスされ、更に切り取りによって撮影時に発生する障害物等に遮りを再現することができる。これに関しても超音波検査でも発生しうる状況であり効果が期待される。

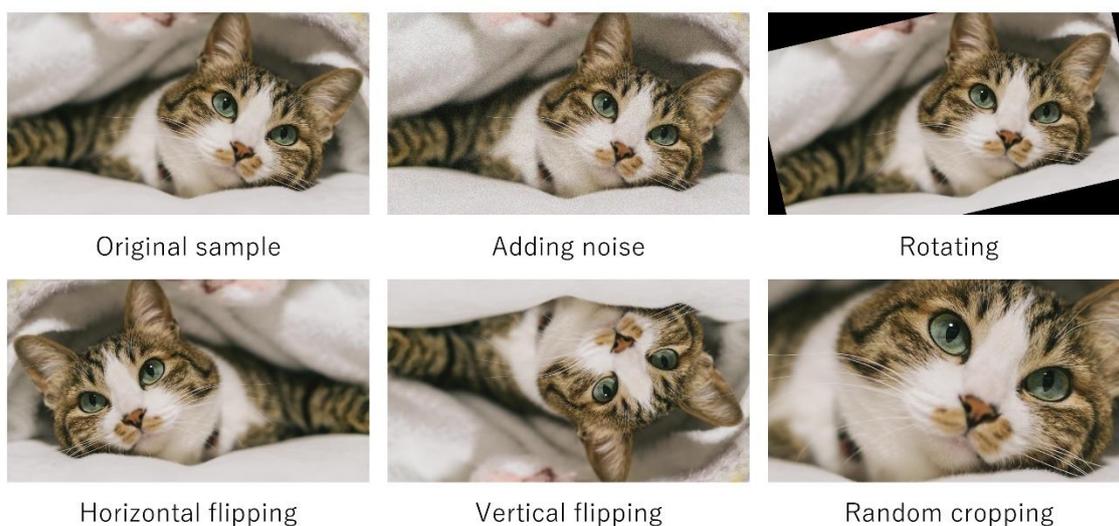


図 2.15 一般的なデータ水増し手法による変化例

2.3.3 診断支援システムとしての評価

最後に診断支援システムを実現する上でその評価方法は重要な要因となる。上記で医師の精度に匹敵した性能を示した先行研究はあくまで評価画像データに対して研究内で構築したモデル自身の性能と医師の診断を比較した結果である。肝臓の超音波検査は検査者がプローブを操作し肝臓内の映像を取得する必要がある。その映像変化の中で検出し、腫瘍発見の補助を実現する必要がある。診断支援システムとしての評価には単純な評価用データに対する検出性能だけでなく、本来の目的である医師の支援を通じた診断を補助に基づいて実際にシステムを利用した場合の有効性について検証が必要となる。検証に関しては超音波検査映像から甲状腺結節の検出研究で ROC 曲線が採用されている他に医用画像診断支援システム（人工知能技術を利用するものを含む）開発ガイドライン[日本医療研究開発 19]で実際の使用状況つまりシステムを利用した場合の診断に対する ROC 曲線の評価が望ましいとされており、一般診療所での利用を想定し肝臓の専門経験が乏しい医師がシステム利用によって腫瘍の認知に関する ROC 曲線と AUC の評価が適切だと考える。

3. 腫瘍・血管検出モデルの実現

3.1 背景・目的

診断支援システムの実現にあたって一つ目の課題である多角的に撮影されることによって発生する複雑性を有する肝臓の超音波検査映像に対する物体検出タスクを実現するため、本章では腫瘍・血管検出モデルを実現するベースモデルを構築する。構築にあたって、映像変化性といった点で複雑な物体検出タスクに対して基礎的な静止画ベース物体検出モデルの Faster R-CNN を 3.3 節で利用する。さらに、映像的变化に着目し Optical Flow によって本来の超音波検査映像による診断に利用される情報を補完し、映像変化に対して検出性能の向上を期待したモデルの導入を提案することで、複雑なタスクへの深層学習技術自体の有効性とそれと比較して Optical Flow による情報補完の超音波検査映像に対する有効性を 3.4 節で明らかにする。

3.2 データセット

深層学習によるタスクの学習には教師データが必要不可欠である。より高精度な識別を実現するには多様性のあるデータが求められる。しかしながら、大人数の患者より肝臓の超音波検査映像の確保とその教師データ作成には大きなコストが掛かり、現実的に困難である。本章でははじめに深層学習技術によって肝臓の超音波検査映像に含まれる主な物体である腫瘍（肝細胞癌）と血管についての検出タスクの実現性を検証することが目的のため、肝細胞癌を有する一人の患者から得た多角的な角度から撮影された超音波検査映像を利用して、教師データに利用する。データは東京大学医学部附属病院 肝胆膵外科・人工臓器移植外科にて得られた約直径 2cm の肝細胞癌を有する患者の術中超音波検査映像を使用した。この術中超音波検査映像は肝切除を行う前の直接肝臓にプローブを当て、肝細胞癌及び、肝臓全体をスクリーニングした際に撮影されたもので、一般的な腹部超音波検査映像よりも多角的な角度で且つ鮮明に肝臓が撮影可能なため、教師データとして選定した。また、超音波検査映像の利用にあたってこの研究は東京大学大学院 医学系研究科・医学部 倫理委員会によって承認されており（No. 2019166NI）、ウェブサイトでのオプトアウト形式としてインフォームドコンセントを得ている。

撮影された映像は合計で 2 分 58 秒の動画となる。この映像をもとに熟練の肝胆膵外科医師が術前に取得した CT 画像、MRI 画像に基づき腫瘍・血管のラベリングを行った。図 3.1 に作成したデータのサンプルを示す。ラベリングデータは超音波検査画像と画像内に含まれる物体の BOX 情報（物体が含まれる最大領域の上下左右の辺で構成される四角形の 4 つ

の頂点座標)とそのラベル情報(肝細胞癌または血管)で構成される。教師データとして72枚、学習評価用データとして19枚、合計91枚のデータを作成した。各画像は2分58秒の動画からいくつかのシーンを選び抽出した画像である。

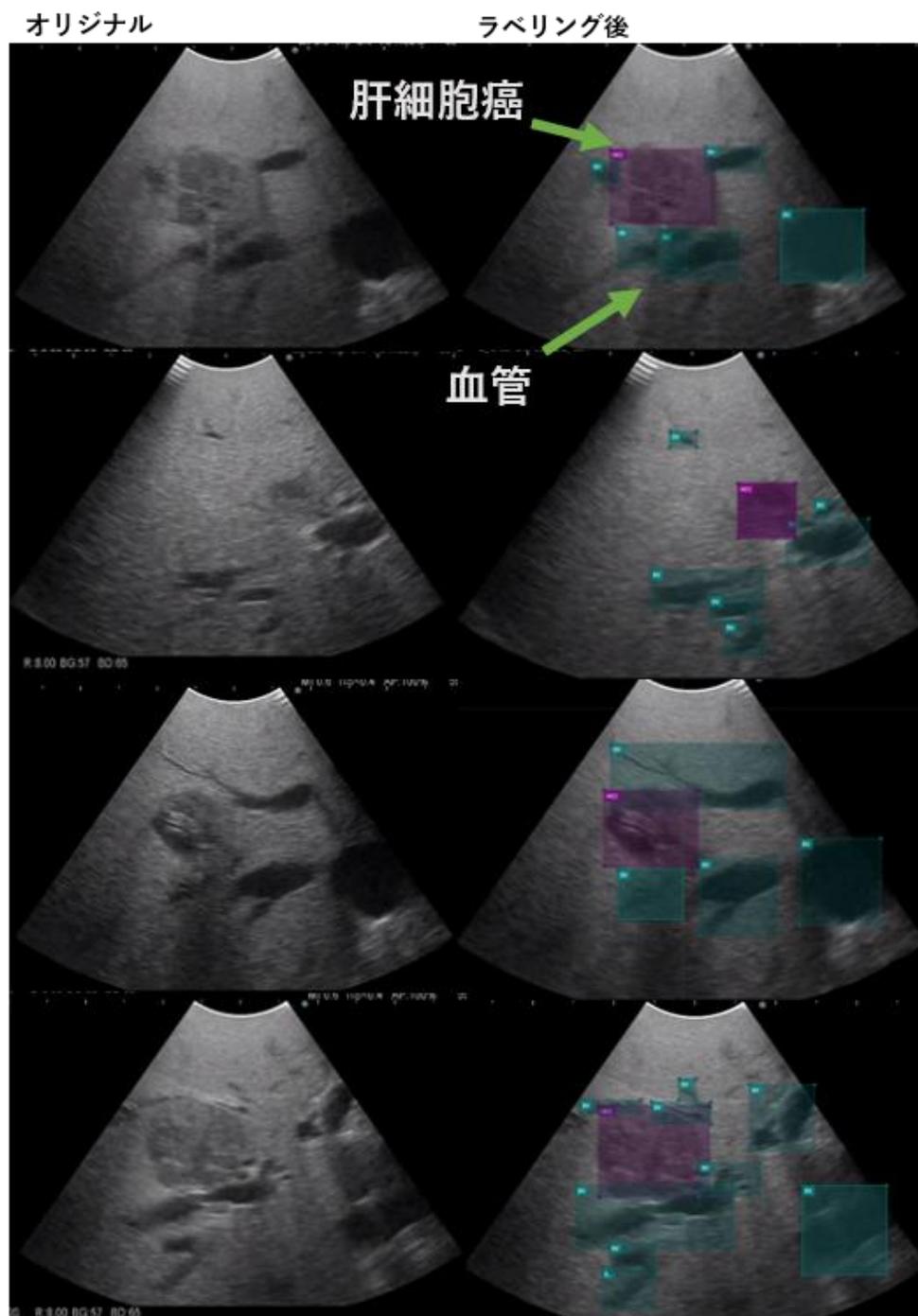


図 3.1 作成した教師データ [Karako 21]

3.3 Faster R-CNN による腫瘍検知タスクの構築

3.3.1 Faster R-CNN 導入目的

肝疾患の診断支援システムを導入し、検査者の診断支援する上で、システムにはリアルタイム性、超音波検査映像からの疾患及びその他部位の検知といった機能が要求される。これを実現するために、本研究では深層学習技術を用いたニューラルネットワークモデルの構築を行う。一般的な深層学習による画像識別・物体検出タスクは光学的に物体を外側から撮影した画像を使用する研究が主であった。超音波検査画像は物体の超音波の反射波を画像化したもの、物体の断面を撮影した画像である。肝臓内には腫瘍だけでなく、グリソン鞘や肝静脈、門脈が複雑に存在しており、これを多角的な角度から撮影された超音波検査画像から、従来の深層学習技術によって物体検出を実現することが難しい可能性が存在する。そこで本章では、物体検出モデルの基礎技術である Faster R-CNN[Ren 15]を用いて肝細胞癌の検知性能について検証を行う。

肝細胞癌の検出を実現する上で、物体検出タスクとして肝細胞癌と肝臓内の血管を検出・識別するモデルの構築を行う。肝細胞癌のような腫瘍は基本的には球体の形状で肝臓内に存在し、血管のような管や管が複数分岐したような形状で存在する。これを超音波検査で撮影した際に、腫瘍は球状のため断面は基本的には円形となるが、血管は撮影する角度によって断面の形状が大きく変化する。このため、肝臓内の主な構成要素である肝細胞癌と血管が正しく識別できるかどうかについて検証を行う。

3.3.2 Faster R-CNN による腫瘍・血管検出モデル構成

Faster R-CNN を用いた物体検出の実現にあたり構成した各ニューラルネットワークの構成を図 3.2 示す。図中の Linear nn は全結合ニューラルネットワークを示し、Linear nn の数字は出力数を示す。構築したモデルは最終的に画像中に含まれる物体の種別情報 (Class) と領域情報 (Box) を出力する。FE には様々な画像認識向けニューラルネットワークが利用できるが、本節では CIFAR-100 で高い認識精度を示した Deep Pyramidal Residual Networks を利用する。また、転移学習と呼ばれる事前に他の画像データを学習したニューラルネットワークのパラメータを利用することで学習データ不足を補う。転移学習用のパラメータには Faster R-CNN を使用して 2017 年度の Common Objects in Context (COCO) 内の Object Detection Task 向けに公開されているデータ[COCO 17b]を使用して学習済みのパラメータを使用した。COCO データセットには 200,000 以上の画像と 80 種類の物体が含まれている。COCO データセット自体は光学的に撮影された写真が含まれたデータセットで腫瘍や血管の画像とは異なるものであるが、画像に含まれる模様パターンの抽出方法をニューラルネットワークに学習させる上で必要なデータとなる。超音波検

査画像から甲状腺結節を検出する先行研究[Wang 19]でも同様に一般的な写真データを学習させた学習済みモデルを利用している。転移学習として FE と RPN に使用されているニューラルネットワークの各パラメータを利用した。また、CN の初期パラメータは He の初期値を使用する。COCO タスクにおいて 80 種類の物体を分類しているが、本研究では腫瘍と血管と背景の 3 種類の分類を行うため、CN の構成が異なりそのまま利用できないため、活性化関数に ReLU 関数を使用する場合に一般的に採用される He の初期値を利用する。

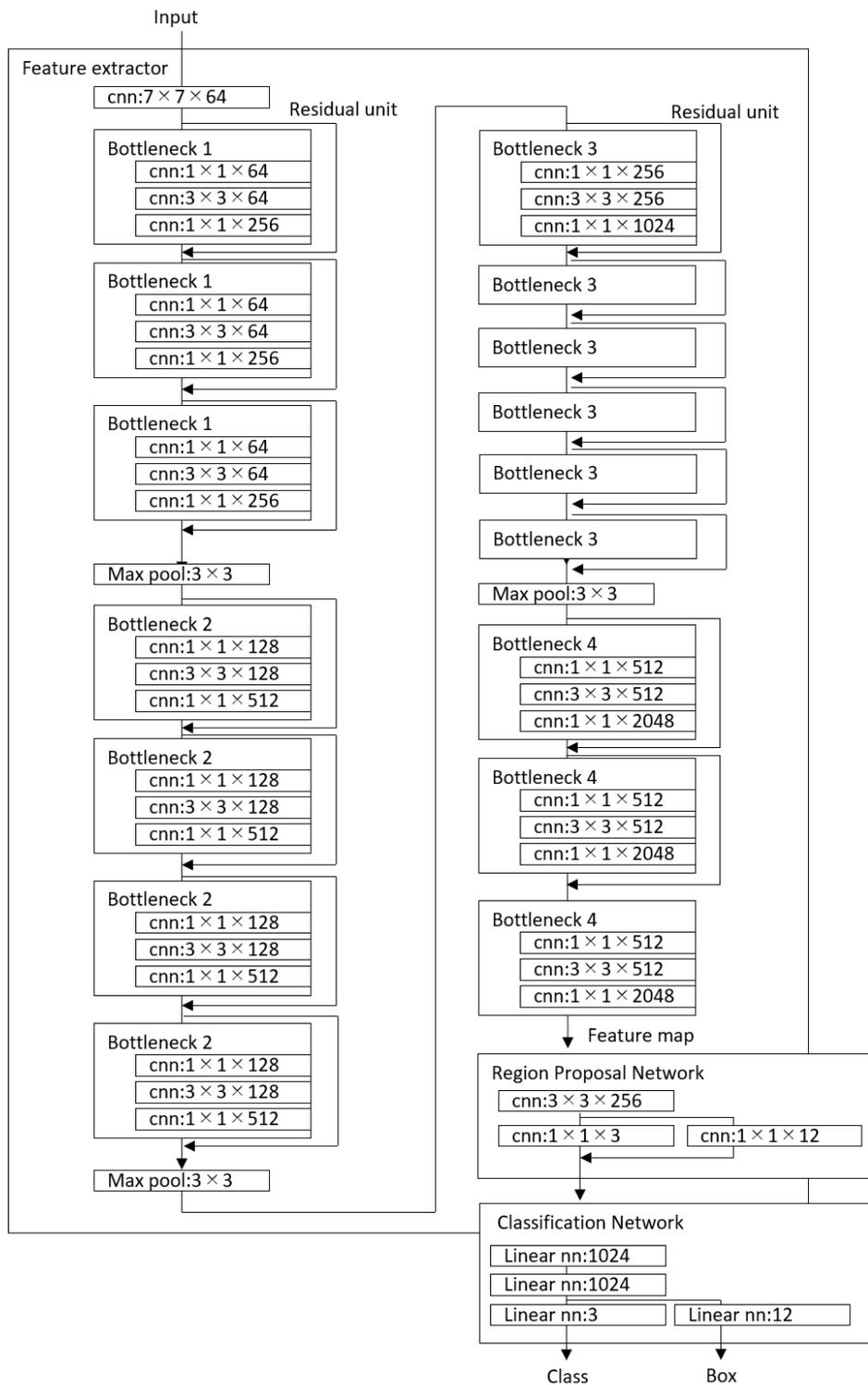


図 3.2 構築した Faster R-CNN のネットワーク構造

3.3.3 Faster R-CNN による腫瘍・血管検出性能の評価

上記で構成した Faster R-CNN に 3.2 節で作成した 72 個の教師データを入力し、各パラメータを調整することで学習を行う。具体的な学習パラメータを表 3.1 に示す。表のバッチサイズとは学習時にデータを入力するひとかたまりを示す。このバッチごとに Faster R-CNN の重みパラメータが一度更新され、これが 72 個の教師データすべてにパラメータ更新が行われる。すべての教師データ入力し、その際の誤差逆伝播法に従いパラメータ更新が更新される。すべての教師データに対して更新を行うことを 1epoch と呼び。この epoch を繰り返すことで誤差関数が小さくなるようにパラメータが更新される。各 epoch における教師データの入力順番はランダムであり、毎 epoch ごとにシャッフルされる。

表 3.1 Faster R-CNN の学習パラメータ

学習パラメータ	
学習 epoch 数	300
バッチサイズ	4
学習率 η	0.001
教師データ数	72

検出精度の評価にあたって、一般的な物体検出評価指標である AP を利用する。上記設定のもと学習したモデルに学習評価用データとして 19 枚を入力した際の AP を表 3.2 に示す[Karako 21]。また合わせて、モデルによって検知された物体のサンプル画像を図 3.3 に示す。厳しい評価指標である IOU=50%から 95%の平均の AP(meanAP)では 0.530 を示した。転移学習に使用したもとの COCO データセットに対する meanAP は 0.370 であり、検出すべき種別が少ない分より高精度に検出が実現できている事がわかる。また、IOU=50%においては 88.1%で物体を正しく検出が実現できた。

表 3.2 評価データに対する Faster R-CNN の検出精度

mean AP	0.530
AP ₅₀	0.881
AP ₇₅	0.625

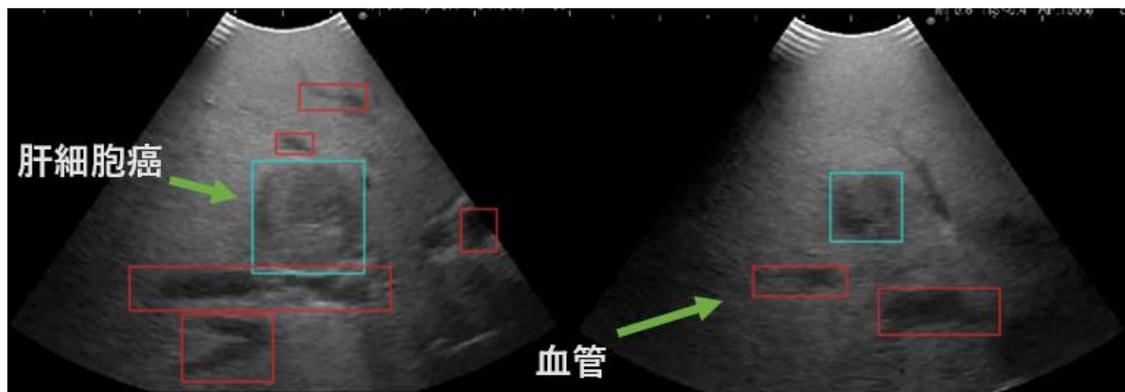


図 3.3 Faster R-CNN による検出結果

3.4 Optical Flow Network 導入による時間変化情報補完モデルの構築

3.4.1 Optical Flow Network 導入目的

Faster R-CNN によって肝臓の超音波検査画像に含む腫瘍・血管の検出が一定以上の精度で実現した。Faster R-CNN は入力された静止画から FE が物体の断面の画像パターンを抽出し、そこで得られた特徴量から物体の存在する場所と種別を推測して検出を行っており、これは一般的な物体検出タスクとして Faster R-CNN を利用したものである。さらなる肝臓の超音波検査に特化したモデルの実現にあたって、本章では医師が肝腫瘍のスクリーニング検査時に利用される情報に着目し Faster R-CNN と組み合わせることで検出性能に与える影響を検証する [Karako 21]。

実際に医師は肝臓の超音波検査を使用した診断では単純に腫瘍の静止画を観察して診断を行っているわけではなく映像の連続的な変化から映像中に含まれる物体の 3 次元的形状や模様パターン、物体とその周辺の色をもとに行っている。3.3 節で使用した Faster R-CNN は単純に 1 枚の画像データを元にそこに含まれる物体の模様パターン、色を抽出し画像に含まれる特徴量を学習している。そのため、Faster R-CNN は実際の観察に必要な情報として映像的变化（立体的な形状）を考慮せず単純に静止画情報のみから検出を行っている。ここにはモデル自身の性能を向上させる余地があると考え、本章では Faster R-CNN の入力情報として静止画だけでなく静止画の前フレームからの変化情報を合わせて入力することで実際に診断に利用する情報を補い、Faster R-CNN の改良及びそれがもたらす影響について検証を行う。

前述のとおり医師による診断プロセスでは、医師が超音波検査機のプローブを動かしながら映像を観察しそこから得られた肝臓の静止画だけでない連続的な断面の変化から物体の 3 次元的形状を理解・認識することで肝臓内の 3 次元的な構造・各物体のつながりを把握し参考にしながら診断を行っている。そこで、本節ではニューラルネットワークが医師

による肝臓の超音波検査診断プロセスで利用される情報である時間による画像変化情報を利用できるようにモデル内部でこの時間による変化情報を抽出及び特徴量の抽出に利用するモデルを新たに構築しその影響について3章の結果と比較する。入力としては1枚の静止画だけでなく、その静止画の1フレーム前に撮影された画像の合計2枚の画像データをニューラルネットワークに入力させ、それに加えて単純な連続した2枚の画像データではその関連性の抽出が困難なため、Optical Flow Network[Dosovitskiy 15]と呼ばれる2枚連続画像から物体の変化を移動量ベクトルとして推測するニューラルネットワークモデルを使用することで超音波検査によって撮影された前後の断面情報と時間経過による変化情報をFaster R-CNNに与える。

3.4.2 学習・評価データセット

構築する改良モデルの超音波検査映像へ適正及びFaster R-CNNと比較を行うため、教師データには3.3節で使用したデータセット（学習用データ72枚、評価用データ19枚）をモデルの学習・評価に使用する。加えて、より正確に2つのモデルに比較を行うため、3章で使用したデータセットに加えて、2分59秒の動画から新たに未使用の30シーンを抽出し、評価データを作成する。1つのシーンから連続した5枚の画像を抽出し、3.2節と同様に腫瘍・血管のラベリングを行った。30シーンから作成したデータは合計150枚で、データには腫瘍が142個、血管が689個含まれる。

3.4.3 Optical Flow Network 導入による腫瘍・血管検出モデル構成

Faster R-CNNをベースにOptical Flow Networkによって生成される2次元変化ベクトルマップ及び連続した2枚の画像を入力するように改良したニューラルネットワークモデルの構成[Karako 21]を図3.4に示す。改良モデルに含まれるFE、RPN、CNは3章で構築したFaster R-CNNと同様のパラメータを持つ。Faster R-CNNに追加された箇所はOptical Flow Network、Summation layer、Aggregation layerである。これら3つはすべてニューラルネットワークで構築されている。改良モデルの処理の流れとしてははじめに超音波検査画像Current frameとその1フレーム前の画像Previous frameを受け取る。FEによって両画像は特徴量が抽出を行う。それと同時にOptical Flow Networkによって両frame間の各ピクセルの移動量ベクトル情報が生成される。Summation layerで生成されたPrevious frameの特徴量と移動量ベクトルが合算を行い、Aggregation layerで合算情報とCurrent frameの特徴量の合成を行い、両frameと移動量ベクトルの3つの情報から抽出された特徴量を生成する。その後は、通常のFaster R-CNNと同様にRPNによって物体が存在する領域の推定、CNによって物体の種別の推定が行われる。

また、追加したニューラルネットワークであるOptical Flow Network、Summation layer、Aggregation layerの詳細構成は下記に示す。

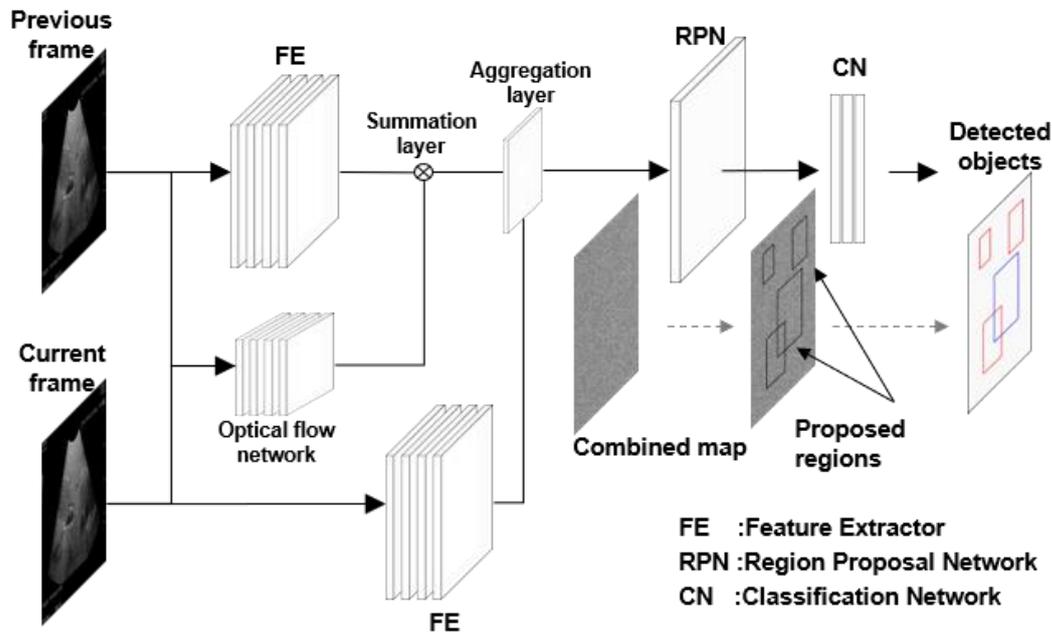


図 3.4 Optical Flow Network を導入した腫瘍検知モデルの構成[Karako 21]

・ Optical Flow Network

Optical Flow Network は CNN をベースに作られた Optical Flow を予測するモデルである。Optical Flow は物体やカメラの移動等によって生じた物体の見かけ上の変化情報であり、連続したフレーム間から 1 フレーム上のあるピクセルが 2 フレーム目にどこに移動したかを 2 次元ベクトルにて表現したものである。この Optical Flow 推定においても CNN を利用した FlowNet[Dosovitskiy 15]、改良版の FlowNet2.0[Ilg 17]といった手法が提案されている。本研究では FlowNet2.0 よりも 1.36 倍高速に実行可能でかつモデルサイズが 30 分の 1 の LiteFlowNet[Hui 18]を採用する。LiteFlowNet のパラメータには Optical Flow 推定のベンチマーク問題として使用される FlyingChairs[Dosovitskiy 15], ChairsSDHom[Ilg 17]のデータを用いて学習済みモデルを使用した。学習済み LiteFlowNet によって肝臓の連続超音波検査画像から推定されたオプティカルフローを図 3.5 示す。オプティカルフローマップの各色はそのピクセルでの 2 次元のオプティカルフローベクトルを示している。オプティカルフローマップの画像化にあたっては青色を 111 に固定し、赤色が横方向、緑色が縦方向を表すように RGB として 2 次元ベクトルを変化させた。111 を超える色の場合、そのピクセルは正の方向へのオプティカルフローを示している。

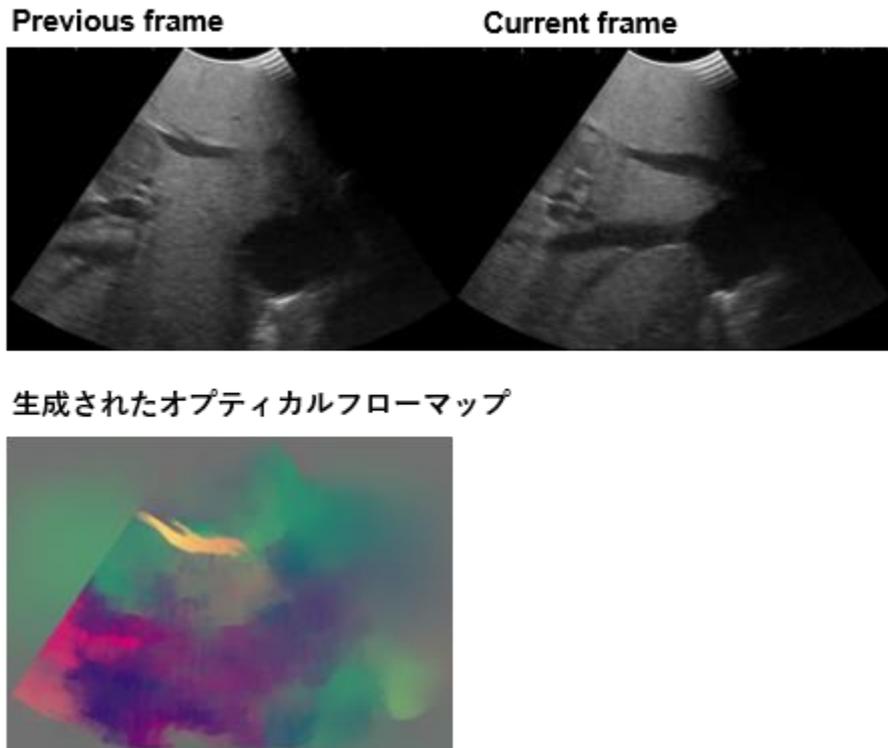


図 3.5 Optical Flow Network によって推定されるオプティカルフローマップ[Karako 21]

・ Summation layer

Summation layer では Optical Flow Network によって生成された 2 次元の移動量ベクトルマップと FE によって抽出された前フレームの特徴量マップを図 3.6 に示すよう合成を行い新たな特徴量マップを生成する。

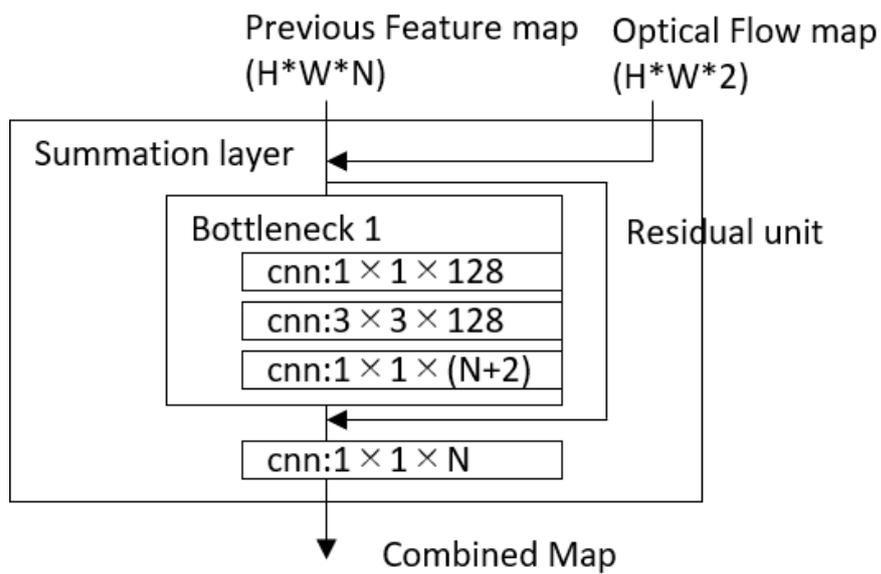


図 3.6 Summation layer の構成

・ Aggregation layer

Aggregation layer では Summation layer によって合成された特徴量マップと現在のフレームから抽出された特徴量マップを合成する。図 3.7 に示すように Aggregation layer によって前フレームと現在フレームの差分を抽出するし、その情報を現在フレームの特徴量に加算する形となっている。

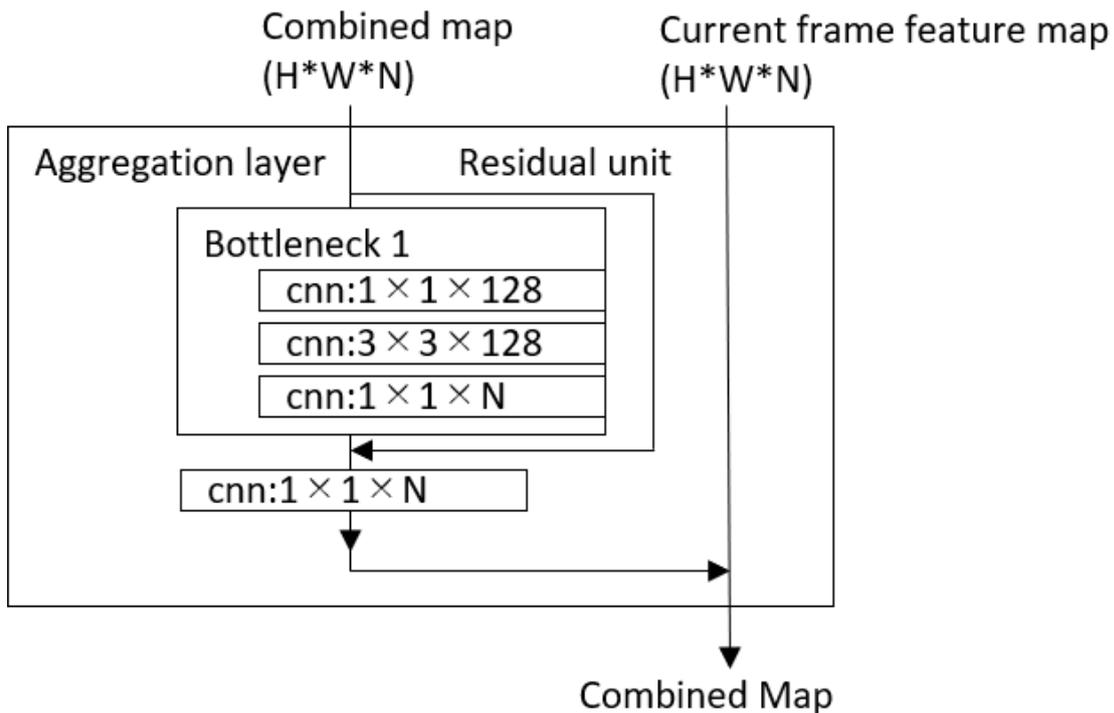


図 3.7 Aggregation layer の構成

3.4.4 Optical Flow Network 導入による腫瘍・血管検出性能の評価

Optical Flow Network を導入した改良モデルの評価及び Faster R-CNN との比較を行うに当たり、3.4 節と同様のデータ・学習設定にて改良モデルの学習を行った。Faster R-CNN との比較するにあたり 3.2 節の評価用データ 19 枚に対する一般的な物体検出評価指標である AP を表 3.3 に示す[Karako 21]。改良モデルは Faster R-CNN より 0.019 だけ優れた meanAP を示した。また、IOU=50%という緩い条件では両モデルの差は 0.006 であるが、IOU=75%では改良モデルが 0.016 だけ優れている。これは改良モデルに物体の認識精度においては大きな差はないが、より正確な物体の領域を予測することができていることがわかる。

表 3.3 Optical Flow Network を導入した改良モデルの検出精度[Karako 21]

精度	Methods	
	Model with Optical Flow Network	Faster R-CNN
mean AP	0.549	0.530
AP ₅₀	0.887	0.881
AP ₇₅	0.641	0.625

加えて、多様性あるデータに対しての腫瘍・血管の検出精度について検証するため 4.2 節の 30 シーンから生成した 150 枚のデータを使用する。各シーンには連続した 5 枚の画像が含まれ、これを改良モデル・Faster R-CNN のそれぞれに入力した際の腫瘍・血管に対する感度を算出する。シーンごとの感度 S は次式によって表現される。

$$S_{Label} = \frac{\text{モデルによって Label として検出された総数}}{\text{シーン中に含まれる Label の総数}} \quad (3.1)$$

予測された領域が正答データの領域の IOU=50% 以上の場合にその物体として正しく検出されたとして計算を行った。30 シーンに対して算出したラベルごとの感度、シーン全体の平均及び対応がある場合の 2 標本の t 検定の値を表 3.4 に示す[Karako 21]。加えて、改良モデルと Faster R-CNN のそれぞれにシーンを入力した際の出力結果を図 4.5 に示す[Karako 21]。

表 3.4 各ラベルに対するシーンの平均感度[Karako 21]

Label	mean Sensitivity per scene		p -value	t -value
	Our model	Faster R-CNN		
Tumor	0.627±0.390	0.456±0.411	0.003	-2.921
Blood Vessel	0.558±0.201	0.576±0.202	0.290	-0.559

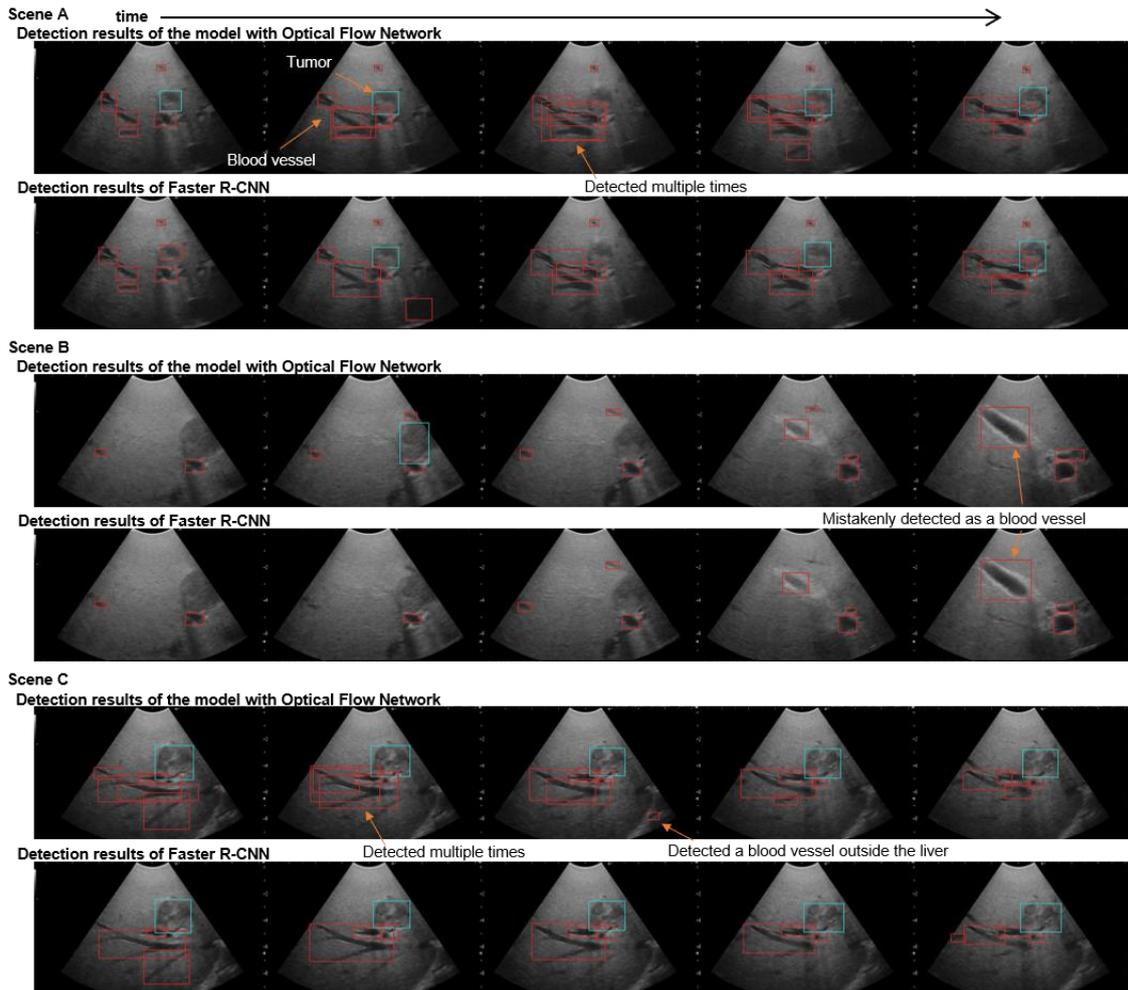


図 3.5 各モデルのシーンに対する検出結果[Karako 21]

血管の検出に関して 2 モデル間で有意差は確認できなかったものの、腫瘍の検出に関してはシーンに対する感度の平均において改良モデルは 0.171 の差を示した。また、それぞれの出力結果を観察すると、Faster R-CNN では血管を正しく一つの物体として検出できているものの、改良モデルでは一つの血管に対して複数個として検出している例が複数存在する。この一つの血管を多重に検出する現象は Faster R-CNN にも一部存在するものの Optical Flow Network 導入モデルで多く観察できた。これはプローブを移動しながら肝臓内を撮影した際に、球状の腫瘍はその断面が基本的に円状でこの円のサイズが縮小・拡大変化が発生する。一方で血管は撮影角度や位置が変化すると血管の形状が大きく変わってしまう。これは例えば特定の断面では小さい楕円形で撮影されるが、血管の流れる方向で撮影されれば管上の長方形のような形となる。また、血管同士はつながっており、ある画像では血管が一つである場合でも次のフレームで隠れて繋がっている血管が表示される場合などである。このような同一物体にもかかわらず、連続フレームで形状が大きく変化する場合には今回導入した Optical Flow Network が悪影響を与え、多重に検出していると思わ

れる。この問題解決には入力データとなる 2 フレームと出力される Box 領域を拡張し、広い空間情報である複数のフレームから物体が存在する立方体のボリュームを推定するモデルを構築することが一つ考えられる。他に、教師データに大きく形状が変化するデータを追加し同一物体であることを学習させることが挙げられる。

3.5 まとめ

本章でははじめに 3.3 節で既存の物体検出モデルである Faster R-CNN を用いて、超音波検査映像から腫瘍及び血管の検出モデルを構築した。学習と評価用に利用したデータ数が一人の患者より得たものであるため今回構築したモデルが汎化性能を獲得したとは言えないが、同一の腫瘍及び血管に関して、学習したデータとことなる映像シーン出会っても IOU=50%においては 88.1%の高い予測精度で腫瘍と血管の検出を実現した。これにより Faster R-CNN 光学的な画像に対する物体検出だけでなく超音波検査画像への応用でも一定以上の精度で物体の検出・識別が可能であることが明らかとなった。

一方で今回構築した Faster R-CNN による腫瘍及び血管検出モデルは既存技術を利用したに過ぎずモデルの改良余地が存在する。また、あくまで本章では超音波検査映像に対して Faster R-CNN モデルが有効か検証するために限定的なデータに対して評価しており、CADe としての腫瘍検出モデルの実現にはより複数人から得たデータから評価が必要である。

次に本章では 3.4 節で実際の肝臓の超音波検査による診断で利用される情報である物体の連続例及び時間に伴う変化情報を補完することで、より肝臓の超音波検査に特化した物体検出モデルを構築する狙いのもと Optical Flow Network を Faster R-CNN に導入した。この導入により、従来では静止画情報のみを検出に利用していたが、連続した 2 フレームとそこから推定される変化の情報を統合し、物体の 3 次元情報を補足した特徴量を検出に利用するモデルを構築し、Faster R-CNN と同様の学習設定・学習データを使用して腫瘍検出タスクの学習を行った。

Optical Flow Network を導入した改良モデルと Faster R-CNN の評価データに対する meanAP では改良モデルがより高精度での領域検出を実現した。加えて、教師データ及び評価データに含まれない映像部分から 5 フレームを 1 シーンとして抽出した 30 シーンに対して腫瘍・血管の検出感度を評価した。改良モデルは腫瘍の検出に関してのみ優れた性能を示した。血管に対しては Faster R-CNN と同程度の性能であるが、血管を多重に検出する事例が複数存在することが確認された。肝臓内では肝静脈やグリソン鞘が複雑に絡み合っており、これらを実際に観察するシステムとして Optical Flow Network 導入モデルを利用することは、医師に誤検出の結果を多数示すこととなり、専門的知識の不足している利用者を混乱させる要因となる可能性がある。

最後に全体をまとめる。本章では複雑性を有する超音波検査映像から腫瘍・血管検出タスクに対する深層学習技術の有効性について確認することが一つの目的であった。それに対して 3.3 節の結果から単純な静止画ベースの Faster R-CNN であっても肝臓の超音波検査映像から腫瘍・血管検出タスクに対して mean AP=0.530 と予測精度が得られた。一方で、この Faster R-CNN は腫瘍の平均感度に対しては表 3.4 で示すように 0.456 となり時間経過に伴う腫瘍の変化に対して検出性能が低い結果となった。この問題は静止画ベースの Faster R-CNN を採用するにあたって予期していた問題であり、本研究ではこれに対して映像変化に対する検出性能向上を期待して時間変化情報である Optical Flow を Faster R-CNN に導入した。その結果、腫瘍に対する検出感度を 0.171 向上させ、平均感度 0.627 の結果を示した。これは多角的に撮影される肝臓の超音波検査において腫瘍支援を行う上で腫瘍の見逃し防止への有効な効果が期待できる。

4. 肝臓内物体の多種分類タスクの実現

4.1 背景・目的

3章では基礎的な物体検出モデルである Faster R-CNN と Optical Flow Network 導入による時間変化情報を補完したモデルを通じて、肝臓の超音波検査映像からの物体検出タスクという複雑なタスクに対する有効性について検証及び、診断支援システムの実現にあたってのベースとなる腫瘍・血管検出モデルの構築をおこなった。これは2種の物体のみを検出するものであった。3章の学習データとして作成した血管のラベルには肝静脈やグリソン鞘等の様々な管状の物体をすべて含めて血管としたものであり、実際に診断を行う際にこれらの詳細な種別情報が、肝臓の観察位置や腫瘍との関係を把握するために使用される情報となる。これら血管は複雑に絡み合い、また観察角度によってその形状や大きさが大きく変化する。これが3章でも述べたタスク自体の複雑性を構成する一つの要因であり、腫瘍・血管といった大きな分類でなくより詳細な種別で分類を実現する上で大きな課題となる。そこで本章ではさらなる詳細な分類を実現するため超音波検査画像から物体の超音波に対する散乱特性・減衰特性を抽出が可能な Power-Law Shot-Noise Model[Kutay 01]によって腫瘍・血管のさらなる細分化された分類が可能になると期待して導入する。また、未導入の場合と比較し超音波検査の多種分類タスクへの深層学習技術と Power-Law Shot-Noise Model の有効性を検証する。

細分化としては腫瘍・血管の2種から血管を細分化し、腫瘍と肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管の6種類に対しての分類タスクを実現する。検証には3章の Faster R-CNN をベースとした分類モデルを使用し、入力データとして既存の超音波検査向け特徴量抽出手法である Power-Law Shot-Noise Model と画像データを組み合わせたものを利用する。Power-Law Shot-Noise Model は超音波に対する組織の散乱特性・減衰率を特徴量の形で抽出する手法であり、単純な画像だけでは識別困難な腫瘍及び各種血管の差異を別の形で補完し、分類タスクに効果があると期待して導入する。Power-Law Shot-Noise Model はもともと乳がんが含まれる超音波検査画像の中で腫瘍と腫瘍外の組織の区別に利用することが可能であると提案された技術であるため導入にあたって Power-Law Shot-Noise Model 自体の適用可能性の検証も合わせて4.3節で行う。4.4節では Power-Law Shot-Noise Model によって生成された特徴量単体と組み合わせたとした際の深層学習による分類モデルと単純な深層学習モデルの性能を比較し、肝臓内の物体を多種類の物体として検出の実現性について検証する。

また、3章でモデル構築に使用したデータが一人の患者より得たデータであり、高性能なモデルの構築には不十分であった。そこで本章ではより多くの超音波検査映像を収集し

データの拡充も図り、腫瘍と肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管の6種類のラベリングしたデータを作成する。

4.2 Power-Law Shot-Noise Model

Power-Law Shot-Noise Model とは[Schottky 18]によって shot noise を表すために導入されたモデルで shot noise の振幅 $s(t)$ は次のようにモデル化される。

$$s(t) = \sum_i h(\Psi_i, t - t_i) \quad (4.1)$$

ここで t は観測される時間、 t_i は発生率 λ で決まる一様ポアソン過程に従って発生するランダムイベントが発生した時とのズレを示す。 Ψ_i はランダムな i 番目に発生したイベントによってパルスの振幅や位相ズレを特徴づけるランダムなベクトルである。 $h(\cdot, \cdot)$ はフィルタ関数となり次の2つ関数を用いる。

$$h_c(\Psi_i, t) = \begin{cases} a_i \cos(\varphi_i) t^{-\nu}, & 0 < t < B \\ 0, & \text{otherwise} \end{cases} \quad (4.2)$$

$$h_s(\Psi_i, t) = \begin{cases} a_i \sin(\varphi_i) t^{-\nu}, & 0 < t < B \\ 0, & \text{otherwise} \end{cases} \quad (4.3)$$

Ψ_i はランダムな振幅 a_i とランダムな位相 φ_i で構成されるベクトルである。また、 B は大きな固定値である。2つの関数は位相のズレがあるものの性質的な同じものである。代表として $h_c(\Psi_i, t)$ を利用した $s(t)$ の自己相関関数 $R_s(\tau)$ は次のように与えられる。

$$R_s(\tau) = E[s(t)s(t + \tau)] = E[s(t)]^2 + \lambda R_h(\tau) \quad (4.4)$$

$$\begin{aligned} R_h(\tau) &= E \left[\int_{-\infty}^{\infty} h_c(\Psi_i, t) h_c(\Psi_i, t + |\tau|) dt \right] \\ &= E \left[\int_{0^+}^{B-|\tau|} a^2 \cos^2(\varphi) t^{-\nu} (t + |\tau|)^{-\nu} dt \right] \\ &= 0.5E[a^2] \int_{0^+}^{B-|\tau|} (t^2 + |\tau|t)^{-\nu} dt \end{aligned}$$

ここで E は期待値を示し、 a と φ は互いに独立した乱数であり $E[\cos^2(\varphi)] = 0.5$ という事実を利用した。更に $s(t)$ のパワースペクトラム $S_s(f)$ は式(5.4)をフーリエ変換することで次のように求まる。

$$\begin{aligned}
S_s(f) = E[s(t)]^2 \delta(f) & \\
& + 0.5\lambda E[a^2] |\Gamma(1 - \nu, j2\pi f 0^+)|^2 \\
& - \Gamma(1 - \nu, j2\pi f B) |^2 (2\pi f)^{-\beta}
\end{aligned} \tag{4.5}$$

ここで $\Gamma(\dots)$ は不完全ガンマ関数で、 $\beta = 2(1 - \nu)$ である。 B が増加するにあたって、 $\Gamma(1 - \nu, j2\pi f B)$ は0に近づくため式(4.5)のパワースペクトラムは次の式に近づく。

$$S_s(f) = E[s(t)]^2 \delta(f) + 0.5\lambda E[a^2] \Gamma^2(\beta/2) (2\pi f)^{-\beta} \tag{4.6}$$

ここで $\Gamma(1 - \nu, j2\pi f 0^+)$ を完全ガンマ関数 $\Gamma^2(\beta/2)$ として近似している。この式よりパワースペクトラムは周波数全域に対して $1/f^\beta$ に従ったグラフとなり。そのためパワースペクトラムは冪乗則にしたがい、スケーリング指数 $-\beta$ によってグラフの形状が決まる。グラフは周波数とパワーに対してそれぞれlogをとった値を図示すると直線となり、このときの傾きが冪乗則指数 $-\beta$ となる。

次に超音波検査時に観測される反射波を Power-Law Shot-Noise Model を使用して表現を行う。反射波を超音波が伝わる領域に存在する散乱体にて非一様ポアソン過程にしたがって反射が発生し、その反射された波の集合として反射波 $r(t)$ は次のようにモデル化する。

$$r(t) = \sum_i h(t, t_i, x_i, \Psi_i) \tag{4.7}$$

ここで t は反射波が観測プローブにて観測される時間を示す。 t_i は発生率 $\lambda(t)$ で決まる非一様ポアソン過程に従って物体内で発生する反射パルス $h(\cdot)$ が観測された時間を示す。 x_i は i 番目に発生した反射地点とプローブとの距離を示す。 Ψ_i はランダムな i 番目に発生した反射地点での反射波の振幅を特徴づけるランダムな特徴ベクトルである。発生率 $\lambda(t)$ は超音波が伝わる空間に存在する物体の密度に関連して決まる値となる。微小空間においては超音波が伝わる物体は静止し、減衰特性や散乱体密度の特性が一定という過程のもこのモデルを次のように簡略化する。

$$r(t) = \sum_i h(t - t_i, \Psi_i) \tag{4.8}$$

各反射点で発生する反射パルス $h(t - t_i, \Psi_i)$ を次のように仮定する。

$$h(t - t_i, \Psi_i) = e(t - t_i, a_i) \cos(\omega_c(t - t_i) + \varphi_i) \quad (4.9)$$

$$= e_c(t - t_i, a_i) \cos(\omega_c t) - e_s(t - t_i, a_i) \sin(\omega_c t)$$

$$e_c(t - t_i, a_i) = e(t - t_i, a_i) \cos(\varphi_i - \omega_c t_i) \quad (4.10)$$

$$e_s(t - t_i, a_i) = e(t - t_i, a_i) \sin(\varphi_i - \omega_c t_i)$$

ここで t は観測された時間で、 t_i は i 番目に発生した反射に関連した時間と t のズレを示す。 ω_c は超音波検査に使用する送受波器によって生成される超音波の中心波長を示す。 $e(t - t_i, a_i)$ は反射パルスのエンベロープである。 a_i と φ_i は反射点の反射特性で決まる乱数である。 $e_c(t - t_i, a_i)$ と $e_s(t - t_i, a_i)$ は反射によって変調されたパルスの同相成分と直行成分となる。さらにエンベロープ $e(t - t_i, a_i)$ を次のように仮定する。

$$e(t - t_i, a_i) = \begin{cases} a_i(t - t_i)^{-\nu}, & 0 < t < B \\ 0, & \text{otherwise} \end{cases} \quad (4.11)$$

この時、式(4.11)を式(4.10)に代入すると式(4.2)、式(4.3)と類似した形となる。そのため、反射波 $r(t)$ の同相成分と直行成分はそれぞれ Power-Law Shot-Noise Model の式(5.6)と同じようにパワースペクトラムは冪乗則に従い、スケーリング指数 β によってグラフの形状が決まる。 $\beta = 2(1 - \nu)$ であり、散乱体の特性によって決まる値であり、 β が腫瘍の外と内で異なることから物体検出の特徴量として利用することが期待される。

この β を求めるには超音波検査によって観測された反射波 $r(t)$ から同相成分 $r_c(t)$ と直行成分 $r_s(t)$ のパワースペクトラムを求める必要がある。同相成分 $r_c(t)$ と直行成分 $r_s(t)$ は ω_c と観測された $r(t)$ を使用して次のように算出できる。

$$r_c(t) = \cos(\omega_c t) r(t) + \sin(\omega_c t) \bar{r}(t) \quad (4.12)$$

$$r_s(t) = \cos(\omega_c t) \bar{r}(t) - \sin(\omega_c t) r(t) \quad (4.13)$$

ここで $\bar{r}(t)$ は $r(t)$ をヒルベルト変換したものである。この2つの式のパワースペクトラムを算出し、両対数スケールで図示した際の傾きを求めることで冪乗則指数 $-\beta$ が求まる。

また、 $r(t)$ の ω_c は物体内を通過する中で物質の特性によって信号が変調され、送受波器によって生成される超音波の中心波長をそのまま利用ができない。そこで次のように物質の特性が同一の微小領域において発生した反射波 $r(t)$ のピリオドグラムを算出し微小領域での変調された中心周波数 \hat{f}_c を求める。

$$\hat{f}_c = \frac{\hat{\omega}_c}{2\pi} = \frac{\int f S_{nb}(f) df}{\int S_{nb}(f) df} \quad (4.14)$$

ここで $\hat{\omega}_c$ は変調された中心波長、 $S_{nb}(f)$ は変調された $r(t)$ のピリオドグラムである。

4.3 データセット

3章で物体検出モデルの構築に利用されていたデータは1名の患者から得られた超音波検査映像から生成されたものであった。このデータに追加で肝細胞癌を有する患者3名と正常な肝臓を有するドナー1名から得た術中超音波検査映像からデータを作成した。データのラベリングとしては今まで腫瘍と血管という大きな分類であったが、血管を細分化し腫瘍、肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管の6種類とした。ここでグリソン鞘は小葉間静脈、小葉間動脈、小葉間胆管、リンパ管、末梢神経を束ねる結合組織である。また、小血管は肝静脈、下大静脈、下行大動脈、グリソン鞘のような管状の他物体として判別が困難な血管を小血管として設定している。また3章では物体の存在する領域の最端を境界線としてBoxの形状としてデータを作成したがより詳細なデータ化及び将来的にMask R-CNNモデルのようなマスク領域での物体検知モデルへの利用を想定し、図5.1に示すマスク状でのラベリングを行った。作成したデータは864枚の画像と各画像に含まれるマスク領域とラベル情報でデータセットには連続したフレームが含まれているデータである。作成したデータのサンプルを図4.1に示す。

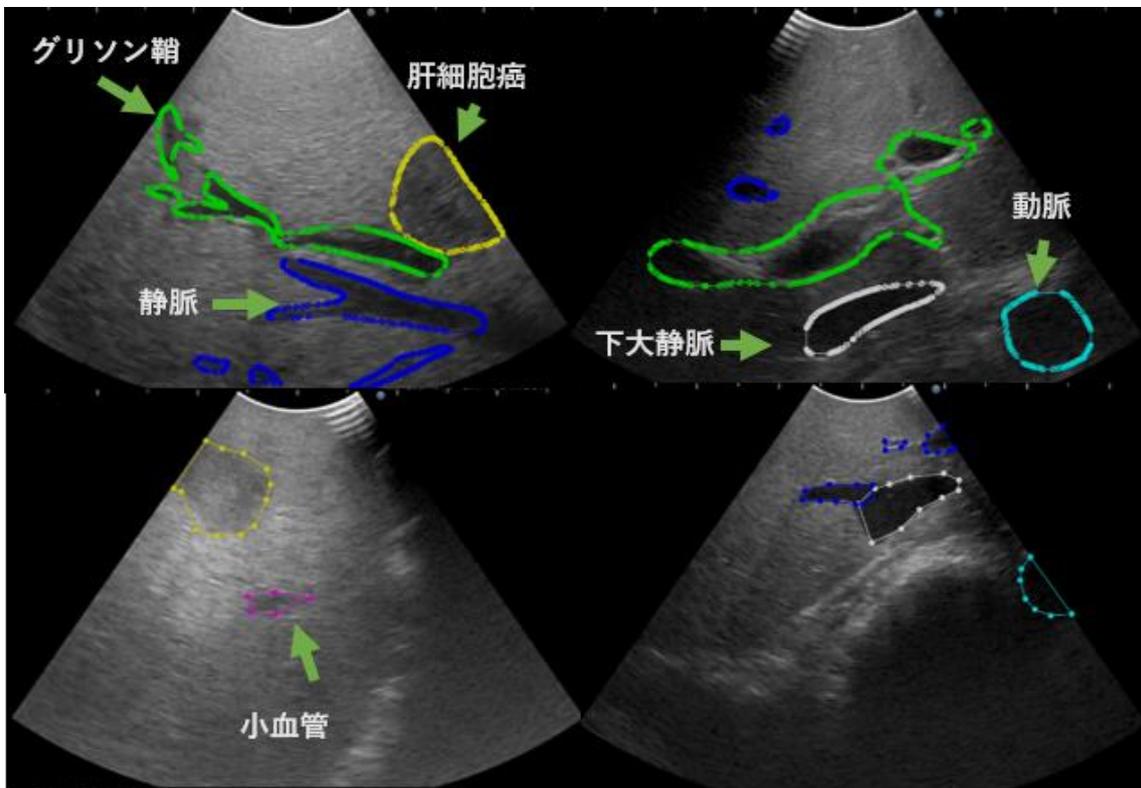


図 4.1 作成したマスク領域情報を含むデータ

このデータセットを Power-Law Shot-Noise Model にて利用する場合、超音波の進行方向の画素情報に容易にアクセスする小音が望ましい。現状は肝臓の超音波検査では扇状の形として画像が得られているため、図 4.2 に示すように扇形を長方形へと写像変換する操作を行い、超音波の進行方向が縦方向と同一になるようにする。画像に適用に変換と同様にラベリングを行ったマスク領域に関し関しても同様に変換を行う。本節ではこの 5 人分から作成した 864 枚のデータセットを利用する。

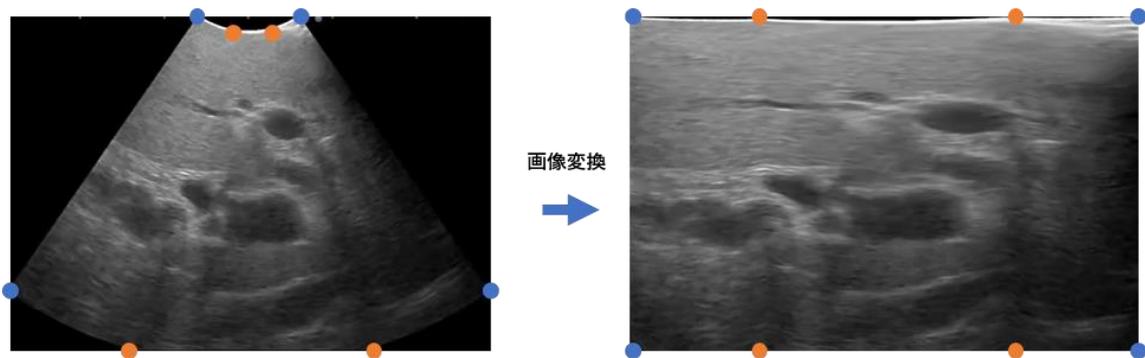


図 4.2 扇形から長方形への画像変換

4.4 Power-Law Shot-Noise Model の肝腫瘍超音波検査画像への適応可能性の検証

Power-Law Shot-Noise Model では画像から算出した超音波の反射信号 $r(t)$ のパワースペクトラムが冪乗則に従う前提のもと、そのパワースペクトラムの冪乗則指数を特徴量として腫瘍領域とその他領域の区別に利用する。本節では肝腫瘍の超音波検査画像に関しても算出した信号 $r(t)$ のパワースペクトラムが冪乗則に従うかを確認し、腫瘍、血管、肝組織における冪乗則指数 $-\beta$ を算出し、各物体の区別が可能かについて比較する。検証にあたっては 5.3 節のデータセットから腫瘍や血管を含む 10 枚の画像を選択し、同相成分 $r_c(t)$ 、直行成分 $r_s(t)$ 、エンベロープ $e(t - t_i, a_i)$ のパワースペクトラムを算出した。各パワースペクトラムの算出には画像中の $(x_1, y_1) = (300, 10)$, $(x_2, y_2) = (350, 410)$ で構成される領域で y 方向を反射信号 $r(t)$ と仮定し x が 300~350 の各ラインに置いてパワースペクトラムを算出した。使用したサンプル画像と抽出領域を図 4.3 に示す。

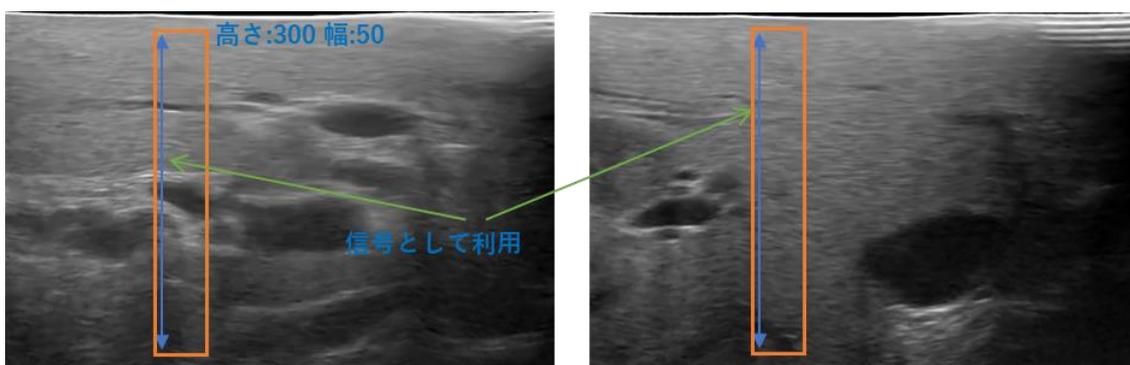


図 4.3 反射信号 $r(t)$ の抽出範囲

上記のもと $x = 300$ にて算出したパワースペクトラムのグラフを図 4.4 に示す。各パワースペクトラムはどれも高周波数領域においてノイズがあるものの 300Hz から 300000Hz において両対数スケールで図示した際に直線の形状を取る。また、10 枚の画像でそれぞれ 300~350 の各ラインで算出したパワースペクトラムをもとに計算した平均の冪乗則指数 $-\beta$ を表 4.1 に示す。画像ごとにばらつきがあるものの、同相成分の冪乗則指数が-1.7、直行成分の冪乗則指数-1.6、エンベロープの冪乗則指数-1.9 に近い値になることがわかる。以上により、Power-Law Shot-Noise Model を肝臓の超音波検査画像に対して適応した場合にもパワースペクトラムが冪乗則従うことがわかった。

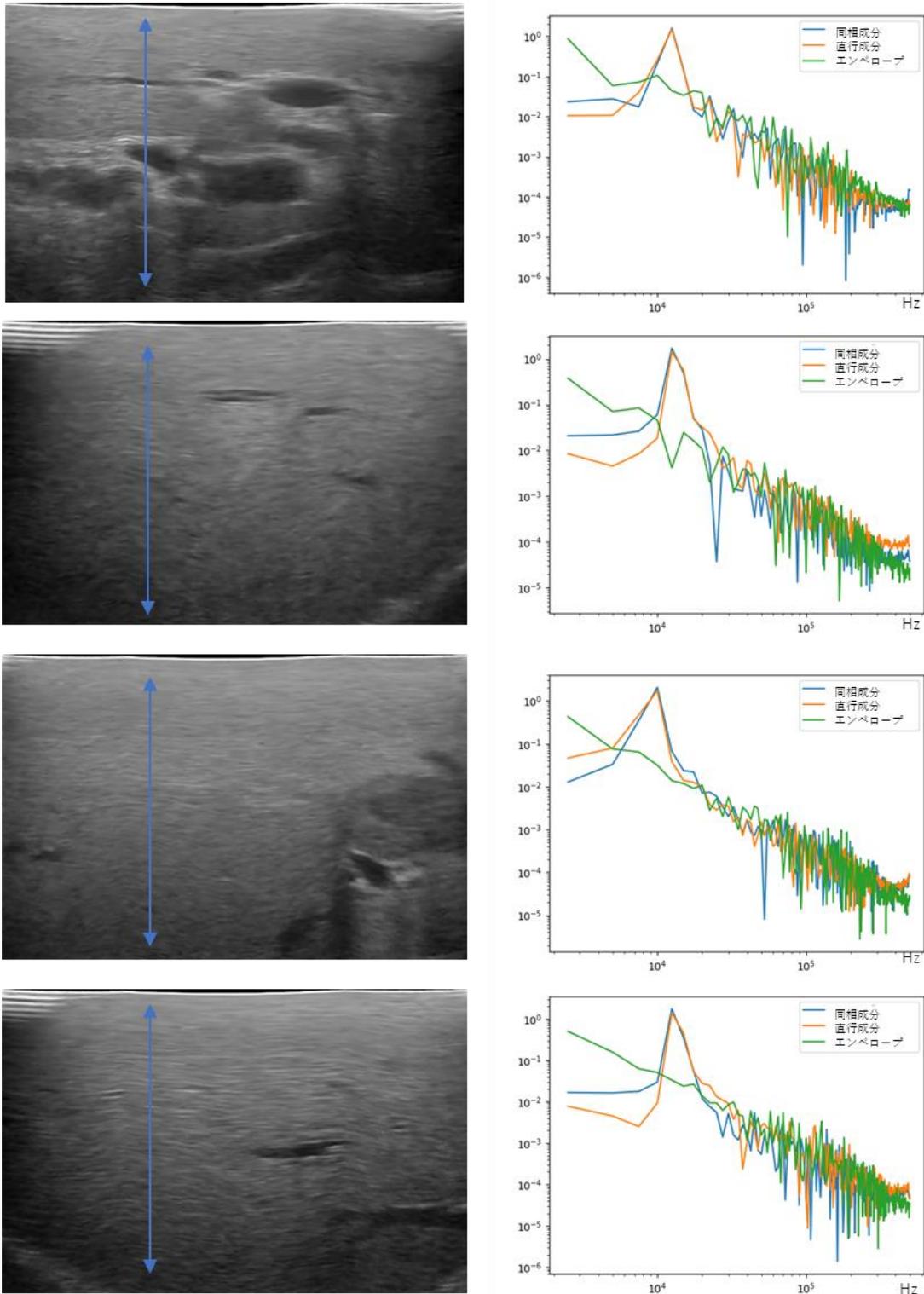


図 4.4 各画像における同相成分、直行成分、エンベロープのパワースペクトラム

表 4.1 各画像の同相成分、直行成分、エンベロープのパワースペクトラムの冪乗則指数

画像番号	同相成分	直行成分	エンベロープ
1	-1.68 ± 0.14	-1.43 ± 0.18	-1.67 ± 0.04
2	-1.68 ± 0.14	-1.42 ± 0.17	-1.73 ± 0.04
3	-1.76 ± 0.13	-1.49 ± 0.22	-1.82 ± 0.08
4	-1.78 ± 0.18	-1.54 ± 0.21	-1.76 ± 0.05
5	-1.69 ± 0.16	-1.41 ± 0.18	-1.73 ± 0.04
6	-1.74 ± 0.14	-1.57 ± 0.21	-1.86 ± 0.04
7	-1.86 ± 0.16	-1.69 ± 0.21	-1.86 ± 0.05
8	-1.77 ± 0.15	-1.62 ± 0.19	-1.92 ± 0.06
9	-1.78 ± 0.18	-1.65 ± 0.25	-2 ± 0.05
10	-1.78 ± 0.14	-1.62 ± 0.22	-2.02 ± 0.04

もともと Power-Law Shot-Noise Model は[Kutay 01]の中で乳がんとその他組織の区別することが可能であることが検証されていた。そこで次に、肝臓超音波検査画像に含まれる腫瘍、血管、肝臓組織の区別にもこの Power-Law Shot-Noise Model が適応可能かどうかを検証する。検証にあたっては図 4.5 示す 3 つの画像に含まれる腫瘍、血管、肝臓組織のそれぞれで同相成分の冪乗則指数、直行成分の冪乗則指数、エンベロープの冪乗則指数を算出する。各冪乗則指数の算出に使用した反射信号 $r(t)$ は図 4.5 に示している領域を使用している。各領域において、 x 座標ごとに算出した冪乗則指数の平均を表 4.2 に示す。

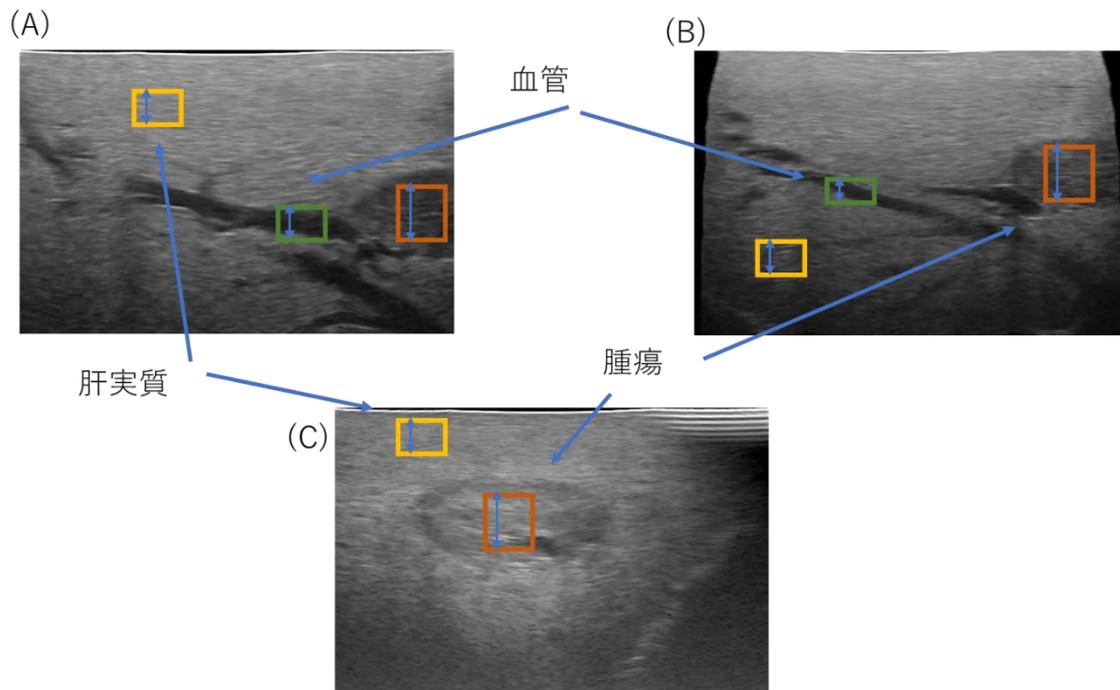


図 4.5 腫瘍、血管、肝実質の抽出領域

表 4.2 各抽出領域における同相成分、直行成分、エンベロープの冪乗則指数

対象物体	同相成分	直行成分	エンベロープ
血管 (A)	-2.75 ± 0.83	-2.11 ± 0.49	-1.84 ± 0.55
血管 (B)	-2.21 ± 0.87	-2.19 ± 0.84	-2.34 ± 0.89
腫瘍 (A)	-2.06 ± 0.66	-1.93 ± 0.87	-2.26 ± 0.47
腫瘍 (B)	-1.94 ± 0.70	-1.97 ± 0.63	-2.00 ± 0.26
腫瘍 (C)	-2.28 ± 0.55	-2.09 ± 0.55	-1.88 ± 0.41
肝実質(A)	-1.32 ± 0.87	-0.99 ± 0.56	-1.28 ± 0.23
肝実質(B)	-1.77 ± 0.67	-1.36 ± 0.67	-2.12 ± 0.39
肝実質(C)	-1.72 ± 0.68	-1.69 ± 0.67	-1.28 ± 0.30

今回、限定的な検証した結果では腫瘍、血管の同相成分の冪乗則指数、直行成分の冪乗則指数は基本的に-1.9より小さい値となった。肝臓組織における同相成分の冪乗則指数、直行成

分の冪乗則指数が-1.8 より大きく、肝臓組織とそれ以外では区別が可能であることがわかった。一方で血管と腫瘍では比較的腫瘍の同相成分・直行成分の冪乗則指数の平均が血管と比べて大きいものの標準偏差が大きく一概に単純な値だけでは区別が困難であることが明らかとなった。

4.5 肝臓内物体の多種分類タスクへの性能評価

4.4 節で Power-Law Shot-Noise Model によって物体の種別区別が可能かどうか検証を行ったが、腫瘍と血管の区別には単純な冪乗則指数の比較だけでは難しいことがわかった。そこで、本節では深層学習が持つ多次元のデータに対する優れた学習性能に期待し、Power-Law Shot-Noise Model によって得られた冪乗則指数を深層学習技術と組み合わせ、冪乗則指数と画像から統合的に特徴量を抽出した際の腫瘍、肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管の 6 種類分類タスクへの影響を検証する。検証として分類タスクを行うモデルは図 4.6 に示したモデルを使用している。このモデルは図 2.13 及び図 3.2 に示している Faster R-CNN をベースとして、そこから物体検出用に必要な処理を担っている RPN を除外し、分類タスクのみの機能を実装してのものである。また、従来の Faster R-CNN では画像データとその Box 領域を入力データとして利用していたが、本節では図 5.5 にあるように画像に加えて各ピクセルで算出した同相成分の冪乗則指数、エンベロープの冪乗則指数を入力として利用する。つまり、入力されるデータは画像のみ（3色を入力として3チャンネル）、冪乗則指数のみ（同相成分とエンベロープの冪乗則指数の2チャンネル）または画像と冪乗則指数を同時に（合計5チャンネル）の形式となる。また、入力される画像及び冪乗則指数は後述の抽出方法によって画像全体でなく物体の領域を含む一部画像のみが入力される。これに加えてノイズ付加またはガウシアンフィルタを適用させ、4.3 節のデータセットを使用し学習を行う。

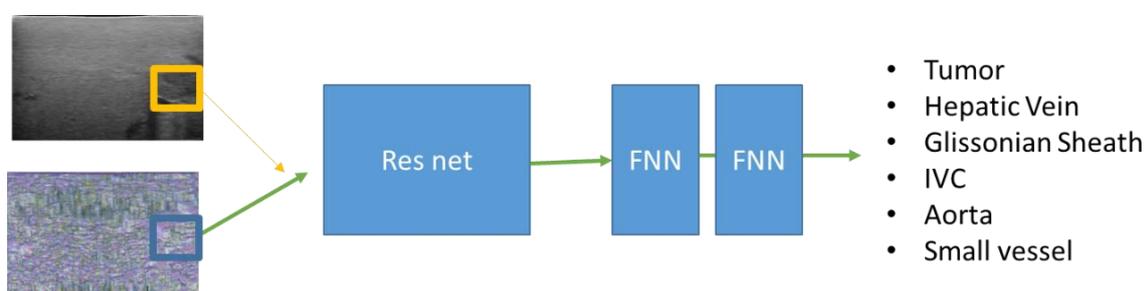


図 4.6 分類タスク向けネットワークモデル

・分類タスク検証用の学習・評価データセット

これまで、4.3 節で用意したデータセットには画像と画像中に含まれる物体の領域・ラベル

情報が含まれている。分類タスクを評価するにあたって、物体の領域部分の画像データとラベル情報のみを取り出し、学習・評価データとして使用する。物体領域の抽出するにあたって、物体の領域に含まれる最小x座標 x_{min} ・最大x座標 x_{max} 、最小y座標 y_{min} ・最大y座標 y_{max} とする場合、左上点 (x_{min}, y_{min}) と右下点 (x_{max}, y_{max}) を物体の Box 領域として使用すると物体のみの情報しか含まれず周辺情報がなく分類が難しくなることから、左上点 $(x_{min} - \frac{x_{min}}{4}, y_{min} - \frac{y_{min}}{4})$ と右下点 $(x_{max} + \frac{x_{min}}{4}, y_{max} + \frac{y_{min}}{4})$ で構成される物体だけでなく周辺情報が含まれた拡張領域を取り出し入力データと使用する。また、画像中の各座標 (x_i, y_j) における同時に同相成分とエンベロップの冪乗則指数の算出には次には (x_i, y_j) から $(x_i + 40, y_j)$ の 40 ピクセル分の画素値を $r(t)$ として式(6.12)を使用し算出したものを使用する。これによって生成されたサンプルデータを図 4.7 に示す。この抽出処理によって 4.3 節のデータセット 864 枚の画像を教師データ 691 枚と評価データ 173 枚をベースとして、実際に含まれる物体を各画像から抽出し、実際には教師データ 10,048 個、評価データ 2311 個が図 4.6 の分類モデルの学習・評価の際にはこれらの物体ごとにデータが利用される。

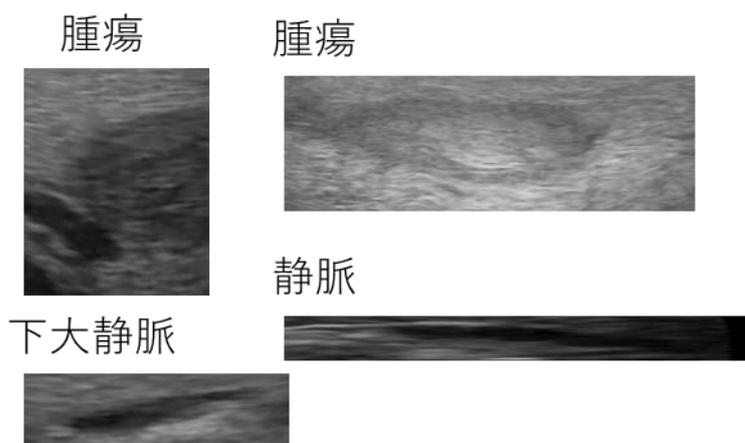


図 4.7 抽出された各物体のサンプル画像

・ノイズ付加

学習時、単純な画像データ・冪乗則指数だけではモデルが過学習する可能性があるため、データ水増し手法として画像・冪乗則指数に対してノイズを付加することで、学習の補助を行う。処理としてはデータ拡張確率 P_{DA} でノイズ付加では座標 (i, j) の画素値 $v_{i,j}$ に対して平均値 m_{noise} 、分散 s_{noise} の分布に従う乱数 $a_{i,j}^{noise}$ を付加する。また、ノイズ付加によって $v_{i,j}$ の各成分が 0 未満または 1 を超える場合には 0 または 1 として更新する。ノイズを付与したサンプルデータを図 4.8 に示す。

ノイズ付加後の教師データ

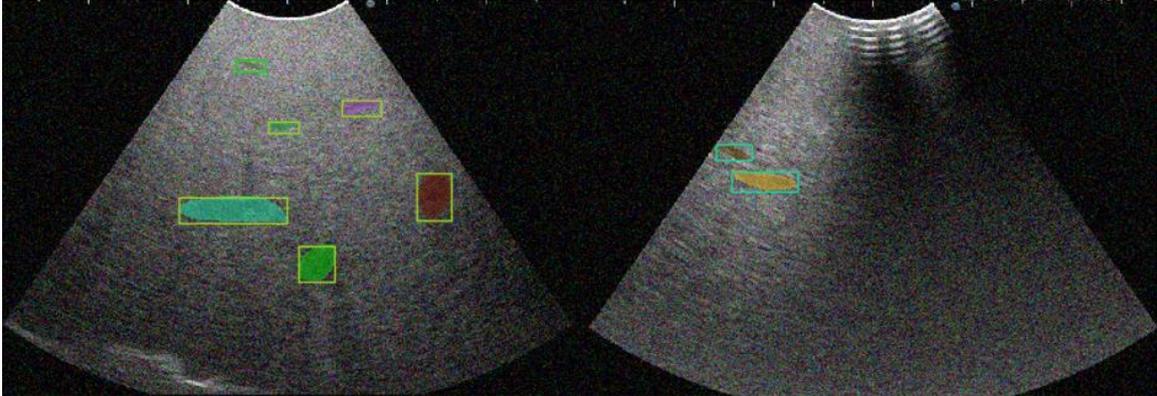


図 4.8 ノイズ付加処理

・ ガウシアンフィルタ

Power-Law Shot-Noise Model によって算出した冪乗則指数はノイズが多く含まれることが表 5.2 の各腫瘍、血管における標準偏差の大きさからわかる。そこで、算出した冪乗則指数のノイズを減らすことで分類タスクが容易になるのではないかと考え、ガウシアンフィルタを適用し入力される冪乗則指数を平滑化する。また、これは冪乗則指数にのみ適用し画像には処理しない。ガウシアンフィルタではある座標 (x_c, y_c) における入力値を次の式 $f(x, y)$ を重みとして、カーネルサイズ N_G として $x_c - \frac{N_G-1}{2} \leq x \leq x_c + \frac{N_G-1}{2}, y_c - \frac{N_G-1}{2} \leq y \leq y_c + \frac{N_G-1}{2}$ の範囲の入力値に対し下記重みを使用して合算させた値がフィルタ後の値となる。

$$f(x, y) = \frac{1}{2\pi\sigma_G^2} \exp\left(-\frac{(x-x_c)^2 + (y-y_c)^2}{2\sigma^2}\right) \quad (4.15)$$

ここで σ_G は平滑パラメータとなる。

座標 (x_c, y_c) における入力値が $I(x_c, y_c)$ とするとガウシアンフィルタを通したあとの出力 $O(x_c, y_c)$ は次式で表される。

$$O(x_c, y_c) = \sum_{x=x_c-\frac{N}{2}}^{x_c+\frac{N}{2}} \sum_{y=y_c-\frac{N}{2}}^{y_c+\frac{N}{2}} f(x, y) I(x, y) \quad (4.16)$$

モデル構築には表 4.3 に示す組み合わせの処理と入力データを使用して行う。また、各組み合わせでは表 4.4 に示されるパラメータを使用している。表 4.3 内のガウシアンフィルタ処理は冪乗則指数にのみ適用しており、画像に対して実施しないまた、ノイズ付与は画像、冪乗則指数の両方に対して処理を行っている。各組み合わせで 400epoch だけ学習させ、その中で得られた最良モデルを使用して算出した各ラベルに対する Accuracy を表 4.5 に示す。

表 4.3 モデル構築に使用した入力データ・処理の組み合わせ

Case	入力データ	水増し手法	画像処理
1	画像	-	-
2	画像	ノイズ付加	-
3	冪乗則指数	-	-
4	冪乗則指数	ノイズ付加	-
5	冪乗則指数	-	ガウシアンフィルタ
6	画像 冪乗則指数	-	-
7	画像 冪乗則指数	ノイズ付加	-
8	画像 冪乗則指数	-	ガウシアンフィルタ

表 4.4 使用した学習パラメータ

学習パラメータ	
学習 epoch 数	400
バッチサイズ	128
学習率 η	0.001
1epoch あたりの学習データ数	500
P_{DA}	0.5
m_{noise}	0.0
s_{noise}	0.1
N_G	5
σ_G	2.5

表 4.5 各ケースで構築したモデルのラベルごとの Accuracy

Case	Accuracy (%)							小血管 を除外	使用 epoch
	下大静脈	下行大動脈	小血管	グリソン鞘	腫瘍	肝静脈	全ラベル		
1	90.91	96.55	42.31	78.03	94.87	84.84	80.00	84.25	210
2	93.18	100	50.00	79.55	93.59	83.75	81.04	84.54	370
3	79.55	96.55	17.95	71.97	78.21	59.57	64.03	69.22	370
4	81.82	96.55	19.23	67.8	84.62	61.37	64.16	69.22	310
5	81.82	96.55	19.25	71.21	84.62	70.04	68.44	73.99	360
6	95.45	93.10	34.62	77.27	96.15	81.59	78.05	82.95	360
7	97.73	100	41.03	77.65	97.44	79.06	78.44	82.66	310
8	90.91	100	34.61	78.03	96.15	82.31	78.57	83.53	350

表 4.5 の結果より画像単体での分類精度は全体で Case2 の画像とノイズ付加の場合が最も高く 81.04%となった。また、冪乗則指数のみを分類タスクに利用した Case3,4,5 の場合、画像単体利用した Case1,2 と比較し全ラベルの Accuracy で 15%ほど精度が悪化した。下行大動脈以外の物体では冪乗則指数だけでは正確に分類できないことがわかった。また、冪乗則指数の平滑化を目的としてガウシアンフィルタを導入した Case5 は冪乗則指数のみ利用の Case3,4,5 の中で最も高い Accuracy を示し、冪乗則指数に含まれるノイズの除去に有効であることがわかった。一方で、画像と冪乗則指数両方利用した Case6,7,8 では画像のみの Case1,2 より全ラベルに対する Accuracy を超えることができなかったがノイズ付加、ガウシアンフィルタ処理は共に Accuracy を改善させる効果があった。しかしながら効果があるものは腫瘍、下行大動脈に関してのみで、肝静脈の分類で悪化し全体的には画像単体の Accuracy に及ばない結果となった。Case5,8 では同じガウシアンフィルタを冪乗則指数に適用しているが、Case8 では下大静脈に関して精度を改善することができていない結果であった。また計算コスト的な観点では冪乗則指数算出にあたってはピクセルごとに冪乗則指数を算出しており、リアルタイムでの画像処理に適していないため、リアルタイム性と性能面から見て Power-Law Shot-Noise Model が肝臓内の物体分類に適していないことが明らかとなった。

4.6 まとめ

本章ではこれまで学習・評価に使用していたデータを更新・拡張し、肝細胞癌患者 4 名・ドナー 1 名から取得した映像をもとにマスクデータを含む 864 枚のデータセットを作

成した。各画像に含まれる物体は腫瘍、肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管としてラベリングされており、合計で 12,359 個の物体が含まれる。

さらに、このデータセットをベースとして Power-Law Shot-Noise Model 導入による 6 種類の分類タスクへの有効性と単純な画像ベースでの分類性能について比較、検証した。検証の結果、2 種から 6 種の細分化されたタスクでは Power-Law Shot-Noise Model から抽出された冪乗則指数だけでは難しく最高でも 68.44%の Accuracy を示す結果となった。画像と冪乗則指数の組み合わせでは 78.57%の Accuracy を達成したが、これは画像単体利用した場合の結果の 81.04%に及ばない結果となった。この要因としては、ガウシアンフィルタによって性能が向上したことから Power-Law Shot-Noise Model によって計算される冪乗則指数自体にノイズが多く含まれており、これがニューラルネットワークのような学習データによって性能が大きく変わるモデルにとって難しいタスクであったと考えられる。ただし、分類の種別ごとに見れば冪乗則指数は下大静脈、腫瘍に関して Accuracy を改善させる効果が得られた。下大静脈、腫瘍はほか物体と比較して大きい物体で、これに有効であったのは大きい物体ほど抽出される冪乗則指数に含まれるノイズが低減するからだと思われる。逆に小さい物体は冪乗則指数の算出に利用されるピクセル値によって大きく変動してしまい、分類が難しくなると思われる、また、処理速度の面から考慮すると Power-Law Shot-Noise Model では各ピクセルでパワースペクトラム、直線回帰による冪乗則指数算出を行う必要があり、単純な画像処理にとどまらない計算コストが発生する。そのため、診断支援システムを実現する上でリアルタイム性の維持が困難であることが明らかになった。

以上の結果、本章では Power-Law Shot-Noise Model 導入による肝臓の超音波検査画像から各種物体の多種分類タスクの実現を目指した。しかしながら、物体の超音波に対する散乱・減衰特性を抽出した冪乗則指数を用いるより、単純に画像のみを利用したほうが優れた Accuracy を実現し、既存の深層学習技術のみでも多種分類が実現できると明らかとなった。ただし、Power-Law Shot-Noise Model は計算コスト面で診断支援システムに利用することが難しいが、冪乗則指数によって超音波検査画像に含まれる腫瘍や下大静脈、下行大動脈といった大きい物体に対する分類性能の向上に有効であることが明らかとなった。

5. 肝臓超音波検査向け扇形回転トリミングによるデータ多 様性の確保

5.1 背景・目的

3,4章では超音波検査映像という複雑性を有するタスクの実現にあたって、深層学習技術及び情報の補完を通じて、腫瘍・血管検出の評価・検証をおこなった。本章ではもう一つの課題となる肝臓の超音波検査映像を学習時のデータ多様性の課題について取り扱う。本来深層学習技術によるデータ学習を通じたモデルの実現には膨大なデータを確保し、それをもって学習データに含まれる物体の多様性を確保することで汎化性能を有するモデルを実現できる。肝臓の超音波検査映像はCT、MRI、X線画像といった特定の角度でしか撮影利用されないものと異なり、検査者の操作によって撮影視点が変化するものである。また、腫瘍の学習にはそのパターンを網羅するだけのデータ収集が要求されるが、データ収集には患者の協力が不可欠で、またデータのラベリングには熟練医師の協力も要求される。このようにデータ作成だけでもコストが要求するタスクとなる。

本章では限られたデータの中で診断支援システムとして利用に叶うモデルの実現のため、超音波検査の操作によって発生する映像変化に着目して肝臓超音波検査映像向けデータ水増し手法として扇形回転トリミングを5.3節で提案し多様性確保を目指す。これと合わせて、5.4節で通常の光学的に撮影された画像向けの一般的なデータ水増し手法を使用してデータ多様性確保を行う。5.5節では扇形回転トリミングと既存の一般的な画像向けデータ水増し手法を比較し提案手法の有効性について評価する。

5.2 データセット

3章で構築してきた腫瘍検知モデルの学習には1名の患者から得られた超音波検査映像を使用していた。本章では4章で作成した5名から作成した肝細胞癌を有する患者3名と正常な肝臓を有するドナーから得た術中超音波検査映像を教師データ・評価データに使用する。各物体のラベリング種別として肝臓内の主な組織を網羅するため、腫瘍、肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管の6種類を新しく検出対象とする。このデータセットは864枚の画像と各画像に含まれるマスク領域とラベル情報でデータセットには連続したフレームが含まれているデータである。腫瘍検知モデルの構築にあたり学習に使用するデータに依存してモデルの検出精度が変化してしまう。また、学習しすぎた場合に特定のデータに対してのみ反応し、逆に検出性能が悪化する可能性がある。このため、連続したフレームが含まれているデータセットを5つパターンの教師データとテストデータにランダム分

割し、各パターンにて学習を行う。また、学習ステップごとにテストデータに対する評価を行い。過学習やデータに対する学習度合いを分析する。各パターンには教師データが 691 個、テストデータが 173 個含まれる。この過程を通して得られて最良モデルに対して教師データが 691 個、テストデータが 173 個に含まれてない映像箇所から作成した 120 枚の評価データに対して性能を評価する。また、未学習データに対する性能評価を行うために上記の 5 名とは異なる 2 名の肝腫瘍患者と 1 名のドナーから得た映像から作成した 120 枚分のデータに対する評価も行う。

5.3. 肝臓超音波検査向け扇形回転トリミングの提案

5.3.1 提案背景

これまで 3,4 章では Optical Flow 導入による深層学習モデルのカスタマイズや Power-Law Shot-Noise Model による特徴量の合成を通じて、検出モデルの性能向上を目指し、性能向上自体は実現したが十分でなかった。そこで本章ではこれまでモデル側のカスタマイズではなく、データ側のカスタマイズを行うことで肝臓の超音波検査映像に特化した腫瘍検出モデルの実現を目指す。

一般的な画像向けに提案・利用されてきた 5 つのデータ水増し手法「左右反転処理、色相変化、ノイズ付加、Random erasing、Random Resize Crop」が存在するが、これらは超音波検査特有の画像変化に対応せず多様性の増加を実現できない懸念が存在する。本章では腫瘍検出の汎化性能向上を目的として、肝臓の超音波検査時の医師の操作に注目し、その操作によって発生する映像的变化を再現することで腫瘍のデータ多様性を向上させる新しいデータ水増し手法を提案する。肝臓に対してスクリーニング検査等によって肝臓全体を撮影する際、すなわち、肝臓全体を観察し腫瘍等の病変を発見しようとする際、超音波検査の撮影視点が超音波検査のプローブを中心に大きく回転変化する。この視点変化による腫瘍または血管の一部隠れる状態のデータを画像処理等によって増やすことが実現できれば、腫瘍だけでなく血管の認識汎化性能の向上ができ、腫瘍検出及びシステム利用者の肝臓内物体の把握に寄与できると考えた。

5.3.2 扇形回転トリミング

本節の性能評価及びモデル学習には 5.2 節で示した 5 パターンの学習データとテストデータ、学習データとテストデータに含まれない領域から作成した評価データ及び、取得元が異なる未学習評価データと同じものを使用する。5 パターンの学習データにてモデルを学習する際、次で示す扇形回転トリミング処理を利用する。

- ・扇形回転トリミングのアルゴリズム

扇形回転トリミング処理では肝臓の超音波検査実施時に行われる医師の操作に着目したものである。基本的に映像は超音波検査機のプローブによってデータが撮影されるが、肝臓全体や腫瘍周辺の形状を把握するため、プローブを基準に映像が変化する。また、映像の撮影範囲が扇形であることにも注目して扇形回転トリミングを構築している。視点変化が左右方向に起こった場合、映像は画像の中心ではなく映像の $-y$ 軸領域に存在するプローブを中心として映像が回転する。回転に伴って映像中の物体は扇形に切り取られる。これを再現することで、肝臓の超音波検査映像特有の映像変化に対応し、より物体の特徴量抽出性能が向上すると期待し手法を実装した。扇形回転トリミング処理の流れとしては 3 ステップに分かれる。

1. 回転処理の中心点 (x_c, y_c) の決定
2. 回転角度 θ の決定
3. 扇形トリミング領域を中心点 (x_c, y_c) 、回転角度 θ に従い変化させ、画像をトリミングする

ステップ 1 では入力画像の幅 W 、高さ H を利用し基本中心点 $(Wx_{c_base}, Hy_{c_base})$ を基準として、中心点変化量 x_d, y_d を利用して下記のように中心点 (x_c, y_c) を決定する。

$$x_c = \text{round}(Wx_{c_base} + Wx_d) \quad (5.1)$$

$$y_c = \text{round}(Hy_{c_base} + Hy_d) \quad (5.2)$$

ここで x_d と y_d は最大変化量 x_{d_max} 、 y_{d_max} によって決まる領域 $-x_{d_max} < x_d < x_{d_max}$ 、 $-y_{d_max} < y_d < y_{d_max}$ から得た一様乱数となる。また round 関数は四捨五入処理であり、中心点 (x_c, y_c) は画像ごとに微小に変化する点座標となる。

ステップ 2 では回転角度 θ を平均値 θ_{ave} 、標準分散 σ に従う確率分布をもとに決定する。

ステップ 3 では回転処理の中心点 (x_c, y_c) と回転角度 θ に従って画像を回転させる。6 点の座標 $(W(0.5 - x_1), Hy_1)$ 、 $(W(0.5 + x_1), Hy_1)$ 、 $(W(0.5 - x_2), Hy_2)$ 、 $(W(0.5 + x_2), Hy_2)$ 、 $(W(0.5 - x_3), Hy_3)$ 、 $(W(0.5 + x_3), Hy_3)$ から決定される扇形（ダイヤモンド型）の領域を切り取り新たなデータとして使用する。この処理は画像だけでなく物体の領域に対しても同様に行う。上記処理による画像処理の流れを図 5.1 に示す。

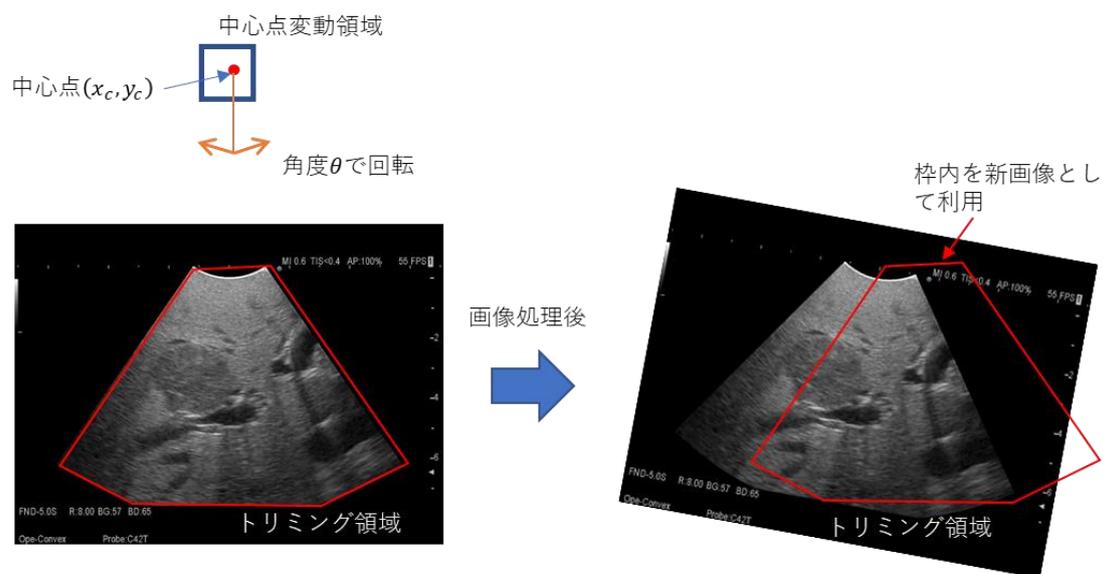


図 5.1 扇形回転トリミングによる画像処理の流れ

5.3.3 Mask R-CNN の構成・学習設定

検出モデルには Mask R-CNN を使用する。Mask R-CNN のベースとなる構成は 3 章で構築した Faster R-CNN と同一である。図 5.2 に示すように Mask R-CNN では RPN で提案された BOX 状領域を受け取り、ピクセルごとのそのピクセル上に物体があるかないかを予測するニューラルネットワークの MP を追加している。また、Mask R-CNN では腫瘍、肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管の 6 種類と背景の合計 7 種類を分類するため CN の最終層にはカーネルサイズ $1 \times 1 \times 7$ の CNN を用いる。MP と CN 以外の各ニューラルネットワークの構成は 3 章の Faster R-CNN と同様である。

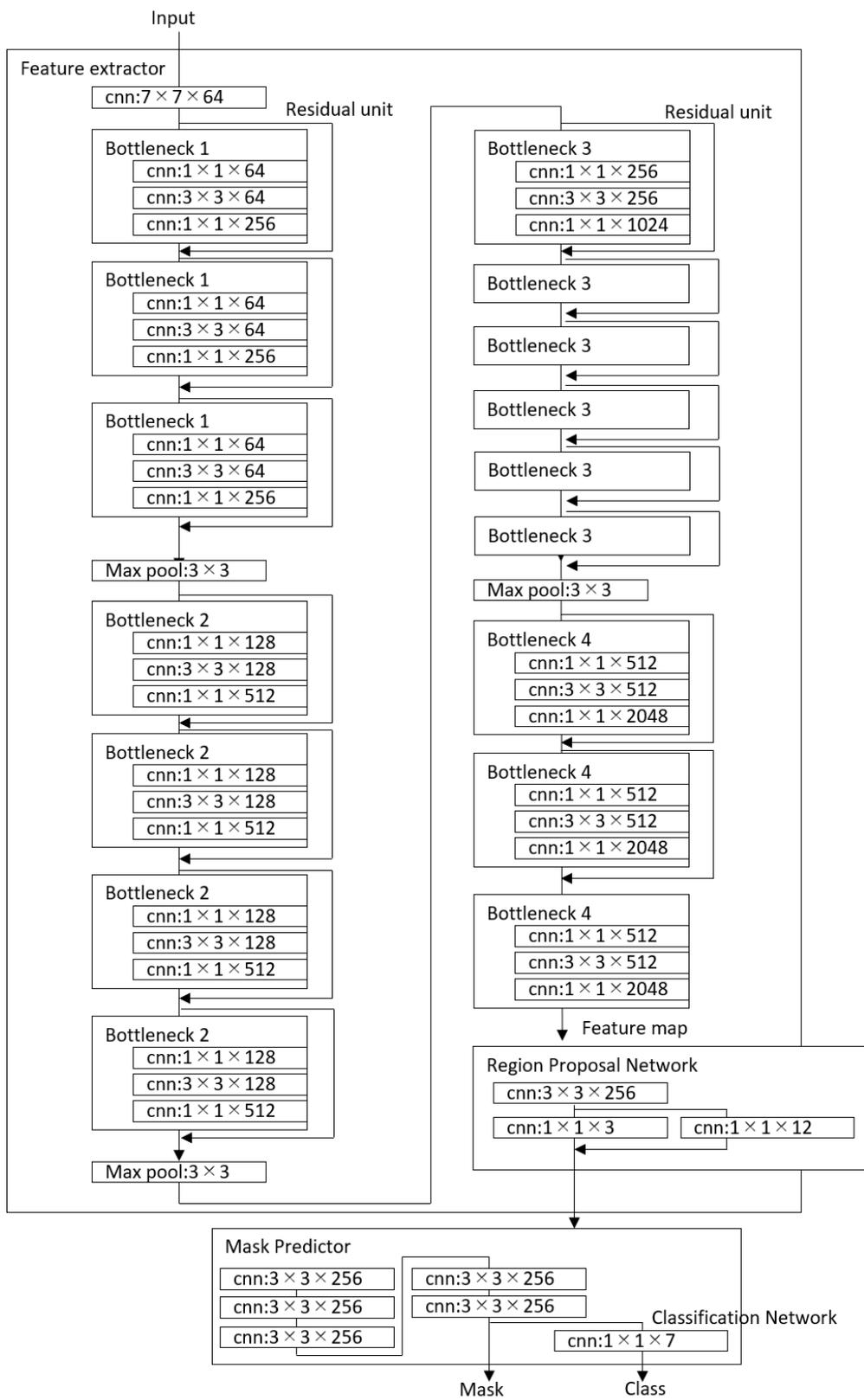


図 5.2 Mask R-CNN のニューラルネットワーク構成

図 5.2 の構成の Mask R-CNN を使用して、新規に提案した扇形回転トリミング手法による性能向上を検証する。学習には 5 パターンの教師データ・テストデータを利用して、表 5.1 に示したパラメータのもと学習を行った。基本的な学習 epoch 及び学習率は表 5.2 に示す値を使用する。今回学習に使用するデータが多いため 1epoch 毎に全教師データを使用して学習を行うと学習がニューラルネットワークの各パラメータの更新が急速に進むことで過学習と呼ばれる現象が発生する。過学習では教師データに対しては優れた性能を示すものの未学習データに対してはうまく認識ができない状態を指す。これを防ぐため、1epoch あたりの学習に使用するデータを制限し、ゆっくり学習を行う。1epoch 中に使用する教師データは epoch 毎に教師データの中からランダムに選択する。また、この扇形回転トリミングは教師データを入力する際にのみ使用しており、テストデータ等の性能評価時は使用していない。また、扇形回転トリミングによって生成された画像に物体が含まれない場合（物体の領域がない場合）オリジナルのデータが教師データとして使用される。

表 5.1 使用した扇形回転トリミングのパラメータ値

パラメータ	
(x_{c_base}, y_{c_base})	(0.5, -0.5)
(x_{d_max}, y_{d_max})	(0.01, 0.01)
θ_{ave}	0
σ	$\pi/6$
(x_1, y_1)	(0.08, 0.12)
(x_2, y_2)	(0.4, 0.75)
(x_3, y_3)	(0.16, 0.9)

表 5.2 使用した学習パラメータ

学習パラメータ	
学習 epoch 数	300
バッチサイズ	8
学習率 η	0.001
1epoch あたりの学習データ数	200

5.3.4 テストデータに対する検出性能

5 パターンのデータセットそれぞれで扇形回転トリミングを適用して学習させた Mask R-CNN を対応するパターンのテストデータに対して適用し epoch 毎に meanAP を算出したものを図 5.3 に示す。加えて、全 epoch 中で最も優れた meanAP を示したモデルの epoch 数とその時の meanAP を表 5.3 に示す。

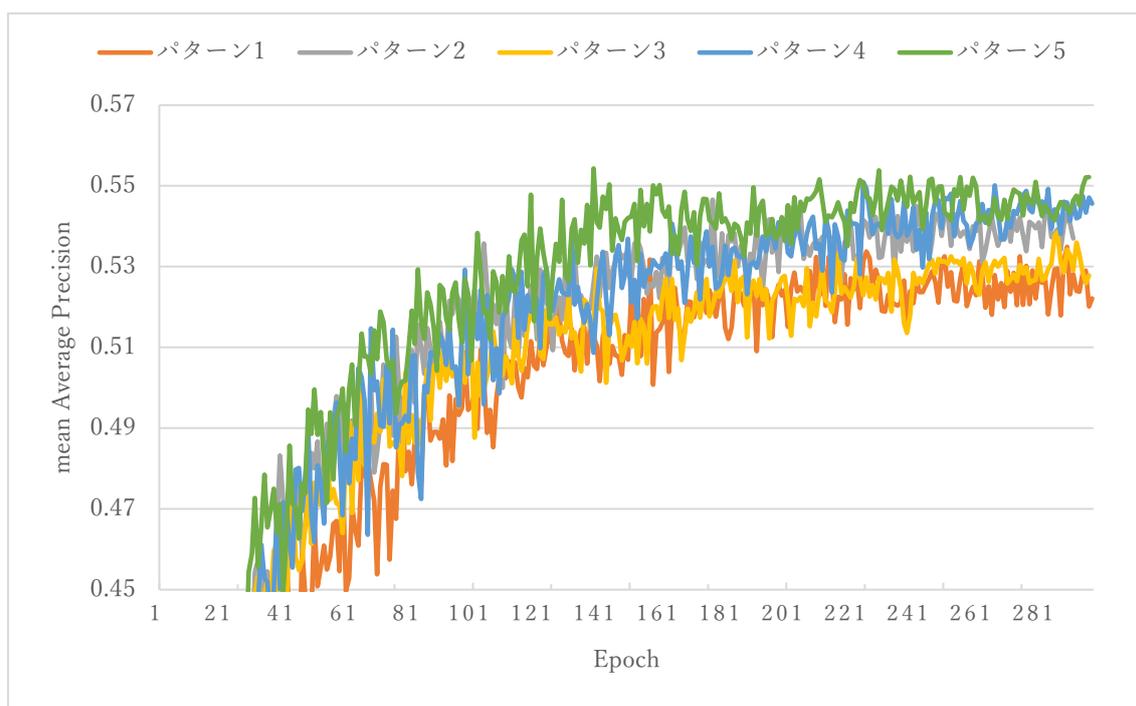


図 5.3 扇形回転トリミング使用時の 5 パターンデータセットに対する学習結果

表 5.3 各パターンで得られた最良モデル

学習データパターン	Epoch	meanAP
パターン 1	291	0.535
パターン 2	180	0.547
パターン 3	289	0.538
パターン 4	225	0.551
パターン 5	140	0.554

5.3.5 評価用データに対する検出性能

5.3.4 節では連続シーンが含む比較的類似したデータであるテストデータに対する検出結果を示した。本節では教師データとテストデータに使用されていない映像箇所から作成した評価データに対する扇形回転トリミング手法によって 5 パターンの学習から得られた最良モデルの性能を評価する。評価データは 5.2 節で示した評価データ 120 枚のデータであり、取得元の個人は教師データとテストデータに使用されている同じ 5 名であるが、異なるシ

ーンから得たデータである。5.3.4 節で扇形回転トリミング処理を導入し構築した各モデルから表 5.3 の各パターンの最良モデルを使用してラベルごとの meanAP、Average Recall を算出したものを表 5.4、表 5.5 に示す。

表 5.4 評価データに対するラベルごとの mean Average Precision

mean Average Precision					
Label	パターン 1	パターン 2	パターン 3	パターン 4	パターン 5
腫瘍	0.681	0.684	0.677	0.675	0.671
肝静脈	0.240	0.230	0.247	0.231	0.240
グリソン鞘	0.219	0.218	0.215	0.219	0.234
小血管	0.357	0.346	0.359	0.338	0.359
下大静脈	0.459	0.456	0.381	0.453	0.460
下行大動脈	0.636	0.647	0.626	0.648	0.670
全体	0.430	0.416	0.411	0.400	0.435
Average Precision IoU=50%					
腫瘍	0.977	0.975	0.969	0.968	0.969
肝静脈	0.621	0.58	0.609	0.603	0.576
グリソン鞘	0.554	0.583	0.557	0.559	0.607
小血管	0.772	0.766	0.772	0.768	0.772
下大静脈	0.872	0.899	0.813	0.87	0.87
下行大動脈	0.919	0.914	0.911	0.934	0.926
全体	0.772	0.757	0.749	0.732	0.765

表 5.5 評価データに対するラベルごとの Average Recall

Average Recall					
Label	パターン 1	パターン 2	パターン 3	パターン 4	パターン 5
腫瘍	0.711	0.721	0.72	0.729	0.707
肝静脈	0.348	0.395	0.382	0.375	0.397
グリソン鞘	0.298	0.341	0.298	0.325	0.34
小血管	0.396	0.393	0.4	0.381	0.396
下大静脈	0.515	0.515	0.43	0.505	0.515
下行大動脈	0.69	0.7	0.685	0.695	0.705
全体	0.505	0.522	0.497	0.513	0.522

表 5.4 から mean Average Precision ではパターン 5、Average Precision IoU=50%ではパターン 1 が最も高い結果を示した。性能については 5.5 節で一般的なデータ水増し手法と比較して評価する。

5.3.6 未学習データに対する検出性能

5.3.4 節及び 5.3.5 節では各モデルを映像取得元が同じ、つまり同一の 5 名から得た映像で学習及び評価を行い、映像変化再現による多様性増加の実現ができた。一方で学習データに含まれない個人から得たデータに対しても映像変化の再現によるデータ拡張による検出性能への影響を評価するため本節では 5.2 節でしめた教師データの 5 名とは異なる 2 名の肝腫瘍患者と 1 名のドナーから得た映像から作成した 120 枚分のデータに対する評価を行う。データはそれぞれ 1 名ごとに 40 枚のデータを作成しておりラベルは従来のもと同様に腫瘍、肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管として行った。表 5.4 で優秀な精度を示したパターン 1 とパターン 5 の教師データに対して扇形回転トリミングを適用して学習したモデルの 2 つを使用し算出した meanAP、Average Recall を表 5.6、5.7 に示す。

表 5.6 未学習データに対するラベルごとの mean Average Precision

mean Average Precision		
Label	パターン 1	パターン 5
腫瘍	0.080	0.065
肝静脈	0.099	0.094
グリソン鞘	0.094	0.097
小血管	0.000	0.002
下大静脈	0.361	0.300
下行大動脈	0.365	0.376
全体	0.161	0.147
Average Precision IoU=50%		
腫瘍	0.201	0.162
肝静脈	0.206	0.195
グリソン鞘	0.290	0.292
小血管	0.000	0.006
下大静脈	0.778	0.633
下行大動脈	0.532	0.539
全体	0.316	0.282

表 5.7 未学習データに対するラベルごとの Average Recall

Average Recall		
Label	パターン 1	パターン 5
腫瘍	0.100	0.084
肝静脈	0.204	0.223
グリソン鞘	0.218	0.225
小血管	0.000	0.031
下大静脈	0.415	0.352
下行大動脈	0.390	0.390
全体	0.226	0.222

表 5.6 のパターン 1、5 では特にパターン 1 が優れた性能を示した。性能については 5.5 節で一般的なデータ水増し手法と比較して評価する。

5.4 比較対象として既存画像データ水増し手法の導入

5.4.1 データ水増し手法導入背景

肝疾患の診断支援システムを構築する上でデータ多様性の確保はモデルの性能に大きく影響する要因となる。本節では一般的なデータ水増し手法の使用を通じて学習データの多様性を増加させ、超音波検査画像に対する効果を評価する。ここで水増し手法とは教師データとなる画像に対して何らかの処理を行い、合成画像であるものの現実においてあり得るようなデータを作成することで、これを新たな教師データとして利用しデータの多様性を拡張するものである。一般的に画像の反転、ノイズ付加といった画像処理がこの水増し手法にあたり、画像向け水増し手法として使用されている。またこの水増し手法はタスクによって向き、不向きが存在し、例えば 1~9 の手書き文字認識タスクでは反転のような文字の形状を全体的に変化させる手法は認識精度と悪化させることが知られている。そのため、本章では肝臓の超音波検査映像からマスク領域での物体検出モデルの構築するにあたって一般的な画像向けデータ水増し手法による肝臓の超音波検査画像に対する有効性について検証を行う。

5.4.2 各種データ水増し手法

教師データを用いたモデルの学習時に画像及びマスクデータに対してデータ水増し手法を適用し、学習データに含まれるデータの多様性を向上させる。超音波検査映像に適用する

にあたり一般的な光学的画像向けに使用される左右反転処理、色相変化、ノイズ付加、Random erasing[Zhong 20]、Random Resize Crop[Pytorch 17]といった画像処理を行い、効果を比較する。

- ・左右反転処理

図 5.4 に左右反転処理を行った教師データを示す。図のように画像を右と左を反転させる処理を行う。腹部超音波検査、術中超音波検査によって取得される映像は図のように扇形の形状で領域内を観察することができ、左右反転によってプローブの向きによる依存を無視し、腫瘍、血管の形状を認識することを期待し水増し手法として採用した。上下反転に関しては超音波検査では現実にならないため、本章では採用しない。また、この左右反転処理は学習時に常に反転が行われるわけではなく、データ拡張確率 $P_{DA} = 0.5$ の 50%の確率で処理が適応される。

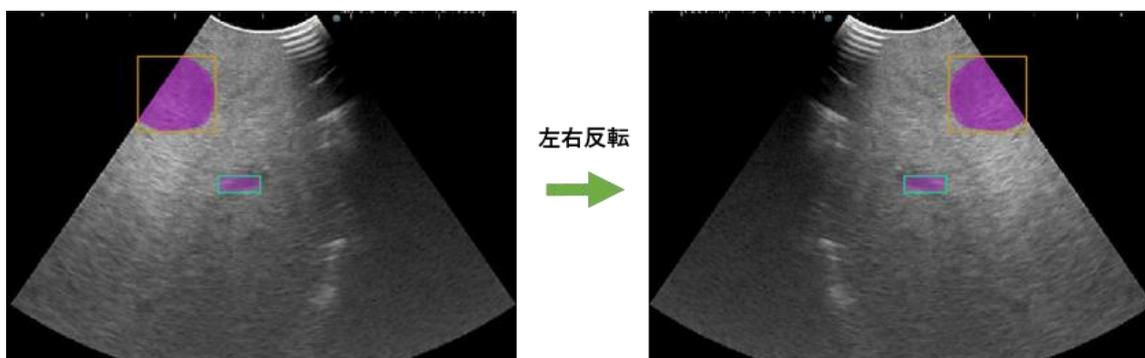


図 5.4 左右反転処理

- ・色相変化

色相変化においては画像の輝度、コントラスト、彩度、色調にたいしてそれぞれ対し微小変化させて新しい画像を作成する。図 5.5 に示すように画像の色調等がランダムに変化する。輝度、コントラスト、彩度の変化は、サイズ $N \times M$ の画像において座標 (i, j) のピクセルに対して操作 $type$ のベース値 $b_{i,j}^{type}$ とオリジナルの画素値 $v_{i,j}$ を使用し、次のように画素値 $v_{i,j}^{new}$ を更新する。

$$v_{i,j}^{new} = b_{i,j}^{type}(1 - \alpha_{type}) + \alpha_{type}v_{i,j} \quad (5.2)$$

$$v_{i,j} = (R_{i,j}, G_{i,j}, B_{i,j}) \quad (5.4)$$

ここで α_{type} は輝度 (brightness)、コントラスト (contrast)、彩度 (saturation) に対応する乱数である。 $R_{i,j}, G_{i,j}, B_{i,j}$ は各座標における $[0, 1]$ の RGB カラーの値である。 α_{type} は各操作の乱数変化幅 w_{type} とすると区間 $1 - w_{type} \leq \alpha_{type} \leq 1 + w_{type}$ の一様乱数となる。各操作のベース値 $b_{i,j}^{type}$ は下記式で表現される。

$$\mathbf{b}_{i,j}^{brightness} = (0, 0, 0) \quad (5.5)$$

$$\mathbf{b}_{i,j}^{contrast} = (b^c, b^c, b^c)$$

$$b^{contrast} = \frac{1}{NM} \sum_{i=0}^N \sum_{j=0}^M \frac{R_{i,j}, G_{i,j}, B_{i,j}}{3} \quad (5.6)$$

$$\mathbf{b}_{i,j}^{saturation} = \left(\frac{R_{i,j}, G_{i,j}, B_{i,j}}{3}, \frac{R_{i,j}, G_{i,j}, B_{i,j}}{3}, \frac{R_{i,j}, G_{i,j}, B_{i,j}}{3} \right) \quad (5.7)$$

色調(hue)の変化操作に関しては RGB フォーマットの画素値($R_{i,j}, G_{i,j}, B_{i,j}$)を HSV 色空間フォーマットの($H_{i,j}, S_{i,j}, V_{i,j}$)へと変換し、 $[0, 255]$ の $H_{i,j}$ に対して下記のように新たな色相情報 $H_{i,j}^{new}$ を合成する。

$$H_{i,j}^{new} = H_{i,j} + 255 \cdot \alpha_{hue} \quad (5.8)$$

ここで α_{hue} は他の操作と異なり区間 $-w_{hue} \leq \alpha_{hue} \leq w_{hue}$ の一様乱数となる。

色相変化後の教師データ

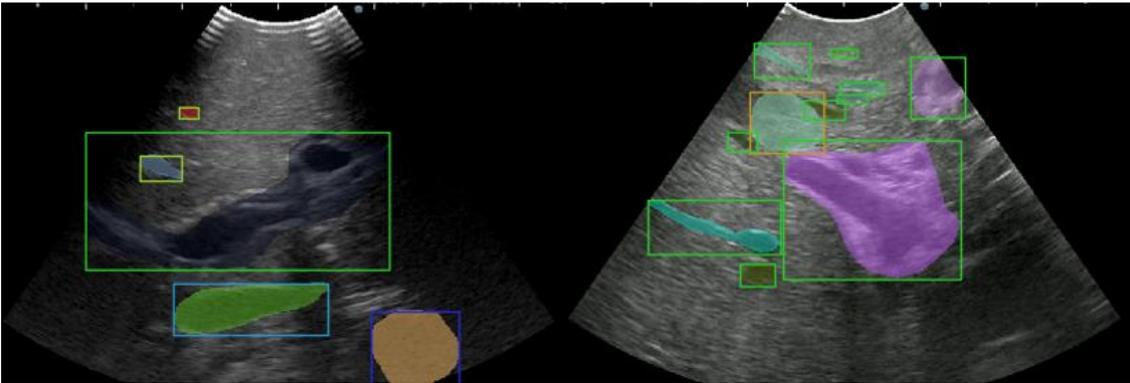


図 5.5 左右反転処理

- ・ノイズ付加

4.4 節で使用したものと同様にデータ拡張確率 P_{DA} でノイズ付加では座標(i, j)の画素値 $\mathbf{v}_{i,j}$ に対して平均値 m_{noise} 、分散 s_{noise} の分布に従う乱数 $\alpha_{i,j}^{noise}$ を付加する。また、ノイズ付加によって $\mathbf{v}_{i,j}$ の各成分が 0 未満または 1 を超える場合には 0 または 1 として更新する。

- ・ Random erasing

Random erasing[Zhong 20]とは画像内の一部の矩形領域を塗りつぶすことで、物体の一部を隠し、より多様性あるデータを作成する手法である。図 5.6 に Random erasing による一部塗りつぶされたサンプルデータを示す。通常のニューラルネットワークでは物体の特

微的なパーツの画像パターンを学習しその物体を認識するが、Random erasing によって物体の一部が隠されるとニューラルネットワークはその隠れた部位以外のパーツに注目し学習が行われる。そのため、物体の主な特徴がある領域が隠れても、代わりとなる特徴も学習しているため、検出が可能となる。Random erasing は画像ごとにデータ拡張確率 P_{DA} で塗りつぶしが行われる。対象となる画像がサイズ $N \times M$ 、画像領域 $S = N \times M$ とすると、消去の対象エリアを決定するパラメータの消去エリア割合 s_e は $[s_{min}, s_{max}]$ の範囲から一様乱数で決定され、消去エリアアスペクト比 r_e は $[r_{min}, r_{max}]$ の範囲から一様乱数で決定される。消去エリアの縦幅 H_e と横幅 W_e 下記のように決定される。

$$H_e = \sqrt{s_e S \times r_e} \quad (5.9)$$

$$W_e = \sqrt{\frac{s_e S}{r_e}} \quad (5.10)$$

また、消去エリアの右上の頂点座標 (x_e, y_e) はそれぞれ $[0, N - W_e]$ 、 $[0, M - H_e]$ の範囲から一様乱数決定される。これによって生成された領域を画素値 0 として塗りつぶす。この処理はマスク領域に対しても同様の消去エリアを使用してマスクを削除する。



図 5.6 Random erasing 処理

・ Random Resize Crop

Random Resize Crop では図 5.7 に示すような画像の一部をランダムに切り取る操作となる。Random erasing とは逆の処理となり、注目したい一部を抽出し、その限定的情報からニューラルネットワークに学習させることで識別対象の特徴的なパーツの学習を促進させる役割を持つ。切り抜きエリアの決定は Random erasing での消去エリア決定と同じで対象となる画像がサイズ $N \times M$ 、画像領域 $S = N \times M$ とすると、切り取りの対象エリアを決定するパラメータの切り取りエリア割合 s_c は $[s_{min}, s_{max}]$ の範囲から一様乱数で決定され、切り取

りエリアアスペクト比 e^{r_c} は $[\log(r_{min}), \log(r_{max})]$ の範囲から一様乱数で決定される。切り取りエリアの縦幅 H_c と横幅 W_c 下記のように決定される。

$$H_c = \sqrt{s_c S \times e^{r_c}} \quad (5.11)$$

$$W_c = \sqrt{\frac{s_c S}{e^{r_c}}} \quad (5.12)$$

加えて、 $H_c \leq M$ かつ $W_c \leq N$ のとき切り取りエリアの右上の頂点座標 (x_c, y_c) はそれぞれ $[0, N - W_c]$ 、 $[0, M - H_c]$ の範囲から一様乱数決定される。 $H_c > M$ または $W_c > N$ のとき、もとの画像サイズ $N \times M$ を超過した辺に対してそれぞれ $H_c = M$ 、 $W_c = N$ として更新し、座標 x_c, y_c は次のように更新する。

$$x_c = \frac{N - W_c}{2} \quad (5.13)$$

$$y_c = \frac{M - H_c}{2} \quad (5.15)$$

これによって得られた切り取りエリアを入力画像サイズ $N_t \times M_t$ に変換したデータを使用する。マスク領域も同様の切り取りエリアを使用して変換する。

Random Resize Crop使用後の教師データ

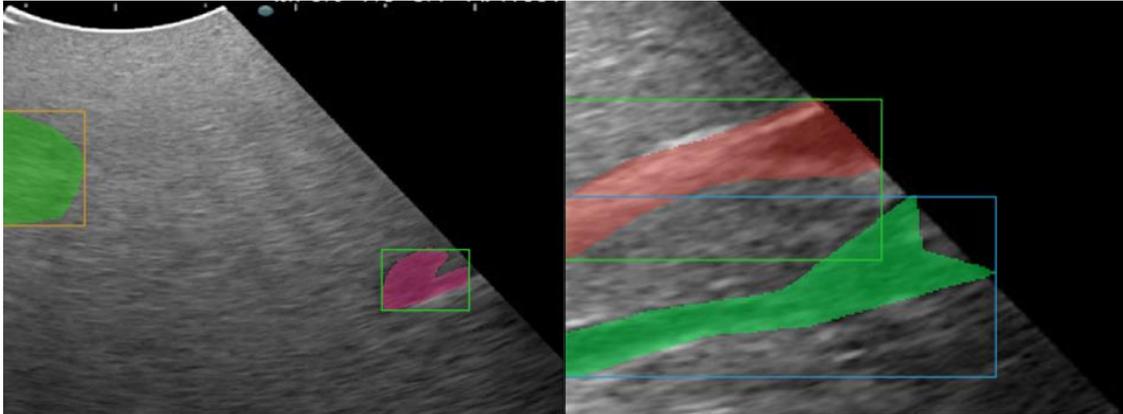


図 5.7 Random erasing 処理

5.4.3 Mask R-CNN の構成・学習設定

5.3.3 節と同様に図 5.2 に示す Mask R-CNN の構成を使用し学習を行う。学習に用いるパラメータも表 5.2 に示したものを使用する。扇形回転トリミングとの比較のため、データ水増し手法なし、左右反転処理、色相変化、ノイズ付加、Random erasing、Random

Resize Crop の 5 手法を実施し、5 パターンのデータセットの学習を行う。このときに使用する各水増し手法のパラメータを表 5.8 に示す。

表 5.8 使用したデータ水増し手法用のパラメータ

データ水増し手法のパラメータ	
P_{DA}	0.5
$w_{brightness}$	0.2
$w_{contrast}$	0.2
$w_{saturation}$	0.2
w_{hue}	0.2
m_{noise}	0.0
s_{noise}	0.1
s_e	[0.02, 0.33]
r_e	[0.3, 3.3]
s_c	[0.08, 1.0]
r_c	[0.75, 1.33333333]

5.4.4 テストデータに対する検出性能

5 パターンのデータセットのそれぞれを学習データ、残りを学習状況判定用のテストデータとして水増し手法なし、左右反転処理、色相変化、ノイズ付加、Random erasing、Random Resize Crop の 6 手法を組み合わせて学習させた。テストデータは epoch 毎に meanAP を算出し、学習状況の把握に利用する。epoch 毎の算出された 5 パターンの meanAP の平均を図 5.8 に示す。加えて、学習が定常的になった 400~380epoch の中で最も優れた meanAP を示したモデルを使用して腫瘍、肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管の 6 種類のラベル毎の meanAP を算出する。これも算出された 5 パターンの meanAP の平均を表 5.9 に示す。また、Random Resize Crop を使用してパターン 1 のデータセットを学習させた際の 400~380epoch の中で最も優れた結果を示したモデルによってテストデータを物体検出させたサンプルを図 5.9 に示す。

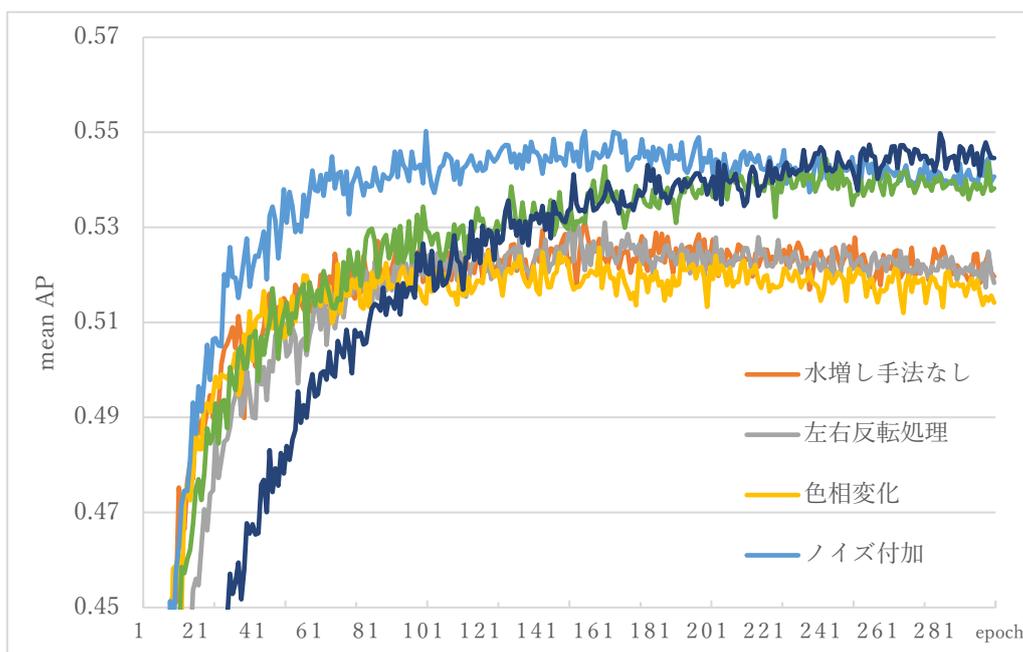


図 5.8 各水増し手法による 5 パターンの学習結果の平均

表 5.9 ラベルごとの検出精度結果

mean Average Precision						
Label	水増し手法 なし	左右反転	色相変化	ノイズ付加	Random erasing	Random Resize Crop
腫瘍	0.733	0.747	0.734	0.756	0.754	0.763
肝静脈	0.381	0.381	0.377	0.374	0.399	0.489
グリソン鞘	0.360	0.362	0.350	0.374	0.375	0.387
小血管	0.287	0.271	0.285	0.262	0.306	0.316
下大静脈	0.657	0.671	0.660	0.713	0.686	0.656
下行大動脈	0.699	0.696	0.672	0.726	0.721	0.693
全体	0.522	0.522	0.519	0.542	0.537	0.543
Average Precision IoU=50%						
腫瘍	0.932	0.951	0.930	0.958	0.963	0.966
肝静脈	0.745	0.756	0.730	0.747	0.786	0.823
グリソン鞘	0.754	0.747	0.727	0.744	0.759	0.801
小血管	0.630	0.616	0.649	0.594	0.686	0.755
下大静脈	0.921	0.935	0.943	0.948	0.961	0.946
下行大動脈	0.910	0.907	0.865	0.933	0.933	0.901
全体	0.816	0.819	0.812	0.833	0.850	0.877

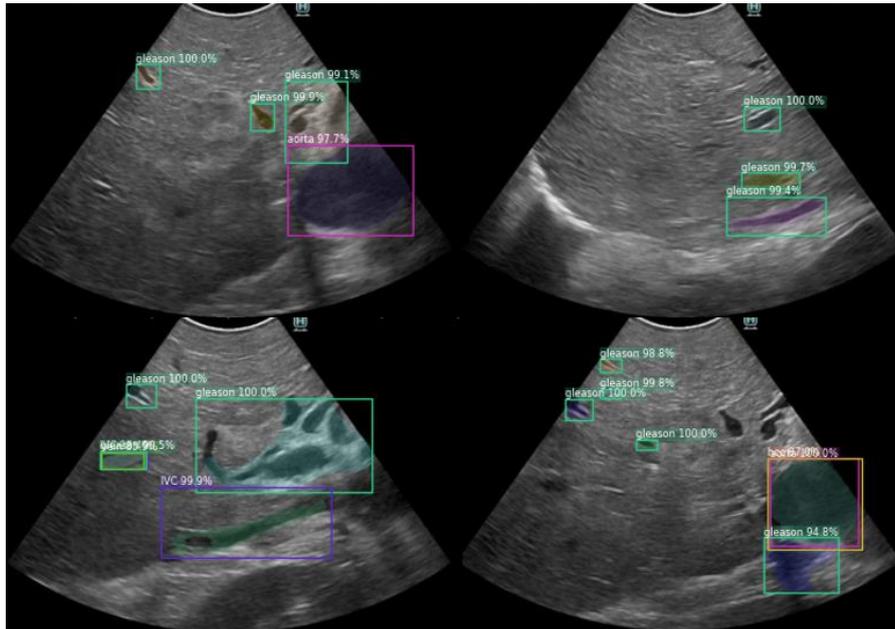


図 5.9 学習モデルによる検出結果

表 5.2 にあるように IoU=50%において水増し手法なしの場合でも $AP_{50}=0.816$ の結果を示した。使用したデータセットと分類ラベル数が異なるため純粹には比較的無いもの、ベースとなった Faster R-CNN が 3 章で示した $AP_{50}=0.881$ と比較すると 0.065 下がっている。これは分類種別が増えたことによるタスクの難化されたことによる影響に加えて、BOX 形状ではなくマスク領域を基準に IoU が算出されたことで下がったと思われる。

水増し手法による認識精度への影響を比較するとノイズ付加、Random erasing、Random Resize Crop の 3 手法では水増し手法なしの場合と比較してそれぞれ優れた結果を示した。物体の種別ごとに比較するとノイズ付加は腫瘍、下大静脈、下行大動脈の 3 つに関して meanAP を向上させ、グリソン鞘、小血管に対しては逆に悪化させた。Random erasing は全部の物体に対する精度を向上させた。Random Resize Crop は特に優れた結果を示し、下行大動脈以外の項目の認識精度を向上させ、全種別に対する Average Precision IoU=50%を 0.071 だけ向上させた。一方で左右反転処理、色相変化では meanAP に大きな変化は現れなかった。また、ラベルごとの検出精度の傾向として腫瘍、下大静脈、下行大動脈の項目は 0.90 以上で検出できるものの肝静脈、グリソン鞘、小血管は 0.80 近辺の結果を示し、大きく差が生じている。腫瘍、下大静脈、下行大動脈は超音波検査によって取得される映像上の形状が円形に近い。逆に肝静脈、グリソン鞘、小血管は管状でかつ腫瘍、下大静脈、下行大動脈と比較して小さい物体となる傾向があり、この影響により検出精度が低いのではと思われる。図 5.8 における右下の検出結果のように円状の形状である下行大動脈を腫瘍、下行大動脈の 2 重に検出する例や肝静脈とグリソン鞘として 2 重に検出される例がいくつか確認された。

5.4.5 評価用データに対する検出性能

5.4.4 節では連続シーンが含み比較的類似したデータであるテストデータに対する検出結果を示した。本節では教師データとテストデータに使用されていない 5 名分の映像箇所から作成した評価データに対する各水増し手法による 5 パターンの学習から得られた最良モデルの性能を評価する。評価データは 5.2 節で示した通り合計 120 枚のデータであり、取得元の個人は教師データとテストデータと同じ 5 名であるが、異なるシーンから得たデータである。評価用データに対する評価には 5.4.4 節で得られたデータ水増し手法によるモデルの中で得られた最良モデルを使用し、その epoch 数と学習データパターン、epoch 数、テストデータに対する meanAP を表 5.10 に示す。ラベルごとの meanAP を算出したものを表 5.11 に示す。また、各ラベルに対する Average Recall を表 5.12 に示す。

表 5.10 評価に使用したモデルの情報

モデル	学習データ	Epoch	meanAP
水増し手法なし	パターン 5	280	0.547113
左右反転	パターン 4	296	0.544043
色相変化	パターン 5	296	0.542194
ノイズ付加	パターン 4	297	0.557452
Random erasing	パターン 5	298	0.565218
Random Resize Crop	パターン 4	294	0.567517

表 5.11 評価データに対するラベルごとの mean Average Precision

mean Average Precision						
Label	水増し手法 なし	左右反転	色相変化	ノイズ付加	Random erasing	Random Resize Crop
腫瘍	0.662	0.655	0.654	0.624	0.656	0.669
肝静脈	0.180	0.202	0.189	0.206	0.164	0.253
グリソン鞘	0.197	0.193	0.195	0.180	0.177	0.216
小血管	0.316	0.284	0.298	0.301	0.316	0.345
下大静脈	0.409	0.435	0.438	0.335	0.451	0.441
下行大動脈	0.683	0.704	0.655	0.669	0.687	0.625
全体	0.375	0.373	0.370	0.384	0.393	0.432
Average Precision IoU=50%						
腫瘍	0.967	0.961	0.944	0.886	0.971	0.947
肝静脈	0.493	0.525	0.523	0.528	0.505	0.645
グリソン鞘	0.541	0.506	0.512	0.445	0.521	0.551
小血管	0.688	0.624	0.727	0.743	0.700	0.772
下大静脈	0.835	0.817	0.827	0.826	0.868	0.816
下行大動脈	0.924	0.926	0.921	0.919	0.931	0.909
全体	0.675	0.656	0.665	0.702	0.708	0.779

表 5.12 評価データに対するラベルごとの Average Recall

Average Recall						
Label	水増し手法 なし	左右反転	色相変化	ノイズ付加	Random erasing	Random Resize Crop
腫瘍	0.710	0.701	0.694	0.668	0.691	0.700
肝静脈	0.295	0.340	0.293	0.260	0.311	0.356
グリソン鞘	0.274	0.265	0.268	0.253	0.269	0.281
小血管	0.352	0.348	0.370	0.333	0.363	0.374
下大静脈	0.435	0.485	0.485	0.415	0.495	0.490
下行大動脈	0.740	0.750	0.715	0.710	0.725	0.700
全体	0.479	0.493	0.483	0.452	0.482	0.496

本節で使用した評価データは教師データと同一人物から得られた映像の異なるシーンから作成したものである。そのため、表 5.9 と表 5.11 を比較すると全体的に meanAP が下がっているものの大きく下がっている物体は肝静脈、グリソン鞘、小血管、下大静脈となった。一方、腫瘍は IoU=50% の Random Resize Crop では 0.019 だけ下がって 0.947 となり、精度の減少幅が小さい。また、下行大動脈に関しては 0.008 だけ増加して 0.909 となり逆に精度が上がっている。腫瘍は 5 人それぞれから得られた映像にそれぞれ一つのみが映像に含まれているため、評価に使用した映像が異なるシーンであっても腫瘍の見え方として近い映像になるため精度が悪化しなかったと思われる。同様に下行大動脈が超音波検査映像に含まれる際の撮影方向が比較的類似しているため、別シーンのデータに対しても検出ができたと思われる。また、Average Recall に関してはすべてのモデルで腫瘍と下行大動脈が 0.7 付近となっておりこの 2 つ特徴を学習がうまく行っているがその他では 0.5 以下と画像に含まれる各物体の検出割合が低い事がわかる。腫瘍が一人に対して一つのものを使用しており、下行大動脈は基本的に肝臓の超音波検査にて映像に出現する箇所が限定的である。この 2 つの物体に関しては学習データのシーンと評価に使用した異なるシーンでの画像が比較的類似していたため、0.7 の数値となったと思われる。逆に血管系の Recall 値が低い結果となった。現状、学習データに含まれる各物体の個数はデータセットで異なるが、データセット 1 では腫瘍が 251 個含まれているのに対して肝静脈は 1,145 個含まれておりデータ数の点では血管系のデータが多く含まれている。データ量の観点で見れば血管系の meanAP、Average Recall とともに良くなるはずだがそうではない。この要因としては血管系の形状が撮影方向によって形状が大きく変化するこの変化に対する学習がうまく行っていないと思われる。

5.4.6 未学習データに対する検出性能

5.4.4 節及び 5.4.5 節では各モデルを映像取得元が同じ、つまり同一の 5 名から得た映像で学習及び評価を行っており、あくまで映像変化によって変化する物体の検出性能に対して評価を行った。本設では 5.2 節の 5 名とは異なる 2 名の肝腫瘍患者と 1 名のドナーから得た映像から作成した 120 枚分のデータに対する評価を行う。データはそれぞれ 1 名ごとに 40 枚のデータを作成しておりラベルは従来のものと同様に腫瘍、肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管として行った。同様に表 5.10 のモデルをそれぞれ使用して、未学習データに対する meanAP を表 5.13、Average Recall を表 5.14 に示す。

表 5.13 未学習データに対するラベルごとの mean Average Precision

mean Average Precision						
Label	水増し手法 なし	左右反転	色相変化	ノイズ付加	Random erasing	Random Resize Crop
腫瘍	0.037	0.056	0.004	0.026	0.019	0.017
肝静脈	0.052	0.064	0.049	0.064	0.054	0.094
グリソン鞘	0.088	0.078	0.084	0.109	0.092	0.080
小血管	0.002	0.006	0.031	0.000	0.000	0.000
下大静脈	0.263	0.348	0.437	0.339	0.309	0.253
下行大動脈	0.445	0.418	0.580	0.438	0.377	0.368
全体	0.143	0.160	0.192	0.159	0.142	0.133
Average Precision IoU=50%						
腫瘍	0.078	0.128	0.026	0.040	0.040	0.078
肝静脈	0.113	0.129	0.105	0.135	0.116	0.188
グリソン鞘	0.254	0.227	0.282	0.351	0.326	0.246
小血管	0.016	0.020	0.149	0.000	0.000	0.000
下大静脈	0.551	0.689	0.868	0.677	0.639	0.484
下行大動脈	0.668	0.629	0.890	0.656	0.560	0.534
全体	0.269	0.294	0.362	0.301	0.279	0.248

表 5.14 未学習データに対するラベルごとの Average Recall

Average Recall						
Label	水増し手法 なし	左右反転	色相変化	ノイズ付加	Random erasing	Random Resize Crop
腫瘍	0.039	0.057	0.007	0.021	0.015	0.024
肝静脈	0.174	0.169	0.180	0.143	0.198	0.203
グリソン鞘	0.199	0.187	0.180	0.203	0.204	0.200
小血管	0.008	0.023	0.108	0.000	0.000	0.000
下大静脈	0.310	0.410	0.522	0.408	0.375	0.286
下行大動脈	0.473	0.460	0.632	0.483	0.398	0.400
全体	0.204	0.222	0.276	0.214	0.202	0.190

評価に使用したデータは異なる人物から得た映像をもとに作成したものであるため、全体的に表 5.4, 5.5 と比べて大きく精度が悪化した。また、ラベルごとに比較すると下大静脈、下行大動脈といったほか物体より大きく、人物が変わっても類似した形状を持つ物体に関しては Average Precision IoU=50%において 0.5 よりも大きい精度で予測できている。また Average Recall に関しても 30% また 40% の割合で検出されている。このことから大きな血管系のデータが十分教師データに含まれており、そこから特徴量の抽出がうまく行ったことで精度が大きく悪化しなかったと思われる。逆に腫瘍の meanAP, Average Recall が著しく低い結果となった。これに関しては腫瘍の形状や位置、肝臓の状態によって、得られる映像パターンや見え方が異なり、それを 5 種のデータ水増し手法にて映像パターンの多様化を図ったものの、腫瘍に対しては効果が現れていない事がわかる。

次に水増し手法に関して比較すると、5.4.5 節で映像の変化を評価した中で最良の手法だった Random Resize Crop が未学習データに対しては meanAP, Average Recall 共に最も低い結果となった。また、5.4.5 節ではデータ水増し手法なしと比べて、meanAP を悪化させていた色相変化が meanAP, Average Recall 共に最も高い結果を得た。特に腫瘍の meanAP が最低値の 0.026 にもかかわらず、下行大動脈、下大静脈で最高値を取得している。これは色相変化によって撮影する際のコントラストや肝臓の状態の違いを教師データに含めることを可能とし、今までは同一映像の異なるシーンであったためにコントラストや肝臓の状態に変化がなく有効的でなかったが、異なる人物から得た映像のパターンを学習データに反映させることで有効だったと思われる。色相変化は画像全体の色相を変化させ、腫瘍やその他の血管系の物体に関係なく変化する。腫瘍に関する色相変化は悪影響しか無いため、色相変化適用は画像全体ではなく物体領域以外、または腫瘍以外の領域に適用することでこれを防ぎつつ汎用性向上が期待できるかもしれない。

この meanAP と Average Recall の結果の他に未学習の肝細胞癌を含む 2 つの映像に対して Random Resize Crop のモデルを適用した結果例を図 5.10 と図 5.11 に示す。未学習映像 1 に対しては腫瘍の検出ができるシーンとできないシーンが発生した。この要因としては今回使用した既存のデータ水増し手法ではデータの多様性確保が不十分であり、微小の映像変化に敏感となって特定のシーンでのみ腫瘍の検出が可能となる結果であった。一方、未学習映像 2 では全く腫瘍の検出ができなかった。これは含まれている腫瘍の形状・模様パターンが学習データに完全に含まれていないものであったためである。



図 5.10 未学習映像 1 に対する検出結果

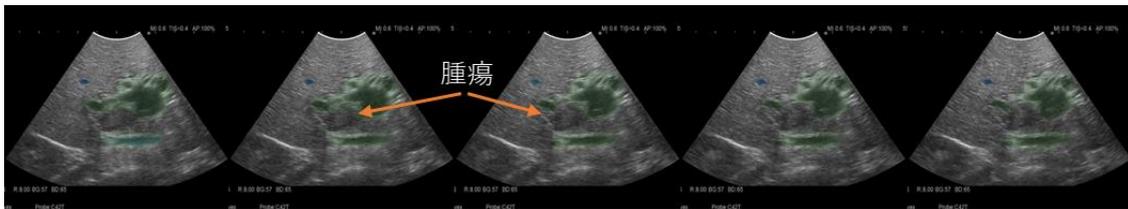


図 5.11 未学習映像 2 に対する検出結果

5.5 扇形回転トリミングと既存データ水増し手法の比較

5.3.5 節、5.4.5 節で行った評価データに対する各水増し手法の結果に関して比較すると扇形回転トリミングは表 5.4 で mean Average Precision: 0.430、 Average Precision IoU=50%: 0.772 の結果となり、これは表 5.11 で Random Resize Crop が示した mean Average Precision: 0.432、 Average Precision IoU=50%: 0.779 に及ばないものの近い値を得た。Random Resize Crop 以外の水増し手法なし、左右反転処理、色相変化、ノイズ付加、Random erasing では Average Precision IoU=50%は 0.708 以下であるため、扇形回転トリミングによる結果は、同一個人異なる映像に対しては Random Resize Crop とほぼ同等の結果を示した。ラベルごとに比較すると Average Precision IoU=50%で扇形回転トリミングを用いたモデルの肝静脈のみが 0.576 で Random Resize Crop より 0.069 だけ悪い結果となったが、それ以外の全てのラベルで扇形回転トリミングが優れた数値を示した。また、Average Recall に関して表 5.5 の結果と表 5.12 を比較すると扇形回転トリミングによって全体に対する値ではパターン 1 で 0.505、パターン 5 で 0.522 であり、表 6.5 中での最高値が 0.496 (Random Resize Crop) であることを考えると大きく改善した事がわかる。微小な回転処理によって各物体の微妙な変化パターンを学習したことで、異なるシーン中に含まれる物体の検出精度が向上したと思われる。扇形回転トリミングの導入によって、本章の一つの目的であった映像変化を再現することで学習データに含まれる多様性を増加させ類似シーンに対しての検出性能 (Recall) を改善できたことがわかった。

一方で、未学習評価データに対する各水増し手法の結果も比較する。扇形回転トリミングの結果である表 5.6, 5.7 と一般的なデータ水増し手法の結果である表 5.13, 5.14 を比較すると mean Average Precision と Average Precision IoU=50%では水増し手法なし、左右反転、色相変化、ノイズ付加、Random erasing、Random Resize Crop、扇形回転トリミングの 7 種の中で扇形回転トリミングは 2 番目に高性能を示したが、腫瘍に関しては Average Precision IoU=50%で 0.201 と最も優れた検出性能を示した。一般的なデータ水増し手法の中で 1 番 Average Precision IoU=50%の数値が高かった色相変化は腫瘍の Average Precision IoU=50%が 0.026 であり、これと比較して扇形回転トリミングは本研究の主目的である腫瘍検出精度に関して大きく性能を向上させる効果が得られた。また、

腫瘍の検出精度に関しては扇形回転トリミングの 0.201、左右反転の 0.128 以外はすべて 0.08 未満であり、扇形回転トリミングによって腫瘍の認識性能に関して汎化性能が大きく向上していることがわかる。

加えて、Average Recall に関して表 5.7 と表 5.14 を比較すると全体の値では色相変化の 0.276 に及ばないものの、腫瘍に関してはパターン 1 で 0.100 と色相変化の 0.007、表 5.7 最高値の左右反転の 0.57 に対して大きく向上している。このことから扇形回転トリミングは映像変化によるデータ増加によって未学習データでも同様に効果を発揮し、既存のデータ水増し手法では特定のシーンでしか検出ができていなかった腫瘍の検出に有効であることが明らかとなった。また、血管系に関しても色相変化に及ばないものの、他の水増し手法と比較して優れており、超音波検査独特の映像変化に適応し、深層学習によるモデル構築に有効であることが明らかとなった。

5.6 まとめ

5.3 では扇形回転トリミングを提案し超音波検査向けにデータ多様性を確保し Mask R-CNN を利用して腫瘍・血管の大きな分類だけでなく肝臓内の主な物体に関してマスク領域として検出するモデルを構築した。同一人物から得た映像から作成した教師データに含まない異なるシーンから作成した評価データに対して扇形回転トリミングは IoU=50%において 0.772 の精度で実現し、腫瘍、下大静脈、下行大動脈の種目に関しては 0.80%以上の精度を示した。従来のデータ水増し手法で最良の結果を示した Random Resize Crop と比較して、Average Precision IoU=50%ではほぼ同等の結果をしめし、Average Recall では従来のデータ水増し手法より優れた 0.522 の値を示した。このことから従来手法と比べて、うまく映像的变化したデータを追加することで学習データの多様性を向上させることができたと考えられる。

また、異なる人物から作成したデータに対しても同様に meanAP と Average Recall を算出した。下大静脈、下行大動脈といった形状が多く、変化が少ない物体に関して、色相変化を使用した場合に IoU=50%において 50%以上の精度で検出が可能であった。これに対して本章で肝臓の超音波検査時に行われるプローブ操作によって発生する映像的变化に着目し提案した扇形回転トリミングでは Average Precision IoU=50%では従来の中で最も優れていた色相変化に勝てず全体で 2 番目の 0.316 であったが、腫瘍に関しては従来の手法から大きく性能を改善することができ 0.201 を達成した。また Average Recall でも同様に全体ラベルでは色相変化に勝てず全体で 2 番目の 0.226 となり、腫瘍に関しては最良の 0.100 を達成した。もともと、扇形回転トリミングの導入目的は映像変化に関してのデータ多様性向上を通しての腫瘍の未学習データに対する性能向上であり、この目的を達成できたと考えられる。まとめると現状では腫瘍に対しては扇形回転トリミングが優れ、血管系には色相変化が優れた結果を示した。この 2 つの手法は画像処理的に共存可能なものであり、同時使

用によるさらなる性能向上が期待できる。ただ、色相変化に関しては腫瘍の Average Precision IoU=50%の中で既存手法中最低値となっており、この点での改善が必要となる。

本章では肝臓の超音波検査映像から腫瘍・血管検出タスクの実現にあたって大きな課題となるデータ多様性の確保について光学的に撮影された画像向けデータ水増し手法と今回超音波検査映像向けに提案した手法である扇形回転トリミングを比較しその有効性を検証した。通常向けデータ水増し手法では学習データと類似したものに対しては優れた性能を Random Resize Crop が示し、血管系検出に関する汎化性能には色相変化が有効であると明らかになった。さらに新規に提案した扇形回転トリミング手法は類似映像に対して Random Resize Crop と同等かつ腫瘍検出に関して優れた汎化性能を得ることができた。これらの検証を通じて、物体ごとに有効な手法や悪影響について明らかにすることができ、超音波検査映像からの腫瘍・血管検出タスクのためのデータ多様性の確保を通して、未学習データに対する優れたモデルを実現した。

6. 肝疾患診断支援システムが実診断に与える影響

6.1 背景と目的

本研究では診断支援システムの実現のためこれまでに深層学習技術を用いた肝臓の超音波検査映像に対して物体検出モデルのモデル面でのカスタマイズや限定的なデータセットの中でデータ多様性の向上を目的とした新規水増し手法の提案を行ってきた。本研究の最終的な目標は支援システムを導入することで医師が行う肝臓の超音波検査による診断の精度を向上させることであり、実際に診断支援システムに効果があるかについて検証が必要である。そこで本章では実際に構築した診断支援システムの性能評価を行う。検証・評価にあたっては医用画像診断支援システム（人工知能技術を利用するものを含む）開発ガイドライン2019（手引き）[日本医療研究開発 19] 及びコンピュータ診断支援装置に関する評価指標[厚生労働省医薬食品局 11]に従って、本研究が構築したシステムが実際に医師による診断に対してどのような影響もたらすかを医師の協力のもと診断支援システムを利用した場合としてない場合で肝腫瘍の認識への影響を計測し、ROC 曲線、AUC にて比較する。

6.2 肝腫瘍診断支援システムの有無による影響の検証方法

診断支援システムによる効果を検証するにあたって、本研究ではシステムを通じて生成された予測領域の情報を付加した映像と付加していない未加工の映像を医師が視聴し、映像中に存在する肝腫瘍の認知への影響を調査する。

・検証に利用する肝腫瘍の超音波検査映像

検証にあたっては肝腫瘍が含まれる異なる 4 人から得た超音波検査映像を使用した。この内、診断支援システムの学習に使用された映像が一部含まれる（学習したことがある映像）対象は 2 名であり、残り 2 名から得た映像は完全に未学習の映像となる。

診断支援システムによる予測領域の表示例を図 6.1 に示す。腫瘍の予測結果は腫瘍領域を赤の Box 形状で表示され、それに合わせて予測の確率が表示される。血管系の種別である肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管に関してはそれぞれ表 6.1 に示されるカラーコードを使用して画像に対して透過度 20%としてオーバーレイする形で表示を行う。また、映像は右側に未加工の映像、左側に上記のアノテーションを追加したものを利用する。また、これらの予測に使用した学習済みモデルは 5.3 節で構築した Random Resize Crop を使用してパターン 5 のデータセットを学習させた際の 400~380epoch の中で最も優れたモデルを利用し、予測された物体の確率が 50%以上より高い領域のみを表示している。

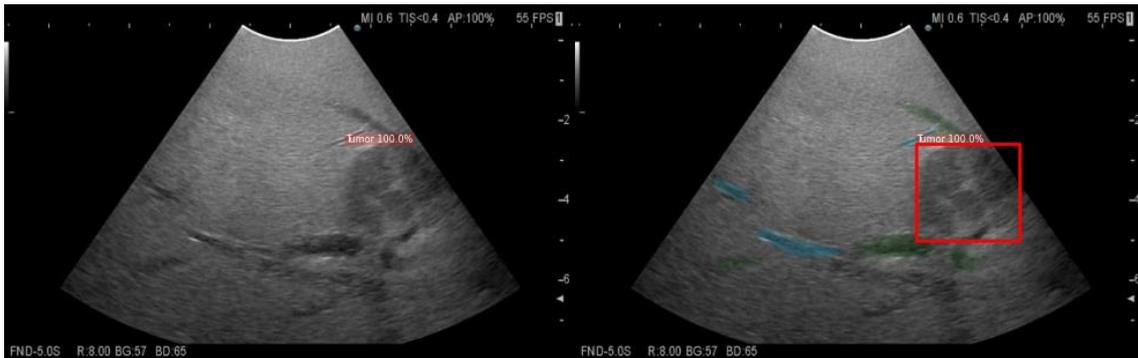


図 6.1 アノテーション有りサンプル画像

表 6.1 アノテーションに使用したラベルごとの色

ラベル	表示用 RGB コード
肝静脈	(0, 128, 255)
下大静脈	(0, 0, 255)
下行大動脈	(51, 255, 51)
グリソン鞘	(51, 128, 51)
小血管	(51, 128, 205)

・腫瘍認知の記録方法

本章では医師が腫瘍を認識しているかどうかを評価する。そこで、映像を視聴しながら、医師が映像中に腫瘍が存在すると認識した時間と存在していない時間を記録しこれをアノテーション付きの映像となしの映像で比較する。記録方法としては図 6.2 に示す記録用アプリケーションを作成した。このアプリでは start ボタンを押下すると Ready ボタンが 5 秒のカウントダウンが始まり、0 秒のタイミングで Tumor ボタンに変化し観察者（医師）が記録対象となる映像を再生する。映像を観察中に、医師が映像中に腫瘍が存在すると認識した場合に Tumor ボタンを押下し続け、存在していない場合は押下を解除することとした。0 秒になった時刻とこの Tumor ボタンが押された時間及び離れた時間を記録し、正答データと比較することで評価する。正答データの作成にあたっては熟練の肝臓専門医の協力の下、1 フレームごとに映像を確認し、映像に腫瘍が含まれている時刻と含まれていない時刻を記録した。

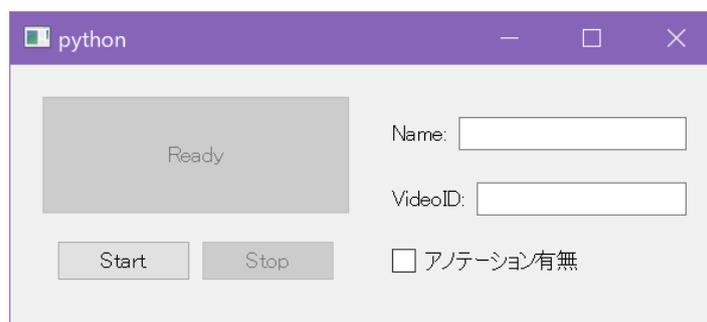


図 6.2 記録用アプリケーション

・映像の観察手順

実際に映像を視聴して腫瘍認知に関して計測するに当たり、観察者が同一の映像から作成したアノテーションあり・なし映像を視聴し観測する場合、同じ位置に腫瘍が存在するため正確な計測ができない。そのため、観測者を A・B グループの 2 つに分割し、それぞれのグループの観測者は同一患者から得た超音波検査映像を視聴しないように表 6.2 に示す組み合わせの映像を視聴することとする。また、映像視聴の順番は表 6.2 で示した数字となる。

表 6.2 各グループの視聴対象となる映像と順番

	アノテーション		
	の有無	A グループ	B グループ
	無	1	-
学習済み映像 1	有	-	2
	無	-	1
学習済み映像 2	有	2	-
未学習	無	3	-
映像 1	有	-	4
未学習	無	-	3
映像 2	有	4	-

・対象となる観測者

本研究では肝臓の専門医でなくともシステムの支援によってある一定以上の腫瘍診断を実現することを目指している。そこで観測者を選定するにあたって、肝臓に対して一定の知識があるものの熟練医師と比較して超音波検査の経験が乏しい肝胆膵外科に所属する研修医を観測者の対象者として協力していただいた。ここで研修医は臨床研修に従事して 2 年未満の医師となる。本章の検証実験では 6 名の研修医に協力いただき、3 名ずつ A グループと B グループに分割し、表 6.2 にある映像の組み合わせと順番にて視聴いただき、記録用アプリケーションを使用して計測した。アノテーションの意味・記録用アプリケーションの説明は動画視聴の直前に行い、表 6.2 に無いサンプル用の映像を使用して記録用アプリケーションの使用に慣れてから表 6.2 の映像に対して計測を行った。

・計測結果の評価方法

本章ではシステムによって生成されたアノテーションによって医師の腫瘍認識への影響を検証するため、映像開始後に計測されたボタンの押下、解除の時刻情報と正答データを比較する。比較するために映像の始まりのタイミングを時刻 $t = 0$ としてそこから単位時間 S_t ごととサンプリングし、ある k 番目のサンプリング点において時刻 $S_t k - 1$ から時刻 $S_t k$ においてボタン押下時間の合計 $T_{on}(k)$ が $T_{on}(k) > \frac{S_t}{2}$ の場合に k 番目を1(腫瘍が存在)、 $T_{on}(k) \leq \frac{S_t}{2}$ の場合に k 番目を0(腫瘍が存在しない)として数列に変換する。また映像中には一瞬腫瘍が消え、また表示されるようなシーンが多数存在する。このような変化の激しいシーンでは医師の認識が追いつかず、評価が難しくなるそのため、正答データ及び、計測データにおいてボタン解除後、次にボタン押下される間隔が T_{gap} 以内の場合、ボタン解除情報を無視し、この期間ではボタンが押下されているとして扱う。

上記にて変換された正答データと計測データの1,0の数列 $G(k), P(k)$ (腫瘍存在情報列)を比較するにあたって、それぞれ k 番目の $G(k), P(k)$ の値を比較するのではなく、腫瘍全体を認識しているかを判定するため、次のように TP,FP,FN,TN をカウントし TPR、FPR、ROC 曲線、AUC を算出する。

変換された正答データの1,0の数列 $G(k)$ から j 番目に0から1または1から0に変化した位置を変化点 k_j とする。このとき、計測データの1,0の数列 $P(k)$ において $k_j + k_d$ から $k_{j+1} - 1 + k_d$ を認識対象となる j 番目のデータとし、それを観測した際に腫瘍であると予測した確率(1である確率) $P_{Tumor}(j)$ を次の式で表現する。

$$P_{Tumor}(j) = \frac{1}{k_{j+1} - k_j} \sum_{k=k_j+k_d}^{k_{j+1}-1+k_d} P(k) \quad (6.1)$$

ここで k_d は正答データと計測データの1,0の数列の間に存在するズレを表している。検証では観測者自身がカウントダウンに従い動画を再生するため、ここに計測開始時間のズレが存在しうる。また、視覚情報から腫瘍を認識し、実際に Tumor ボタンを押下、解除には遅れが存在しうるため、これらのズレを表すために k_d を導入した。この予測確率の算出イメージを図 8.3 に示す。予測確率 $P_{Tumor}(j)$ がしきい値 P_{th} と比較して $P_{Tumor}(j) \geq P_{th}$ の場合に j 番目のデータ区間に腫瘍が存在する(Positive: 1)と予測し、 $P_{Tumor}(j) < P_{th}$ の場合に j 番目のデータ区間に腫瘍が存在しない(Negative: 0)と予測したとカウントする。これらの予測と $G(k_j)$ の値と比較し同一ならば True、そうでなければ False として動画ごとにすべて変化点で TP,FP,FN,TN をカウントする。その後、式(2.24)、式(2.27)をもとにしきい値 P_{th} にお

る $TPR(P_{th})$ と $FPR(P_{th})$ を算出する。更に、ROC 曲線の算出にはしきい値 P_{th} を $0 \leq P_{th} \leq 2$ まで変化させた際の TPR と FPR をプロットする。また AUC の算出は以下の式のように TPR を縦軸、 FPR を横軸にしている ROC 曲線の面積を算出する。

$$AUC = \int_0^1 TPR(FPR) dFPR \quad (6.2)$$

また、評価にあたって使用したパラメータ値を表 6.3 に示す。

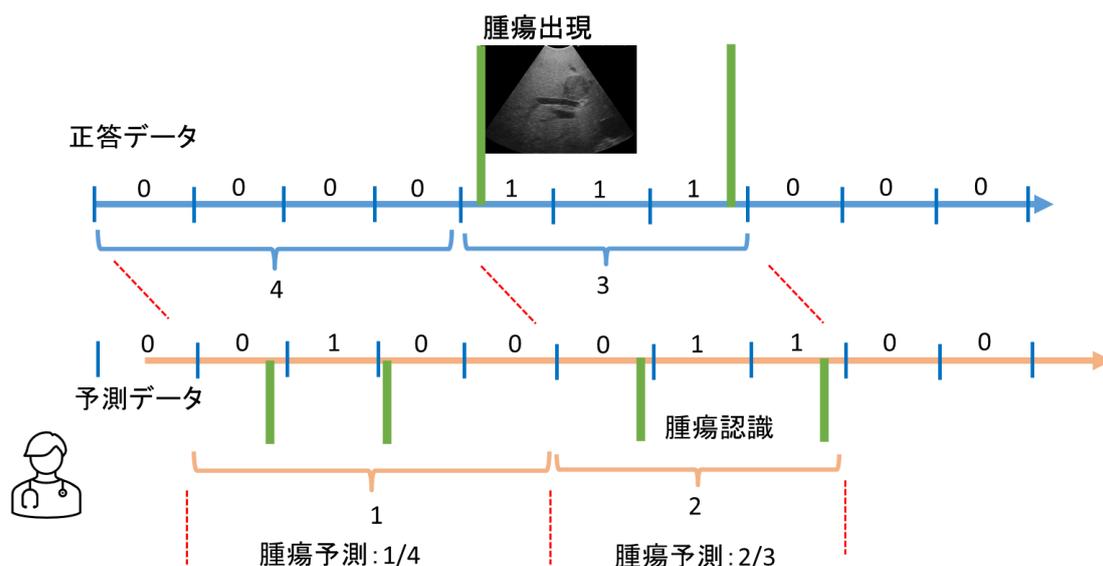


図 6.3 観察者による腫瘍の予測イメージ

表 6.3 評価値算出時に利用するパラメータ値

評価用パラメータ	
S_t	0.25 [s]
T_{gap}	0.25 [s]
k_d	3

6.3 肝腫瘍診断支援システムの有無による影響の検証結果

6.2 節の検証方法にもとづき研修医 6 名の協力のもと、表 6.2 にあるパターンにて腫瘍の存在の予測を行った。アノテーション有・無と学習・未学習の合計 8 つの動画それぞれに対して 3 名分の計測結果が得られた。その際の動画ごとの ROC 曲線を図 6.4-6.11 に、動画ごとの AUC とその平均を表 6.4 に示す。

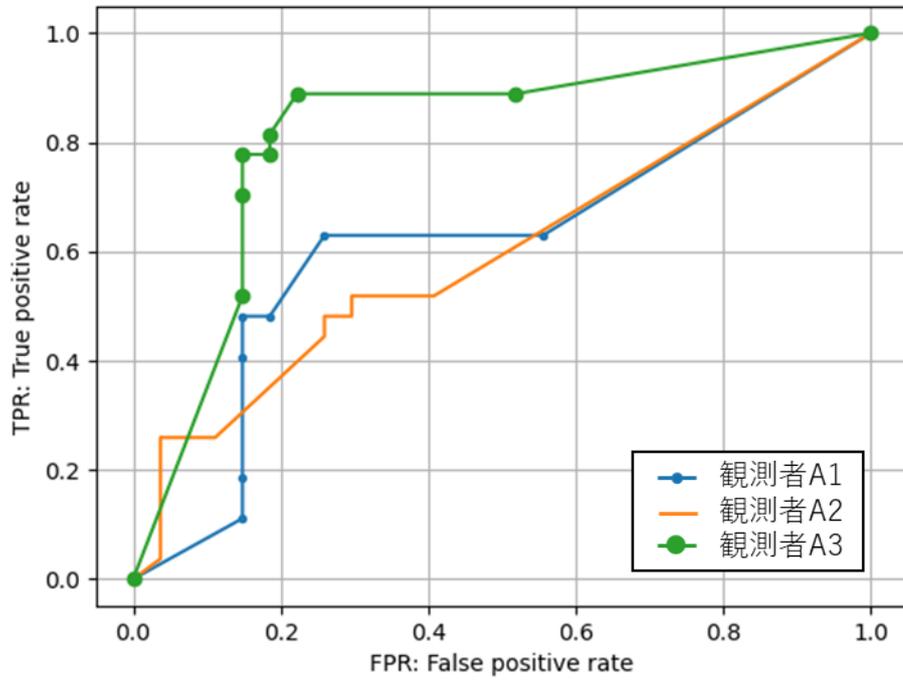


図 6.4 学習済み映像 1 (アノテーション無) の ROC 曲線

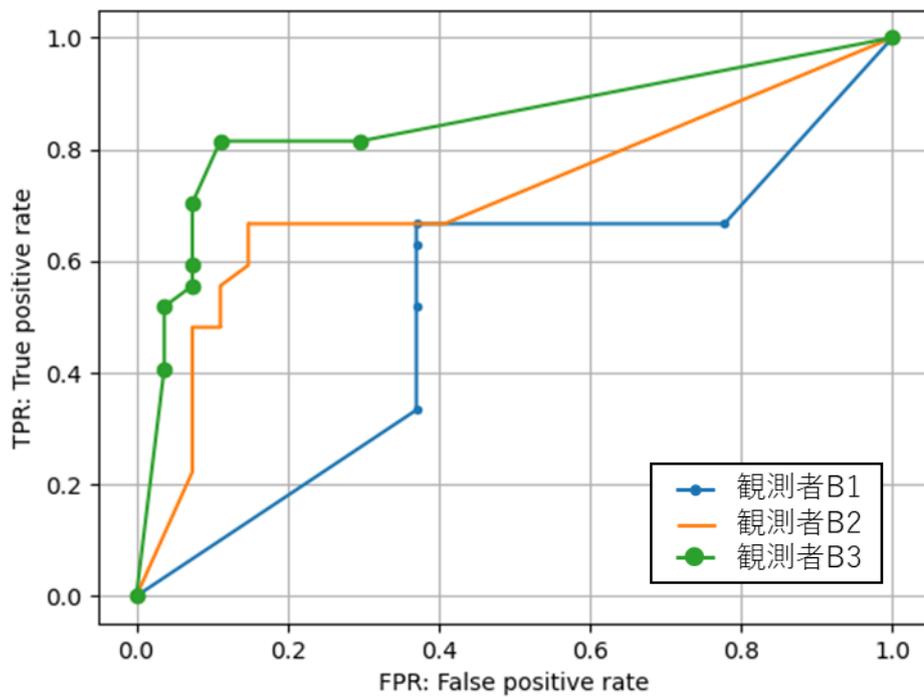


図 6.5 学習済み映像 1 (アノテーション有) の ROC 曲線

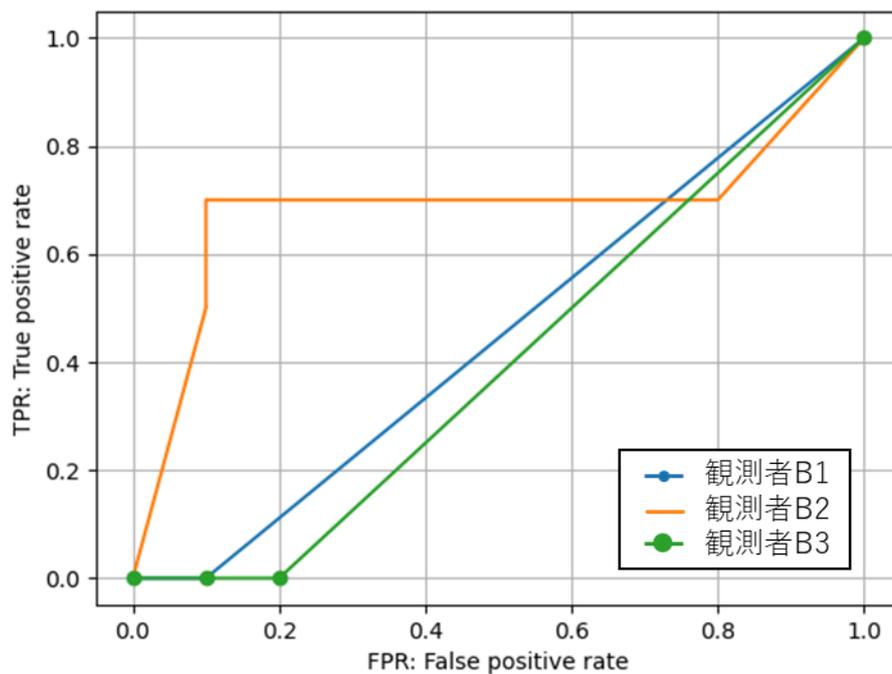


図 6.6 学習済み映像 2 (アノテーション無) の ROC 曲線

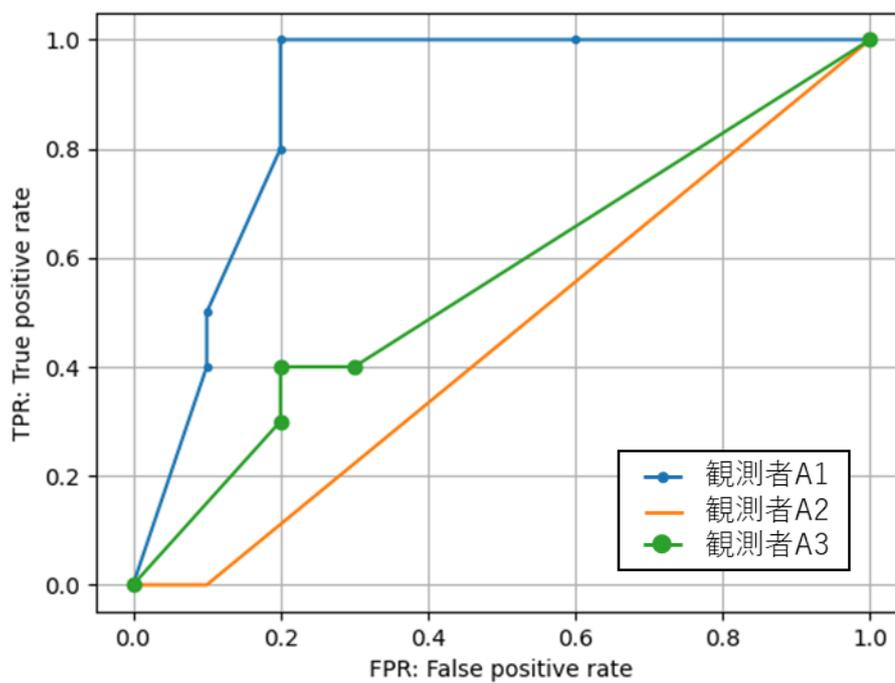


図 6.7 学習済み映像 2 (アノテーション有) の ROC 曲線

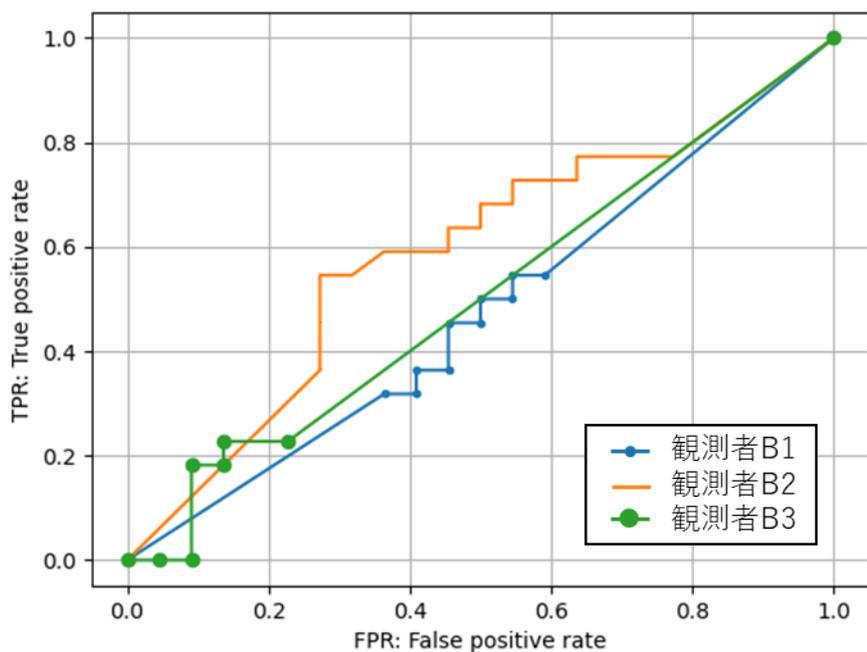


図 6.10 未学習映像 2 (アノテーション無) の ROC 曲線

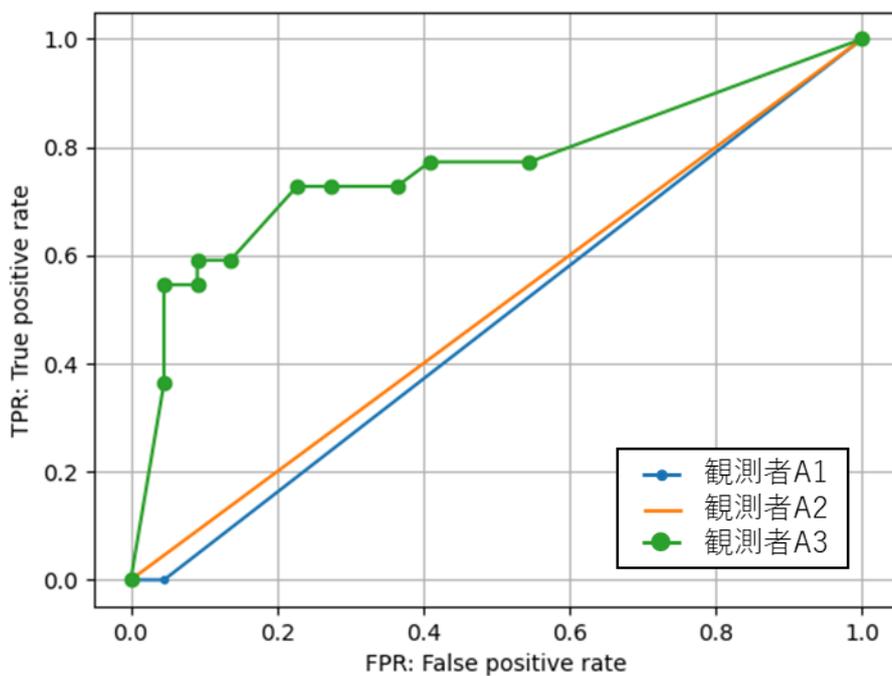


図 6.11 未学習映像 2 (アノテーション有) の ROC 曲線

表 6.4 各映像に対する AUC

対象	アノテーション	AUC						平均
		観測者 A1	観測者 A2	観測者 A3	観測者 B1	観測者 B2	観測者 B3	
学習済 1	無	0.616	0.597	0.817	-	-	-	0.677
	有	-	-	-	0.519	0.714	0.845	0.693
学習済 2	無	-	-	-	0.450	0.685	0.400	0.512
	有	0.885	0.450	0.560	-	-	-	0.632
未学習 1	無	0.650	0.500	0.683	-	-	-	0.611
	有	-	-	-	0.600	0.587	0.745	0.644
未学習 2	無	-	-	-	0.473	0.587	0.503	0.521
	有	0.477	0.500	0.761	-	-	-	0.580

表 6.4 の結果からアノテーション有と無で AUC の平均をそれぞれの対象で比較すると平均 AUC の差の平均は 0.057 となり微小ながら診断支援システムを使用した映像を利用した場合に腫瘍認知に関して優れた結果を得られた。また、未学習データである未学習 1,2 を含む全てにおいてアノテーション有のほうが観測者 3 人分の平均 AUC がアノテーション無と比べ優れた結果となった。5.3 節で構築した Mask R-CNN は学習済みの映像に対しては優れた性能を示しているが、未学習映像に対しては腫瘍の検出がうまく行かない場合やデータによっては完全に検出できない場合がある。これは図 6.12-6.15 に示す各映像での検出映像例からわかる。図 6.12, 6.13 の学習済み映像では腫瘍を高精度で検出できているが、図 6.14 の未学習映像 1 では特定の画像でのみ検出できている。また、図 6.15 の未学習映像 2 では腫瘍検出自体ができていない。これにも関わらず、すべての映像に対して診断支援システムの利用時に優れた結果を示した。この要因として血管系のアノテーションによる効果が考えられる。血管系は腫瘍と比べ、学習データに含まれる個数が多く未学習データに対しても高精度で予測・検出できている。このアノテーションされた血管の信頼度は高く、腫瘍を探す上で観察者は血管以外の箇所注目することが可能である。この効果が図 6.15 の未学習映像 2 で発生し、アノテーションが存在しない領域に存在する円状物体の腫瘍の発見に効果があったと思われる。一方では、図 6.14 の未学習映像 1 では一部の腫瘍自体は検出できるため、一瞬検出された腫瘍のアノテーション情報をもとに観察者がそこに注目することでうまく腫瘍を認識できたと思われる。

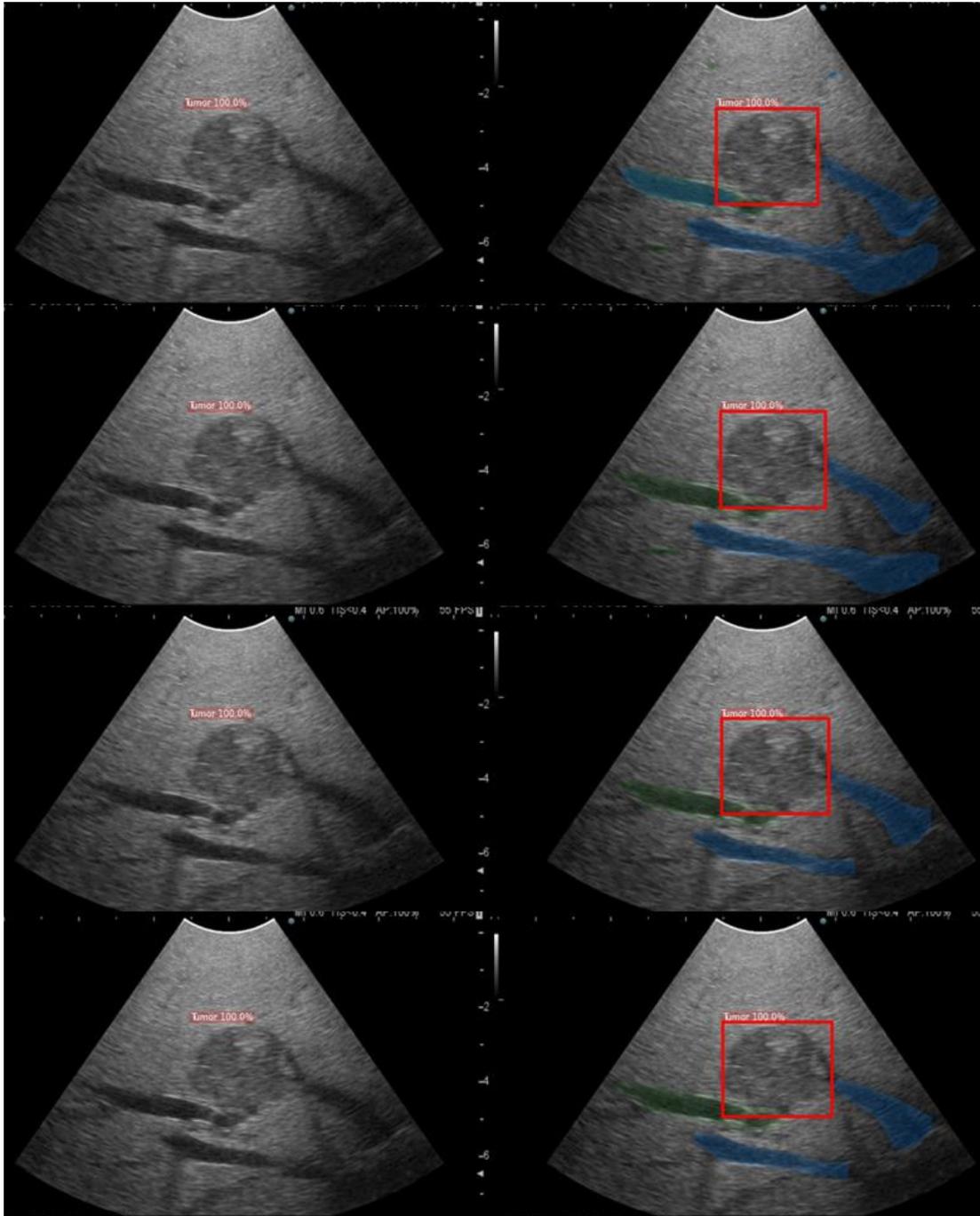


図 6.12 学習済み映像 1 に含まれる連続シーンに対するアノテーション

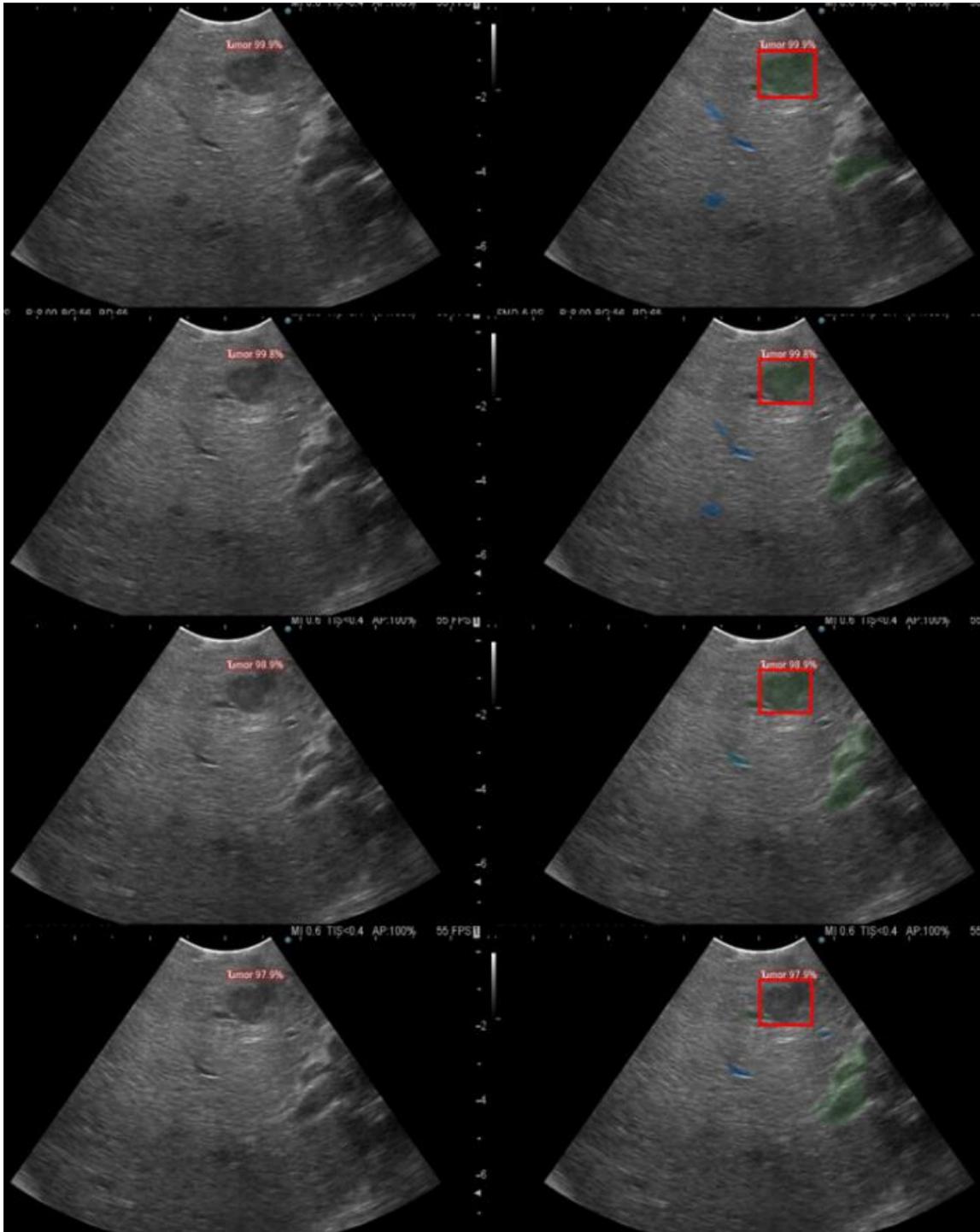


図 6.13 学習済み映像 2 に含まれる連続シーンに対するアノテーション

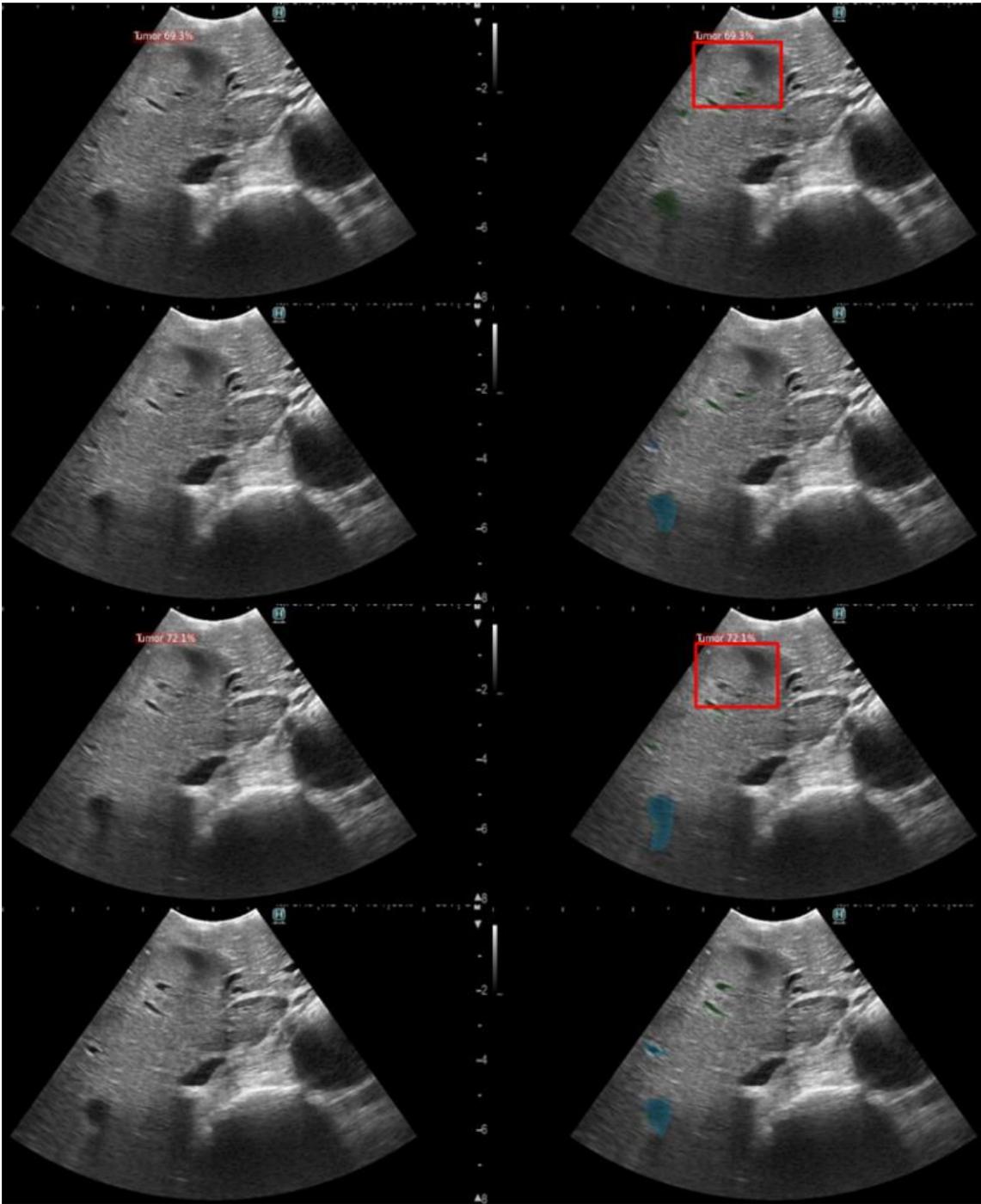


図 6.14 未学習映像 1 に含まれる連続シーンに対するアノテーション

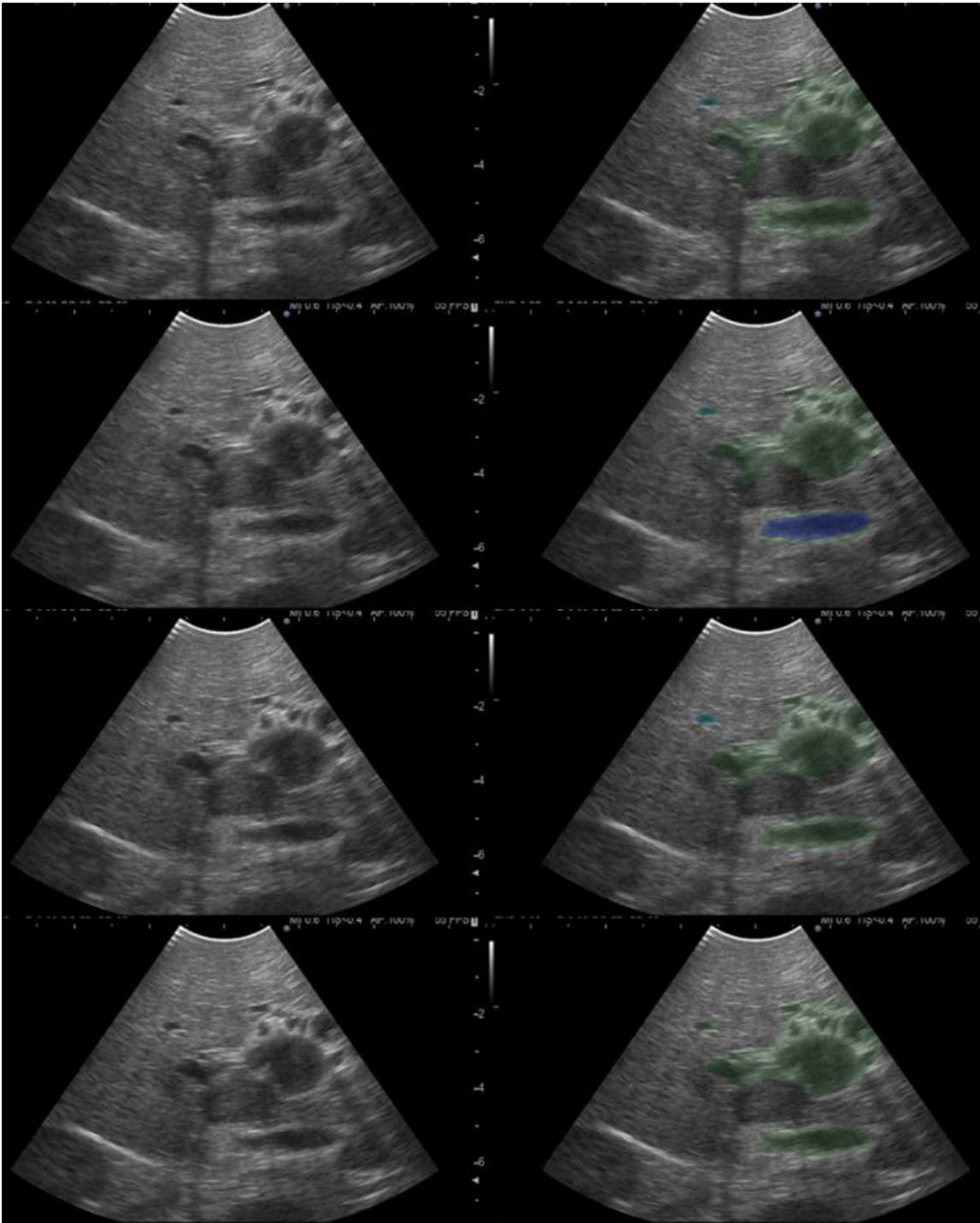


図 6.15 未学習映像 2 に含まれる連続シーンに対するアノテーション

6.4 まとめ

5 章までは診断支援システムとしての利用を想定し、システム構築上での課題を解決するため深層学習技術を用いた肝臓の超音波検査映像に対する物体検出モデルを構築し、

モデルのカスタマイズ及び学習データの多様性確保を通してよりタスクにより適したモデルを実現した。本章ではこのモデルを使用した場合の腫瘍認識に与える影響について検証を行った。検証にあたっては肝胆膵外科の研修医 6 名の協力のもと診断支援システムによるアノテーション有り・無しの映像を視聴した際の腫瘍認識状況を計測した。その結果、学習済み映像と未学習映像のいずれでも支援有りが優れた AUC を示した。支援有と無の差は平均 AUC0.057 で微小ながらも優れた結果を示した。また、支援無の通常映像に対する平均 AUC は 0.58 であり、もともと肝臓の超音波検査映像から腫瘍を認識・把握することは経験年数が少ない研修医にとって難しい事がわかる。また、難しかった要因として医師本人が超音波検査を操作して得られた映像でないことも挙げられる。システム利用によって劇的な認識精度向上は実現できなかったものの学習済み映像の一つに関しては 0.12 だけ性能向上を実現できている。本検証を通じて、映像内の腫瘍検出が一部のみ、検出できない状態であった未学習データに対しても診断支援システムが有効であったことが明らかとなった。これは今後診断支援システムの改良を行っていく上で重要な知見となる。

7. 結論と今後の展望

7.1 研究内容のまとめ

本研究ではこれまで医学的知識だけでなく熟練の経験が要求されていた肝臓の超音波検査に関して、診断支援システムによるサポートにより専門医のいない一般診療所での超音波検査による診断を実現するため、深層学習技術を用いたモデル構築及び、データ多様性の確保、モデルの評価を行った。

はじめに深層学習技術による肝臓の超音波検査映像に対する有効性とそれに特化したモデル構造の探索のため、3章では一般的な物体検出モデルとして使用される Faster R-CNN とそれをベースにし超音波検査映像の映像変化に伴う物体の変化情報を補うことを目的として Optical Flow 情報を利用するモデルを構築した。肝臓内の物体を腫瘍と血管の2つの大きな分類としてデータ学習を行った。両モデルはそれぞれ $AP_{50}=0.881$, $AP_{50}=0.887$ の精度で検出を実現した。Optical Flow 導入モデルでは腫瘍に対する感度で Faster R-CNN と有意差が見られたものの、血管に対しては多重に検出される現象が見られた。また、学習と評価に利用したデータが1名から得られたデータであり限定的なものでもあった。超音波検査映像に着目しモデル構造のカスタマイズを行ったが腫瘍の感度の増加は微小であり、血管検出に悪影響があること、また利用データ限定的であったことも考慮し次の4章では別方向での改良を目指した。4章ではデータ量が限定的だったものを追加作成し4名の肝細胞癌患者及び1名分から864枚のデータを作成した。また、これまで腫瘍と血管の2種だけであったものを腫瘍、肝静脈、下大静脈、下行大動脈、グリソン鞘、小血管の6種類として作成した。これには合計で12359個の物体が含まれている。物体検出タスクを実現する前にこのデータに対して深層学習を利用した正確な分類が実現可能かについて検証するため、Faster R-CNN から物体検出に必要な RPN を除外したモデルを利用し、6種分類タスクを学習させた。また3章と同様により超音波検査映像に特化したモデルの実現を目指して、既存の超音波検査映像から腫瘍の特徴量抽出可能な Power-Law Shot-Noise Model を導入しこれによって生成される冪乗則指数と画像データを組み合わせたデータを入力データとして利用した。検証の結果、冪乗則指数単体学習しただけでは単純に画像データを学習したモデルに及ばない結果となり、冪乗則指数と画像を融合した場合には下大静脈と腫瘍に関する分類精度が上昇するものの全体的には Accuracy が 1.95%悪化した。また、冪乗則指数の計算はピクセルごとに計算が必要となり、リアルタイムでの処理は現実的でないことが明らかとなった。

5章では3章で利用した Faster R-CNN の発展型である Mask R-CNN を使用し、Box 上ではなくマスク状での物体検出モデルを構築する。また、使用する864枚のデータには12359個も物体が含まれているが、超音波検査映像に対しての検出には不足していると考え、学習

データの多様性を増加させるため肝臓の超音波検査時の操作に着目して操作の結果変化する映像パターンを再現するデータ水増し手法として扇形回転トリミングを新規に提案し、合わせて既存のデータ水増し手法として左右反転、色相変化、ノイズ付加、Random erasing、Random Resize Crop と比較した。これらのモデル評価には学習データと同じ5名から作成した評価データと完全に異なる肝細胞癌患者2名とドナー1名から作成した未学習評価データに対して評価を行った。評価データに対する性能では Random Resize Crop が最も優れた $AP_{50}=0.779$ の結果を示し、映像的变化に対しては Random Resize Crop が優秀であることが明らかとなった。一方で未学習評価データに対しては色相変化が $AP_{50}=0.362$ の値で最も優れていた。色相変化特に血管系の検出を高精度で実現した結果、高い AP_{50} を実現している、腫瘍に関しては未学習評価データに対して全データ水増し手法で腫瘍に関する AP_{50} が評価データとの結果と比べて大きく下がり最高でも左右反転の $AP_{50}=0.128$ となっている。実際に検出結果を観察したところ、ある腫瘍に対しては特定のシーンは検出できるがその前後で検出できない事例と別の腫瘍では腫瘍の検出自体が困難な場合あるとわかった。腫瘍に関する多様性の増加を目指し、扇形回転トリミングは評価データに対して Random Resize Crop の結果とほぼ同等の $AP_{50}=0.772$ を得られた。また、未学習評価データに対しては全体に対して $AP_{50}=0.316$ 、腫瘍に対しては $AP_{50}=0.201$ の結果となった。全体の AP_{50} は色相変化に及ばないものの全データ水増し手法中の2番目に優れた性能であり、腫瘍に関しては従来と比べて大きく性能を向上させることを実現した。扇形回転トリミングが腫瘍の検出性能を向上させた要因として次の事が考えられる。腫瘍の形状は球状であり撮影方向によって断面形状は大きく変化せず円状となる。この際、扇形回転トリミングによって変更されたデータは腫瘍を別方向から観察したデータと近いものとなると予測される。一方で、血管系は管状の物体のため撮影方向によって断面の形状が楕円や細長い物体となる。画像の回転とトリミングだけでは血管の形状変化に対応したデータを生成することができず、腫瘍に対して性能向上した結果となったと思われる。

最後に診断支援システムと使用する検出器の性能ではなく実際に診断支援システムとして利用した場合の有効性についても評価を行った。肝臓の超音波検査に関して経験が浅い研修医6名の協力の下、システムによって生成されたアノテーション（腫瘍及びその他物体の位置情報）が付与された映像と通常の超音波検査映像で腫瘍認知に関する比較を行った。映像中に含まれる腫瘍をどれだけ認識できるかについて正答データと比較して評価を行った。視聴する映像としては検出器が学習済みの映像2個と未学習の映像2個それぞれでアノテーション有の8個の映像に対して同じ映像を2度視聴しないように6名と映像8個を組み合わせ、映像ごとに3名による認識結果を計測した。その結果、学習済み・未学習映像に関わらずすべての場合でアノテーション有が優れた AUC を示した。アノテーションによる AUC 増加値は 0.057 と小さいながらも、限定的なデータを学習したモデルであっても未学習データに対して効果があることが明らかとなった。この要因としては特定のシーンだけでも腫瘍が検出できることで検出時の腫瘍形状を参考にその後観察できるこ

と腫瘍検出されなくとも血管系の検出領域外に存在する異常な領域を注目できることの2つが挙げられる。この結果から、本研究では学習データが限定的である状況ながらも専門的経験が乏しい研修医に対して腫瘍認知を支援可能なシステムの構築を実現した。

7.2 結論

本研究では肝臓超音波検査に対する診断支援システムの構築を通して、実現する上で課題となる「肝臓超音波検査映像に対する物体検出タスクの複雑性」、「データの多様性確保の困難さ」、「診断支援システム自体の有効性」の解決及び検証を行った。

複雑性に対しては3章にて画像ベースの物体検出モデルとそれに加えて Optical Flow の導入モデルを提案評価し、基礎的な物体検出モデルである Faster R-CNN であっても腫瘍・血管の検出が可能であることを明らかにした。加えて Optical Flow 導入による時間変化情報の補完によって腫瘍に対して検出感度を向上させ、血管に対しては多重検出の悪影響があることを明らかにした。また、4章で複雑性を有する多種分類タスクの実現性について検証した。深層学習技術単体であっても80%の Accuracy で実現可能であること、物体の散乱特性・減衰特性を抽出した Power-Law Shot-Noise Model の導入により、腫瘍、下大静脈といった他の微小な血管類と比べて大きい物体に対して分類精度向上を実現したが計算コスト的な観点からリアルタイム処理が求められる診断支援システムへの利用には難しい結果となった。上記を通じて、これまで肝臓の超音波検査映像からの物体検出という研究例の少ないタスクにおいても深層学習の有効性と各種情報補完による効果を明らかにした。

データ多様性の確保に対しては既存データ水増し手法及び超音波検査映像向けに提案した扇形回転トリミング手法による各種法の有効性を検証した。本研究で新規提案した扇形回転トリミングは未学習データの腫瘍検出に対して他手法と比較して優れた結果である $AP_{50}=0.201$ が得られた。限られたデータセットの中で多様性確保を実現し、超音波検査映像に適したデータ水増し手法を明らかにした。超音波検査映像に対する応用例が少ない状況と同様に超音波検査向けのデータ水増し手法に関する研究例はなく、本研究で検証した各手法がもたらす各物体への影響と扇形回転トリミングは今後さらなる肝臓の超音波検査映像からの物体検出を改善する上で大きな知見となる。

最後に診断支援システムの有効性について、実際にシステム利用状況を想定し、医師が実際に利用した際の腫瘍認識に対する影響を調査・評価を行った。システムによって AUC を 0.057 増加させることが実現し、診断支援システム自体の有効性を明らかにした。

7.3 今後の展望

腫瘍認知に効果があるシステムを実現したものの、AUC 増加値は 0.057 と小さい。さらなる支援性能向上や腫瘍見逃しを防止する上では、特定シーンでしか腫瘍が検出できないことや検出自体ができない問題について解決する必要がある。これを解決する単純な方法としては腫瘍に関するデータを増やすことである。しかしながら、腫瘍の形状やパターンは個人によって異なるためそれを網羅するには数十名以上のデータが必要と思われる。ただデータ収集及び作成のコストは大きく、時間も要するものである。そこで単純に腫瘍データを増やすのではなく、腫瘍自体の多様性を増やすようなデータ水増し手法の実現を行う。本研究では映像的变化に着目した扇形回転トリミングであったが、より腫瘍にフォーカスし、その形状自体の変化や腫瘍内部のエコーパターン、腫瘍の個数といった変化を行うことで多様性を増やすことが期待できる。また、腫瘍検出性能が悪く、血管に特価して優秀だった色相変化は扇形回転トリミングや腫瘍のパターン増加処理と同時に使用が可能と思われるため、腫瘍に適用しないように組み合わせることで、肝臓内の物体全部に対して多様性増加が期待でき、実現していきたい。また、画像処理ベースのデータ拡張の他に深層学習技術の Generative Adversarial Networks (GAN)[Goodfellow 20]をベースとした合成画像生成も肝臓の超音波検査向けデータ水増し手法として期待ができる。GAN の応用例として入力された顔の輪郭、目や口等の表情の情報をもとに顔の表情を合成することが可能な研究例[Qian 19] [Wu 20]が存在し、これを肝臓内の腫瘍、血管の輪郭（マスク情報）をベースに超音波検査画像を合成する応用も期待できる。ただ、これはあくまで映像的变化や物体の位置、個数、形状の変化に対応できるが腫瘍内部や肝実質のエコーパターンに対する多様性増加が難しいと思われ、上記で述べた腫瘍内部のエコーパターン増加、色相変化との組み合わせることも重要と思われる。

加えて、本研究では肝臓内部の物体を検出することで支援を行うシステムを構築したが、腫瘍を検知することのみが支援につながるわけではなく、血管等の通常存在する物体の情報を提示し、逆にアノテーションが存在しない領域を注目させることも支援につながる事が明らかとなっており、この知見から通常肝臓内に存在しないようなエコーパターンを検出（異常検出）するモデルの実現も有効と思われる。また、この異常検出は学習データ作成の効率化としても利用が期待でき、診断支援システムで検出できずかつ異常と判断された領域を新規学習データの候補として使うことで、より検出モデルに不足していたパターンを持つデータを確保できる。そのため、画像処理だけでなく GAN といった深層学習のデータ生成手法及び異常検出と組み合わせた半自動のデータ追加システムの実現を目指したい。

参考文献

1. [国立がんセンター 21] 国立がんセンター(2021)、最新がん統計：[国立がんセンターがん登録・統計]、https://ganjoho.jp/reg_stat/statistics/stat/summary.html (accessed April 19, 2021)
2. [日本肝臓学会 17] 日本肝臓学会(2017)、肝臓診療ガイドライン 2017年版、金原出版
3. [工藤 21] 工藤正俊, 泉並木, 久保正二, 國土典宏, 坂元亨宇, 椎名秀一朗, ... & 上妻智子. (2021). 第 22 回 全国原発性肝臓追跡調査報告 (2012~ 2013) 日本肝臓研究会追跡調査委員会. 肝臓, 62(5), 251-299.
4. [McCulloch 43] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
5. [Rosenblatt 58] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
6. [Hinton 06] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786), 504-507.
7. [ImageNet 12] ImageNet. (2012). Large Scale Visual Recognition Challenge 2012 (ILSVRC2012). <http://image-net.org/challenges/LSVRC/2012/results.html> (accessed April 19, 2021)
8. [ImageNet 15] ImageNet. (2015). Large Scale Visual Recognition Challenge 2015 (ILSVRC2015). <http://image-net.org/challenges/LSVRC/2015/results> (accessed April 19, 2021)
9. [He 15] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).
10. [Huang 17] Huang, X., Shan, J., & Vaidya, V. (2017, April). Lung nodule detection in CT using 3D convolutional neural networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 379-383). IEEE.
11. [Zhou 17] Zhou, X., Takayama, R., Wang, S., Hara, T., & Fujita, H. (2017). Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. *Medical physics*, 44(10), 5221-5233.
12. [Gerard 18] Gerard, S. E., Patton, T. J., Christensen, G. E., Bayouth, J. E., & Reinhardt, J. M. (2018). FissureNet: a deep learning approach for pulmonary fissure detection in CT images. *IEEE transactions on medical imaging*, 38(1), 156-166.

13. [Wang 21] Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., ... & Xu, B. (2021). A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *European radiology*, 1-9.
14. [Pereira 16] Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE transactions on medical imaging*, 35(5), 1240-1251.
15. [Suk 16] Suk, H. I., Wee, C. Y., Lee, S. W., & Shen, D. (2016). State-space model with deep learning for functional dynamics estimation in resting-state fMRI. *NeuroImage*, 129, 292-307.
16. [Moeskops 16] Moeskops, P., Viergever, M. A., Mendrik, A. M., De Vries, L. S., Benders, M. J., & Išgum, I. (2016). Automatic segmentation of MR brain images with a convolutional neural network. *IEEE transactions on medical imaging*, 35(5), 1252-1261.
17. [Ismael 20] Ismael, S. A. A., Mohammed, A., & Hefny, H. (2020). An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artificial intelligence in medicine*, 102, 101779.
18. [Wang 17] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2097-2106).
19. [Rajpurkar 17] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.
20. [Hemdan 20] Hemdan, E. E. D., Shouman, M. A., & Karar, M. E. (2020). Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:2003.11055*.
21. [Rajkomar 17] Rajkomar, A., Lingam, S., Taylor, A. G., Blum, M., & Mongan, J. (2017). High-throughput classification of radiographs using deep convolutional neural networks. *Journal of digital imaging*, 30(1), 95-101.
22. [Esteva 17] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639), 115-118.
23. [Sakai 18] Sakai, Y., Takemoto, S., Hori, K., Nishimura, M., Ikematsu, H., Yano, T., & Yokota, H. (2018, July). Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network. In *2018 40th Annual*

International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 4138-4141). IEEE.

24. [Rezaeilouyeh 16] Rezaeilouyeh, H., Mollahosseini, A., & Mahoor, M. H. (2016). Microscopic medical image classification framework via deep learning and shearlet transform. *Journal of Medical Imaging*, 3(4), 044501.
25. [Liu 17] Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G. E., Kohlberger, T., Boyko, A., ... & Stumpe, M. C. (2017). Detecting cancer metastases on gigapixel pathology images. arXiv preprint arXiv:1703.02442.
26. [Reddy 18] Reddy, D. S., Bharath, R., & Rajalakshmi, P. (2018). Classification of nonalcoholic fatty liver texture using convolution neural networks. In 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom) (pp. 1-5). IEEE.
27. [Xu 17] Xu, Z., Liu, X., Cheng, X. E., Song, J. L., & Zhang, J. Q. (2017). Diagnosis of cirrhosis stage via deep neural network. In 2017 IEEE International conference on bioinformatics and biomedicine (BIBM) (pp. 745-749). IEEE.
28. [Carneiro 13] Carneiro, G., & Nascimento, J. C. (2013). Combining multiple dynamic models and deep learning architectures for tracking the left ventricle endocardium in ultrasound data. *IEEE transactions on pattern analysis and machine intelligence*, 35(11), 2592-2607.
29. [Xu 19] Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., & Carson, P. L. (2019). Medical breast ultrasound image segmentation by machine learning. *Ultrasonics*, 91, 1-9.
30. [Wang 19] Wang, L., Yang, S., Yang, S., Zhao, C., Tian, G., Gao, Y., ... & Lu, Y. (2019). Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. *World journal of surgical oncology*, 17(1), 1-9.
31. [Kutay 01] Kutay, M. A., Petropulu, A. P., & Piccoli, C. W. (2001). On modeling biomedical ultrasound RF echoes using a power-law shot-noise model. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 48(4), 953-968.
32. [Kutay 03] Kutay, M. A., Petropulu, A. P., & Piccoli, C. W. (2003). Breast tissue characterization based on modeling of ultrasonic echoes using the power-law shot noise model. *Pattern recognition letters*, 24(4-5), 741-756.
33. [Shorten 19] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 1-48.
34. [Zhong 20] Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2020). Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 34, No. 07, pp. 13001-13008).

35. [Pytorch 17] Pytorch.(2017). TORCHVISION.TRANSFORMS
<https://pytorch.org/vision/stable/transforms.html> (accessed April 19, 2021)
36. [Takahashi 19] Takahashi, R., Matsubara, T., & Uehara, K. (2019). Data augmentation using random image cropping and patching for deep CNNs. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(9), 2917-2931.
37. [Goodfellow 20] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144.
38. [Han 18] Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., ... & Nakayama, H. (2018, April). GAN-based synthetic brain MR image generation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (pp. 734-738). IEEE.
39. [Sun 20] Sun, Y., Yuan, P., & Sun, Y. (2020, August). MM-GAN: 3D MRI Data Augmentation for Medical Image Segmentation via Generative Adversarial Networks. In *2020 IEEE International Conference on Knowledge Graph (ICKG)* (pp. 227-234). IEEE.
40. [Islam 20] Islam, J., & Zhang, Y. (2020). GAN-based synthetic brain PET image generation. *Brain informatics*, 7(1), 1-12.
41. [Frid-Adar 18] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018, April). Synthetic data augmentation using GAN for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)* (pp. 289-293). IEEE.
42. [日本医療研究開発 19] 国立研究開発法人日本医療研究開発 (2019)、医用画像診断支援システム (人工知能技術を利用するものを含む) 開発ガイドライン 2019 (手引き)、
https://www.meti.go.jp/policy/mono_info_service/healthcare/iryuu/downloadfiles/pdf/47_guideline.pdf (accessed July 14, 2021)
43. [厚生労働省医薬食品局 11] 厚生労働省医薬食品局 (2011)、コンピュータ診断支援装置に関する評価指標 (薬食機発 1207 第 1 号 (別添 3))、
<https://www.pmda.go.jp/files/000161584.pdf> (accessed July 14, 2021)
44. [Rumelhart 86] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533-536.
45. [LeCun 89] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4), 541-551.

46. [Glorot 11] Glorot, X., Bordes, A., & Bengio, Y. (2011, June). Deep sparse rectifier neural networks. In Proceedings of the fourteenth international conference on artificial intelligence and statistics (pp. 315-323). JMLR Workshop and Conference Proceedings.
47. [Zeiler 14] Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.
48. [He 15] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034).
49. [Han 17] Han, D., Kim, J., & Kim, J. (2017). Deep pyramidal residual networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5927-5935).
50. [Krizhevsky 09] Krizhevsky, A. (2009). CIFAR-10 and CIFAR-100 datasets. <https://www.cs.toronto.edu/~kriz/cifar.html> (accessed April 19, 2021)
51. [He 16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
52. [Veit 16] Veit, A., Wilber, M., & Belongie, S. (2016). Residual networks behave like ensembles of relatively shallow networks. arXiv preprint arXiv:1605.06431.
53. [Ren 15] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497.
54. [He 17] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).
55. [COCO 17a] Common Objects in Context(COCO). (2017). Detection Evaluation. <https://cocodataset.org/#detection-eval> (accessed April 19, 2021)
56. [COCO 17b] Common Objects in Context(COCO). (2017). COCO 2017 Object Detection Task. <https://cocodataset.org/#detection-2017> (accessed April 19, 2021)
57. [Szegedy 16] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).
58. [Krizhevsky 12] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.

59. [Szegedy 15] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).
60. [Simonyan 14] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
61. [Redmon 17] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).
62. [Karako 21] Karako, K., Mihara, Y., Arita, J., Ichida, A., Bae, S., Kawaguchi, Y., Ichizawa, T., Akamatsu, N., Kaneko, J., Hasegawa, K., & Chen, Y. (2021). Automated liver tumor detection in abdominal ultrasonography with a modified Faster R-CNN architecture. *HepatoBiliary Surgery and Nutrition*, (in press).
63. [Dosovitskiy 15] Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., ... & Brox, T. (2015). Flownet: Learning optical flow with convolutional networks. In Proceedings of the IEEE international conference on computer vision (pp. 2758-2766).
64. [Ilg 17] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., & Brox, T. (2017). Flownet 2.0: Evolution of optical flow estimation with deep networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2462-2470).
65. [Hui 18] Hui, T. W., Tang, X., & Loy, C. C. (2018). Liteflownet: A lightweight convolutional neural network for optical flow estimation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8981-8989).
66. [Schottky 18] Schottky, W. (1918). Uber spontane stromschwankungen in verschiedenen elektrizitatsleitern. *Annalen der Physik* (Vol. 57, pp. 541-567).
67. [Qian 19] Qian, S., Lin, K. Y., Wu, W., Liu, Y., Wang, Q., Shen, F., ... & He, R. (2019). Make a face: Towards arbitrary high fidelity face manipulation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 10033-10042).
68. [Wu 20] Wu, R., Zhang, G., Lu, S., & Chen, T. (2020). Cascade ef-gan: Progressive facial expression editing with local focuses. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5021-5030).

謝辞

本研究にて使用された超音波検査データは三原裕一郎先生、長谷川潔先生及び東京大学医学部附属病院 肝胆膵外科・人工臓器移植外科の協力のもと収集・作成が行われました。また、診断支援システムの構築、評価・今後の展望に関して、とても参考になる意見を頂いき割澤先生、稗方先生、二瓶先生、長谷川先生、三原先生に感謝の意を表します。最後に指導教員の陳先生に厚く御礼申し上げます。

補論 研究内で使用される超音波検査データについて

超音波検査映像は東京大学医学部附属病院 肝胆膵外科・人工臓器移植外科の協力のもとデータの収集・作成が行われた。超音波検査映像の利用にあたってこの研究は東京大学大学院 医学系研究科・医学部 倫理委員会によって承認されており (No. 2019166NI)、ウェブサイトでのオプトアウト形式としてインフォームドコンセントを得ている。