

深層学習に基づく画像化された肺音内の異常の検出 Deep Learning to Detect Abnormalities in Imaged Lung Sounds

松本 章吾[†]
Shogo Matsumoto

島川 博光[†]
Hiromitsu Shimakawa

1. はじめに

近年、呼吸器系疾患を診断できる医師の不足が問題視されている。呼吸器疾患の診察において最も有効な検査は聴診であり、呼吸器の専門医は多くの症状を聴診で聞き分けて診断している。しかし、呼吸器系の疾患は重なって発症することも多く、肺音に含まれる異常音である副雑音は病気の種類、病変の部位、病態等によって異なる。さらに、異常音は、年齢、性別、体格等によっても異なる。

別の問題として、聴診による診断は主観的な判断を必要とすることも挙げられる。肺音に異常音が含まれているかどうかは、経験によって培われた主観的な判断基準により診断されるからである。専門医が不足する今日においては、非専門医でも高い精度で診断できるように、情報技術を用いて客観的な診断を可能とすることが求められている。

聴診音の中に疾病特有の音が含まれているのであれば、肺音のスペクトル画像中にも、特有の特徴が現れるはずである。本研究では、聴診器から収集した肺音を画像化し、画像から呼吸器系疾病の特徴を示す部位を検知することによって、非専門医が呼吸器疾患の診察をできるようにする。

短時間フーリエ変換は、聴診にて収集された肺音をスペクトル画像に変換する。本研究では、肺音を専門医が診断している音の領域を強調して表示するスペクトル画像である mel スペクトログラムに変換する。提案手法は mel スペクトログラム画像中に現れる、疾病に特有の特徴を深層学習に基づいて指定された特徴をもつ物体を検出する手法、YOLO で検知する。深層学習で画像を客観的に分析すれば、判断基準は、診断者の技量に依存しない。本研究は、呼吸器の非専門医が、呼吸器疾患を診察することを可能にし、医者不足を解消する。

2. 対象とする聴診音の画像化

2.1 対象とする疾病とその特徴

ヒトの肺音は呼吸音と副雑音で構成される。呼吸音は正常な呼吸にともなって発生する音であり、副雑音は疾病に罹患しているさい、呼吸運動にともなって発生する異常な音である。

副雑音は音の発生する部位によって異なる特徴音が聴取され、連続性と断続性が存在する。本論文では、間質性肺炎の特徴である捻髪音を検知することを考える。捻髪音は、細かい断続性のファイン・クラックルである。捻髪音は、持続時間の短い不連続な音であり、音を時間軸上の波形として観測したさいに休止期間が存在する。

2.2 mel-スペクトログラム

本研究では、デジタル聴診器を用いて採集した聴診音データに含まれる副雑音の周波数特徴を、機械学習を用いて抽出する。非専門医への支援を前提とし、聴診音データを

視覚的に表現し、疾病の根拠を客観的に提示する。短時間フーリエ変換を利用して、肺音を、時間、周波数、強度の 3 次元で表現するスペクトル画像に変換する。スペクトル画像では、経過時間が横軸で、周波数が縦軸で、強度が色で表現される。

スペクトル画像は、観察を開始してから時間 t が経過したときに、周波数 ω の音が、どれほどの強度を持っているかを、時間 t 上の色の並びで示す。

デジタル聴診器を用いて採集した聴診音データは、アナログ信号から離散信号のデジタル信号に変換される。この変換は、聴診音データの周波数特徴を保持するようにサンプリングを実施する。聴診音に含まれる呼吸や副雑音の周波数に現れる特徴は約 2,000Hz 以下の音である。短時間フーリエ変換で周波数特徴を抽出するためには、サンプリング定理に基づき、約 4,000Hz 以上の周波数でサンプリングを実施する。本研究では 10,000Hz のサンプリング周波数でデジタル信号にサンプリングするデジタル聴診器を用いて聴診音データを採集する。これにより短時間フーリエ変換で聴診音データに含まれる約 5,000Hz 以下の周波数特徴を抽出することが可能となる。

3. 提案手法

本研究では、肺音画像データを物体認識モデル YOLO を用いて認識する。本研究の提案手法の概要図以下図 1 に示す。本研究では、データの前処理として、デジタル聴診器を用いて採集した聴診音データを、短時間フーリエ変換を用いて時間、周波数、強度の 3 次元で表現する mel スペクトラム画像データに変換し、短い時間幅で分割する。

次に聴診音データから算出された mel スペクトラム画像データから、深層学習法である YOLO を用いて、異常な肺音に相当する部分を検知する。物体検出とは、画像の中から特定の物体の位置、個数、ラベルを認識する技術である。間質性肺炎に特有の捻髪音が肺音に含まれる場合、その肺音を変換した mel スペクトラム画像データに、捻髪音の発生した時刻に、特有の周波数と強度に相当する領域があると考えられる。

YOLO は、この領域の特徴を予め学習させた検知器を使って、mel スペクトラム画像データに含まれる捻髪音に相当する領域を囲う形で検出する。YOLO は バウンディング

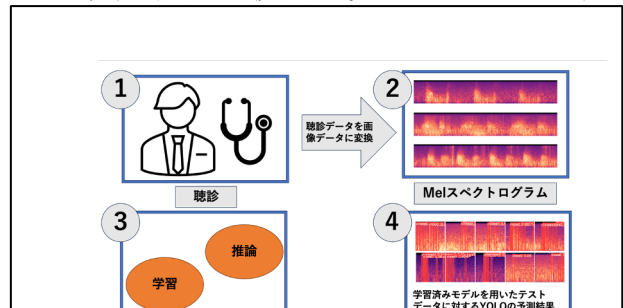


図 1 手法概要図

[†]立命館大学 大学院情報理工学研究所, Graduate School of Information Science and Engineering, Ritsumeikan University

ボックスによって検知した領域を示す。

YOLO には様々なバージョンがあり、今回は YOLOv5s を用いて捻髪音の領域を検知する。YOLO の構造は、畳み込みネットワークと YOLO ネットワークの 2 つのネットワークで構成される。畳み込みネットワークでは入力された画像のさまざまな特徴が抽出される。YOLO ネットワークでは、抽出された特徴から、予め登録された物体が画像内のどこにあるかが計算され、バウンディングボックスの座標情報、それが対応する領域が登録物体である信頼度を出力する。

本研究では、YOLO に mel-スペクトラム画像データを入力し、異常音に相当する領域を検知する。YOLO は、その領域が捻髪音を表す Confidence Score とバウンディングボックスの座標を出力する。この出力を使って、捻髪音が発生している領域をバウンディングボックスで囲んで視覚的に示すことができる。

本研究では評価指標として mAP を採用する。mAP は、適合率と再現率から求められる AP を、すべての登録物体について平均した評価指標で、0 から 1 の値をとる。

4. 実験と結果

本研究で扱う YOLO のパラメータの詳細は以下の表 1 のとおりである。

| Image Size | Train Batch | Val Batch | Epochs | Conf Level | Early Stopping |
|------------|-------------|-----------|--------|------------|----------------|
| 640 | 16 | 16 | 1000 | 0.25 | 100 |

| Model | Layers | Parameters | GFLOPs |
|----------|--------|------------|--------|
| <i>n</i> | 213 | 1760518 | 4.1 |
| <i>s</i> | 213 | 7012822 | 15.8 |
| <i>m</i> | 290 | 20852934 | 47.9 |
| <i>l</i> | 367 | 46108278 | 107.6 |
| <i>x</i> | 444 | 86173414 | 203.8 |

表 1 YOLO のパラメータ

本研究では YOLO v5 の中でも、model *s* を採用する。データは訓練データ 40 枚、検証データ 18 枚、テストデータ 10 枚とし、学習と推論を行う。適合率が 0.769、再現率が 0.714、mAP 0.5 が 0.779 という結果が得られた。図 2 は、テストデータでの検出結果である。

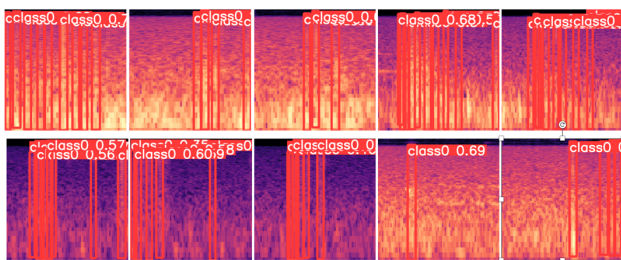


図 2 テストデータの検出結果

図 2 より、捻髪音の聞こえる範囲にバウンディングボックスが表示されていることがわかる。class 0 はそこに捻髪音があることを表し、その右隣にある値は Confidence Score を表す。

5. 考察

本研究の実験結果から、適合率が 0.769、再現率が 0.714、mAP が 0.779 という結果が得られた。適合率、再現率が共に 0.7 以上であり、悪くない精度で yolov5 が捻髪音を検

知したことがわかる。テストデータの推論結果を視覚化した図 2 が示すように、バウンディング・ボックスが捻髪音を囲うことにより、聴覚による主観的評価ではなく、視覚化により、非専門医や患者にもわかりやすく捻髪音が発生していることを客観的に示せる。

6. おわりに

本研究は、人の肺音データを画像化することにより、非専門医や患者にも、捻髪音が発生していることを視覚的に示すことが可能であることを示す。その精度は mAP で 0.779 であった。今後様々な患者や健康者から得られたより多くのデータを用いて同様の実験を行い、これらの実験での検知結果の精度を、今回の実験結果と比較しながら検証する必要がある。

参考文献

- [1] Aviles-Solis JC, Storvoll I, Vanbelle S, Melbye H. The use of spectrograms improves the classification of wheezes and crackles in an educational setting. *Sci Rep.* 2020 May 21;10(1):8461. doi: 10.1038/s41598-020-65354-w. PMID: 32440001; PMCID: PMC7242373.
- [2] Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. *N Engl J Med.* 2014 Feb 20;370(8):744-51. doi: 10.1056/NEJMra1302901. PMID: 24552321.
- [3] Gurung A, Scrafford CG, Tielsch JM, Levine OS, Checkley W. Computerized lung sound analysis as diagnostic aid for the detection of abnormal lung sounds: a systematic review and meta-analysis. *Respir Med.* 2011 Sep;105(9):1396-403. doi: 10.1016/j.rmed.2011.05.007. Epub 2011 Jun 14. PMID: 21676606; PMCID: PMC3227538.
- [4] Pramono RXA, Bowyer S, Rodriguez-Villegas E. Automatic adventitious respiratory sound analysis: A systematic review. *PLoS One.* 2017 May 26;12(5):e0177926. doi: 10.1371/journal.pone.0177926. PMID: 28552969; PMCID: PMC5446130.
- [5] Lakhe A, Sodhi I, Warriar J, Sinha V. Development of digital stethoscope for telemedicine. *J Med Eng Technol.* 2016;40(1):20-4. doi: 10.3109/03091902.2015.1116633. Epub 2016 Jan 5. PMID: 26728637.
- [6] Sarkar M, Madabhavi I, Niranjana N, Dogra M. Auscultation of the respiratory system. *Ann Thorac Med.* 2015 Jul-Sep;10(3):158-68. doi: 10.4103/1817-1737.160831. PMID: 26229557; PMCID: PMC4518345.
- [7] Zulfiqar R, Majeed F, Irfan R, Rauf HT, Benkhelifa E, Belkacem AN. Abnormal Respiratory Sounds Classification Using Deep CNN Through Artificial Noise Addition. *Front Med (Lausanne).* 2021 Nov 17;8:714811. doi: 10.3389/fmed.2021.714811. PMID: 34869413; PMCID: PMC8635523.
- [8] Diwan, T., Anirudh, G. & Tembhrne, J.V. Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed Tools Appl* 82, 9243–9275 (2023). <https://doi.org/10.1007/s11042-022-13644-y>
- [9] Gulzar, Hafsa and Li, Jiyun and Manzoor, Arslan and Rehmat, Sadaf and Amjad, Usman and Jalil Khan, Hadiqa, Transfer Learning Based Diagnosis and Analysis of Lung Sound Aberrations (March 15, 2023). International conference on Health Informatics (HEIN 2023), Available at SSRN: <https://ssrn.com/abstract=4389141> or <http://dx.doi.org/10.2139/ssrn.4389141>
- [10] You Only Look Once: Unified, Real-Time Object Detection, Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788