

メルスペクトログラムとCNNによる肺音における異常音有無の分類

背景

在宅療養中の肺炎、COPD(慢性閉塞性肺疾患)、肺がん患者は高齢化等の理由によって増加している。

また現在、コロナウイルスの影響により病院の受診控えも増え、医療機関への受診・相談が遅れる場合も多く、受診時には病状がすでに進行してしまっていることがある。

目的

在宅でも肺に異常があれば分かるようなシステムを構築したい。
そのための基礎研究として、本研究では肺音の異常音分類を行う。

肺音に異常音(Crackling, Wheezing)が含まれていた場合、高い精度でノイズ有りとして分類することを目標とする。

そのために、音声データである肺音をメルスペクトログラムとして視覚化し、そのスペクトログラムに対し、画像データに対して定評があるCNNを使用して、肺音をノイズ有りとノイズ無しに分類する。

使用するデータ

ICBHI 2017 Challenge
Respiratory Sound Database
<https://bhichallenge.med.auth.gr/>

使用するデータ

肺音に関わるものたち

- ・肺音のデータ(.wav)
 - 肺音を聴取した部位
 - * 気管 左肺前方 右肺前方 左肺後方 右肺後方 左肺側面 右肺側面
 - 聴取モード
 - * single channel
 - * multi channel
 - 使用した機材
 - * 各種メーカー
 - 呼吸周期の始まり[s]と終わり[s]
 - (crackling)破裂音の有無
 - (wheezing)ホイッスル音の有無



使用するデータ

患者にかかわるものたち

- ・患者データ

- id

- 年齢

- 性別

- BMI [kg/m²] (大人の場合)

- 体重 [kg] (子供の場合)

- 身長 [cm] (子供の場合)

- 病名

- * 健康

- * COPD(慢性閉塞性肺疾患)

- * LRTI(下気道感染症)

- * URTI(上気道感染症)

- * Bronchiectasis(気管支拡張症)

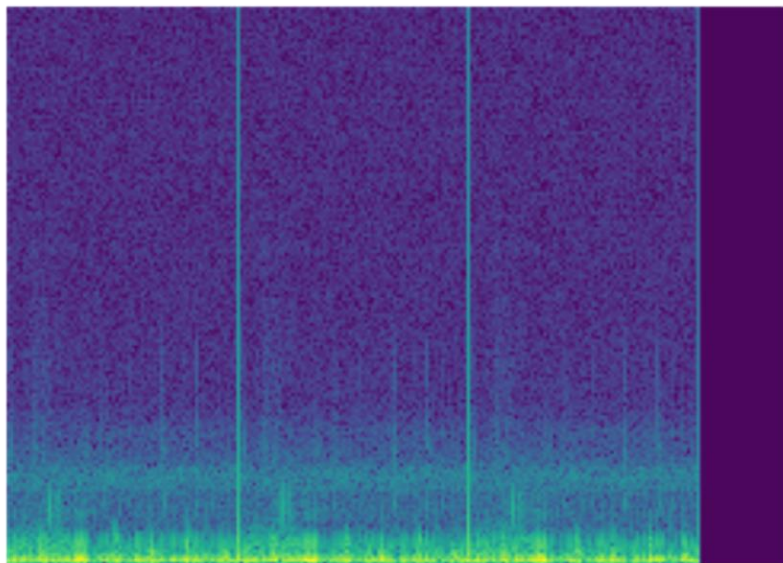
- * Pneumonia(肺炎)



肺音データの加工

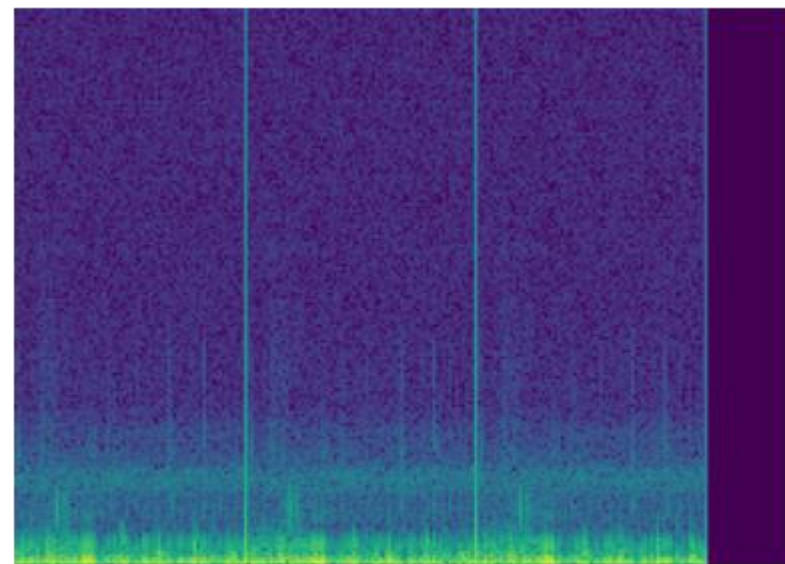
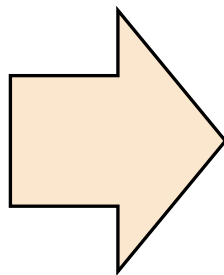
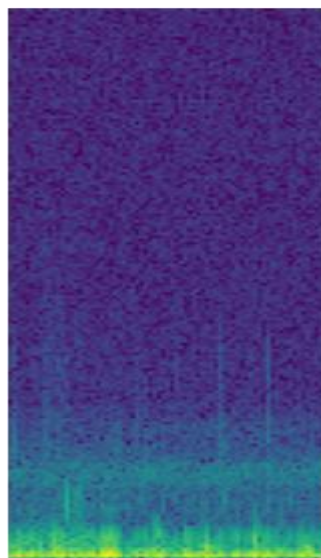
横軸が時間、縦軸が周波数Hz、色が振幅dbを表すスペクトログラムを作成する。なお、周波数Hzをメル尺度melに置換したメルスペクトログラムを本研究では扱う。

メル尺度は人間が感じる周波数の高低を反映した尺度であり、周波数が1000Hz-2000Hz-3000Hzと変化しても、メル尺度では1000mel-1521mel-1876melしか変化しない。



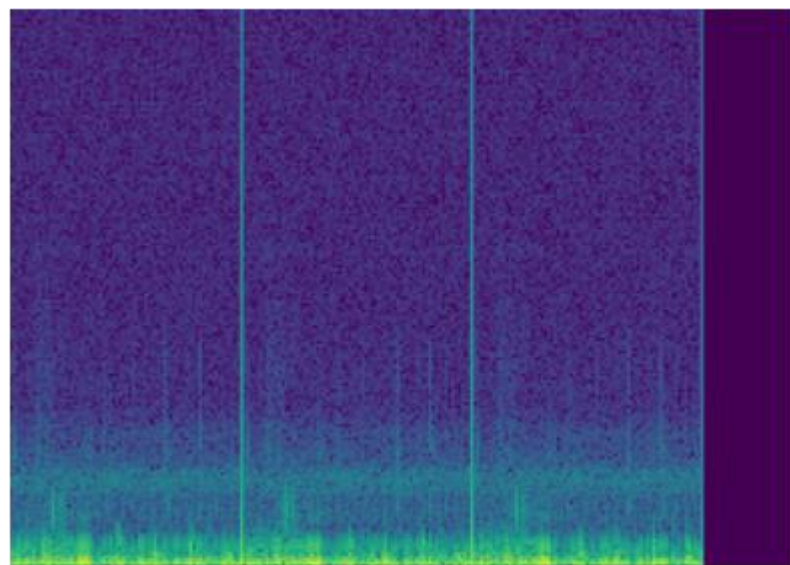
肺音データの加工

肺音データを5秒間のデータに編集する。2.5秒未満のデータは5秒を超えない限り、コピーして5秒に近づける。

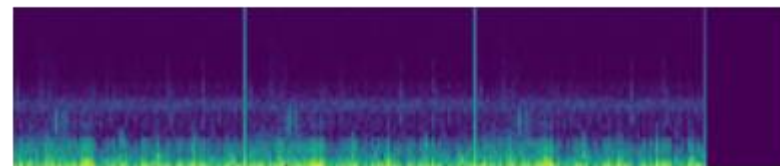
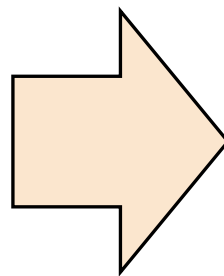


肺音データの加工

情報圧縮と特徴抽出のためにMFCC(振幅に三角窓をかけて対数をとる)という手法を用いてスペクトログラムを変換する。



(257, 245, 1)



(50, 245, 1)

患者データの加工

病名のデータは肺音の異常から判断されるものなので、使用しない。

BMI、体重、身長データの欠損が多かったため、今回は使用しない。

よって今回は、年齢と性別のデータを使用する。

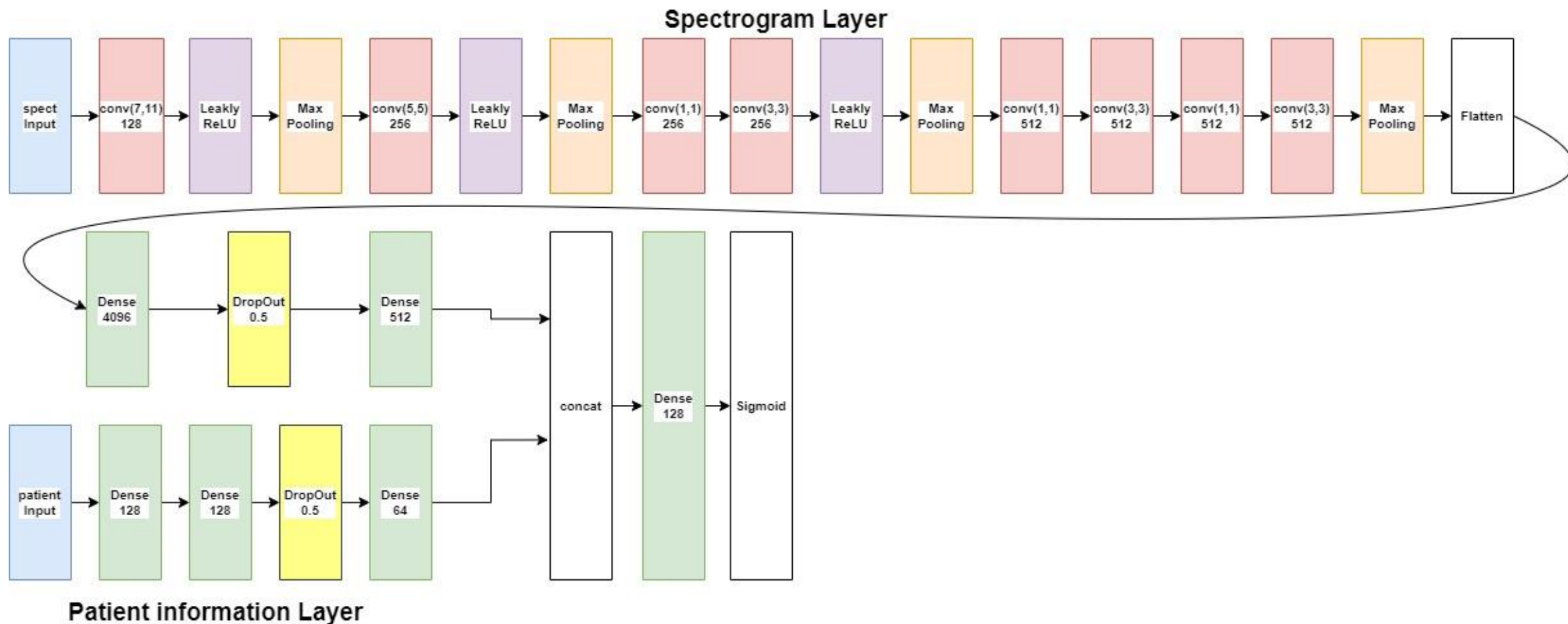
ニューラルネットワークに与える前に、性別は0・1に変換し、年齢は0～1に正規化する。

ニューラルネットワークの構成

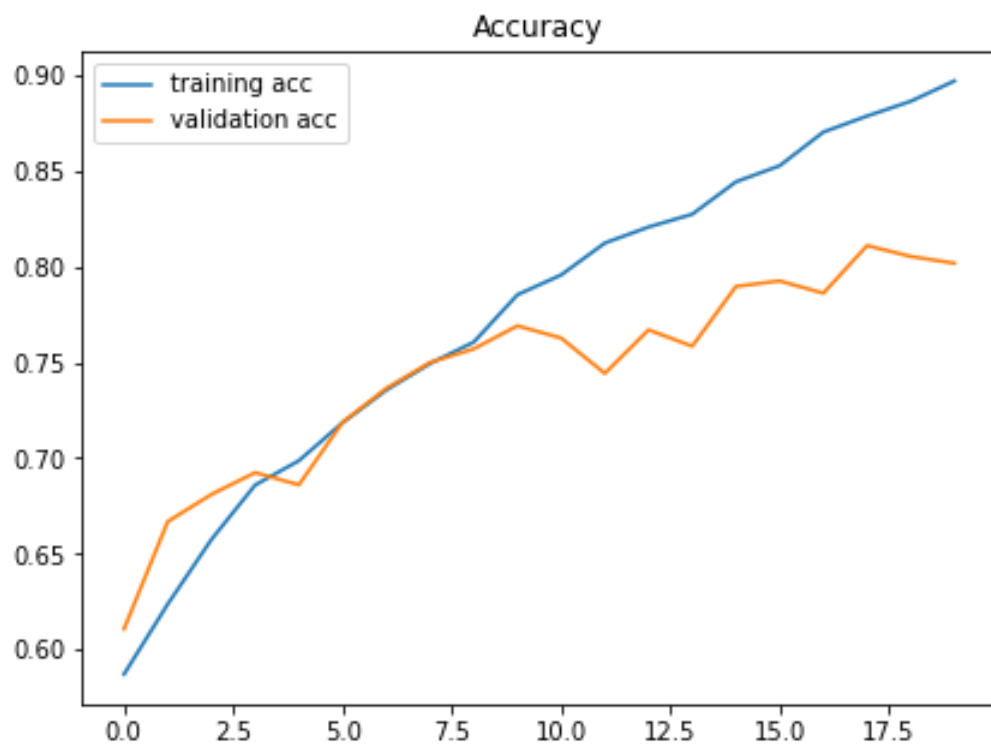
- ・メルスペクトログラムと患者データをネットワークに与える



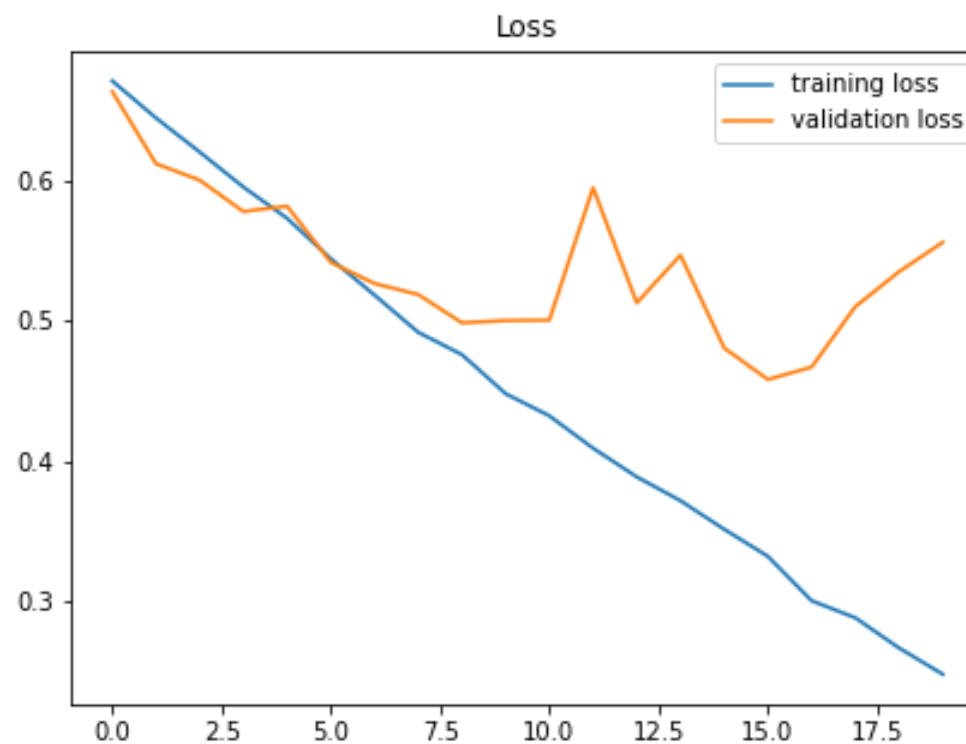
ニューラルネットワークの構成



訓練結果



最良は0.81



最良は0.45

テストデータでの評価

予め訓練データと分けておいた、正常な肺音データ(none)758個とcracklingとwheezingを含む肺音データ(noise)667個をテストデータとしてモデルに適用し、結果を評価する。結果は以下の表の通りである。

テストラベル/予測ラベル	none	noise
none	562	196
noise	88	579

正解率	適合率	再現率	F1値
0.80	0.74	0.86	0.80

考察・正解率と患者の背景データ

結果は適合率が低く、再現率が高かった。これは偽陽性が多く、偽陰性が少ないことを意味し、正解率、F1スコアとしては8割程度となった。これは参考にした先行研究のノイズ無しのF1スコアとほとんど変わらない。

先行研究が今回の研究と異なる点は、先行研究がnone, crackling, wheezing, bothの4クラス分類である点と、患者の背景データである年齢と性別を使用していない点である。

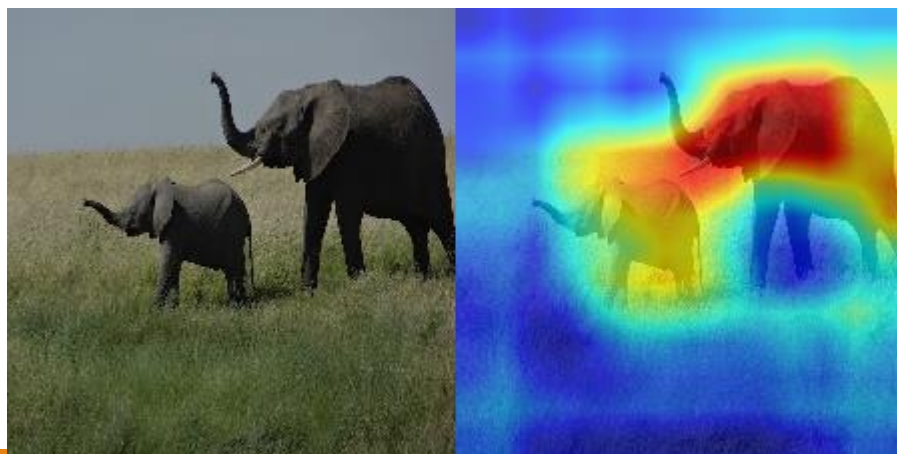
使える情報が年齢と性別の分だけ増えているため正答率が上昇することが期待されるが、今回の結果だけで判断するとそうはなっていない。あくまでも今回の研究の結果だけで判断すると、肺の異常音の判別には性別や年齢はあまり重要ではないと思われる。

Grad-CAMでの評価

ネットワークが何を重要視しているのかを理解する手がかりとしてGrad-CAMによるネットワークの可視化を考える。

Grad-CAMとはRamprasaath R.Selvarajuらが提案した多クラス分類CNNモデルにおける「視覚的な説明」手法である。

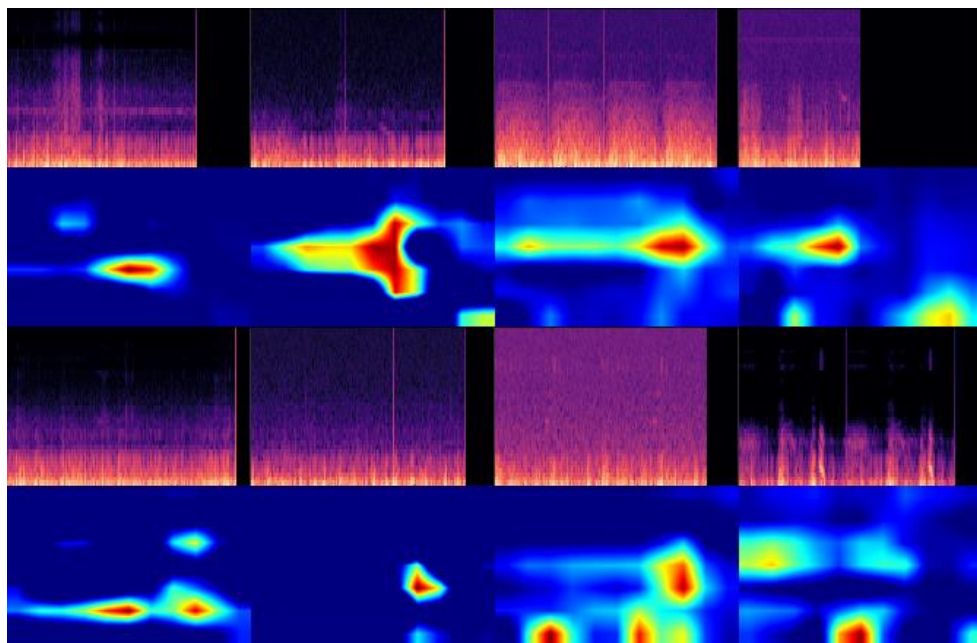
損失関数の勾配とレイヤーの出力を掛け合わせることで、勾配が急峻なレイヤーの出力が強調される。それをヒートマップとして変換した後にテスト画像と重ね合わせれば、テスト画像のどの部分が損失値減少に寄与したのか分かる。



Grad-CAMでの評価

ヒートマップを重ね合わせるにはデータを画像として扱わなければならないので、チャンネル数3の画像をネットワークに与える。そのためにネットワークにも手を加えた。

実装する上で、前の例とスペクトログラムの形が異なってしまったが、表現としては変わらない。以下は、ノイズ有りのスペクトログラムに対してヒートマップを重ね合わせたものである。



考察・ネットワークの可視化

モデルがCNNを通じて画像からどのような学習をしているのかを把握すべく、Grad-CAMを使用してヒートマップを作成した。

ヒートマップを見ると、高周波数部分や低周波数部分に赤い部分が広がっており、特に規則性が無いように思われる。

私は肺音診察の専門家ではないので、専門家が見れば有意な部分が赤くマッピングされているのかもしれないが、いくつかのヒートマップでは右半分の黒い部分が赤くなっている。

今回の研究では全ての肺音を5秒の音声に編集しており、5秒に満たない音声では無音な部分が生じる。その無音な部分がスペクトログラム右半分の黒い部分である。その場合、異常音有無の判別の根拠としてその部分が赤くマッピングされるのは適当ではないのは明らかである。

今後の方針検討

- RNN-CNNアーキテクチャを使用したモデルで精度が向上したとの論文を見つけたので、実装する。
- 使用していない患者データを使用する。
- 短時間フーリエ変換や、その他の関数のパラメータを変更し、訓練する。
- メル尺度やMFCCを使用せず、そのままに訓練する。
- ゼロパディングではなく、何らかの音声を加える。
- データ数をもっと増やす。
- 誤分類したデータを吟味する。

参考文献

- Deep Neural Network for Respiratory Sound Classification in Wearable Devices Enabled by Patient Specific Model Tuning (<https://arxiv.org/pdf/2004.08287.pdf>)
- CNN: Detection of wheezes and crackles (<https://www.kaggle.com/eatmygoose/cnn-detection-of-wheezes-and-crackles>)
- Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization (<https://arxiv.org/pdf/1610.02391.pdf>)