

カスケード識別器の出力を入力ベクトルとする フュージョン識別器の開発と胸部 X 線 CT 画像による評価

阪本正治^{†1} 中野宏毅^{†2}

概要: 近年, 低線量で高精細な X 線 Computed Tomography 画像をスクリーニングし, 早期に肺癌を発見し死亡率を下げる取り組みが行われている. しかし, 大量の胸部 X 線 CT 画像を読影し結節を判別することは放射線科医の負担が大きい. 筆者らは, 選択的なサブサンプリングを用いたカスケード・ニューラルネットワークによって, 結節候補画像から精度良く結節を識別できることを示した. この方法では, カスケードの段数に応じて複数の識別モデルが得られる. 本論文では, それら複数の識別モデルが出力する確率値をベクトル化し, それを入力として新たな識別器を学習し, 高性能な識別器を構成する方法を提案する. 本手法を胸部 X 線 CT 画像上の肺結節の識別に適用したところ, スキャンあたり 4 個と 8 個の偽陽性を許容するとき, それぞれ 94.4%と 95.9%の感度を達成した.

キーワード: 畳み込みニューラルネットワーク, コンピュータ支援診断, CT 画像, 結節, アンサンブル学習, FROC

Developments of Fusion Classifier Using Input Vectors Generated by Cascaded Neural Networks and Its Evaluations by Lung X-ray CT Images

Masaharu Sakamoto^{†1} Hiroki Nakano^{†1}

Abstract: Lung cancer occupies a high percentage in the mortality rates of cancer on a worldwide basis. Early detection is one of the most promising strategies to reduce lung cancer mortality. However, to distinguish nodules from the vast amount of CT scan data is a laborious work for radiologists. We showed that cascaded convolutional neural networks can classify the nodule candidates precisely in our previous study. In this paper, we propose Fusion classifier trained by vectorized nodule probabilities produced by the multiple models generated from the cascaded convolutional neural networks. The proposed method achieved the sensitivity of 94.4% and 95.9% at 4 and 8 false positives per scan in Free Receiver Operating Characteristics (FROC) curve analysis, respectively.

Keywords: convolutional neural network, computer-aided diagnosis, X-ray CT, lung nodule, ensemble learning, FROC

1. はじめに

日本における, 部位別にみた癌の死亡数は肺癌が最も多く, しかも死亡率は年々上昇している[1]. また, その傾向は世界的である[2]. 一方, 肺癌は早期に発見することで死亡率が下がることも知られている[3]. 近年, 胸部 X 線 CT (Computed Tomography) 画像をスクリーニングし, 早期に肺癌を発見する取り組みが行われている. 最新の低線量 X 線 CT 装置は, 高精細な画像(例えばスライス厚 1mm 間隔)を短時間に撮影でき, 1 回のスキャンあたり 300 枚程度のスライス画像を生成できる. しか

し, それらをくまなく読影し, 大きさが数ミリ程の結節(円形の陰影)を見つけ出すことは, 放射線科医の負担が大きく, 胸部 X 線 CT 画像から結節を自動的にスクリーニングするコンピュータ支援診断システム(CAD: Computer-aided Diagnosis system)の開発が進んでいる[6].

肺結節のスクリーニングにおいて, 放射線科医は胸部 X 線 CT 画像のなかに肺結節に特徴的な形や大きさの陰影があるかどうか, またその陰影が特徴的な位置にあるかどうかなどを診る. そのような放射線科医の知識をコンピュータアルゴリズムに実装したのが肺結節スクリーニングのための CAD である. 胸部 X 線 CT 画像から肺結節を自動的に識

提出日: 2017 年 08 月 22 日

^{†1}: sakamoto@jp.ibm.com, 東京ソフトウェア&システム開発研究所, ワトソンヘルス開発, ファンデーション・サービス開発 (Watson Health Development, Tokyo Software & Systems Development Laboratory (TSDL))

^{†2}: 東京ソフトウェア&システム開発研究所, ワトソンヘルス開発 (Watson Health Development, Tokyo Software & Systems Development Laboratory (TSDL))

別するには、まず、肺結節の候補を検出し、次に、その候補を肺結節とそうでないものに識別する。

2009 年に行われた肺結節の自動スクリーニングアルゴリズムの比較調査[3]では、当時の代表的な 6 つの手法がとりあげられている。それらは放射線科医が肺結節を見つけるために使っている画像上の特徴を特徴量として肺結節候補を検出し、次に統計処理に基づくマッチング、あるいは k 近傍法や 2 層ニューラルネットなど機械学習を使ったパターンマッチングを用いて、肺結節を識別している。

しかし、従来の検出手法では血管や肋骨などの陰影(これを正常例と呼ぶ)が結節候補として大量に検出される。一方、結節(これを異常例と呼ぶ)が検出されるのはごく稀である。そのため、検出された候補画像を結節とそうでないものに分ける識別器を構成しようとする時、学習は不均衡学習(Imbalanced learning)となり、識別精度が低下することが知られている[12]。

筆者らは、肺結節の識別における不均衡学習問題に対処するために、畳み込みニューラルネットワーク(Convolutional Neural Network, 以下 CNN)[10]をカスケード接続し、明らかに正常であるサンプルを取り除きながら学習とテストを進める選択的なサブサンプリングを提案し、結節候補画像から結節を精度良く識別できることを示した[4]。この手法では、カスケードの段数に応じて複数の異なる特性を持つ識別モデルが得られる。

今回、筆者らは、それらのモデルから得られる識別結果を融合させることによって、さらに精度の良い識別モデルを構成できるのではないかと考え、各識別モデルに結節候補画像を入力し、得られた複数の結節確率から確率ベクトルを生成し、それらを入力とする新たな識別器を構成する方法を開発した。筆者らはその識別器をフュージョン識別器と名付けた。

フュージョン識別器を用いることによって、選択的なサブサンプリングを用いたカスケード・ニューラルネットワークよりも高い性能が得られることを確認した。

2. 先行研究

2.1. アンサンブル学習

別々に学習した複数の学習器を融合させて汎化能力(未学習データに対する予測能力)を向上させた一つの学習器を作成することをアンサンブル学習[7]と呼ぶ。アンサンブル学習には次のようなバリエーションがある。

1. バギング

各識別器を独立に学習し、テスト入力を各識別器に入力し結果を得て、それらの結果の多数決あるいは平均を最終結果とする。

2. ブースティング

各識別器は、逐次的に依存関係を持って学習される。顔検出で用いられるアダブースト[5]はブースティングの一種である。弱学習器(ランダムな選択よりはましな程度)をカスケードし、前段の識別器で誤ったサンプルを重視して、学習を進める。筆者らが開発したカスケード・ニューラルネットワークもブースティングの一種である。

3. 混合エキスパート

データ空間を複数の識別器で領域分担する。各識別器が得意な領域を分担する。

フュージョン識別器は、バギングに近いが、多数決や平均を使わず、新たな学習モデルを用いることが一般的なバギングとは異なる点である。

2.2. 肺結節の識別におけるアンサンブル学習

Dou ら[11]は異なるレベルの文脈情報を組み込んだ 3 つの 3D 畳み込みネットワークを設計し、それらのネットワークが生成する結節確率の線形和として最終的な結節確率を求めている。これは、典型的なバギングである。胸部 X 線 CT によるスキャンあたり 4 個と 8 個の偽陽性を許容するとき、それぞれ 90.7%と 92.2%の感度を得ている。

Setio ら[8]は、図1に示すように 9 種類のスライス画像を並列に CNN に入力し、最終段で全結合ニューラルネットと接続し、それらの結果を統合することによって、複数の CNN モデルを統合して結節確率を導出している。この手法は、胸部 X 線 CT によるスキャンあたり 4 個と 8 個の偽陽性を許容すると

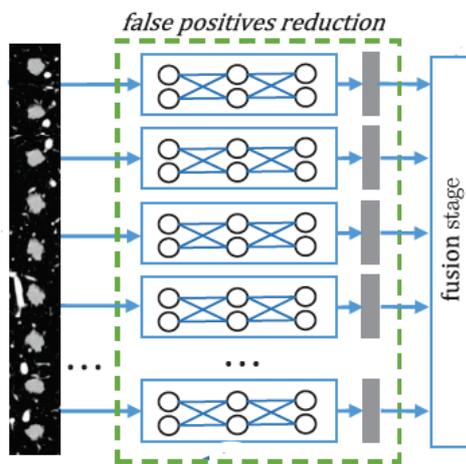


図1 Multi-View Convolutional Networks [8]

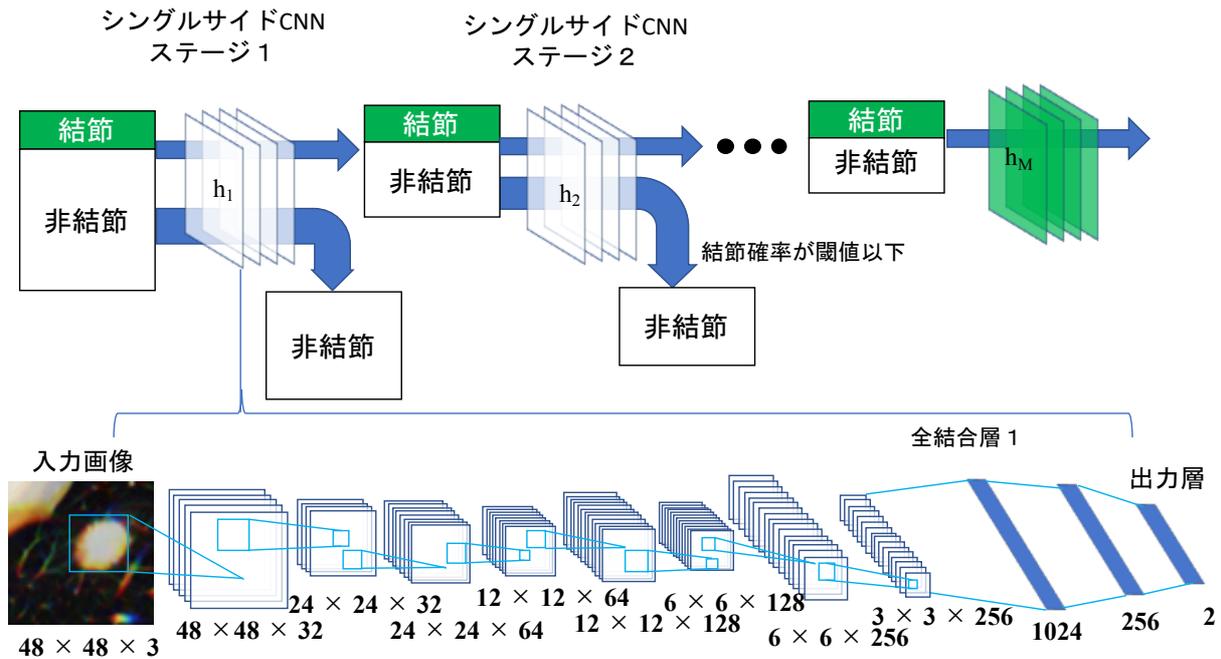


図 2 カスケード・ニューラルネットワークによるテスト過程の模式図[4]

き、それぞれ 85.4%と 90.1%の感度を得ている。この手法も一般的なアンサンブル学習とは異なる。また、ニューラルネットワークモデル内にフュージョンステージを持つ点は我々の手法とも異なる。

3. カスケード・ニューラルネットワーク

筆者らは、これまでの研究において選択的なサブサンプリングを用いたカスケード・ニューラルネットワークを開発し、結節候補画像から精度良く結節を識別することを示した[4]。図 2 にテスト過程の模式図を示す。図中のシングルサイド CNN ステージは、CNN とフィルタプロセスから成る。CNN は結節候補画像の結節確率を計算し、フィルタプロセスは結節確率が閾値より低いテストサンプルを正常例と判定し、結節確率として零を割り当てる。シングルサイド CNN は、異常例のサンプル数を正常例に比して極端に多くした不均衡な学習データを用いて学習した CNN である。シングルサイド CNN ステージは従続接続され、各ステージで明らかに正常例であるサンプルを取り除いていく。そして最終段では異常例と見分けがつきにくい正常例と異常例とが残される。最終段では同数の正常例と異常例で学習した CNN を用いる。本論文ではバランス CNN と呼ぶ。各段のニューラルネットワークの構造は全て同じであり、図中の下部にその構成を示している。

学習時にもステージを経る毎に結節確率の低いサンプルが取り除かれ、異常例と見分けがつきにくい正常例と異常例とが後段に渡される。最終段で

は、正常例と異常例を同数サンプリングし CNN モデルを学習する。したがって、各ステージのシングルサイド CNN は異なる識別特性を有し、バランス CNN も前段のシングルサイド CNN の段数に応じて異なる特性を有数する。例えば、1 段のシングルサイド CNN ステージをバランス CNN につなげる場合、2 段のシングルサイド CNN をバランス CNN につなげる場合などである。最大 5 段のシングルサイド CNN をバランス CNN に接続した場合、5 つの特性の異なるバランス CNN が得られる。

4. フュージョン識別器

フュージョン識別器では、複数の CNN モデルから得られる結節確率をベクトル化し、4 層ニューラルネットワークに入力する。このニューラルネットワークの出力を最終的な結節確率とする。

4.1. 確率ベクトルの生成

図 3 に結節候補の確率ベクトルを計算する過程を示す。胸部 X 線 CT 画像から切り出した結節候補画像を各モデルに入力し、複数の結節確率を計算する。一つの結節候補画像を M 個のモデルに入力し、長さ M の結節確率ベクトルを得る。全結節候補画像数を N とすると N 個の確率ベクトルが得られる。

4.2. フュージョン識別器の構造

図 4 にフュージョン識別器で用いる 4 層ニューラルネットワークを示す。入力は長さ M の確率ベクトル

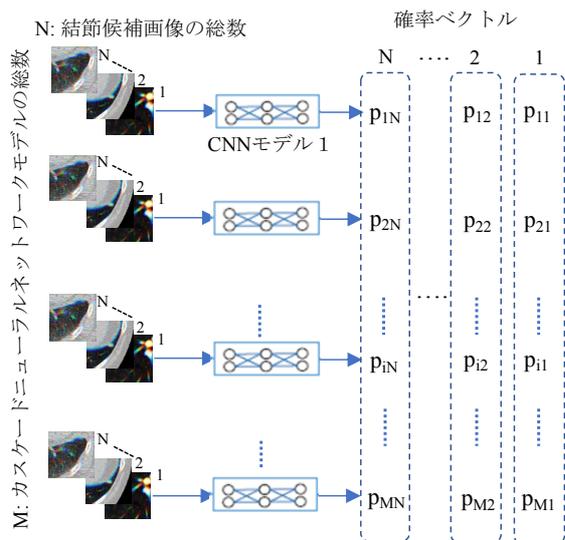


図3 結節候補の確率ベクトルの計算

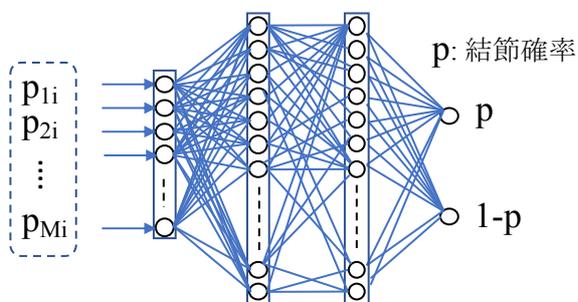


図4 フュージョン識別器による結節確率の計算

である。中間層は全結合のニューラルネットワークであり、出力層のノード数は2である。

4.3. 画像資料

本実験では, Setio ら[8], Dou ら[11]の実験と同様に, 肺結節の自動識別アルゴリズムのベンチマークを目的とした Lung Nodule Analysis 2016[9](以下, LUNA16)で提供されている胸部 X 線 CT 画像データを用いた。

LUNA16 では 888 人分の胸部 X 線 CT 画像がデータセットとして提供される。各スキャンは 4 人の熟練した放射線科医によってアノテーションが付与されている。4 人が疑いなく正常と判断されたものを正常データとし, 4 人のうち 3 人が肺結節であると判定し, かつ大きさが 3mm 以上のものを異常例としている。888 スキャンの中の肺結節候補として 551,065 箇所が調べられており, そのうち 1,348 箇所が異常例と判定されている。また, データセットは交差検証のために 10 セットに分割されている。結節候補には中心位置座標 (x, y, z) が付与されている。この座標を, x 座標と y 座標はピクセル番号に,

z 座標はスライス番号に変換し, 中心位置に対して x 方向と y 方向に前後 64 ピクセル, スライス方向に前後 1 スライス計 3 枚の画像を抽出し, 3 枚の画像を RGB の各チャネルに対応させ png 形式の画像を作成する。

図 5 に png に変換した正常例と異常例を示す。この画像の中心に結節候補が存在するが, 正常例では, それは肋骨の縁や血管あるいは過去に罹患した病気の痕跡などであり, 異常例では, それは結節の陰影である。

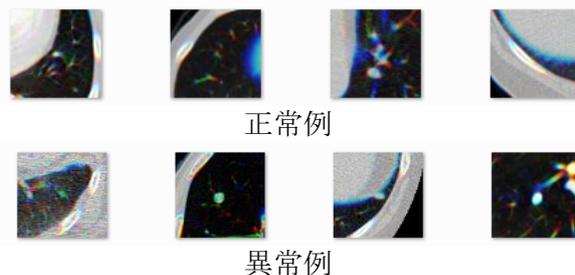


図5 結節候補画像

64 ピクセル×64 ピクセル×3 スライス

なお, 異常例をアップサンプリングするために, 画像に回転操作を施す。その時 64 ピクセル×64 ピクセルの画像から 48 ピクセル×48 ピクセルの画像を抽出するので, CNN に入力される画像サイズは 48 ピクセル×48 ピクセルとなることを付記する。

4.4. 実験環境

実験は IBM Cloud が提供する bare metal server 上で行った。OS は Red Hat Enterprise Linux Server release 7.3, ハードウェアスペックは表 1 のとおりである。

ニューラルネットワークを実装するためのツールキットがオープンソースとして各社から提供されている。それらを使えば, 畳み込みニューラルネットの実装は容易である。本研究では TensorFlow[14]を利用した。

表 1 ハードウェアスペック

CPU	Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz, 4 CPUs / 24 cores
Memory	64GB
HDD	1TB

4.5. 評価指標

結節の有無を判別するような 2 値分類器の分類結果は, 異常例を異常と判別する真陽性(True Positive), 正常と判別する偽陰性(False Negative), 正常例を異常と判別する偽陽性(False Positive),

正常と判別する真陰性(True Negative)の4通りに分類される. ここで, 真陽性の個数を N_{TP} , 偽陽性の個数を N_{FP} とすると, 感度 TP_{rate} は以下の式で計算される.

$$TP_{rate} = \frac{N_{TP}}{(N_{TP} + N_{FP})}$$

どんなサンプルも陽性であると判定する識別器は感度が高いが, 偽陽性の数も多くなる. したがっ

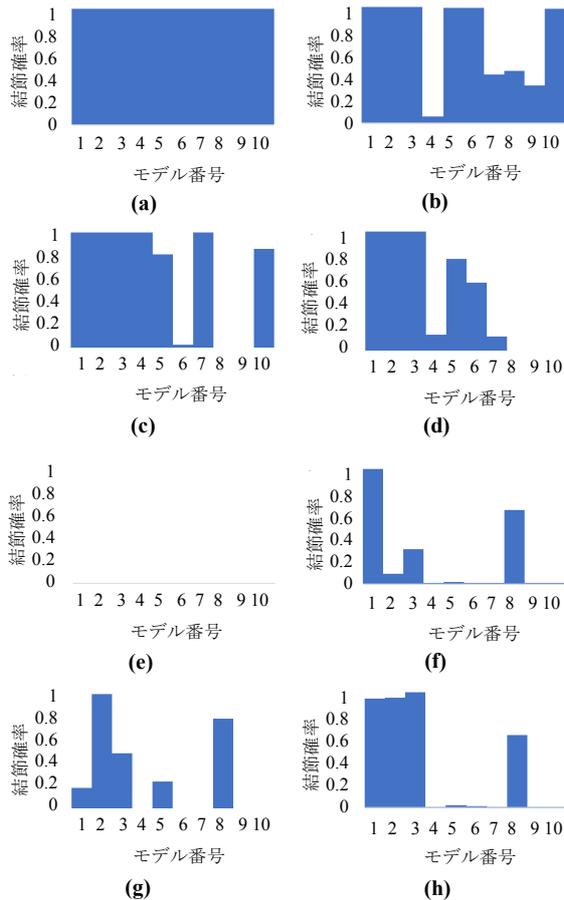


図6 確率ベクトルの実例 (a)から(d)は異常例, (e)から(h)は正常例

て感度が高く, かつ偽陽性の数が少ない識別器が性能の良い識別器であると言える. 本論文で用いる性能評価手法は FROC (Free Receiver Operation Characteristic) [13] と呼ばれ, 1 スキャン(一人分)の胸部 X 線 CT 画像中に含まれる偽陽性の数を平均 1/8 から 8 個許すように結節確率の閾値を変化させた時の感度を指標としている.

5. フュージョン識別器による識別実験

5.1. 確率ベクトル

カスケード・ニューラルネットワークから 10 個のモデルを選択した. そのうち 5 つがシングルサイド

CNN であり, 残り 5 つが, 前段に 1 個から 5 個のシングルサイド CNN を持つ最終段のバランス CNN である.

図 6 に確率ベクトルの例を示す. 1 から 5 のモデルがシングルサイド CNN であり, 6 から 10 は最終段のバランス CNN モデルである. 異常例の場合, 多くは(a)のように 10 個の確率値がほぼ 1.0 になるが, (b), (c), (d) のような分布を示すものもある. 正常例の場合も同様に多くは(e)のように 10 個の確率値がほぼ零になるが, (f), (g), (h) のような分布を示すものもある. 4 層ニューラルネットワークは, これらの多様な分布を学習する.

5.2. モデルの学習と識別実験

長さ 10 の確率ベクトルをフュージョン識別器に入力した. 入力層のノード数 M は 10, 中間層のノード数は 70 と 20, 中間層の活性化関数にはランプ関数(ReLU: Rectified Linear Unit), 出力層の活性化関数にはソフトマックス関数を用いている. 過学習を防ぐため, ドロップアウトを 0.5 としている. 学習データとして, 異常例の画像に回転及び拡大操作を施し, 異常例のサンプル数を 9 倍にアップサンプリングした. また, 正常例はアップサンプリングした異常例と同数になるように, 無作為にダウンサンプリングし, 均衡学習を行なった. ニューラルネットの学習は 300 エポックで停止した.

5.1 で説明したように, データセットは 10 セットに分割されているので, そのうち 9 セットを学習データとして用い, 1 セットをテストデータとして用いて, 10 通りの組み合わせで交差検証を行なった.

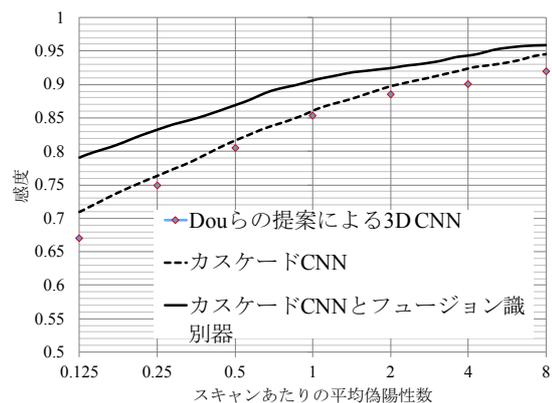


図7 識別性能. (グラフを見やすくするため FROC カーブの上限と下限は省略している.) Douらの結果は参考文献[11]から転記

5.3. 結果と評価

結果の評価には LUNA16 [9] で提供される

FROC 計算プログラムを利用した. Dou ら[11]も同じ計算プログラムを利用している. また, 先に述べたように, 結節のデータも LUNA16 で提供されるものを使っているため性能比較が可能である.

図 7 に FROC を示す. マークで示しているのが Dou らの結果である. 点線はカスケード・ニューラルネットワークによる FROC カーブ, 実線は本提案手法である. カスケード・ニューラルネットワークの性能は, すでに Dou らの結果を上回っているが, カスケード・ニューラルネットワークのモデルを使ってフュージョン識別器を構成することによって, さらに性能が向上している.

6. おわりに

今後, この成果を識別だけでなく, 胸部 X 線 CT 画像全体からの結節の検出にも利用する予定である. 従来の肺結節候補の検出では大量の偽陽性を含むことが問題であった. しかし, 本手法を応用すれば肺結節である可能性が高いものだけを検出できるので, 偽陽性を大幅に削減できる可能性があり, かつ同時に識別もできる.

例えば, 今回開発したフュージョン識別器を利用して 48×48 の画像ブロックをスライドしながら結節確率を計算し, 肺結節を検出した例を図 8 に示す. 黄色い領域が肺結節と判定された箇所であり, 放射線科医の診断と一致する. このように, 本手法を適用することにより, 放射線科医と同等以上の検出精度が得られることが期待される.



図 8 肺結節の検出例

謝辞

本研究に当たって, ご助言をいただいた横浜南共済病院放射線科医長の荻原浩明氏に深謝いたします.

参考文献

- [1] 厚生労働省 平成 26 年我が国の人口動態, Retrieved July 31, 2017, <http://www.mhlw.go.jp/toukei/list/dl/81-1a2.pdf>
- [2] Cancer Research: Lung cancer mortality statistics, Retrieved July 31, 2017, <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/mortality>
- [3] Ginneken, B. V.: Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study, Medical Image Analysis 14, pp. 707–722 (2010).
- [4] Sakamoto, M., Nakano, H., Zhao, K. and Sekiyama, T.: Multi-stage neural networks with single-sided classifiers for false positive reduction and its evaluation using Lung X-ray CT Images, Image Analysis and Processing - ICIAP 2017, pp. 370–379 (2017).
- [5] Viola, P., Jones, M.: Fast and robust classification using asymmetric adaboost and a detector cascade, in: Proceedings of the Neural Information Processing Systems Conference, Vancouver, pp. 1311–1318 (2001).
- [6] El-Baz, A., et al.: Computer-aided diagnosis systems for lung cancer: challenges and methodologies, International journal of biomedical imaging 2013 (2013).
- [7] 上田修功: アンサンブル学習, 情報処理学会, コンピュータビジョンとイメージメディア, vol. 46, No. SIG 15(CVIM 12), Oct. (2005).
- [8] Setio A. A. A., et al.: Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks, IEEE Trans. on Medical Imaging, Vol. 35, No. 5, pp. 1160–1169 (2016).
- [9] Lung Nodule Analysis 2016. Retrieved July 31, 2017, <http://luna16.grand-challenge.org/>
- [10] 岡谷貴之: ディープラーニングによる画像認識-畳み込みネットワークの能力と限界, 情報処理 Vol. 56, No. 7, pp. 198–204 (2015).
- [11] Dou, Q., Chen, H., Yu, L., Qin, J. and Heng, P. A.: Multi-level Contextual 3D CNNs for False Positive Reduction in Pulmonary Nodule Detection. IEEE Transactions on Biomedical Engineering, 64(7), pp. 1558–1567 (2016).
- [12] Japkowicz, N.: Learning from Imbalanced Data Sets: A Comparison of Various Strategies, AAAI Technical Report WS-00-05 (2000).
- [13] Chakraborty, P., Winter, H.: Free-Response Methodology: Alternative Analysis and a New Observer-Performance Experiment. Radiology, vol. 74, pp. 873–881 (1990).
- [14] TensorFlow: Retrieved July 31, 2017, <https://www.tensorflow.org/> Dou, Q., Chen, H., Yu, L., Qin, J. and Heng, P. A.: Multi-level Contextual 3D CNNs for False Positive Reduction in Pulmonary Nodule Detection. IEEE Transactions on Biomedical Engineering, 64(7), pp. 1558–1567 (2016)