

奈良先端大において取り組みたい研究テーマについて

試験区分：情報科学区分

氏名：榎田修慎

希望研究室：生体医用画像研究室

現在の専門：深層学習

1 これまでの修学内容

学部では1年次に線形代数学・解析学など、専門的な学習を始めるために必要な基礎を学び、2年次からは人工知能基礎・応用、データベース、自然言語処理など専門的な講義が開講され、積極的に受講した。そしてこの時期から、将来取り組みたい分野を見つけるきっかけ作りとして、資格取得に向けて学習をはじめ、基本情報技術者・応用情報技術者を取得した。3年次には、KaggleやSIGNATEなどでデータサイエンスに興味を持ち、特にKaggleの胸部X線画像のカテテルクラス分類コンペティションへの参加は医用画像工学に興味をもつきっかけとなった。4年次の研究室配属では、脳型情報処理や人工知能を専門としている研究室に配属され、Vision Transformerを用いた画像解析の研究をしている。現在はKaggleの植物の病気を診断するコンペティションで、Vision Transformerモデルを使った実装をしており、今後はモデルの評価をもとに改良を加えたり、他のモデルとの比較を行うつもりだ。

2 取り組みたい研究テーマ

私が、奈良先端大において取り組みたい研究テーマは「Self Attentionを用いた生体医用画像のマルチモーダルセグメンテーション」である。本稿では、この研究テーマの研究背景、研究課題、提案手法、期待される結果について述べる。

3 研究の背景

近年、ディープラーニング技術は実務レベルの利用が急速に拡大しており、医療分野にも多く応用されている。中でも「診断支援」の領域では、ディープラーニングの持つ特徴抽出能力を用いることで、生体医用画像から病変のある箇所を抽出する技術が注目されており、コンピュータ診断支援 (Computer-Aided Diagnosis : CAD) と呼ばれている。この技術は「医師への負担を

緩和させ、病変の見落としリスクを低減させる」ことが目的であるが、病変箇所と解剖学的構造物が重なった場合や、病変の規模が小さい場合は抽出することが困難である。そのため、誤診断や医師による二重チェックなどが発生し、根本的な解決には至っていない。

病変を抽出する手法としてセグメンテーションが用いられることが多く、最近ではU-netを利用した研究が頻繁に行われている^[5]。U-netでは、畳み込み層を重ねることで失われていく画像構成要素の位置情報を保持するために、Skip Connection機構が採用されている。また、U-netはさまざまな派生手法が提案されており、Zongwei Zhouらの研究によるとU-netを改良したU-net++^[4]は元のモデルより、肺小結節を正確に抽出できることがわかっている。しかし、U-netをはじめとするセグメンテーション手法は、セグメンテーションしたい対象が重なっている場合や複数ある場合に効果を発揮しづらい。その問題を解決するために、インスタンスごとにセグメンテーションを行うモデルが開発され、Mask R-CNNと呼ばれている^[1]。Mask R-CNNは同一クラスに属する物体をインスタンスごとにセグメンテーションできるので、重なった同一病変一つあたりの形や大きさを正確に抽出することが可能になる。例えば、肺小結節は、画像の濃さや大きさから処置が必要か否かを判断するため、個々のインスタンスに対するセグメンテーションは大きな意味を持つ。このことから、Mask R-CNNの病変抽出精度を向上させることができれば、過剰抽出や欠損が無いロバスト性の高いモデルができると予想される。

4 研究課題

上述の通り画像による診断支援を実務で利用するには、病変抽出精度の向上を目指すべきである。そのためには先に記した通り、解剖学的構造物同士の重なりや、病変部位との重なりによる過剰抽出や欠損を減らし、誤診断をどのように解決するかを考える必要がある。

5 提案手法

私は二つの視点から医用画像を解析する手法 (図 1) を提案する。一つ目は自然言語処理の Self Attention^[3] の視点。二つ目は画像処理の Mask R-CNN の視点である。これらを組み合わせ、文章と画像双方からアプローチを行うことにより、病変の正確なセグメンテーションができると私は考える。

図 1 は Mask R-CNN のアーキテクチャの一部を抜粋し改変したものである。ここでは、画像と文章の入力から、すでに提案されている Mask R-CNN のアーキテクチャに接続するまでの処理を説明する。初めに、医用画像から想定される病変や、病変の位置を文章として与える。その文章を単語埋め込み^[6]により固有のベクトルに落とし込み、位置エンコーディングにより位置情報を付加する。これにより得られたベクトルを特徴量として、CNN で得られた特徴量と組み合わせ、マルチモーダル特徴量として扱う。次はマルチモーダル特徴量を元に Self Attention を求める。図 1 の例では、入力文章として「右肺の中葉に結節の疑い」を与え、CNN によって得られた特徴量と組み合わせて Self Attention を求めている。この場合、入力文章の Attention は「右肺」・「中葉」・「結節」で大きくなり、「の」・「に」・「疑い」で小さくなると考えられる。以上により得られた Attention を元の特徴量に掛け合わせて重み付きの特徴量を算出。次に、RPN^[2]によって得られた物体候補領域を重み付き特徴量と共に ROI Align に入力する、という流れだ。その後、通常の Mask R-CNN と同様に、マスク推定ブランチと全結合層を経てセグメンテーションされた出力が得られる。

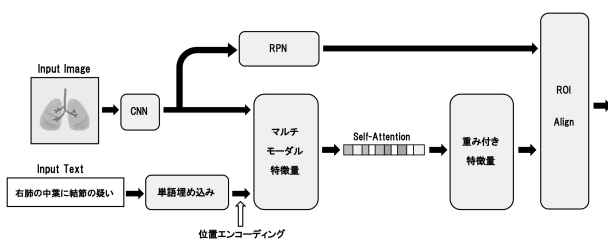


図 1: マルチモーダルなセグメンテーションの流れ

6 期待される結果

私が提案する Self Attention を用いた手法は、CNN 単体で特徴量を抽出した場合に比べ、病気の場所や種類に関する情報を多く保持しているので、過剰抽出や欠損の少ない結果が安定して得られると考える。また、Attention を求めていることから、画像・文章のどこに注目しているかを可視化できるので、患者に対する病症の説明も比較的容易になると予想される。

7 最後に

本稿では、私が奈良先端大で取り組みたい研究テーマである「Self Attention を用いた生体医用画像のマルチモーダルセグメンテーション」の研究概要について述べた。奈良先端大は、医用画像工学を学べる数少ない大学院の一つであり、学部を持たずに、様々なバックグラウンドをもった人間を受け入れており、そのサポートも充実している。また、全ての講義が英語で行われており、早期からグローバルに活躍するための力を養うことができ、同時に高度な技術力を身につけられる環境が備わっている。そのため、私は貴大学院を強く志望するのである。

参考文献

- [1] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
- [2] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*, 2015.
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.
- [4] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 3–11. Springer, 2018.
- [5] 近藤堅司, 小澤順, 清野正樹, 藤本真一, 田中雅人, 安達登志樹, 伊藤春海, 木村浩彦. U-net を用いた胸部 x 線画像からの解剖学的構造の領域抽出. 人工知能学会全国大会論文集 第 32 回全国大会 (2018), pp. 2J404–2J404. 一般社団法人 人工知能学会, 2018.
- [6] 堅山耀太郎. Word embedding モデル再訪. *オペレーションズ・リサーチ*, Vol. 11, pp. 717–724, 2017.