

カプセルネットワークを用いた CT 画像からの脂肪領域抽出 Segmentation of Fat Region on CT Images using Capsule Network

竹下 智章
Tomoaki TAKESHITA

1. 序論

医用画像に対する深層学習を用いた診断法において、NN を用いて特定の臓器や器官などに対応する領域を画像上で色分けする目的で領域抽出 (セグメンテーション) の機械学習手法が用いられている。医用画像の領域抽出は、関心領域提示による医師への診断補助や、自動鑑別における特定領域の自動抽出においても必要であるため、様々な手法が研究されている。

医用画像の診断を NN 等の深層学習によって行う上でしばしば問題になるのが、組織の形状や位置などの医用画像間での特徴差である。身体断面の画像である CT や MR の医用画像において、部位によっては、撮影する身体上の位置がわずかに移動するだけで画像内の組織の特徴が大きく変化する。このような部位における深層学習を用いた画像診断では、組織の形状位置に特徴差を持った医用画像という条件下で全ての画像に対し同一の高い精度で抽出しないしは検出を行うことが求められる。

特徴差の大きな医用画像の深層学習による診断の分野ではデータ量の確保が重要となるが、症例数の少なさや、専門家による注釈付きのデータを大量に得るコストなど観点から必ずしも多くのデータを得られるとは限らないのが実情である。

本研究では、撮影断層位置の差や撮影タイミングで組織の形状が大きく変化する、かつ組織の位置関係に個人差のみられる部位の医用画像に対して、高精度の領域抽出が可能なモデルを導出することを目的とする。また、そのモデルが、学習データ量が少量の段階においても特徴を学習可能であることを示す。画面内の物体の構造や位置関係の認識と学習に優れた性質を持つカプセルネットワークを用いたモデルを適用する。本モデルが空間的な特徴の認識において、ピクセルレベルでの従来型の学習手法と比較して高い精度での領域抽出を可能とすることを示す。

2. カプセルネットワークを用いた脂肪領域抽出

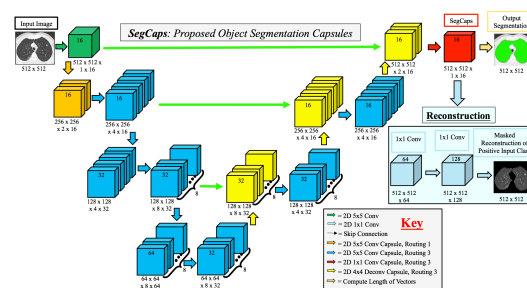


図 1: SegCaps の構造

カプセルネットワークとは、2017 年に登場したニューラルネットワークの新手法である。通常のニューラルネットワークのニューロンをカプセルという新しい概念に置き換えることで、(1) 空間的特徴の抽出に優れる、(2) データのクラス不均衡に頑健、(3) データ量の少なさに耐性がある、などの性質を持つとされる。SegCaps はカプセルネットワークをもとに考案されたモデルである。カプセル層ではこの多次元ベクトルが単位となり、通常のニューラルネットワークにおけるニューロンに代わってカプセルという概念として導入される。カプセル内のベクトル成分は、画像中に存在する特定のエンティティの様々な特性を表す。これらには、位置、サイズ、方向、色相、テクスチャなど、さまざまな種類のパラメータが含まれる。カプセル層は通常の完全結合層のように層を深くすることが可能で、入力側に近いほど物体の大きさや形状などの下位の概念を表し、出力側に近いほどそれらの位置関係や比率などの空間の全体的な特徴を表すようになる。このカプセル層によってピクセルレベルの学習よりも上位の概念となる空間的特徴の認識と、単純な画素の画面内

位置への依存の防止を同時に満足することが可能となる。

SegCaps^{*1}は、カプセルによる完全結合層を畳み込み層に置き換えることで、大幅にパラメータ数を少なくしながらも空間的特徴の抽出を可能としている。この畳み込みカプセル層を U-Net の畳み込み層に対応する位置に配置した構造になっている。512 × 512 の入力の場合、同一の深度を持つ U-Net よりもパラメータ数が 1/30 ほどに抑えられる。

ここで、少量のデータを学習に用いた際に生じる問題がある。本研究は、学習用データ数が少ないという条件下で行っている。少量の画像データで学習をする場合、バリデーションデータの枚数は数枚とかなり少数になる。バリデーションデータの選び方によっては画像間の特徴が似たようなものに偏りやすく、指標として正しく機能しない。そこで本研究では、通常、バリデーションデータには行わないデータ拡張を施すことで擬似的にバリデーションデータに多様性を持たせる。これにより、未学習データへの精度が過大評価されることを抑制し、過学習防止の指標として正しく機能することを狙った。具体的には、バリデーションステップごとにランダムで選出されたバリデーション画像に、さらにランダムで回転や平行移動、拡大等のアフィン変換を施し、バリデーションに用いた。

3. 評価実験

3.1. データセットと実験設定

評価実験には、組織の形状や位置に差のある医用画像として、骨盤領域を撮影した CT 画像 5 人分 (48 枚) に対し、脂肪領域を塗り分けたマスク画像のセットを用いた。各画像の身体上における撮影間隔は 2.5mm で、1 人分は最少で 8 枚、最多で 11 枚の画像で構成される。

tiramisu と比較し評価する。tiramisu は、SegCaps と同様に U-Net を派生させたモデルで、パラメータ数を抑えながら高精度な抽出が行えるように改良したもので、パラメータ数が SegCaps の 1.5 倍程度であるため、類似の抽出手法として比較対象とした。形状類似度である Dice 係数、Jaccard 係数を評価指標とし、領域抽出の精度を比較する。トレインデータは 3 人分、テストデータは 2 人分、Split によってそれらの組み合わせを変えて平均をとる。

^{*1}Rodney LaLonde, Ulas Bagci :“Capsules for Object Segmentation”, 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), 2018.

表 1: 実験結果

	Split	Dice (%)	Jaccard(%)
SegCaps	Split-1	92.10	85.42
	Split-2	86.52	76.48
	Split-3	90.63	83.56
	平均	89.75	81.82
tiramisu	Split-1	71.15	56.27
	Split-2	75.17	60.94
	Split-3	74.59	59.48
	平均	73.63	58.89

表 2: 各モデルのパラメータ数 (入力 512 × 512)

Model	Pram.
SegCaps	1,425,649
tiramisu	2,314,208

3.2. 実験結果と考察

Dice 係数、Jaccard 係数ともに SegCaps の精度が高いことが示された。画素特徴を学習したと考えられ、明領域と暗領域が両方含まれる脂肪領域を区別できていない tiramisu に対し、SegCaps は位置関係から正しく領域抽出を学習している。

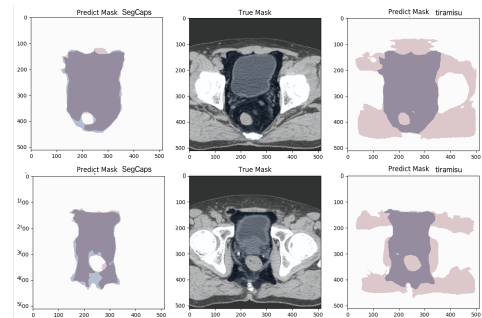


図 2: 抽出サンプル

4. 結論

カプセルネットワークを用いたモデルにより、画像間で組織の形状や位置に差異があり、かつデータ量が少量の場合において、ピクセルレベルの学習モデルと比較して高い精度での脂肪領域抽出が可能であることが示された。