

教師なし学習によるがん検出

平田 凌歩[†] 井上 浩孝^{††}

† 呉工業高等専門学校専攻科 プロジェクトデザイン工学科 †† 呉工業高等専門学校 電気情報工学科

1. はじめに

近年、ディープラーニングはコンピュータービジョンにおいて優れた成績を残し、今後も多くの分野での活用が期待されている。特に、画像分野における発展は目覚ましく、医療分野での利用が提案されているものもある。しかし、教師あり学習はその特性上、モデルを学習する際にラベリングされた学習用のデータを多く必要とする。このため、学習用データセットを作成するコストが大きく、これらを活用したい医療現場では、昨今の感染症の対策の必要もあるため、作成する余裕がない場合が多い。したがって、ラベリングの必要がなく、画像そのものから学習できる教師なし学習の利用が望まれているが、教師あり学習に比べて研究例は少ない。本研究は、昨年時に教師なし学習を用いて皮膚がんをクラスタリングする手法の提案を行った、本学校の卒業生の研究を引き継いだものである。

2. データセット

本研究には、HAM10000[1]のデータセットを利用する。このデータセットは色素性皮膚病変の皮膚鏡画像からなるデータセットである。このHAM10000 データセットからメラノーマ画像 1113 枚、ランダムに抽出したメラノサイト性母斑画像 1113 枚、計 2226 枚を用いて学習を行う。また、それらの画像を訓練用データセットと検証用データセットの2つに分けて、利用する。

3. Invariant Information Clustering (IIC)

IIC とは、クラスタリングやセマンティックセグメンテーションを教師なしで可能とする手法である。モデルの基本構造は畳み込み、Batch Normalization, ReLU 関数を用いた一般的な CNN モデルとなっている。このモデルは、入力画像とランダム変換により変換された入力画像の二つを入力し、モデルから出力された確率分布の相互情報量を最大化することで表現を学習することができる。

4. 実験方法

構築した IIC モデルについて、以下に示す様々な条件下で学習と検証を行った。入力の種類として、画像に対してグレイスケール画像に変換したもの、RGB カラー画像、RGB カラー+グレイスケール画像の3種類、それ

ぞれ1次元、3次元、4次元画像に変換したものをを用いた。IICの内部モデルにはResNet-18を用いた。転移学習については、RGB カラー画像に対して ResNet-34, ResNet-50 に内部モデルを変更して実験を行った。Data Augmentation についてランダム返還前の RGB カラー画像に対して Random Crop, Random Horizontal Flip, Random Vertical Flip を行ったものを用いて実験を行った。内部モデルには ResNet-18 を用いた。最後に精度がよかった条件を組み合わせモデルを構築し、教師なし学習を行った。また、このモデルに対して全体の訓練データ数の1%に当たる20枚のラベリングされたデータを用いて半教師あり学習を行った。これらのモデルは最適化手法 Adam を用いて 300Epoch 学習させた。

5. 結果とまとめ

本研究では、IICによる教師なし学習での皮膚がんの分類を行い、80.9%の精度で分類することができた。実験結果より、グレイスケール画像にすることで肌の色などにロバストになり、精度が向上すると考えられる。また、一般的な画像と比較して分類が難しい医用画像などに対しても、ImageNet による事前学習や Data Augmentation は有効であると分かった。IICでは教師なし学習でセマンティックセグメンテーションも行えるため、IICによるクラスタリングとの併用は精度向上に有用であると考えられる。

半教師あり学習を用いて皮膚がんの分類を行ったところ85.4%の精度で分類できた。実験結果より、教師あり学習であるResNet-18の87.6%と比較しても高い精度で分類できていることが分かる。これは半教師あり学習を用いることで、教師なし学習で得た特徴量に対し、分類に有用な特徴量が選択できるため、少ない教師データで呼応精度を達成できたと考えられる。また、半教師あり学習においてはSoTAな半教師あり学習手法simCLRが有用であると考えられるため、今後はこの手法を組み込むなどして研究に取り組んでいきたい。

参考文献

- [1] The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions.
<https://doi.org/10.7910/DVN/DBW86T>