

説明可能なメラノーマ自動診断のための 多様な皮膚病変画像の生成と病変特徴の移植

小尾 佳純 村林誠也 彌富 仁

法政大学大学院 理工学研究科 応用情報工学専攻

概要

皮膚がんの一種であるメラノーマは非常に悪性度が高く、早期発見が必要不可欠である。熟練の専門医においても診断が難しいメラノーマの診断精度向上を目的として、深層学習を用いた自動診断システムによる診断支援の研究が進められている。深層学習モデルの訓練には教師ラベルが付与された大規模なデータセットが必要であるが、医療分野におけるラベルの作成コストは非常に高い。本研究では、診断が確定した多数の良性腫瘍画像から GAN を応用したネットワークを用いて悪性病変のもつ特徴をスタイル変換させることで、多様な悪性特徴をもつ病変画像を生成する手法を提案する。これらの生成された画像の品質、およびそれらを用いて深層学習モデルで評価した結果について検証した。

1 はじめに

メラノーマ (悪性黒色腫) は世界的に増加傾向にあり、2018 年の全世界における患者数は 239,700 人ほどと推定されており、早期発見が重要である [1]。がんの進行度によって 5 年生存率が大きく異なり、初期ステージである I 期ではおよそ 90% 以上であるのに対し V 期では 10% 未満と生存率は格段に悪化する [2]。しかし、初期のメラノーマは良性腫瘍である母斑と視覚的に酷似しており、熟練の皮膚科専門医による皮膚病変画像を用いた拡大鏡での診断でさえも、初期ステージである I 期では診断精度はおおよそ 75%~88% に留まる。このような背景を受け、メラノーマの病変を早期発見するための高精度な識別を目的とした、画像認識による自動診断システムの構築が進められている。

先行研究では、画像認識分野において主流である深層学習によるアプローチが提案されており、高い精度を記録している。特に、最先端の結果を実現している

深層学習による事前学習済みモデルと半教師あり学習に加え、体毛を擬似的に付与する data augmentation によって精度 87.2% を達成している [3]。また、こうした深層学習手法は予測の判断根拠が不明瞭なため、臨床において医師が皮膚病変に対して視診する際、定量的な判断を下すために用いられる 7-point checklist を応用した研究が行われている。具体的には、皮膚科医 4 名によってアノテーションされた 7-point checklist を用いて識別器を構築することで、これらの指標に基づく解釈性の高い診断システムを実現している [4]。

こうした深層学習のアプローチは高い精度を記録している一方で、多量のデータが必要である。しかしながら医師によるアノテーションコストは極めて高く、十分な量のデータを確保するのは簡単ではない。一般物体認識では、generative adversarial network (GAN) による data augmentation 手法が複数提案されている。しかし GAN は多量のデータを要するため医療分野では直接適用するのは困難である。

そこで本研究では、比較的データの多い良性腫瘍画像を症例数の少ない悪性の特徴を持つ画像にスタイル変換することで data augmentation を行った。特に、7-point checklist の主要項目の一つで悪性度の高い Atypical pigment network (不規則な網構造) に着目し、スタイル変換により同特徴の良性腫瘍への付与を試みた。また、このときの悪性・良性の識別結果への影響について評価した。

2 手法

2.1 悪性のスタイルを持つ皮膚腫瘍画像の生成

敵対的生成ネットワークの学習モデルの 1 つである CycleGAN [5] を用いて比較的データ数の多い良性腫瘍画像を、7-point checklist の中でも特に悪性の特徴として診断されている Atypical pigment network (不規則な網構造) にスタイル変換し、症例の少ないデータの生成を試みた。

2.2 生成画像の良性・悪性推定

皮膚腫瘍画像により事前学習済みの ResNet50 [6] ベースのモデルを用い、2.1 において Atypical pigment

Generation of various pigmented skin lesion images and transferring dermoscopic structures towards explainable melanoma diagnosis

Kasumi OBI and Seiya MURABAYASHI and Hitoshi IYATOMI

Applied Informatics, Graduate School of Science and Engineering, Hosei University

184-8584, Koganei, Tokyo, Japan

{kasumi.obi.6e@stu., iyatomi@}hosei.ac.jp

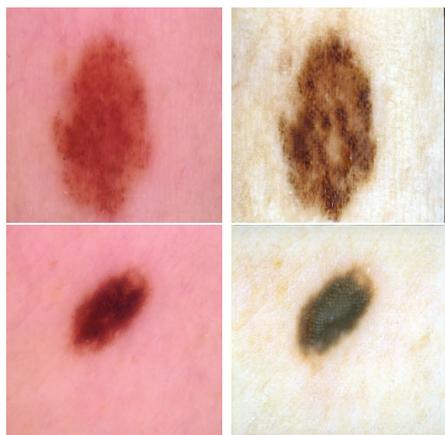


図 1: 左: 元画像 (良性) 右: スタイル変換後画像

表 1: 悪性度比較結果

	悪性診断数 (枚)
元画像	130
付与画像	269

network の特徴付与による良性・悪性推定結果への影響を評価した。

HAM10000 のデータセットのうち良性腫瘍画像 6,705 枚と悪性画像 1,113 枚を 100epoch 学習した ResNet50 [6] ベースの事前学習モデルを用いて良性・悪性の二値分類を行う。

3 実験

3.1 使用したデータセットの詳細

本実験では、良性腫瘍画像として HAM10000[7] から 6,705 枚、悪性腫瘍画像として 7-point checklist のラベルが付与された 7-point criteria evaluation database[8] から Atypical pigment network の特徴を持つ画像 611 枚を使用した。

3.2 評価方法

良性と診断された皮膚病変画像が、Atypical pigment network の特徴を付与された際、上記の識別器により悪性と予測された症例数の変化の有無を検証する。比較対象は、HAM10000 の良性腫瘍画像 6,705 枚と、これに Atypical pigment network の特徴を付与した同数の生成画像であり、上記の条件で悪性を示す画像数の変化を確認する。

4 結果と考察

良性画像と、良性画像を元に Atypical pigment network の特徴を持つ画像にスタイル変換した画像の例を図 1 に示す。目視による評価は専門医の確認が必要だ

が、スタイル変換した画像は、Atypical pigment network 特有の不規則な網構造が付与されている事を確認できた。CycleGAN で生成された画像には不規則な網構造が付与されていた。良性画像に特徴を転移するという点では、多様な画像が生成できたと捉えられるが、定量的な評価ができる指標を導入することが課題である。

HAM10000 のデータセット及びスタイル変換画像、各 6,705 枚を識別した際の悪性度比較結果を表 1 に示す。悪性と診断される画像数は生成画像の方が増加したが僅かであった。Atypical pigment network の特徴は 7-point checklist の中でも比較的悪性度が高い診断指数であるが、総合的な診断では他の項目も考慮した上で診断が下されるため、悪性診断数が大幅に増加するような結果にならなかったと考えられる。

5 おわりに

生成画像におけるデータの水増しは一定の効果が見られたが改善の余地があることがわかった。今後の目標として、皮膚病変診断における説明性に着目し、7-point checklist のアノテーション済みのデータを用いて、今回着目した特徴以外の特徴を持つ画像を生成する試みを進めていきたい。また元画像とスタイル変換後の画像を比較すると、腫瘍領域以外の部分である肌の色調も変化していることがわかる。今後は領域に注目したスタイル変換を比較することが望まれる。

参考文献

- [1] A. C. Society, “Global cancer facts & figures 4th edition,” <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/global-cancer-facts-and-figures/global-cancer-facts-and-figures-4th-edition.pdf>.
- [2] F. L. Meyskens Jr, D. H. Berdeaux, B. Parks, T. Tong, L. Loesch, and T. E. Moon, “Cutaneous malignant melanoma (arizona cancer center experience). i. natural history and prognostic factors influencing survival in patients with stage i disease,” *Cancer*, vol. 62, no. 6, pp. 1207–1214, 1988.
- [3] S. Kitada and H. Iyatomi, “Skin lesion classification with ensemble of squeeze-and-excitation networks and semi-supervised learning,” *CoRR preprint arXiv:1809.02568*, 2018.
- [4] S. Murabayashi and P. Hitoshi Iyatomi, “Towards explainable melanoma diagnosis: Prediction of clinical indicators using semi-supervised and multi-task learning.”
- [5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proc. of IEEE CVPR*, 2017, pp. 2223–2232.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [7] P. Tschandl, C. Rosendahl, and H. Kittler, “The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” *Scientific data*, vol. 5, p. 180161, 2018.
- [8] J. Kawahara, S. Daneshvar, G. Argenziano, and G. Hamarneh, “Seven-point checklist and skin lesion classification using multitask multimodal neural nets,” *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 538–546, 2018.