

CNN を用いた自然画像からの文字状パターンの検出

Image Detection of Character-like Patterns from Natural Phenomena Using CNN

松崎 大雅

指導教員 菊池 眞之

東京工科大学大学院 バイオ・情報メディア研究科 コンピュータサイエンス専攻

キーワード：CNN, 自然画像, 文字, 形状, パターン認識

1. はじめに

近年のデザイン分野における AI の利用は、自動彩色やデザインの評価、生成系 AI などその種類も多岐に渡る。一方で、自然現象などの偶発的に発生するオブジェクトをデザインの着想として考え、検出するような AI の研究は行われていない。そこで本研究では、フォントデザインの着想となる CNN を用いた検出によって図 1 に示すような自然画像中から文字状パターンを獲得する。

従来の画像認識における CNN は、人間と同様の識別能力を発揮するようなものが存在する。一方で、その中でも人間の主観とは異なるラベル付けをするような結果が見られる。その CNN の振る舞いには局所的特徴を捉える傾向があり、テキストチャバイアス[1]や特定のノイズ[2]下での誤認識が見られる。そこで、テキストチャやノイズによらず、形状を評価し、その概形が大域的特徴において典型的であると見なせるパターンであるならば、人間の主観と一致した判断が可能であると仮定する。本研究では、大域的特徴を捉えるために従来の CNN の構造に対して、その特徴抽出機構から得られるエンコードされた深層の特徴マップによってオブジェクトの概形を捉える。これにより、典型的であると評価されるオブジェクトの形状が人間の主観として形状を判断することができるとして、CNN での文字状パターンの獲得を行う。

なお、本研究にて扱う自然画像のデータなどの提供、テーマの考案には、東京工科大学デザイン学

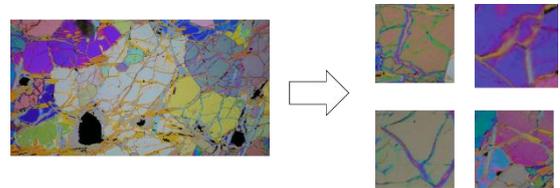


図 1 自然画像

部准教授の中島健太氏の協力を頂いている。

2. 手法

本研究では、大域的特徴による形状を捉えるために、高次層の特徴マップを典型的パターンとしてみなし、ネットワーク中の分布を比較することで文字状パターンの獲得を検討する。

形状を判断する大域的特徴を捕捉するために、高次層の特徴マップを評価することで文字状パターンの獲得を目指す。CNN における畳み込み層から得られる特徴マップを高次元空間上の高次元ベクトルとみなし、特徴マップ同士の類似性を測ることで典型性評価として、人間の主観に一致した検出を行う。

その上で、典型性の評価には、典型的パターンとする画像の入力から得られた特徴マップを鋳型として考える。このとき、典型的パターンとして見なせるデータを CNN に入力した際に、同一ラベル内の反応分布における相加平均を求める。その結果が典型的パターンを持つ特徴マップとする。これによって自然画像を入力したときの反応分布(特

徴マップ)と鋳型の特徴マップを比較することで人間の主観に沿った文字として判断できるかどうかの典型性評価とする。

3. モデルとデータセット

モデルには、物体認識として高い精度を持つYOLOシリーズにおけるYOLOv8[4]を使用する。YOLOにはインスタンスセグメンテーションの拡張もあり、これも併用することで、明示的に位置を示す検出を行う。

学習と典型的パターンとして扱うデータセットには、手書き文字であるNISTが提供しているNIST Special Database 19[3]を利用する。本研究では、その中から、大文字のアルファベットと小文字のアルファベットのクラスにのみ絞り、学習データとする。

4. 特徴マップとの比較手法

相加平均として算出した典型的パターンによる特徴マップと自然画像を入力した際の反応である特徴マップを比較する。このときの距離計算の手法をユークリッド距離が大きい数値を示したデータによって決定する。距離が遠いとされる典型的パターンから算出される全てのラベルにおけるそれぞれの距離手法での結果の中で正しいものを選択する。

5. 結果

ユークリッド距離において大きい数値を示していた正解ラベルがAという文字に対しては、マハラノビス距離以外では、最も近い距離がA以外のラベルを示す結果となっていた。この結果を鑑みて、マハラノビス距離がより良い典型性評価が見込めると考え、この手法での距離の比較を行う。

6. 考察

本研究の手法においては、より良い手法の改善が考えられる。典型性評価を行うデータは、YOLOの出力における矩形領域に従い切り出した画像を固定の画像サイズに変換し、そのデータの反応分布

としての特徴マップを典型的パターンによる特徴マップと比較していた。この画像には、画像全体とその中心に文字が位置しているとみなせる画像である。そのため、検出対象の位置と大きさが画像の比率に対してあらかじめ大きいものだと断定できているのであれば、YOLOのマルチスケール段階における最小の特徴マップ、即ち大域的特徴を内包していると考えられる特徴マップを比較対象とすることで、典型性評価として良い結果が得られることが期待されると考えられる。

一方で、データセットには、単純な文字画像を教師データとしていたが、これらのデータセット依存によることが原因で、自然画像に対する典型性評価が行えていない可能性がある。そのために、より検出対象に合わせて調整した処理によるバリエーションを増やす必要性があると考えられる。

参考文献

- [1] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, Wieland Brendel, "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness", ICLR 2019, 2019.
- [2] Ian Goodfellow, Nicolas Papernot, Sandy Huang, Yan Duan, Pieter Abbeel, Jack Clark, "Attacking machine learning with adversarial examples", OpenAI, <https://openai.com/research/attacking-machine-learning-with-adversarial-examples>.
- [3] National Institute of Standards and Technology (NIST), NIST Special Database 19, <https://www.nist.gov/srd/nist-special-database-19>.
- [4] Ultralytics, yolov8, <https://github.com/ultralytics/ultralytics>.