横浜国立大学 大学院環境情報学府
博士学位論文

画像分類器の説明性向上に関する研究

A Study on Improving the Interpretability of Image Classification Models

情報環境専攻　情報学プログラム

小林　雅幸
Masayuki KOBAYASHI

請求学位　　博士（情報学）
責任指導教官　　長尾　智晴　教授

提出年月日　　令和 3 年 1 月 13 日
請求年度　　令和 3 年度 3 月修了

あらまし

近年，深層学習をはじめとした機械学習による処理の自動化に関する技術が注目されており，様々な分野で研究が活発に行われている．画像分類を例に挙げると，深層学習モデルの一つである Convolutional neural network (CNN) は様々な画像分類タスクにおいて高精度な手法として知られている．しかし，これらの構築された処理は複雑で解析は困難であり，処理過程は未だにブラックボックス化されている場合がほとんどである．CNN などの獲得された処理の多くは社会や産業応用の現場で使用されることが期待されているが，これまでは精度のみが注目され，構築された処理を利用する人の立場にたって注目していることは少ないと考えられる．人間は自身の知的で複雑な脳内の働きを論理的に説明することが困難であるにもかかわらず，機械が構築した処理に対してはその処理の論理的な説明を求めたがる．特に人の命に関わる医療分野や車載カメラを用いた認識などでその傾向が強い．そのため，機械学習を用いた処理が今後，社会や産業で利用されるようにするためには処理の精度だけではなく，処理の可読性の根本的な改革やアルゴリズムの可視化が求められる．そこで本論文では分類精度と可読性の両方に優れる画像分類器の提案を行う．

本論文ではまず，分類器の中でも比較的に人間が理解しやすい手法とされる If-then ルールで分類を行う分類器の精度と可読性を向上させる方法の提案を行う．提案手法では分類に用いられている特徴量分布を考慮したヒートマップの作成と分類器内の各条件分岐における特徴量の可視化を行うことで，分類器が分類の際に画像のどの領域に注目して分類を行ったかを利用者に提示することが可能になる．実験では提案手法を一般画像分類に適用し，獲得した可視化画像の有効性の検証を行う．

次に，学習済み深層学習モデルの可視化手法の提案を行う．先行研究では学習済み CNN の特定ユニットの活性化と自然画像への変換を同時に満たすための制約を人手で設計した上で生成を行うが，提案手法では Generative adversarial networks (GAN) の枠組みを用いることで，自然画像への制約を用いることなく End-to-end で画像生成器を学習することが可能になる．実験では提案手法を学習済み AlexNet に適用し，モデルの内部で獲得された特徴量の可視化を行う．また，任意のクラスに対して強く発火する特徴量の可視化を行うことで，モデル内部で階層的な特徴量が獲得されていることを示す．

最後に, 高い分類精度と可読性をもつ新しい深層学習モデルである Evolutionary generative contribution mappings (EGCM) の提案を行う．先行研究では精度と可読性のトレードオフの問題のため，可読性向上の機構を導入することで分類精度の低下が見られる．しかし，提案手法は可読性の高いモデルに適したモデル構造を進化計算法で自動最適化することで，高い精度と高い可読性をもつ分類器を自動獲得することが可能になる．実験では提案手法を複数のデータセットに適用し，有名な深層学習モデルの性能比較を行うことで，EGCM が高い可読性を保ちながら先行研究と同等精度以上の性能もつことを示す．

Abstract

In recent years, machine learning techniques are becoming popular and have shown good performance in a variety of tasks. Especially, convolutional neural networks (CNNs) have continued to show significant improvement in a variety of computer vision tasks over the past few years. Despite their enormous development, their black-box nature has become increasingly problematic. This black-box nature leaves only two questions: *why and how did they reach their decision?* In addition, their uninterpretable nature may diminish users' trust and can be a barrier to their adoption in applications. For instance, in applications where interpretability is important, users do not want to rely on black-box systems. In order to build trust with users, we believe that classification systems should be able to provide a comprehensive reason for their decision. Therefore, it is important to review work on machine learning interpretability. In this paper, we tackle this machine learning interpretability problem. To this end, we propose methods for designing more interpretable training methods, gaining an insight into how model works, and learning more interpretable models.

Firstly, we propose a new technique for visualizing the feature distribution of rule-based classifiers. This technique allows us to gain a better understanding of classifications and intuitive interpretation. We applied our method on several benchmarks and found our visualizations intuitive.

Secondly, we introduce a visualization method based on generative adversarial networks (GAN), one of the most powerful generative models, to gain an insight into how a CNN works. We applied our method to AlexNet and we visualized their neuron activations. Our method produced comparatively interpretable visualizations, we found that our method is efficient.

Lastly, we take a closer look at CNN interpretability and propose a new method called evolutionary generative contribution mappings (EGCM) for achieving a high classification performance together with high-level interpretability. In EGCM, the networks incorporate both a classification and an interpreting mechanism in an end-to-end manner. We applied the EGCM on several datasets and empirically demonstrate that the EGCM maintains high-level interpretability without sacrificing classification performance.

目 次

図 目 次

表 目 次

第1章　序論

近年，深層学習をはじめとした機械学習による処理の自動化に関する技術が注目されており，様々な分野で研究が活発に行われている．画像分類を例に挙げると，深層学習モデルの一つである Convolutional neural network (CNN) は様々な画像分類タスクにおいて高精度な手法として知られている．しかし，これらの構築された処理は複雑で解析は困難であり，処理過程は未だにブラックボックス化されている場合がほとんどである．CNN などの獲得された処理の多くは社会や産業応用の現場で使用されることが期待されているが，これまでは精度のみが注目され，構築された処理を利用する人の立場にたって注目していることは少ないと考えられる．人間は自身の知的で複雑な脳内の働きを論理的に説明することが困難であるにもかかわらず，機械が構築した処理に対してはその処理の論理的な説明を求めたがる．特に人の命に関わる医療分野や車載カメラを用いた認識などでその傾向が強い．そのため，機械学習を用いた処理が今後，社会や産業で利用されるようにするためには処理の精度だけではなく，処理の可読性の根本的な改革やアルゴリズムの可視化が求められる．ここで処理の自動化に関する研究は様々な分野で研究が活発に行われているが，特に可読性の研究が行われていて提案手法の評価が行いやすい画像分類の分野に注目し，本論文では高精度かつ利用者が安心して利用可能な可読性の高い画像分類器の提案を行う．

本論文で目標とする画像分類器の概略図とその解決方策を図 1.1 に示す．ここでの画像分類器の可読性については様々な定義の仕方が考えられるが，本論文において画像分類器の可読性とは「モデルから判定以外の情報を利用者に提示すること」と定義する．そして，これらの情報を提示することができない状態であることを画像分類器のブラックボックス化と呼ぶこととする．これら分類器から提示する情報の中でも本論文では次の 2 点に着目して手法の提案を行う．

1. 任意の入力画像に対する出力に対して，その出力に至った根拠の説明
2. 分類器が学習した特徴量の可視化による動作原理の説明

以上のように分類器の可読性の意味は広い定義がされているため，各章における可読性の意味が異なる場合が考えられるが，本論文ではこれらすべてを可読性向上と定義することにする．そして，これら可読性の向上によって分類器の動作原理や分類の根拠の解析につながり，ブラックボックスな画像分類器に対する安心感が得られることが期待される．また，構築した分類器の理論的な解析が進むことも期待できる．

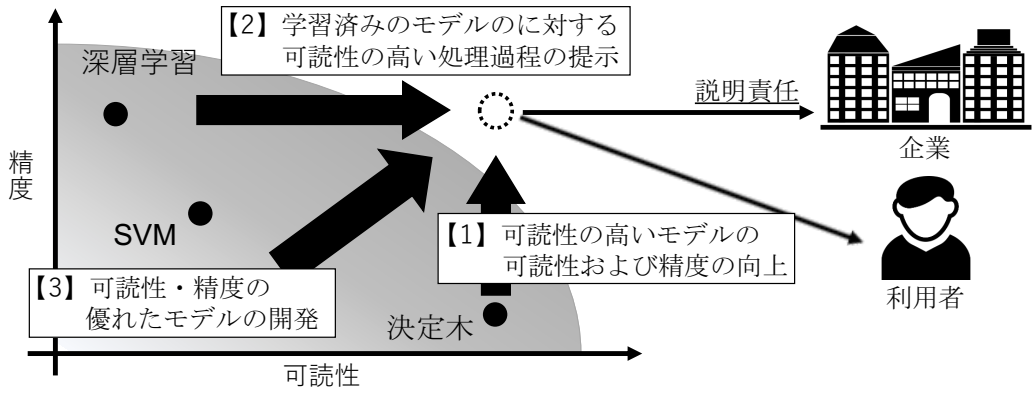


図 1.1: 本論文で目標とする精度と可読性の高い分類器の概略図とその解決方策

1.1 本論文の構成

本論文の構成は次のとおりである．まず第 2 章で上述した画像分類器の可読性向上に関する先行研究について述べる．第 3 章では，比較的に人間が理解しやすいとされる If-then ルールを用いた分類器の精度と可読性向上の提案を行う．ここでは提案手法を一般画像分類に適用し，獲得した可視化画像の有効性の検証を行う．第 4 章では，学習済み深層学習モデルの特徴量の可視化手法の提案を行う．提案手法では Generative adversarial networks (GAN) の枠組みを拡張することで，従来手法で提案されてきた自然画像に関する事前知識や正則化を用いることなく視認性の高い可視化を行う生成器を End-to-end で学習することが可能となる．第 5 章では，高精度かつ高い可読性をもつ新しい深層学習モデルである Evolutionary generative contribution mappings (EGCM) の提案を行う．ここでは提案手法を複数の画像分類問題に適用し，有名な深層学習モデルの性能比較を行うことで，EGCM が高い可読性を保ちながら先行研究と同等精度以上の性能もつことを示す．最後に，第 6 章で本論文のまとめと今後の課題について述べる．

第2章　関連研究

本章では本研究と関連の深い画像分類器の可読性向上に関する先行研究について説明する．まず，比較的に人間が理解しやすい手法とされる決定木などの If-then ルールで分類を行う分類器の可読性向上手法について述べる．特に本章では先行研究として進化的条件判断ネットワークとその分類過程の文章化に関する先行研究について述べる．

続いて，Convolutional neural networks (CNNs) の解析に関する先行研究について説明する．ここでは学習済み CNN を用いた画像分類において分類に有効な領域の特定と特定のユニットを活性化させる画像生成に関する先行研究について述べる．

最後に，高精度かつ高い精度をもつ深層学習に関する先行研究に関する先行研究について説明する．具体的にここでは深層学習の可読性を高めるためのモデル構造に関する先行研究について述べる．

2.1　進化的条件判断ネットワークの分類過程の文章化に関する先行研究

2.1.1　Evolutionary Decision Network (EDEN)

中山らは決定木を拡張し，高精度かつコンパクトで可読性の高い構造の分類器である進化的条件判断ネットワーク (Evolutionary decision network；EDEN) を提案している［4］．EDEN は入力画像の特徴量の大きさで条件分岐を行うノードをネットワーク状に配置した

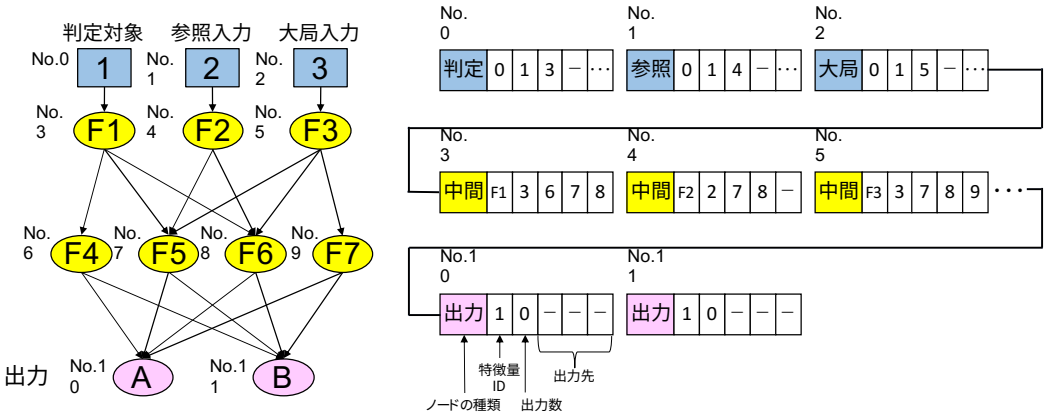
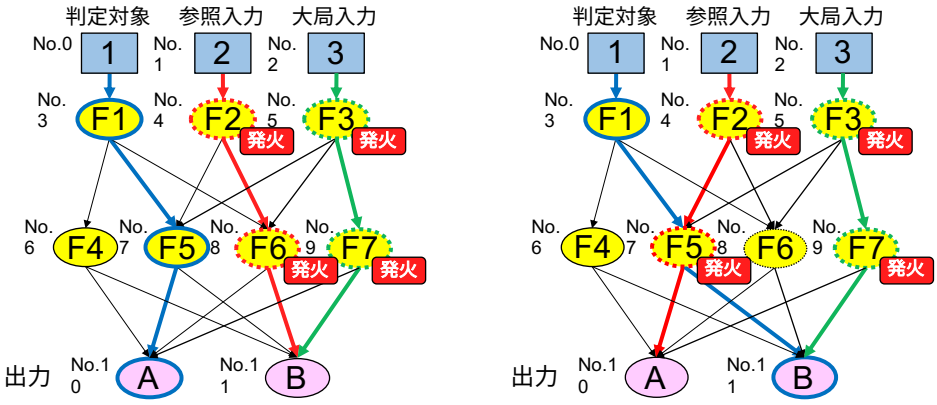


図 2.1: EDEN の遺伝子型と表現型の例



(a) 参照入力による変更がない場合　　(b) 参照入力による変更がある場合

図 2.2: 参照入力による処理の変更例

分類器であり，C4.5 [5]で構築された決定木よりも高精度かつ少ないノード数で可読性の高い分類が可能であることが示されている．

EDEN を用いた画像分類では決定木と同様に画像から算出した特徴量を入力ノードに入力し，中間ノードの条件分岐によって分岐先を決定する．この操作を出力ノードに到達するまで繰り返し行い，最終的に到達した出力ノードに対応するクラスに分類を行う．EDEN はフィードフォーワード型のネットワーク構造で，このネットワーク構造を 1 次元の文字列で表現する．この文字列に対して交叉や突然変異などの遺伝操作を適用することでネットワークの最適化を行う．染色体はネットワークの各ノードに対応しており，ノードの種類や比較に用いる特徴量の種類，ノードの出力先，特徴量のしきい値などのパラメータを保持する．EDEN の表現型であるネットワーク構造とそれに対応する遺伝子型の例を図 2.1 に示す．

また，EDEN は分類を行う判定対象の入力に加えて，判定対象データに関連するデータと画像全体からのデータをそれぞれ参照入力と大局入力として別の入力ノードから入力することで，少ないノードで高精度な分類を可能にしている．具体的には EDEN では中間ノードの発火という状態が提案されており，参照入力が通過した中間ノードは発火状態となり，発火していないノードとは異なる処理が行われる．これによって認識対象の入力だけではなく分類対象の周辺の入力も考慮することができ，それらの相互作用によって少ないノード数で複雑なネットワークの処理を表現することが可能となる．発火による処理の変更例を図 2.2 に示す．この例ではノードの発火時には条件分岐で用いるしきい値が変更されている．

2.1.2　分類過程の言語化

崎津らは，決定木や EDEN などの If-then ルールを用いた分類器の分類過程を文章で説明する手法を提案している [6]．この手法では，あらかじめ分類に使用する特徴量やしきい値に対応する語句を定義した辞書を用意しておき，分類に用いた特徴量やしきい値を辞書の語句と対応付けることで説明文を生成する．分類過程の言語化の流れを図 2.3 に示す．

図 2.3: EDEN の遺伝子型と表現型の例

この手法では 2 つのパラメータを導入することで生成される説明の粒度を変更することができ，分類器の利用者によって異なる粒度に合わせた説明文を生成することが可能である．本論文ではこれらのパラメータで指定した分類過程 (パス) を説明パスと呼び，このパスに対して説明を行うこととする．

まず 1 つ目は説明するパスの数 S である．説明文を生成する際に，分類器に入力したデータの中で通過頻度が高いパスほど分類において重要なパスと考えることができる．そこで，構築した分類器に入力したデータの通過頻度を求め，通過頻度の高いパスから順に S 個のパスについて説明文の生成を行う．この S の個数を 1 とすることで生成される説明を分類器を通過するデータ全体ではなく，その分類器において重要なパスのみに制限することが可能となる．

2 つ目は生成する文章の長さ s である．C4.5 などの手法で構築した決定木はルートノードから情報量が減少するようにノードが配置されるため，ルートノードの近くに配置されるノードがより重要なノードであると考えることができる．そこで，説明文を生成する際には，ルートノードから語句に変換するノードの数 s を指定し，s 個分のノードの説明を行うこととする．この s の個数を説明パス上の最大ノード数 max とすることでルートノードから出力ノードをすべて説明する文章を生成することが可能になる．EDEN の場合は必ずしもルートノードに近いノードほど重要であるとは一概にいうことができないが，崎津らの手法では決定木と同様にルートノードから順に説明文を生成することとしている．

2.2 学習済み深層学習の解析に関する先行研究

ここでは，本研究と関連の深い学習済みの畳み込みネットワーク (CNN) の解析に関する先行研究について述べる．学習済み CNN の解析に関する先行研究は任意の入力画像に対して分類に有効な領域の特定を行う手法と，学習済み CNN の特定ユニットおよび層を活性化させる画像生成の手法の 2 つに大別される．

2.2.1 任意の画像を用いた学習済み CNN の解析方法

CNN を用いたクラス分類の結果の解析を行うために，入力画像においてクラス分類に重要な領域を可視化する手法が数多く提案されている．これらの手法では，画像を学習済み CNN に入力した際の出力から入力画像に対する勾配を求めることで分類の判定において重要な領域の特定を行う．これらの手法における可視化の概略図を図 2.4 に示す．

Simonyan らは入力画像の画素値ごとに勾配を算出することで，クラス分類を行う際の重要領域を示す顕著性マップを生成する手法を提案している［7］．同様に Zeiler らは任意の入力画像に対して画像内のどの画素がネットワークのユニットの活性化に起因しているか提示する手法を提案している［8］．この手法では Deconvnet と呼ばれる機構を導入し，各ユニットの出力を入力画像空間に写像することで画像内の分類に有効な領域の特定を行う．また，彼らは入力画像の一部をマスキングし，マスクの有無で生じる出力差を観測することで分類に有効な領域を特定する手法も提案している．そして，Selvaraju らはネットワークの出力に対して勾配で重み付けを行う Grad-CAM を提案している［9］．この手法では，全結合層における対象クラスの出力を特徴量マップについて偏微分することで対象クラスにおける重要度の可視化を行う．また，彼らは Guided backpropagation の出力と Grad CAM の出力を画素ごとに掛け合わせることで可視化を行う Guided-Grad-CAM と呼ばれる手法も提案している．Bach らは Layer-wise relevarance propagation (LRP) と呼ばれる CNN の解析方法を提案している［10］．これはネットワークの各層間の逆伝搬させることで入力に対する重要度を算出する手法であり，算出された画素ごとの重要度をヒートマップ等で提示することでクラス分類に重要な入力画像の領域を特定することが可能となる．

これらの手法は任意の入力画像に対して画像内のどの領域が分類に重要であるかを特定する上で非常に有効である．一方これらの手法は実際のデータに対する可視化であるため，実際の CNN ネットワーク内部の処理やどのような特徴量が中間層で獲得されたかを解釈することは困難である．

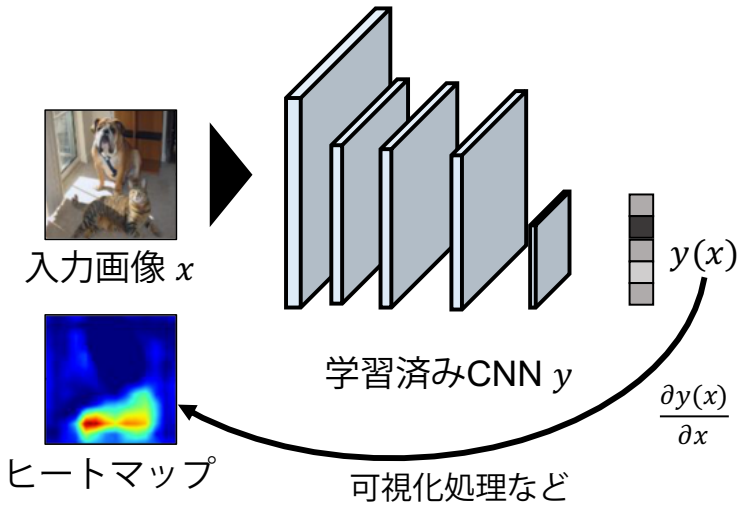


図 2.4: 任意の画像を用いた学習済み CNN 解析方法の概要図

2.2.2　学習済み CNN の特定ユニットを活性化させる画像作成

学習済み CNN の特定ユニットを活性化させる画像生成は，そのユニットの出力に対する最適化問題と考えることができる．つまり与えられた学習済み CNN に対して特定ユニットを活性化させる画像の探索を行う．その際，一般的なニューラルネットワークは入力に対して微分可能であるため，ネットワークの出力に対する勾配を用いて入力画像を再帰的に変化させることで探索を行うことが可能である．

これらのことから，Erhan らは Gradient ascent を用いた学習済み CNN の特定ユニットを活性化させる画像生成の手法を提案している［11］．Gradient ascent を用いた可視化画像の概要図を図 2.5 に示す．この手法では任意の画像を入力した際の出力とその勾配を用いて入力画像の画素を再帰的に更新し，学習済み CNN の特定ユニットを活性化させる画像を生成する．具体的には，乱数を用いて初期化した画像 x の画素値の更新を考える．まず，画像 x を学習済み CNN の入力として特定ユニットの出力 $y(x)$ を算出する．次に誤差逆伝播法を用いて算出された出力 y(x) に対する勾配を算出する．最後に式 (2.1) に示すように，算出された $\frac{\partial y(x)}{\partial x}$ を画像 x に足し合わせることで学習済み CNN の特定ユニットを活性化させる画像を生成する．

$$x = x + \alpha \frac{\partial y(x)}{\partial x} \tag{2.1}$$

ここで α は学習率であり，一回の更新で変化させる画素値の大きさの調節を行うパラメータである．この手法は比較的に簡単な操作で画像生成が可能であるが，この方法で生成される画像は高周波なノイズを多く含む可読性が低い画像が生成されることが知られている［1］．そのため，生成画像の可読性を高めるために Gradient ascent で生成される画像をより自然画像に近づけるための事前知識や画像に対する正則化方法が数多く提案されている．正則化の有無の違いによる生成画像の可読性の違いを図 2.6 に示す．

Simonyan らは Gradient ascent による画像生成に加えて，L2 正則化を導入することで生成画像の可読性を向上させる手法を提案している［7］．これにより獲得された画像に対象クラスの特徴を捉えることが可能となった．Mahendran らは Total variation と呼ばれる近

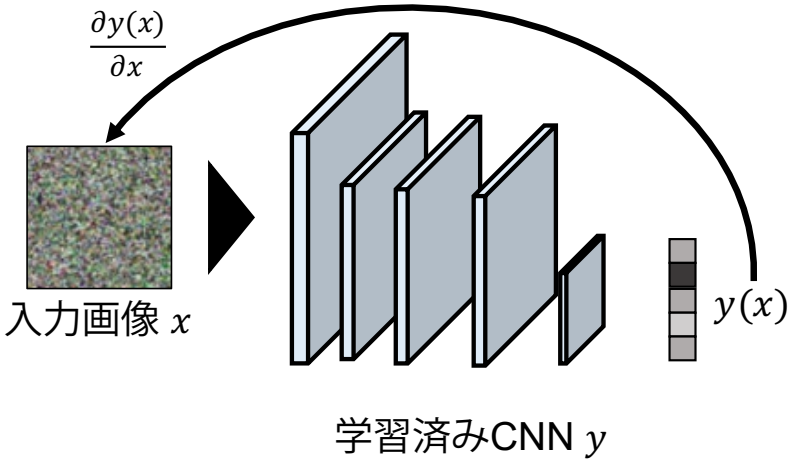
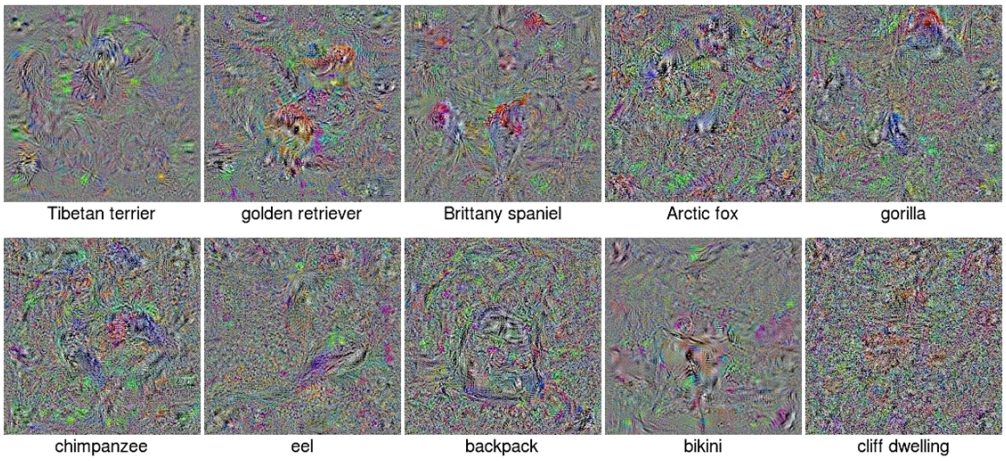
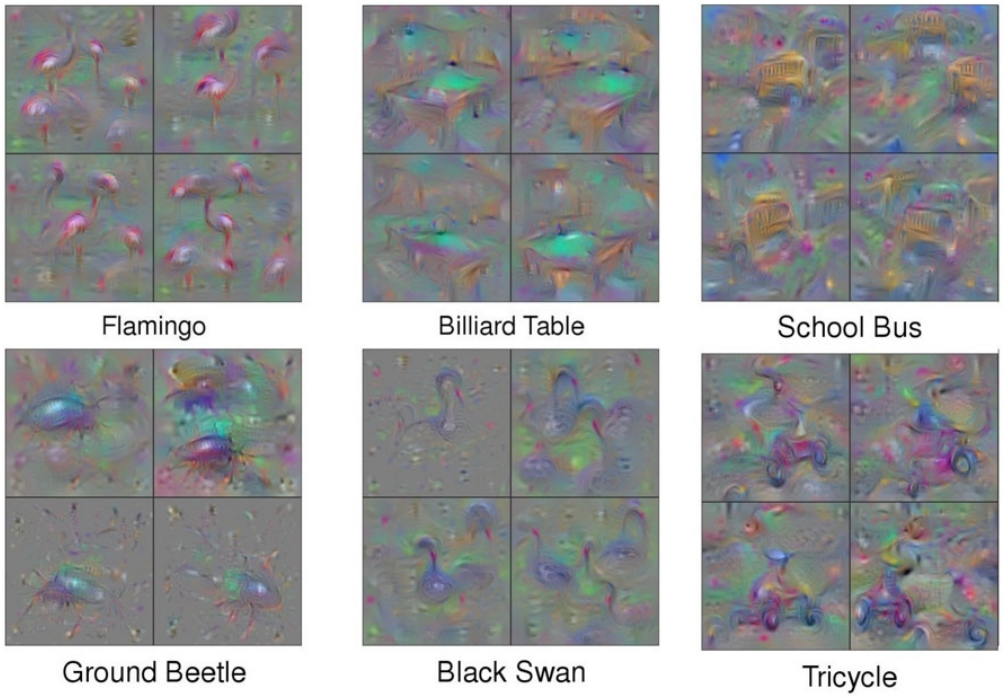


図 2.5: Gradient ascent を用いた可視化画像生成の概要図

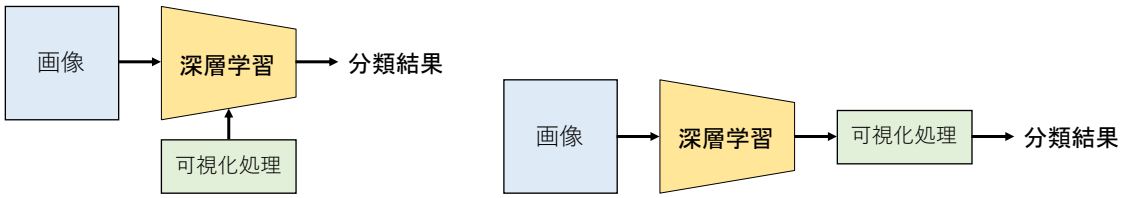


(a) 正則化を用いない場合の可視化画像例



(b) 正則化を用いた場合の可視化画像例

図 2.6: Gradient ascent を用いた学習済み CNN の特徴量の可視化（画像は文献 [1] から引用）



(a) 学習済みのモデルの解析手法の概要図　(b) 可視化機構を組み込んだモデルの概要図

図 2.7: 可読性向上を目的とした従来手法の概要図

傍画素に対する正則化手法を提案し，学習済み CNN の特定の層の特徴量から入力画像の再構築を行っている［12］．Yosinski らは Gaussian blur と呼ばれる画像処理を画素値の更新の際に適用し，生成画像の各画素の近傍に対する制約を加えることで可読性の高い画像生成を可能としている［13］．Wei らは Data-driven patch prior と呼ばれる事前知識を提案し，出力画像のパッチとあらかじめ用意した画像のパッチとの距離の最小化を行うことで生成画像の色空間に対して正則化を行っている［14］．また，Nguyen らは任意の特徴量から入力画像を再構築する Variable autoencoder (VAE) を用いて，そのネットワークの入力値の最適化を行うことで学習済み CNN の特定ユニットの活性化を行っている．

これらの研究で提案されてきた様々な事前知識や正規化の導入により，生成画像の可読性は向上してきた．しかし，これらの手法で生成される画像の可読性は，可視化の際に用いる事前知識に大きく依存してしまう．また，人手で設計される事前知識による精度向上には限界があると考えられる

2.3　高精度かつ高い可読性をもつ深層学習に関する先行研究

前節で述べた深層学習の特徴量の可視化に関する研究に加えて，可読性を高めるための深層学習のモデル構造の提案も数多く行われてきた．特徴量の可視化では可視化処理を画像分類器の学習と独立して適用するのに対し，モデル構造を提案する研究では予めに可視化機構をモデル内部に組み込むことで可読性の向上を行う．これらの手法における可視化の流れの違いを図 2.7 に示す．

Zhou らは Class activation mapping (CAM) と呼ばれる深層学習のモデル構造の提案している［2］．図 2.8 のモデル構造が示すとおり，この手法ではモデル内の全結合層を Global average pooling に置き換えることでクラス分類に有効な領域の可視化を可能としている．この手法で生成される可視化結果は重要領域の特定に有効であるが，解像度が低くなることで視認性が低下する問題点がある．

このような背景から，荒井らは画像分類において重要な領域を直感的に可視化することができる深層学習モデル Generative contribution mappings (GCM) を提案している［3］．図 2.9 のモデル構造が示すとおり，この手法では Encoder-eecoder ベースのネットワーク構造を用いることで入力画像と同じ大きさの高解像度の可視化を可能としている．一方，可読性を向上させるための機構をネットワークに導入することで，他の深層学習モデルと比較して分類精度が低下してしまう問題点がある．

これらの手法はクラス分類に有効な領域を利用者に直感的に提示することが可能である．しかし，Selvaraju らが指摘するとおり，精度と可読性のトレードオフの問題から，可読性

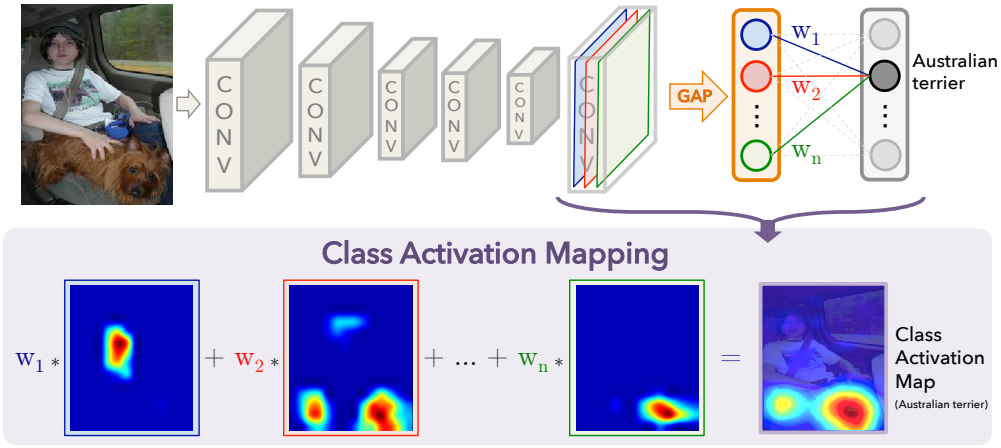


図 2.8: Class activation mapping（CAM）の構造例（画像は文献 [2] から引用）

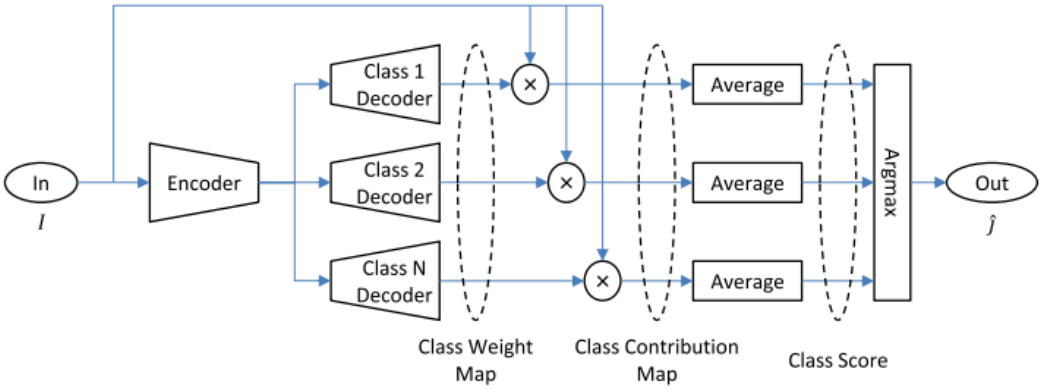


図 2.9: Generative Contribution Mappings（GCM）の構造例（画像は文献 [3] から引用）

向上のためのモデル構造の変更はモデルの分類精度を低下させる可能性が考えられる [9].

2.4 まとめ

本章では本研究に関連する先行研究として，進化的条件判断ネットワークの分類過程の文章化，学習済み CNN の処理の解析，高精度かつ高い可読性をもつ深層学習に関する先行研究に関する研究について述べた.

If-then ルールを用いた分類器の可読性の向上では，構築した分類器の分類過程を文章で表現する手法について説明した．分類過程を文章で説明することで，これまでブラックボックスであった分類過程を人間が理解しやすい形で提示することが可能となった．一方で，語句への変換が困難である特徴量を用いた分類器の説明では説明文の可読性が低下してしまうといった課題がある.

学習済み CNN の処理の解析では，任意の入力画像の分類問題において分類に有効な領

域の特定と特定のユニットを活性化させる画像生成に関する先行研究について述べた．これらの手法は学習済み CNN が学習した特徴量の解析や分類器内部処理の理解に有効である．一方で，可視化の際に可読性を高めるための事前知識や制約人手で設計する必要があり，生成する可視化画像の視認性の向上には限界があると考えられる．

最後に，深層学習の可読性向上のためのモデル構造に関する関連研究について述べた．これらの手法では判定の根拠提示を行う機構を組み込んだ上でモデルを構築するため，モデルの出力自体を判定の説明と考えることができる．一方で可読性と精度のトレードオフの問題により精度が低下してしまう問題が存在する．

第3章　If-thenルールを用いた分類器の精度と可読性の向上

3.1　はじめに

本章では進化計算法で構築した画像分類器の精度と可読性の向上を目的とし，進化的条件判断ネットワークの分類過程を画像を用いて説明する手法を提案する．進化的条件判断ネットワークの可読性向の従来手法として，構築した分類器の分類過程を文章で説明する手法があり，分類過程を文章で説明することでブラックボックスである分類器の分類過程を人間が理解しやすい形で提示することが可能となる．しかし，この手法では対応する語句への変換が困難な特徴量を用いた分類器の説明に対して，説明文の可読性が低下してしまうといった課題がある．

そこで本章では，直感的な分類過程の提示を行うために分類に用いられている特徴量分布を考慮したヒートマップの作成と特徴量の可視化の 2 つの可視化手法を提案する．そして，提案手法を一般画像分類問題に適用して手法の有効性を検証する．以下，提案手法について詳しく述べていく．

3.2　提案手法の画像分類方法

提案モデルの概要を図 3.1 に示す．提案手法では 1 枚の画像からオーバーラップありでパッチを取得・判定し，それらのパッチ判定の多数決により最も多く判定されたクラスに分類を行う．1 枚の画像から取得するパッチ数はすべての画像で同じ枚数とし，取得するパッチの縦横の大きさを画像ごとに算出する．縦横の大きさ H，W の画像からそれぞれ

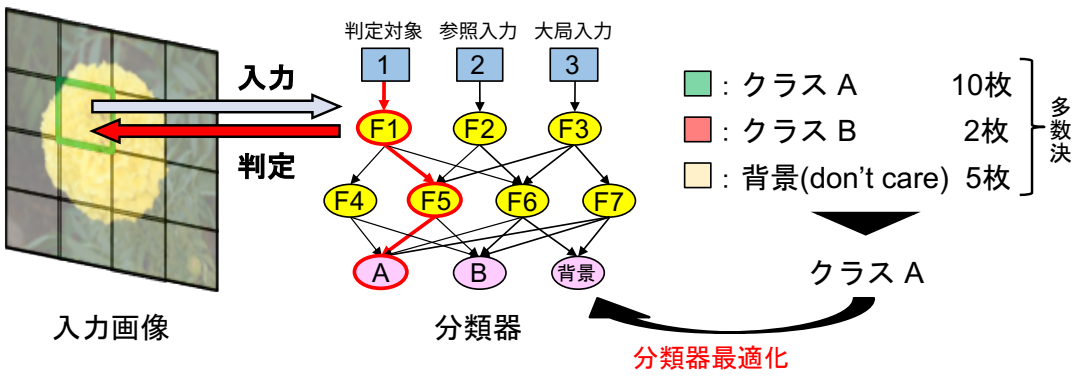


図 3.1: 提案手法の概要図

h_o，w_o だけオーバラップさせて $n \times n$ 枚のパッチをそれぞれ取得したときのパッチの大きさ w，h はそれぞれ次式で算出される．

$$w = \frac{W + (n+1)w_o}{n}, \quad h = \frac{H + (n-1)h_o}{n} \tag{3.1}$$

また，EDEN の入力ノードは判定対象と参照入力，大局入力の 3 種類から構成されるものを使用する．その際，画像から取得したパッチを判定対象，判定対象を中心とした縦横 2 倍の領域のうち判定対象を除いた領域を参照入力，画像全体から算出した特徴量を大局入力とする．

3.3 進化的条件判断ネットワークの画像分類過程の可視化

上記の方法で分類を行う分類器に対して，提案手法では次の 2 種類の可視化方法の提案を行う．

(1) 説明パスのデータ集合の性質を表す領域の可視化

(2) 各分岐ノードでの特徴量の可視化

提案モデルの分類は画像内から取得されるパッチ単位での判定を行うため，パッチ単位でこれらの可視化を行うことで分類に重要な領域の特定を行うことが可能となる．

3.3.1 説明パスのデータ集合の性質を表す領域の可視化

本手法では，説明パスに対して説明パスを通過したデータ集合のクラスらしさを表すヒートマップを作成する．ここで生成されるヒートマップは，対象である分類器の説明パスを通過するデータの分布を考慮して算出するものとする．その際のデータ分布の推定方法として EM アルゴリズムなどの方法も考えられるが，本論文では各クラスの特徴量空間の重心をもってクラスらしさを定義する．クラスらしさの算出の流れとしては，はじめに説明パスを通過するデータ集合に対してパス内のノードごとに条件分岐に用いられる特徴量のスカラ値を用いてクラスタリングを行う．そして，各ノードと対応する特徴量のスカラ値ごとに重心を算出する．

このときのクラスタリング手法として，データに応じて適切なクラスタ数を自動で推定する x-means 法 [15]を用いる．このとき，構築されたクラスタの中心に近く，データ数の多いクラスタに属するデータほどパスを通過するデータの中で代表的なデータ (クラスらしいデータ) であると考えることができる．これらのことから，本論文では通過パスの各特徴量における特徴量分布から構築された分類器のクラスらしさを算出し，クラスらしさのヒートマップを生成する．各分岐ノードでの特徴量分布とクラスらしさの関係を図 3.2 に示す．

本論文ではパスを通過したデータのクラスらしさは，パス内の各特徴量のスカラ値ごとにクラスタリングをしてできたクラスタの大きさと中間距離をもとに算出する．構築された EDEN の各ノードにおける特徴量分布を確認した際，特徴量空間上で重心を中心におよ

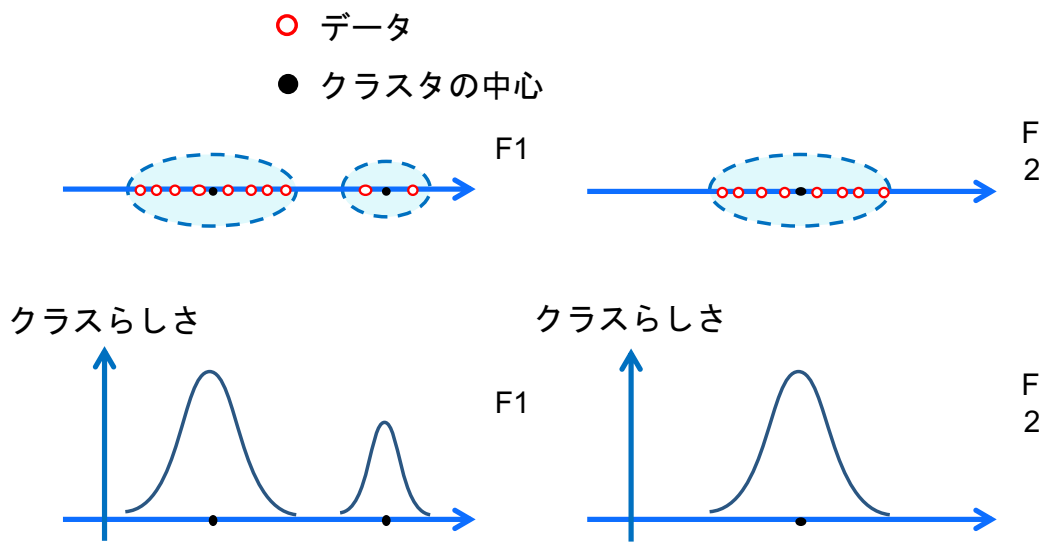


図 3.2: 各分岐ノードでの特徴量分布とクラスらしさの関係

そ釣鐘型の特徴量分布をもつことが分かったため，本論文ではクラスらしさが特徴量空間上の重心から正規分布に従うものとし，データ i のクラスらしさ p_i を次式で算出する．

$$p_i = \frac{1}{N_{\text{node}}} \sum_{\text{J}=1}^{N_{\text{node}}} \left\{ \exp(-\frac{Z_{ij}^2}{2}) \times \frac{n_{ij}}{n_{\max_j}} \right\} \tag{3.2}$$

ここで，N_{node} は説明パスのノード数であり，n_{ij} はノード j についてデータ i が属するクラスタのデータ数，$n_{\max_j}$ は属するデータ数が最も多いクラスタのデータ数を表す．Z_{ij} は通過するデータ i とクラスタ j の中心間距離に関する正規化項であり，次式で表すことができる．

$$Z_{ij} = \frac{\alpha(c_{ij} - a_{ij})}{\sigma_{ij}} \tag{3.3}$$

ここで，c_{ij} はノード j についてデータ i が属するクラスタの中心の値，a_{ij} はデータ i のノード j に関する特徴量のスカラ値，σ_{ij} はデータ i が属するクラスタのノード j に関する特徴量の標準偏差を示す．また，α は利用者の要求度に合わせて変更が可能な説明のパラメータであり，ヒートマップの色の変化を調整することが可能である．α を小さくすることでヒートマップの変化が激しくなり，クラスらしさが高い領域のみを表示することが可能になる．また，パッチのオーバーラップ部は，重なるパッチ同士のクラスらしさの値の平均値とすることにした．ヒートマップはクラスらしさが高いほど赤色に，クラスらしさが引くほど青色で表される．これによって，構築された分類器が画像内のどの領域に注目して分類を行っているか直感的に提示することが可能となる．なお，このクラスらしさは構築された分類器の説明の際に使用することを目的としており，分類の判定には影響しない指標である．

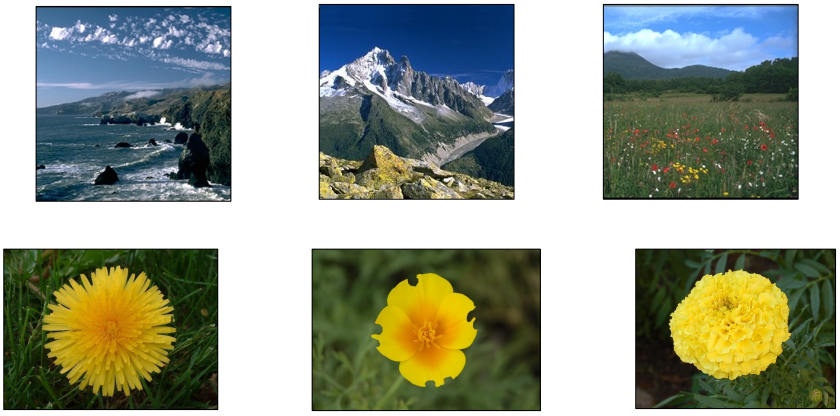


図 3.3: 実験に使用した画像例

3.3.2　各分岐ノードでの特徴量の可視化

本手法では，画像内の各画素を中心に特徴量を算出して説明パスにおける各ノードでの特徴量のスカラ値による条件分岐との対応関係を出力することで説明パス内の各ノードの動きを直感的に提示する．まず，判定に用いたパッチと同サイズのウィンドウを走査し，画像内のすべての画素に対して，ある画素を中心にしたときの特徴量を算出する．この際，画像の縁などでウィンドウがはみ出してしまう場合は，はみ出していない領域を特徴量の算出領域とする．次に，各画素で算出した特徴量のスカラ値と説明パスの条件分岐の対応関係を出力する．本実験では分岐条件と一致している画素を赤色で示し，反対に分岐条件と一致していない画素を青色で示す．この操作を画像内のすべての画像に適用し，各分岐ノードにおける特徴量のスカラ値との対応関係を領域として示すことで特徴量の可視化を行う．これによって，説明パス中の各特徴量の分岐条件が画像内のどの領域と対応するものであるかを直感的に提示することが可能となる．なお，この際に特徴量を算出する領域は EDEN が判定で用いた領域と異なっている場合もあるが，可視化の際の特徴量の算出領域が EDEN の判定に用いたパッチと共通する領域を多く含むため，EDEN の判定における特徴量ごとの分岐と画像内の領域の対応関係を見る上では妥当な操作と考える．

3.4　一般画像分類実験

3.4.1　データセット

一般画像データセットである Urban and Nature Scene Categories［16］と 102 Category Flower Dataset［17］に適用し，提案手法の有効性の検証を行った．Urban and Natural Scene Categories から Coast & Beach と Open Country と Mountain クラス，102 Category Flower Dataset から Common Dandelion と California Poppy と Marigold のそれぞれ 3 クラスの画像を用いて分類器を構築し，構築した分類器に対して分類過程の可視化を行った．これらのデータセットは，分類対象の構造が複雑で色が類似している点で崎津らの手法で使われていた色などの語句との対応付けがある特徴量だけでは分類が困難であることから選択した．実験では各クラスから 50 枚を学習画像とし，画像から縦横それぞれオーバーラップ

表 3.1: EDEN のパラメータ

世代交代モデル	MGG［18］
世代数	50,000
個体数	100
子個体数	30
交叉率	0.9
一様交叉率	0.1
突然変異率	0.05
入力ノード	3
中間ノード	100
出力ノード	分類クラス数+1

ありで 15 枚取得して一枚の画像から 125 枚のパッチを取得した．そして，これらのパッチを分類器に入力した際の判定の多数決によって最も多く判定されたクラスに分類を行った．使用した画像例を図 3.3 に示す．

3.4.2 実験設定

本実験で用いる EDEN では分類精度を高めると同時にネットワークのノード数が少なくなるように適応度を設定し，進化計算法を用いて最適化した．具体的には次式で表される評価関数を用いて EDEN の最適化を行った．

$$Fitness = Accuracy + \beta \times \frac{1}{N_{\text{node}}} \tag{3.4}$$

ここで，$Accuracy$ は学習データセットに対する分類精度であり，N_{node} は構築されたネットワークのノードの総数を示す．β はネットワーク規模の項の影響を調整するパラメータであり，本実験では $\beta = 0.001$ を使用した．この評価関数を用いることで，分類精度の向上と同時にネットワークの規模を考慮した最適化を行うことが可能となる．分類精度 ($Accuracy$) はデータの総数を N，正しく分類されたデータ数を N_{correct} とすると次式で示すことができる．

$$Accuracy = \frac{N_{\text{correct}}}{N} \tag{3.5}$$

そして，本実験で用いた EDEN の設定を表 3.1 に示す．説明文生成のパラメータは $S = 1$，$s = \max$ として，最も通過頻度の高いパスに対してルートノードから順にすべてのノードを語句に置き換えるようにした．そして，ヒートマップ作成のパラメータは $\alpha = 1.5$ とした．

3.4.3 EDEN で使用する画像特徴量

本実験で使用した特徴量は語句と対応付けられた HSL やエッジに関する特徴量 16 種類と語句での表現が困難な特徴量 34 種類を用いた．本実験で使用した特徴量と語句の対応関

表 3.2: 特徴量と語句の対応表

特徴量	対応する語句
0.0 ≤ L < 0.2 または 0.0 ≤ S < 0.2 のとき：	
L が 0.2 以下の画素の割合	黒色っぽい部分
L が 0.8 以上の画素の割合	白色っぽい部分
それ以外のとき：	
H が 30° 未満または 330° 以上の画素の割合	赤色っぽい部分
H が 30° 以上 90° 未満の画素の割合	黄色っぽい部分
H が 90° 以上 150° 未満の画素の割合	緑色っぽい部分
H が 150° 以上 210° 未満の画素の割合	水色っぽい部分
H が 210° 以上 270° 未満の画素の割合	青色っぽい部分
H が 270° 以上 330° 未満の画素の割合	紫色っぽい部分
S の平均	鮮やかさ
L の平均	明るさ
H の標準偏差	色数
S の標準偏差	鮮やかさの差
L の標準偏差	明るさの差
ある程度のエッジ強度があり：	
水平方向 ±10° のエッジをもつ画素の割合	横線
垂直方向 ±10° のエッジをもつ画素の割合	縦線
斜め方向 ±10° のエッジをもつ画素の割合	斜め線

係を表 3.2 に示す．語句での表現が困難な特徴量として ULBP 特徴量［19］($P = 10$, $R = 2$) と 6 種類の色成分 (RGB，L*a*b*) にフィルタ処理 (sobel フィルタ，mean フィルタ) を施してから算出した統計特徴量 (平均，標準偏差) を用いた．また，各ノードのしきい値に対応する語句は用意する辞書の定義の仕方によって説明の粒度を変更することができるが，本実験では多いか少ないかどうかの 2 種類の表現のみに制約をして説明文を生成し，その説明文に対する分類過程の可視化を行った．

3.4.4 実験結果

それぞれのデータセットの未知画像における EDEN の分類結果を表 3.3 に示す．ここで未知画像の分類精度は構築した分類器に対して，Urban and Natural Scene Categories は各クラス 50 枚，102 Category Flower Dataset は各クラス 15 枚を未知画像として適用した際の分類精度である．それぞれのデータセットで構築された分類器の構造例を図 3.4，図 3.5 に示す．図のノード間の色は $S = 1$ のときの各クラスのフローを表しており，複数のクラスが同じパスを流れている場合は色を変更して表示している．従来手法［6］で生成された説明文例を次に示す．そのうち，提案手法を Urban and Natural Scene Categories に適用したときの生成画像を図 3.6 と図 3.7 に示す．

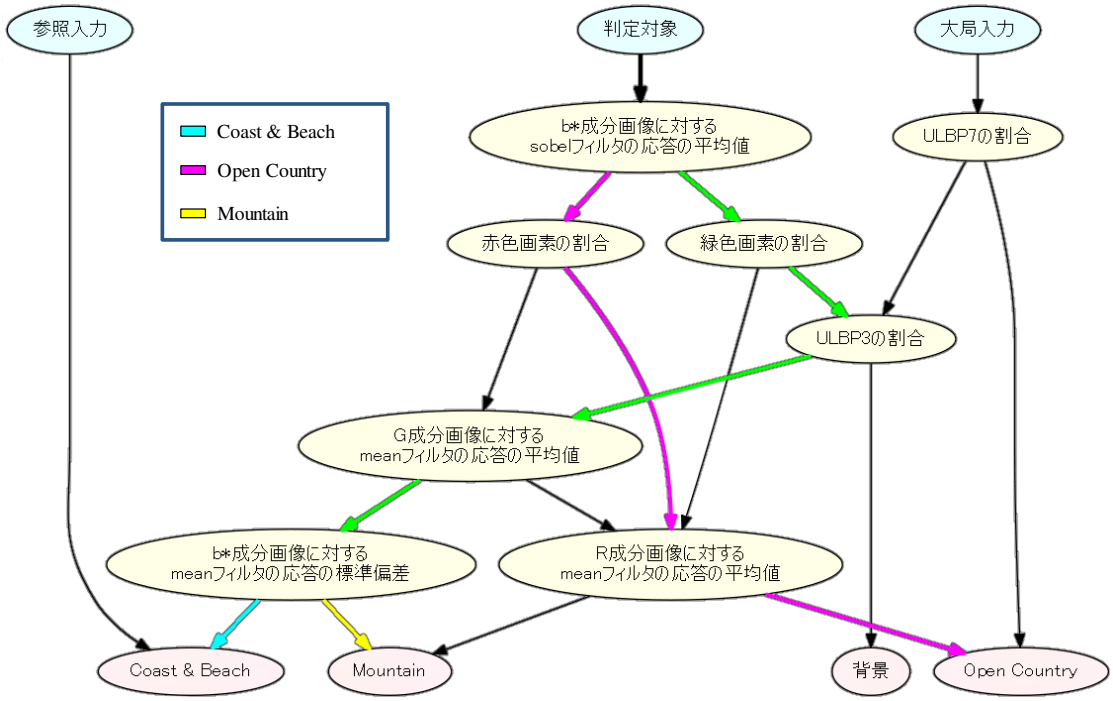


図 3.4: Urban and Natural Scene Categories を分類する EDEN のネットワーク例

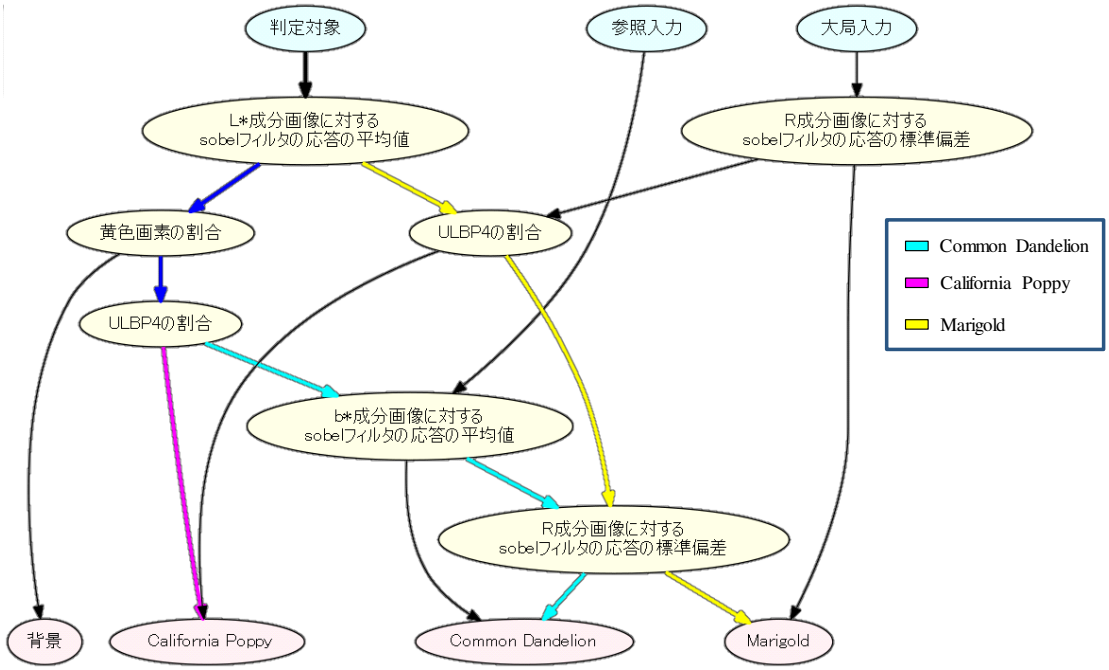
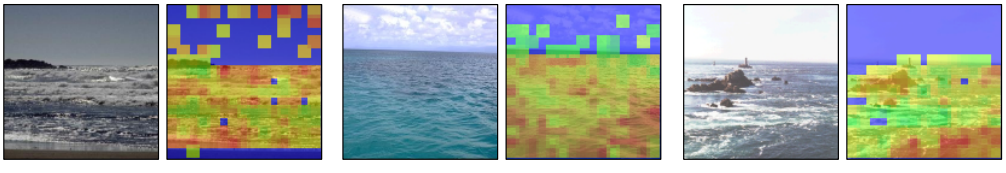
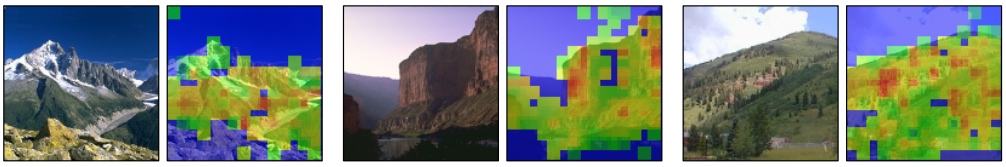


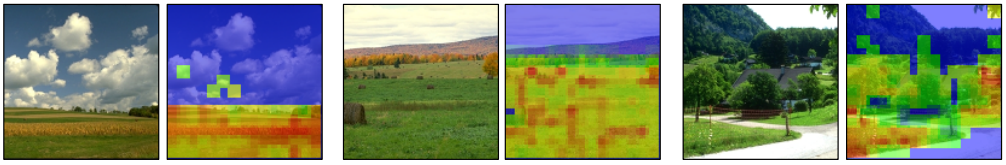
図 3.5: 102 Category Flower Dataset を分類する EDEN のネットワーク例



(a) Coast & Beach

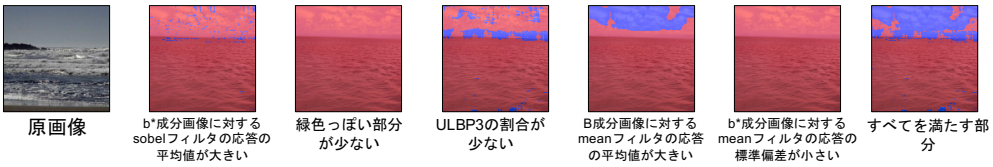


(b) Mountain



(c) Open Country

図 3.6: Urban and Natural Scene Categories のクラスらしさのヒートマップ可視化例



(a) Coast & Beach

(b) Mountain

(c) Open Country

図 3.7: Urban and Natural Scene Categories の特徴量の可視化

表 3.3: 一般画像分類の正解率

	15 試行平均	
	精度 (学習)	精度 (未知)
Urban and Natural Scene Categories	91.7%	71.5%
102 Category Flower Dataset	95.6%	84.4%

- Coast & Beach と Open Country，Mountain クラスの分類過程の説明文
  - b*成分画像に対する sobel フィルタの応答の平均値が大きく，緑色っぽい部分が少ない．また，ULBP3 の割合が少なくて，G 成分画像に対する mean フィルタの応答の平均値が大きい．そして，b*成分画像に対する mean フィルタの応答の標準偏差が小さいため，Coast & Beach である．
  - b*成分画像に対する sobel フィルタの応答の平均値が大きく，緑色っぽい部分が少ない．また，ULBP3 の割合が少なくて，G 成分画像に対する mean フィルタの応答の平均値が大きい．そして，b*成分画像に対する mean フィルタの応答の標準偏差が大きいため，Mountain である．
  - b*成分画像に対する sobel フィルタの応答の平均値が小さい，赤色っぽい部分が少ない．また，R 成分画像に対する sobel フィルタの応答の平均値が大きいため，Open Country である．

生成された図 3.6 のヒートマップを見ると，Coast & Beach は海面，Mountain は山肌，Open Country は草木の領域がそれぞれ強く反応しており，画像内の分類対象に対して正しく注目していることが分かる．また，図 3.7 の各分岐ノードにおける特徴量の可視化結果を見ると，b*成分画像に sobel フィルタを施した後のパッチ全体の平均値が Open Country の草木領域への反応に大きく寄与していることが分かる．これは草木の領域はエッジ成分が小さく，他のクラスの対象領域と分離する上で重要な特徴量であるためと考えられる．

次に従来手法を 102 Category Flower Dataset に適用した際に生成された説明文例を次に示す．そして，提案手法によって生成された可視化画像を図 3.6 と図 3.7 に示す．

- Common Dandelion と California Poppy，Marigold の分類過程の説明文
  - L*成分画像に対する sobel フィルタの応答の平均値が大きく，黄色っぽい部分が多い．また，ULBP4 の割合が少なくて，b*成分画像に対する sobel フィルタの応答の平均値が小さい．そして，R 成分画像に対する sobel フィルタの応答値の標準偏差が大きいため Common Dandelion である．
  - L*成分画像に対する sobel フィルタの応答の平均値が大きく，黄色っぽい部分が多い．また，ULBP4 の割合が多いため，California Poppy である．
  - L*成分画像に対する sobel フィルタの応答の平均値が小さく，ULBP4 の割合が少ない．また，R 成分画像に対する sobel フィルタの応答値の標準偏差が小さいため，Marigold である．

表 3.4: motorbike の学習・未知の再現率

学習 (5 試行平均)		未知 (5 試行平均)	
再現率 C_{white}	再現率 $C_{\overline{white}}$	再現率 C_{white}	再現率 $C_{\overline{white}}$
0.968	0.988	0.057	0.488

生成された図 3.6 のヒートマップを見ると，花びらの領域がそれぞれ強く反応しており，画像内の分類対象に対して正しく注目していることが分かる．また，図 3.7 の各分岐ノードにおける特徴量の可視化結果を見ると，Marigold は L*成分画像に sobel フィルタを施した後の平均値でおおよそ注目する領域が決定されていることが分かる．これは Marigold の花びらの形状を認識する上で重要な特徴量であるためと考えられる．

3.4.5 文章で説明を行う従来手法との比較

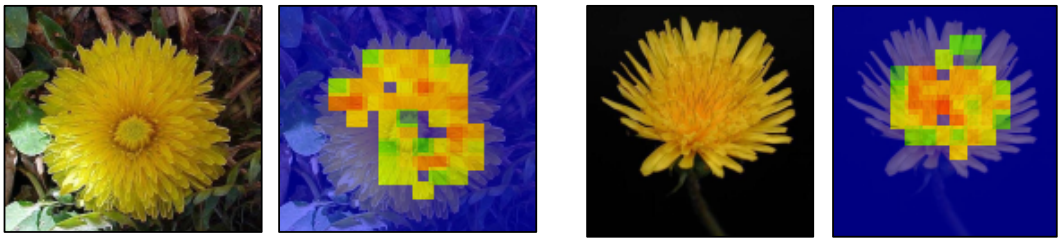
それぞれのデータセットで構築された分類器の説明パス内について各分岐ノードで算出されるクラスらしさの値を確認すると，ULBP 特徴量や色成分にフィルタ処理を施した統計特徴量がクラスらしさに大きく寄与することが分かった．例えば，Urban and Natural Scene Categories データセットの Mountain クラスでは ULBP3 や b*成分画像に対する統計特徴量などが山肌領域の反応に寄与することが分かる．このような説明文への変換が困難な特徴量に対しても利用者に視覚的に提示できる点で従来手法よりも有用性があると考える．

また本実験では語句での表現が困難な ULBP 特徴量などを使用しているため，従来手法で生成される説明文は文章が長くなると同時に説明文自体の可読性は低く，利用者にとって分かりやすい説明であるとはいえない．また，図 3.4，図 3.5 で示すネットワークよりも複雑なネットワーク構造の分類器に適用した際にはさらに複雑な説明文が生成されることが考えられる．これに対して，提案手法を用いて分類器の分類過程を可視化し画像として利用者に提示することで，分類過程を文章で説明するよりも直感的で理解しやすい提示が可能であるといえる．

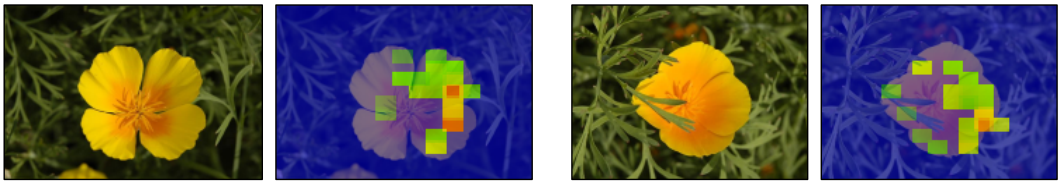
3.4.6 可視化結果を用いた分類器の妥当性に関する考察

構築した分類器の信頼性や妥当性は分類精度によって評価されることが一般的である．しかし，直感的な分類過程の提示も構築した分類器の評価の上で重要な指標であると考える．分類過程の提示が信頼性の評価に有効であることを示すため，Caltech-256 [20] から Airplane と Motorbike，Helicopter の 3 クラス分類における Motorbike の可視化を行った結果について考察を行う．

今回の実験では，Motorbike クラスの画像の大半 (45 / 50 枚) は背景が白いものを学習に用いた．これらの画像を用いて構築した分類器に対して提案手法を適用した場合，画像の背景部分が主に反応する分類器 C_{white} と分類対象である Motorbike が強く反応する分類器 $C_{\overline{white}}$ の 2 種類が主に構築された．その時のそれぞれの分類器における可視化例を図 3.10 に示す．これらの分類器の学習における再現率と構築した分類器に対して Motorbike の背



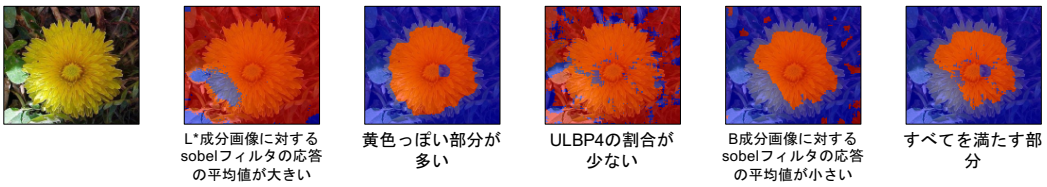
(a) Coast & Beach



(b) Mountain

(c) Open Country

図 3.8: Urban and Natural Scene Categories のクラスらしさのヒートマップ可視化例



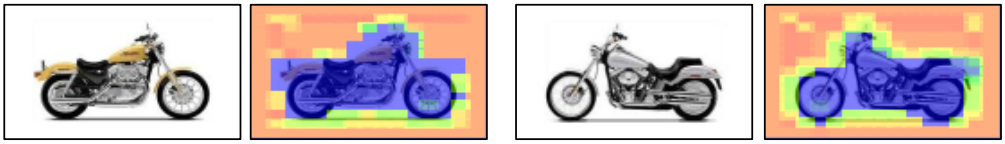
(a) Coast & Beach

(b) Mountain

(c) Open Country

図 3.9: Urban and Natural Scene Categories の特徴量の可視化

(a) 分類対象が反応する分類器 $C_{\overline{white}}$



(b) 背景が反応する分類器 C_{white}

図 3.10: Motorbike のクラスらしさのヒートマップの可視化例

景に草木などの物体が写っている画像を未知画像として適用した際の再現率を表 3.4 に示す．表 3.4 が示すように，分類器 C_{white} は背景に物体が写っている未知画像に対しては全く分類ができていないことが分かる．一方，分類器 $C_{\overline{white}}$ は背景に反応している分類器よりも高い精度が獲得されていることが分かる．このように，進化計算法や機械学習を用いて構築した分類器は学習画像と未知画像の違いや分類器の学習のさせ方の違いによって分類器の性能や汎用性が大きく左右される．そのため，提案手法を用いて分類過程の可視化を行うことで分類器の分類精度だけではなく分類過程の提示をすることが可能となり，構築した分類器の理解につなげることができると考える．

3.5 まとめ

本章では If-then ルールで分類を可視化する手法の提案を行った．特に進化的条件判断ネットワークについて，分類に用いられている特徴量分布を考慮したヒートマップの作成と特徴量の可視化の 2 つの可視化手法の提案を行った．そして，提案手法を一般画像分類問題に適用し，獲得した可視化画像の有効性の検証を行った．結果として，構築した分類器の分類過程を直感的に利用者に示すことが可能となり，分類器の信頼性の向上につなげることができたといえる．

今後の課題としては，利用者の要求度も合わせて分類クラスの違いを明確に提示することが可能な説明方法の検討が挙げられる．これは医用画像などの産業応用では，判定結果に対するすべての説明を行う必要はなく，分類クラスの違いが明確になればよい場合があるためである．そのため，多クラス分類の分類過程を説明する際にクラス間の違いが明確に分かるような説明方法の検討が必要であると考える．また，生成された説明の分かりやすさを数値などで評価し，提案手法の有効性を客観的に示すことも今後の課題である．

第4章　学習済み深層学習モデルの特徴量の可視化

4.1　はじめに

近年，画像認識分野において畳み込みニューラルネットワーク (Convolutional neural networks; CNNs) が高精度な手法として注目をされており，多くの画像認識の問題において優れた性能を示している．様々な構造のネットワークや学習アルゴリズムの提案により精度の向上が図られてきた一方，学習した CNN の内部でどのような処理が行われているかは十分に理解されていない．一般的にユーラルネットワークの処理はブラックボックス化されてしまうことが知られており，内部で行われている処理および獲得された特徴量の解析は困難である．構築した CNN の処理を今後の社会現場で活用するためには，高精度な処理であると同時に内部でどのような処理が行われているかを利用者へ提示することが求められる．

第 2 章で述べたように，学習済み CNN の内部処理の解析を目的とした画像生成の研究は数多く行われてきた．従来の一般的な方法は次の 2 つの制約を満たす画像生成を行う．

(1) 特定ユニットの活性化

(2) 自然画像の性質の付加

これらの制約を満たすために従来手法では人手で設計した事前知識を用いて可視化画像の生成を行っているが，その制約が最適であるとは限らない．また，医用画像などの画像に対しては事前に自然画像の制約を設計することは困難であると考えられる．

そこで本章では，事前に明示的な自然画像の制約を与えることなく特定ユニットを活性化させる自然画像の生成モデルを提案する．提案手法では，高精度な画像生成モデルとして知られている Generative adversarial networks (GAN) の枠組みを拡張することで学習済み CNN の特定ユニットを活性化させる可視化画像の生成を行う．

4.2　Generative Adversarial Networks を用いた学習済み Convolutional Networks の可視化

本手法では，近年高精度な画像生成手法として知られている Generative adversarial networks (GAN) の枠組みを用いる．一般的な GAN は Generator と Discriminator の 2 つのネットワークから構成されており，それぞれのネットワークを相互に学習することで自然画像に近い画像を生成することが可能となる．そこで本手法では一般的な GAN の枠組みによる学習に加えて，学習済み CNN の特定ユニットの活性化を目的とした Generator の学習を

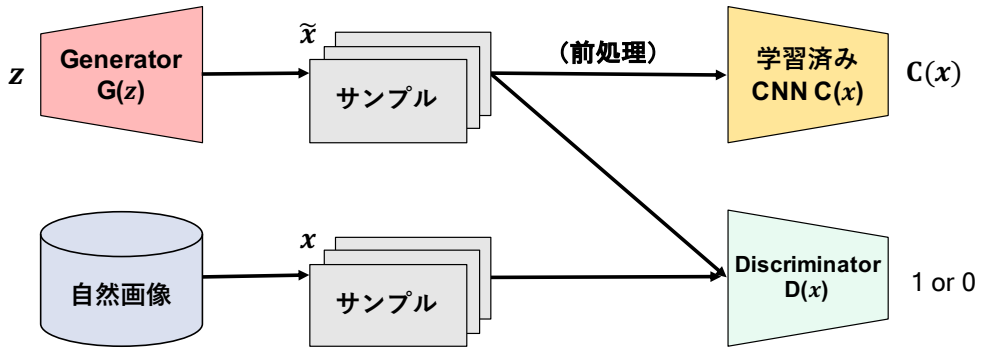


図 4.1: 提案手法の可視化モデルの構造．Generator と Discriminator，学習済み CNN の 3 つのネットワークから構成される．

行う．これにより事前に明示的な画像の制約を与えることなく学習済み CNN の任意のユニットを活性化させる自然画像を生成することが可能となり，CNN の各ユニットの機能の解析につなげることが可能となる．提案手法の可視化モデル構造を図 4.1 に示す．提案手法の可視化モデルは，Generator と Discriminator，学習済み CNN の 3 つのネットワークから構成される．

4.3 ネットワーク構造

4.3.1 Generator

Generator は任意の分布からサンプリングされた乱数ベクトル z を入力として画像空間上に写像するネットワークである．Generator は複数の逆畳み込み層 (Deconvolutional 層) によって構成されており，出力画像のサイズなどを考慮して問題に合わせて適切な構造を設計する必要がある．本論文では次に示す基本構造をもとにした Generator を使用することで学習済み CNN の特定ユニットを活性化させる画像の生成を行う．

Generator の基本構造：

- Upsampling 処理にはストライド幅 2 の Deconvolutional 層を用いる．
- ネットワークの出力層を除くすべての層で Batch normalization［21］を用いる．
- ネットワークの出力層を除くすべての層で Rectified linear unit (ReLU)［22］関数を用いる．
- ネットワークの出力層では Tanh 関数を用いる．

4.3.2 Discriminator

Discriminator は入力データが学習データセット由来か Generator によって生成されたデータかを判定するネットワークであり，複数の畳み込み層と全結合層によって構成されている．Discriminator も Generator と同様に問題に合わせてネットワーク構造を設定する必要があり，本論文では次の基本構造から構成される Disicriminator を使用する．

Discriminator の基本構造：

- ネットワークのすべての層で Weight normalization［23］を用いる．
- ネットワークの出力層を除くすべての層で Leaky rectified linear unit (Leaky ReLU)［24］関数を用いる．
- ネットワークの出力層で Sigmoid 関数を用いる．

4.4 提案モデルの学習方法

提案手法では (1) 特定ユニットの活性化，(2) 可読性の高い画像生成の 2 つの制約のもと，GAN の枠組みを用いてそれぞれ定義される目的関数に対して誤差逆伝播を行うことで Generator の学習を行う．このように複数の制約のもとでネットワークの学習を行うため，通常の GAN 以上に学習が不安定になることが考えられる．そこで，GAN の問題である学習の安定性について，伝播させる勾配のノルムの調整を行う層を導入することで学習の安定性の向上をさせる．

4.4.1 特定ユニットの活性化

Gradient ascent などの従来の一般的な可視化手法では，生成画像 x に対して再帰的に画素値の更新を行うことで学習済み CNN の特定ユニット h を活性化させる画像を獲得する．一方，提案手法では Generator が特定ユニット h を活性化させる画像を生成するようにネットワークの学習を行う．つまり，学習済み CNN の特定ユニット h の出力 C_h を活性化させる画像を生成するように Generator のパラメータ θG の更新を行う．このとき，出力の最大化問題は負の出力の最小化問題として考えることで通常の GAN の学習の枠組みの中で最適化を行うことが可能になる．

具体的には Generator を G，学習済み CNN の特定ユニット h の出力を C_h，入力する乱数ベクトルを z としたとき，C_h を活性化させるために次式の確率的勾配を用いて Generator のパラメータ更新を行う．

$$\nabla_{\theta G} \frac{1}{n} \sum_{i=1}^{n} C_h \left(G \left(z \right) \right) \tag{4.1}$$

Algorithm 1 学習済み CNN の特定ユニットを活性化させる GAN 学習

Input: D: Discriminator, G: Generator, C: 学習済み CNN

Procedure:

1: **for** number of epochs **do**

2: 事前分布 $p(z)$ から乱数ベクトル z を取得

3: 学習データセットから画像 x を取得

4: 確率的勾配降下法を用いて Generator，G のパラメータを更新：

$$\nabla_{\theta G} \frac{1}{n} \sum_{i=1}^{n} C_h(G(z))$$

5: Generator，G と Discriminator，D を確率的勾配降下法によって，それぞれのパラメータを更新：

$$\min_G \max_D L_{GAN} = \mathbf{E}_{x \sim p_{real}}[\log D(\boldsymbol{x})] + \mathbf{E}_{x \sim p_z}[\log(1 - D(G(z)))]$$

6: **end for**

4.4.2 可読性の高い画像生成

提案手法では一般的な GAN と同様に Mini-max game の枠組みを用いて画像生成を行う．つまり，Generator，G は乱数ベクトル z を入力として実空間に写像し，Discriminator，D は入力データが自然データの分布から得られた画像か Generator によって生成された画像かを判定する．このときのネットワークの目的関数 L_{GAN} を次式で示す．

$$\min_G \max_D L_{GAN} = \mathbf{E}_{x \sim p_{real}}[\log D(\boldsymbol{x})] + \mathbf{E}_{x \sim p_z}[\log(1 - D(G(z)))] \tag{4.2}$$

4.4.3 学習の安定性の向上

提案手法は 3 つのネットワークの出力を用いて GAN の学習を行うため，通常の GAN よりも学習が不安定になる可能性がある．そこで提案手法では伝播させる勾配のノルムを調整する層を学習済み CNN の入力層の手前に挿入することで，学習済み CNN の出力に影響を与えずに GAN の学習の安定化を行う．事前に行った実験により，学習済み CNN から伝播される勾配のノルムが Discriminator から伝播される勾配よりも大きくなる傾向が確認されたため，学習済み CNN から伝播される勾配のノルムに対して正規化処理を行う．本論文ではこの層を Gradient-buffering 層と呼ぶこととする．Gradient-buffering 層の概要図を図 4.2 に，具体的な処理の流れをアルゴリズム 2 にそれぞれ示す．

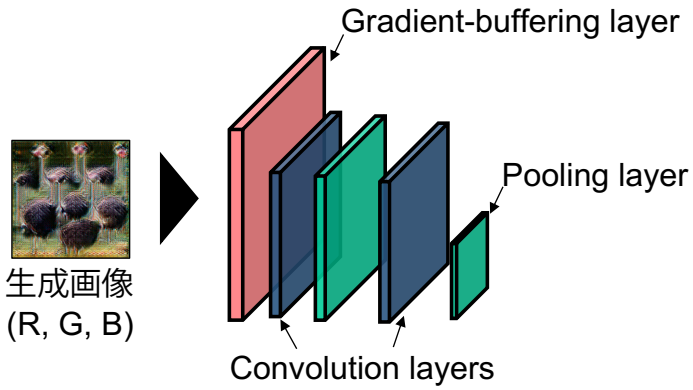


図 4.2: Gradient-buffering 層の概要図．Gradient-buffering 層を学習済み CNN に挿入することで学習の安定化をさせることが可能となる．

Algorithm 2 Gradient-buffering 層の処理の流れ

Input: $\frac{\partial E_{gen}}{\partial x}$: Generator からの勾配, $\frac{\partial E_{pre}}{\partial x}$: 学習済み CNN からの勾配

1: $\boldsymbol{g_{dis}} \leftarrow \frac{\partial E_{dis}}{\partial \theta}$

2: $\boldsymbol{g_{pre}} \leftarrow \frac{\partial E_{pre}}{\partial x}$

3: **if** $\|\boldsymbol{g_{pre}}\| > \|\boldsymbol{g_{dis}}\|$ **then**

4: 勾配 $\boldsymbol{g_{pre}}$ のノルムの正規化：

$$\boldsymbol{g_{pre}} \leftarrow \frac{\boldsymbol{g_{pre}}}{\|\boldsymbol{g_{pre}}\|} \cdot \|\boldsymbol{g_{dis}}\|$$

5: **end if**

4.5 学習済み CNN の特徴量の可視化実験

4.5.1 実験設定

本論文では，学習済みの AlexNet [25] に提案手法を適用することで有効性の検証を行った．AlexNet は深層学習において代表的なネットワークの 1 つであり，5 つの畳み込み層と 2 つの全結合層から構成される．本実験では ImageNet2012 [26] のベンチマークで学習された AlexNet (Caffe model zoo) [1]を用いた．本実験で用いた Generator と Discriminator のネットワーク構造をそれぞれ表 4.1，表 4.2 に示す．それぞれのネットワークの重みは $[-0.01, 0.01]$ の一様分布の乱数を用いて初期化を行った．そして，ネットワークの学習には Adam 法 [27] (学習率 $\alpha = 0.0002$, $\beta_1 = 0.5$, $\beta_2 = 0.999$) を用い，100 エポックの学習を行った．各エポックで 50,000 枚の ImageNet の検証用データセットからランダムに 1,000 枚の画像を選択し，これらを GAN の学習のための自然画像とした．また，ミニバッチサイズは 64 とした．学習済み CNN に画像を入力する際はモデルに対して適当な前処理を入力画像に行う．本論文で用いた AlexNet に対しては Generator の出力を 255 倍した後，画素ごとの平均差分を用いた．

[1] `http://dl.caffe.berkeleyvision.org/bvlc_alexnet.caffemodel`

表 4.1: Generator のネットワーク構造

層種	カーネル	ストライド	出力
z	-	-	$1 \times 1 \times 100$
l1+bn	-	-	$14 \times 14 \times 256$
conv1	3×3	1	$14 \times 14 \times 256$
deconv1	4×4	2	$28 \times 28 \times 128$
conv2	3×3	1	$28 \times 28 \times 128$
deconv2+bn	4×4	2	$56 \times 56 \times 64$
conv3	3×3	1	$56 \times 56 \times 64$
deconv3+bn	4×4	2	$113 \times 113 \times 32$
conv4	3×3	1	$113 \times 113 \times 32$
deconv4+bn	5×5	2	$227 \times 227 \times 16$
conv5	3×3	1	$227 \times 227 \times 3$

表 4.2: Discriminator のネットワーク構造

層種	カーネル	ストライド	出力
data	-	-	$227 \times 227 \times 3$
conv1+wn	7×7	2	$111 \times 111 \times 32$
conv2+wn	5×5	1	$107 \times 107 \times 64$
conv3+wn	3×3	2	$53 \times 53 \times 128$
conv4+wn	3×3	1	$51 \times 51 \times 256$
conv5+wn	3×3	2	$25 \times 25 \times 512$
GAP	-	-	$1 \times 1 \times 256$
fc1	-	-	$1 \times 1 \times 1$

4.5.2 学習済み AlexNet の各層の特徴量の可視化

提案手法を用いて学習済み AlexNet の各層の特徴量を可視化した際の可視化結果を図 4.3 に示す．それぞれの生成画像は学習済み AlexNet の特定ユニットを強く活性化させる画像であり，画像の大きさはユニットの活性化に寄与する領域 (受容野) の大きさと等しい．この可視化結果から，CNN の各層で段階的な特徴量が獲得されていることが分かる．具体的には，CNN の入力層付近の Layer 1 や Layer 2 ではエッジや格子などの基本的な特徴量が獲得されている．また，Layer 3 や Layer 4，Layer 5 ではより複雑な特徴量が獲得され，全結合層である Layer 6 や Layer 7 では様々なクラスの複合的な特徴量が獲得されている．そして，出力層である Layer 8 では特定のクラスを示す可視化画像が獲得されていることが確認できる．

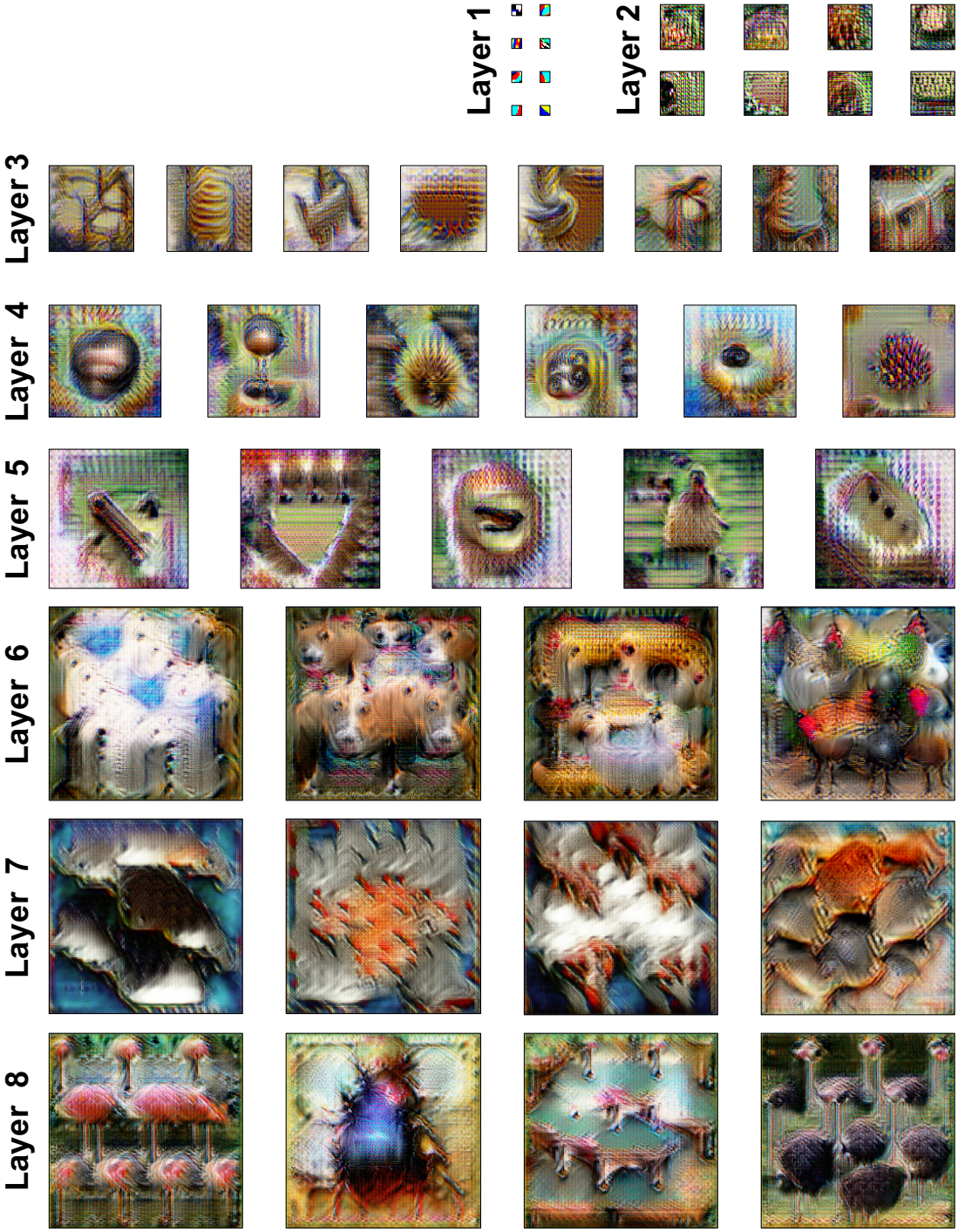


図 4.3: AlexNet の各層の特徴量およびクラスの可視化結果．それぞれの可視化結果は実際のユニットの活性化に影響する領域 (受容野) を示している．

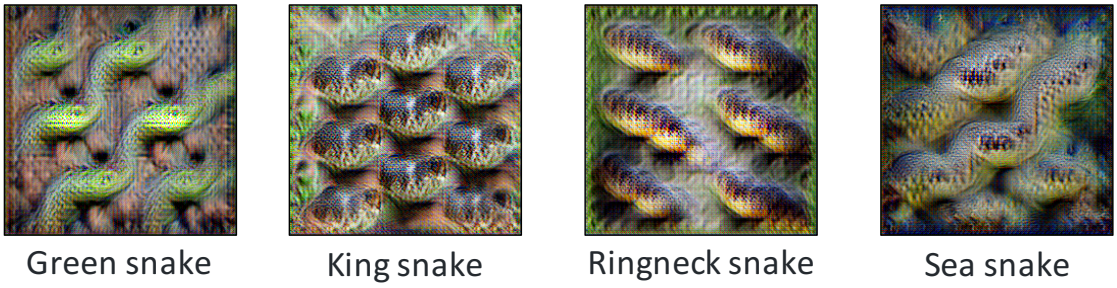


図 4.4: ImageNet 内の類似クラスに対する可視化結果

4.5.3 ImageNet 内の類似クラスの可視化

提案手法で生成される画像の視認性を検証するために ImageNet 内の類似クラスに対して提案手法を適用し，そのクラスの出力を最大化させる画像の生成を行った．図 4.4 に ImageNet の異なる 8 種類の蛇クラスの可視化結果を示す．これらの可視化結果から分かるとおり，生成される可視化結果には各クラスの特徴が表れており，各クラスの違いを明確に識別することが可能である．具体的に今回得られた可視化結果では，蛇の肌の色や模様，身体の太さなど，そのクラス特有の特徴が可視化画像に強く表れている．この結果から，提案手法で生成される可視化画像は視認性が高く，判定クラスに対してモデルがどのような特徴量を学習したかを確認することが可能であることが分かる．

4.5.4 学習済み CNN の階層的な特徴量の可視化

提案手法によって獲得された可視化結果から，モデル内部で階層的な特徴量が獲得されていることが分かる．そこで，任意のクラスに対して強く発火する特徴量に対して提案手法を用いて可視化を行い，分類に寄与する階層的な特徴量の解析を行った．具体的には学習済み AlexNet に対して任意のクラスの活性化を行い，その際に強く発火する中間ユニットに対して提案手法を用いてユニットを活性化させる画像の生成を行った．図 4.5 に Billiard と Teddy bear クラスのクラス分類において，AlexNet のネットワーク内部で強く活性化されるユニットの可視化結果を示す．ここでは AlexNet の Layer4 から Layer6 の中間層の特徴量の可視化結果を示す．これらの可視化結果から，学習済み CNN がクラスに特化した特徴量を学習していることが分かる．例えば Billiard クラスに対して特徴量の可視化を行った場合，ビリヤードボードの角やビリヤードボールの特徴量が獲得されている．また，Teddy bear クラスに対しては毛皮のようはテクスチャ系の特徴量が獲得されていることが分かる．このように提案手法を用いて学習済み CNN の特徴量の可視化を行うことで，画像分類の判定においてどのような特徴量が寄与しているかの解析に利用することができることが分かる．

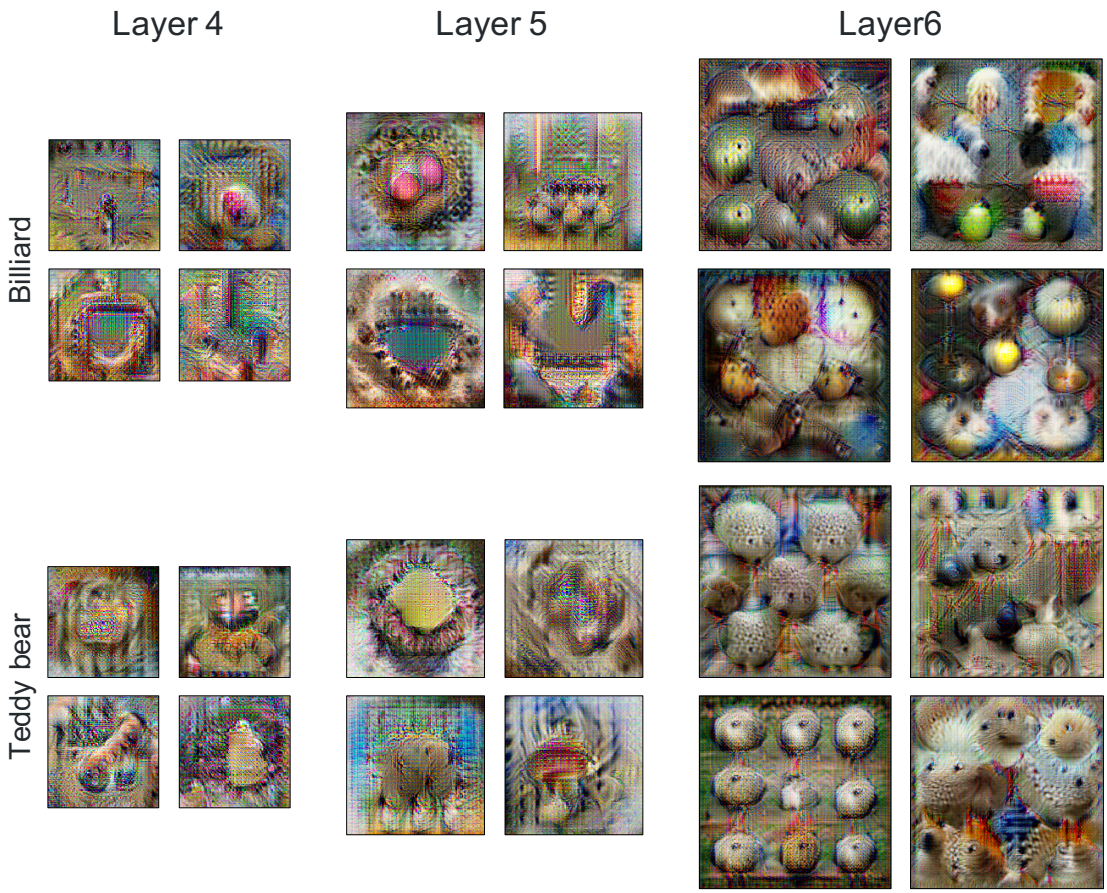


図 4.5: ImageNet の BIlliard クラスと Teddy bear クラスに対する階層的な特徴量の可視化結果

4.6 まとめ

本章では GAN の枠組みを用いた学習済み CNN の特定ユニットを活性化させる画像生成方法を提案し，その有効性の検証を行った．提案モデルは GAN の枠組みを拡張し，学習済み CNN の特定ユニットを活性化させる画像の生成を行う．先行研究では，学習済み CNN の特定ユニットの活性化と自然画像への変換を同時に満たすための制約を人手で設計した上で生成を行うことが一般的である．一方，提案手法は GAN の枠組みを用いることで，先行研究で提案されてきた制約を用いることなく可視化画像の生成が可能である．提案手法を学習済み AlexNet に適用し各層の特定ユニットを活性化させる画像生成を行った結果，分類に有効な特徴量がネットワークの内部で段階的に獲得されていることが確認された．また，階層的な特徴量の可視化を行うと，クラス分類に有効な特徴量が獲得されていることが分かった．

今後の課題として多様性のある可視化画像の生成が挙げられる．提案手法を用いることで学習済み CNN の任意のユニットを活性化させる自然画像を生成することができるが，Generator に入力する乱数ベクトル z を変化させた際の生成画像に多様性は見られない．一方で施行間においては多様的な画像が生成される場合もある．今後は一施行の中で多様的

な画像生成が可能なネットワークの学習方法の検討を行う必要がある．また，一般画像だけでなく，医用画像などの実画像に提案手法を適用し，有効性の検証を行っていく必要もある．

第5章　高精度かつ高い可読性をもつ深層学習

5.1　はじめに

深層学習モデルの可読性を高めるためのモデル構造の提案が数多く行われてきた．一方で精度と可読性のトレードオフの問題から，可読性向上のためのモデルの変更はモデルの分類精度を低下させる可能性がある．そこで本章では，提案手法である Evolutionary generative contribution mappings (EGCM) について述べる．EGCM では (1) クラス分類において分類に有効な領域を示す Class contribution map の生成，(2) 進化計算法を用いたネットワーク構造最適化の 2 つの考え方のもとで精度と可読性がともに高い分類モデルの獲得を行う．以下，提案手法について詳しく述べていく．

5.2　Evolutionary Generative Contribution Mappings

EGCM は画像分類のための処理と解析のための処理を保持し，それぞれを End-to-end で学習させる．従来の CNN の処理はネットワークを $\phi_{\mathrm{CNN}}(\cdot)$，入力画像を x_0 としたとき，出力 y は次式で算出される．

$$y = \phi_{\mathrm{CNN}}(x_0) \tag{5.1}$$

一方，EGCM は重みマップ W を生成し，入力画像 x_0 と掛け合わせることによって三次元マップ $M \in \mathbb{R}^{m \times n \times v}$ を算出する．

$$M^c = x_0 \odot W^c \quad s.t.\, W = \phi_{\mathrm{EGCM}}(x_0) \tag{5.2}$$

ここで，演算子 $\odot$ は要素ごとの積，$c \in \{0, \ldots, C-1\}$ は分類クラスのインデックス，$\phi_{\mathrm{EGCM}}(\cdot)$ は畳み込みおよび逆畳み込み層から構成されるネットワークとする．重みマップ W は Class weight map (CWM) と呼び，クラス分類において重要な領域を画素単位で表す．また，三次元マップ M は Class contribution map (CCM) と呼び，クラス分類の根拠の提示だけではなくクラス分類に直接使用することが可能である．

最後に，得られた CCM の空間およびチャネル軸に対して平均 (Global average pooling) と Softmax 関数を適用することでクラス分類を行う．

$$y_c = \underbrace{\frac{1}{mnv} \sum_i \sum_j \sum_k M^c_{i,j,k}}_{\text{global average pooling}} \tag{5.3}$$

$$y = \mathrm{softmax}(y) \tag{5.4}$$

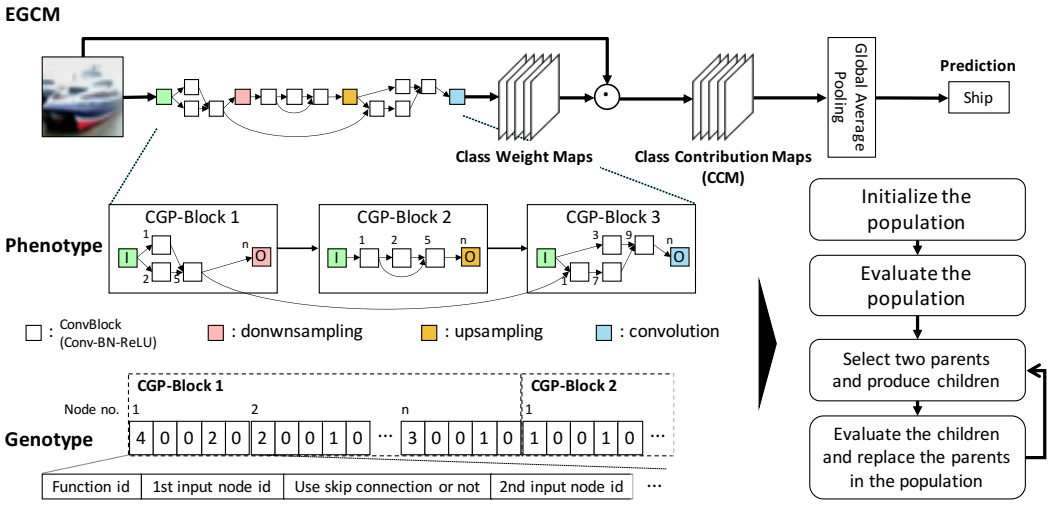


図 5.1: EGCM の表現型と遺伝子型の例

5.3 進化計算法を用いた構造探索

提案手法ではネットワーク構造を 2 次元の有向グラフで表現し，Cartesian genetic programming (CGP) によってその構造の最適化を行う．ここで EGCM のネットワークを 1 次元の文字列である遺伝子型で記述し，それを 2 次元のグリッド構造に変換することでネットワーク構造を記述する．図 5.1 に EGCM の表現型であるネットワーク構造とそれに対応する遺伝子型の例を示す．

5.3.1 EGCM のネットワークの構造表現

EGCM で生成される Class contribution maps (CCM) は式 (5.2) で示すとおり，Class weight maps (CWM) と入力画像の画素ごとの乗算によって算出される．そのため，CWM と入力画像の大きさは同じである必要がある．一方で一般的な CNN は複数の畳み込み層やプーリング層から構成されており，分類に有効な特徴量が階層的に獲得される．EGCM のネットワークにおいても出力の大きさを変えずにプーリング処理で有効な特徴量を獲得するため，ネットワーク表現を複数のブロックへの分割を行う．これを本論文では CGP-block と呼ぶ．

CGP-block は CGP グラフにおける部分グラフであると考えることができる．CGP-block の各ノードは特定の畳み込み処理を表し，同じブロック内のノード間でのみと接続される．この探索空間の制約により，一般的な CNN と同様の畳み込み処理による階層的な特徴量の獲得が可能となる．また，近年の高精度な深層学習モデルに用いられているスキップ接続やショートカット接続などの機構を導入するため，個体の初期化時に一定確率 r_{skip} で CGP-block 間を超えたノードの接続を許容するようにする．この確率を本論文では Skip probability と呼ぶ．この確率の値は構築されるネットワークのノード接続関係と個体の初期化時のネットワークの深さに影響を与えるパラメータであり，構造探索の方策を調節することが可能になる．具体的には Skip probability の値を大きく設定することで規模の小

さいネットワークから構造探索を行うことができ，小さな値を設定することで規模の大きなネットワークから構造探索を行うことができる．また，各 CGP-block の出力ノードを Transition node と呼び，ダウンサンプリング処理またはアップサンプリング処理を適用する．ただし出力チャネル数を調節するため，最後の CGP-block の Transition node のみ通常の畳み込み処理を適用する．

5.3.2 EGCM で用いるノード関数

CGP を用いて最適化を行う場合，事前にグラフ内で使用するノード関数を定義する必要がある．本論文では近年の CNN で用いられる処理を考慮し，次のノード関数セットを選択した．

Basic function set： Basic function set は次の関数から構成される：Convolution モジュール，Concatenation，Summation．Convolution モジュールは特定の畳み込み処理を行う関数である．それぞれの畳み込み処理は任意のチャネル数と出力数をもつが，畳み込み処理を行う際にゼロパディングを行うため処理の前後で特徴量マップのサイズは変わらない．Concatenation は 2 つの特徴量マップをチャンネル方向で結合する関数である．Summation は 2 つの特徴量マップを要素ごとに加算処理を行う関数である．チャネル数が異なる特徴量マップ同士に適用する場合は，チャネル数が少ない特徴量マップに対してゼロパディングを行いチャネル数をあわせた上で加算処理を行う．

Downsampling function set： Downsampling function set は次の関数から構成される：2×2 max pooling，2×2 average pooling，ストライド幅 2 の Convolution モジュールから構成される．

Upsampling function set： 本論文ではストライド幅 2 の逆畳み込み処理を適用することでアップサンプリング処理を行う．この関数を DeconvBlock と呼び，Deconvolution と Batch normalization，Rectified linear units (ReLU) 関数を順に適用する．DeconvBlock の畳み込み処理も ConvBlock と同様に任意のチャネル数と大きさのフィルタ処理が適用される．

5.3.3 Convoluton モジュール

本論文では，ConvBlock と ShakeShakeBlock の 2 つの Convolution モジュールを定義する．ConvBlock は畳み込み処理，Batch normalization，ReLU 関数から構成される処理であり，ShakeShakeBlock は Gastaldi によって提案された強力な深層学習の処理の 1 つである［28］．本論文では，これらのモジュールの機能の違いによって分類精度にどの程度影響を与えるか比較を行う．

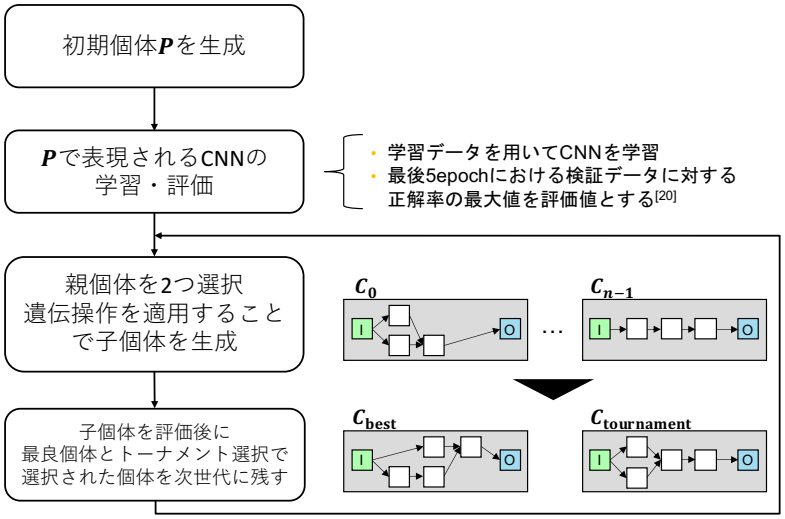


図 5.2: EGCM の最適化の流れ

5.3.4 表現型

EGCM の表現型は行数 N_r，列数 N_c 列の部分グラフ CGP-block の集合として表現される．このとき，各 CGP-block は $N_r \times N_c$ 個の中間ノードと 1 つの Transition node から構成される．また，各 CGP-block のノードは Level-back と呼ばれるパラメータ L によって接続関係が制御され，c 列目のノードは $(c-L)$ 列目から $(c-1)$ 列目のノードのみと接続することができる．この接続関係の制約により，CGP の固定長の遺伝子型で可変長のネットワークを表現することが可能となる．

5.3.5 遺伝型

EGCM の遺伝子型は 1 次元の文字列で記述される．遺伝子は各ノードで適用するフィルタ処理の種類や接続関係，スキップ接続の有無を整数文字列によって表現する．遺伝子の長さは構成される CGP-block の最大ブロック数に依存し，CGP-block のブロック数は入力画像の大きさから決定する．スキップ接続については，遺伝子の初期化時に Skip probability の確率に従い初期化を行う．

5.3.6 進化計算法

進化計算法の世代交代モデルには Minimal generation gap (MGG) を使用した．MGG は多様性保持に優れており，実数値 GA などで広く使用されている世代交代モデルである．このモデルでは各世代に個体集団から親個体を選択し，交叉や突然変異などの遺伝操作を適用することで λ の子個体を生成する．各個体は学習データセット $\mathcal{T}$ で学習を行い，検証データセット $\mathcal{V}$ を用いて評価する．このときの検証データセットに対する正解率を評価値とし，この評価値が最大となるように解の探索を行う．EGCM の最適化の流れとアルゴリズムの詳細をそれぞれ図 5.2 とアルゴリズム 3 に示す．

Algorithm 3 進化計算法を用いた構造探索

Input: 個体集団 $\mathcal{P}$, 子個体 C, 個体数 N_{pop}, 個体数 N_{child}, 世代数 N_{gen}, 学習データセット $\mathcal{T}$, 検証データセット $\mathcal{V}$,

Procedure:

for $i = 1$ to N_{pop} **do** ▷ 個体集団の初期化
  $\mathcal{P}_{i.gene} \leftarrow$ GENERATE( )
  $\mathcal{P}_{i.fitness} \leftarrow$ TRAINANDEVAL($\mathcal{P}_i; \mathcal{T}, \mathcal{V}$)
end for
for $g = 1$ to N_{gen} **do**
  $\mathcal{P}_m, \mathcal{P}_n (m \neq n) \leftarrow$ SELECTPARENTS($\mathcal{P}$)
  $C_1, C_2 \leftarrow \mathcal{P}_m, \mathcal{P}_n$
  for $i = 1$ to N_{child} **do**
    $C_{i+2.gene} \leftarrow$ CROSSOVER($\mathcal{P}_{m.gene}, \mathcal{P}_{n.gene}$)
    $C_{i+2.gene} \leftarrow$ MUTATE($C_{i+2.gene}$)
    $C_{i+2.fitness} \leftarrow$ TRAINANDEVAL($C_{i+2}; \mathcal{T}, \mathcal{V}$)
  end for
  $best =$ ELITISM($C_{.fitness}$) ▷ エリート個体の選択
  $select =$ TOURNAMENTSELECTION($C_{.fitness}$)
  $\mathcal{P}_m, \mathcal{P}_n \leftarrow C_{best}, C_{select}$
end for

5.4 画像分類実験

本節では，提案手法を複数の画像分類データセットに適用することで有効性の検証を行う．また，新たなデータセット Two-digit MNIST を提案し，EGCM で生成された可視化画像の性能評価を行う．また，生成された可視化画像に対して Sanity check [29] を行い，提案手法の有効性を定量的に評価した．提案手法は `PyTorch 0.4.1` [30] で実装し，NVIDIA GTX 1080ti を 2 枚搭載した計算機を用いて実験を行った．

5.4.1 データセット

CIFAR-10

CIFAR-10 [31] は 10 クラスの画像データセットであり，学習画像 50,000 枚とテスト画像 10,000 枚で構成される．従来手法 [32, 33] の分割方法に従って，学習画像からランダムに選択した 45,000 枚の画像を学習データ，残りの 5,000 枚を適用度算出のための検証データとした．

Street View House Number

Street View House Number (SVHN) データセット [34] は CIFAR-10 と同様に 10 クラスの画像分類用データセットであり，学習画像 73,257 枚とテスト画像 26,032 枚，追加の学習

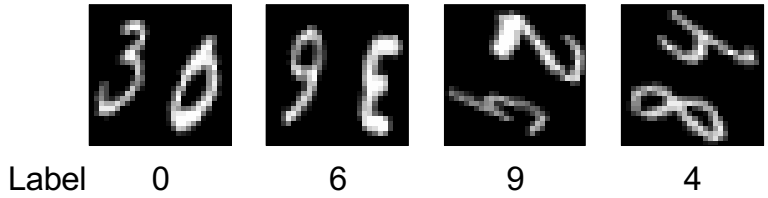


図 5.3: Two-digit MNIST の画像例

画像 531,131 枚から構成される．CIFAR-10 と同様に，5,000 枚のランダムに選択された画像を検証データ，残りの画像をネットワークの学習データとした．

Two-digit MNIST

EGCM は入力画像に対してクラス分類に有効な領域の可視化を行う．そのため，画像内に複数の物体が存在する場合であっても分類に有効な領域の可視化が可能であると考えられる．この性能を検証するため，本論文では新しいデータセット Two-digit MNIST を提案し，EGCM で生成された可視化結果の性能評価を行った．提案する Two-digit MNIST は 2 桁の数値の 1 桁目を正解する数値データセットであり，MNIST データセット [35] からランダムに選択された 2 枚の画像を横に並べてリサイズを行うことで生成した．図 5.3 に Two-digit MNIST の画像例を示す．他のデータセットと同様に，学習画像からランダムに選択した 5,000 枚の画像を検証データ，残りの画像をネットワークの学習データとした．

5.4.2 学習設定

ネットワークの学習には Nesterov の加速勾配法を使用し，CIFAR-10 と SVHN，Two-digit MNIST のデータセットに対して 50，25，25 エポックの学習を行った．Nesterov の加速勾配法 [36] の慣性項は 0.9，ミニバッチサイズは 64，Weight decay の値は 1.0×10^{-4} とした．学習率の初期値は 0.1 とし，全エポック数の 50% と 75% 経過時に学習率を 0.1 倍した．

提案手法による精度への影響を明確にするため，本論文では前処理は適用せず，画素値を 255 で除算するのみとした．また，入力画像に対して 4 画素のゼロパディングを行った後，32×32 画素領域を切り出すデータ拡張を行った．さらに CIFAR-10 に対しては，ランダムに水平方向へ反転を行った．各ネットワークを学習データを用いて学習し，最後 5 エポックの検証データにおける分類精度の最大値を適応度とした．この適応度が最大になるように構造探索を行った．

CGP の探索終了後，最も適応度の高かったネットワークを学習データすべてを用いて再学習を行い，テスト画像に対する精度を報告した．この際，SVHN については Extra データセットも使用して学習を行った．また，再学習の学習率等の設定は学習時と同じ設定で行った．CIFAR-10 と SVHN，Two-digit MNIST に対して，310，45，45 エポックの学習を行い，Cosine annealing [37] を用いて学習率の調節を行った．

5.4.3 探索空間および遺伝的アルゴリズムの設定

CGP-block に関する設定として，表現型の列数は $N_r = 3$，行数は $N_c = 5$，Level-back は $L = 3$ とした．また，CIFAR-10 と SVHN，Two-digit MNIST のそれぞれのデータセットにおける CGP-block の最大ブロック数は 9，9，5 とした．各 CGP-block 内のノードは特定の関数を表し，中間ノードは Basic function set から，Transition node は Downsampling function set および Upsampling function set からそれぞれ選択した．また，本論文で用いられる畳み込み層のフィルタ数 F とカーネルサイズ k はそれぞれ $F \in \{32, 64, 128\}$ と $k \in \{1, 3, 5\}$ から選択されるものとした．

構造探索に用いる進化計算法の世代交代モデルには Minimal generation gap（MGG）[18] を使用した．MGG の設定として，最大世代数 $N_{\text{gen}} = 250$，個体数 $N_{\text{pop}} = 10$，ここ大数 $N_{\text{child}} = 4$，トーナメント数 $T = 2$，skip probability $r_{rmskip} = 0.8$，突然変異率 $\mu = 0.05$ を使用した．

5.4.4 実験結果

比較手法との分類精度の比較

提案手法と比較手法の分類誤差率および構築されたネットワークのパラメータ数の結果を表 5.1 に示す．表の結果から提案手法は他の比較手法と同等以上の性能を示していることが分かる．特に EGCM（ConvBlock）はパラメータ数に対して高い分類精度を達成していることが分かる．DenseNet や ResNet と比較すると分類精度が劣ってしまっているが，これらのモデルは ResBlock や DenseBlock などの非常に高機能なモジュールを使用している．一方，提案手法は畳み込み処理やスキップ接続などの基本的な処理のみでネットワークを構築し，同等程度の分類精度を達成している．EGCM（ShakeShakeBlock）は EGCM（ConvBlock）と比較して高い精度を達成していることが分かる．このモデルは約 10M 程度のパラメータ数であるが，パラメータ数が 30.0M 以上の FractalNet や Wide ResNet などの高精度なモデルよりも高い分類精度を示していることが分かる．

また，構造探索を行わない GCM [3] の CIFAR-10 と SVHN における分類誤差率はそれぞれ 9.57% と 3.81% であったことからも，構造最適化を行うことで任意のデータセットに対して適切なネットワークが自動構築されて分類精度が大幅に向上していることが分かる．

5.4.5 生成された可視化画像の解析

図 5.4 に CIFAR-10 における提案手法の可視化結果を示す．この可視化結果から，提案手法が画像分類問題において有効な領域を可視化していることが分かる．例えば airplane クラスの分類において，空の領域ではなく飛行機の領域が強く反応していることが分かる．また，truck クラスの分類においては automobile クラスにも反応が見られることが分かる．このとき，車のボンネット領域はそれぞれの可視化画像で反応しているが，荷台の領域は truck クラスの可視化画像のみで確認された．このことから truck クラスと automobile クラスの分類を行う上では荷台の領域が有効であるということを考察することが可能となる．

表 5.1: CIFAR-10 と Street View House Number データセットにおける分類誤差率およびパラメータ数の比較．提案手法の精度は 3 施行における分類精度誤差率を示す．

Model	**#Params ($\times 10^6$)**	**CIFAR-10**	**SVHN**
Maxout [38]	–	9.34	2.47
Network in Network [39]	–	8.81	2.35
ResNet (referred from [40])	1.7	6.61	2.01
FractalNet [41]	38.6	5.22	2.01
Wide ResNet [42]	36.5	4.00	–
DenseNet ($k = 40$) [42]	27.2	3.74	1.59
DenseNet + BC ($k = 40$)	25.6	**3.46**	–
EGCM (ConvBlock)	2.23	4.93 (5.05 ± 0.16)	–
	1.80	–	1.63 (1.70 ± 0.09)
EGCM (ShakeShakeBlock)	9.27	3.84 (3.96 ± 0.17)	–
	10.2	–	**1.49** (1.58 ± 0.08)

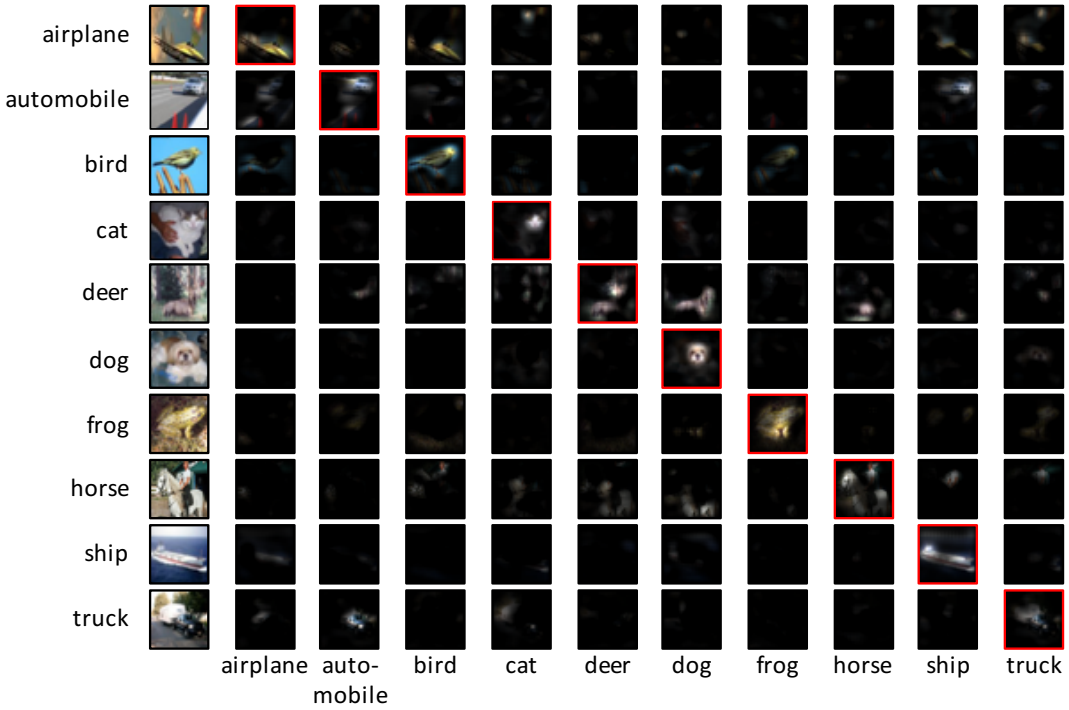


図 5.4: CIFAR-10 において提案手法で生成された可視化結果例

次に誤分類した CIFAR-10 画像に対する可視化結果例を図 5.5 に示す．この画像から構築した分類器が誤分類した原因を考察することができる．例えば，図 5.5 において分類器は deer 画像を horse クラスと誤分類している．しかし，可視化画像では deer の身体領域のみが反応しており，horse と誤分類した原因の解析につなげることができる．

図 5.5: CIFAR-10 において誤分類した画像に対する可視化結果例

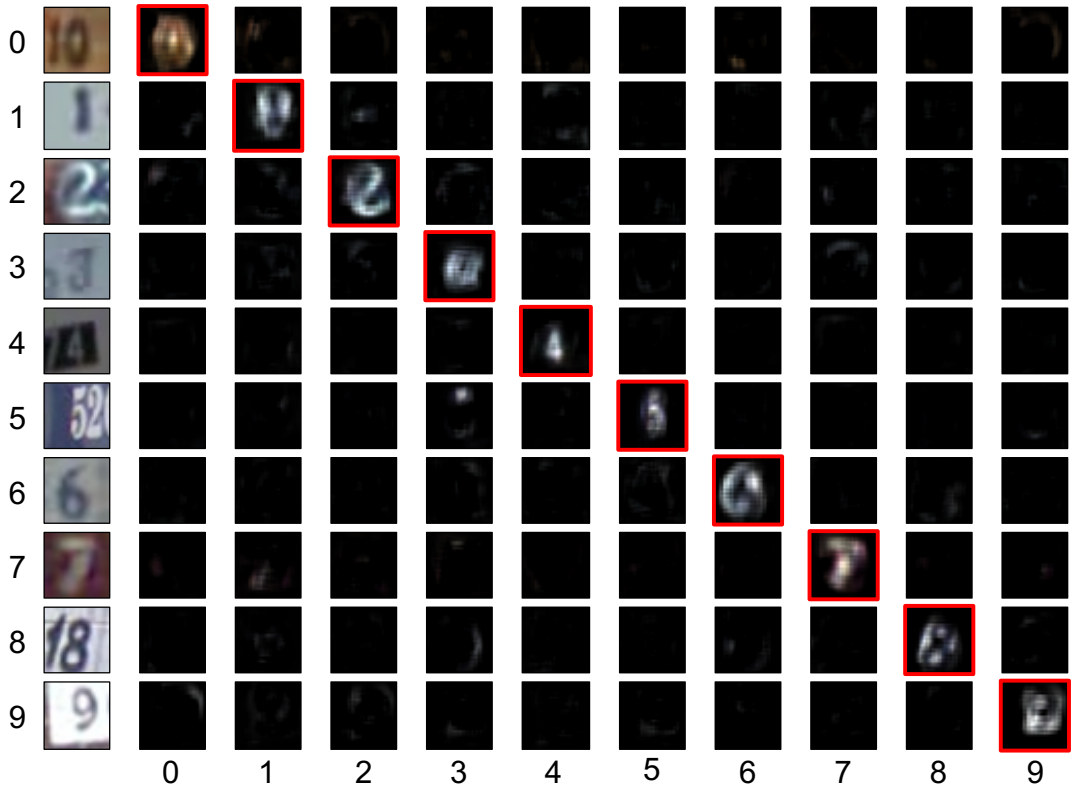


図 5.6: Street View House Number において提案手法で生成された可視化結果例

最後に SVHN における提案手法の可視化結果を図 5.6 に示す．この可視化結果からも，提案手法で構築した分類器が画像内のクラス分類に有効な領域を正しく反応させていることが分かる．

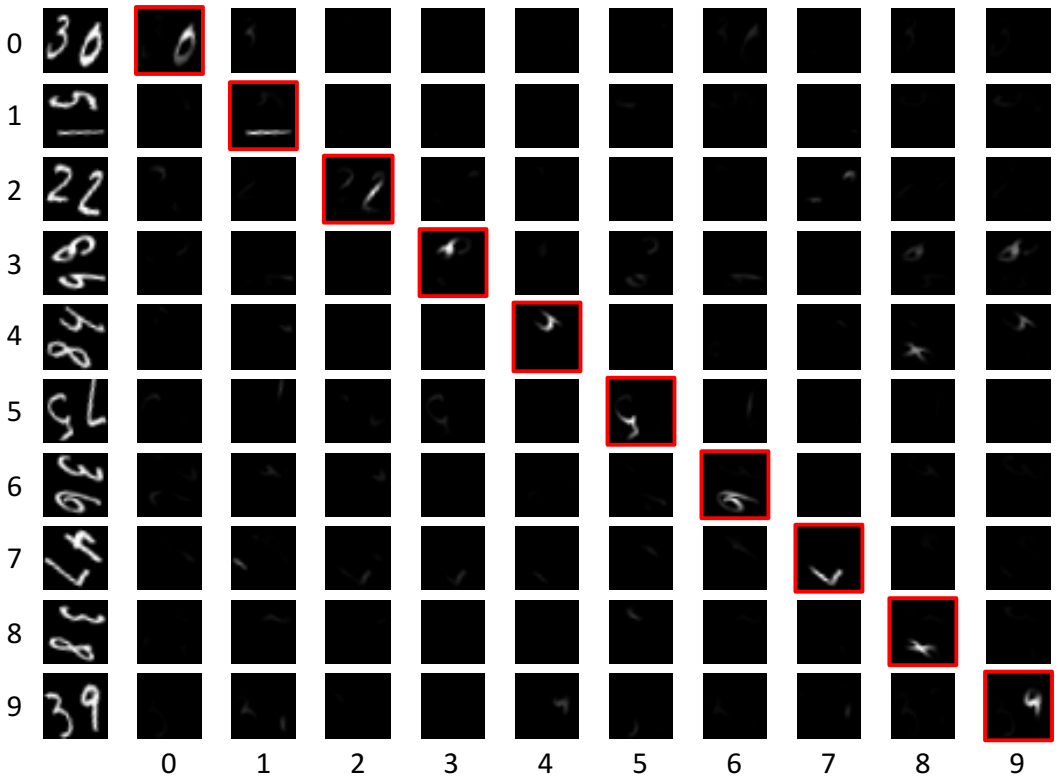


図 5.7: Two-digit MNIST において提案手法で生成された可視化結果例

図 5.8: Two-digit MNIST において判定が困難な画像に対する可視化結果例

5.4.6 Two-digit MNIST を用いた画像分類実験結果

Two-digit MNIST において，提案手法は 1.99% の分類誤差率であった．このときの可視化結果を図 5.7 に示す．可視化結果から，画像内の 1 桁目の数字が反応していることが分かる．Two-digit MNIST は 1 桁目の数字を判定する分類問題であるため，提案手法が画像を分類する上で重要な領域を正しく注目していることが分かる．また，これらの可視化結果を用いることで判定に対する考察を行うことも可能である．図 5.8 にその例を示す．Two-digit MNIST において，‘98’ は ‘86’ と誤分類する可能性のある画像である．今回の例においては分類機は正しい判定をすることができたが，可視化結果では ‘8’ と ‘9’ の両方の数字に対して反応していることが分かる．この可視化結果から分かるとおり，提案手法を用いることで構築した分類器がどのように判定を行ったかを利用者に提示することも可能である．

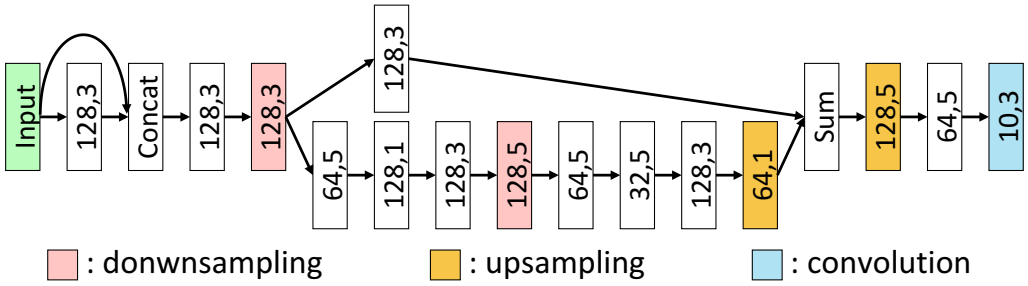


図 5.9: CIFAR-10 において提案手法によって獲得されたネットワーク構造例．各ノードはフィルター数 F とカーネルサイズ k を示す．

5.4.7　獲得されたネットワーク構造の解析

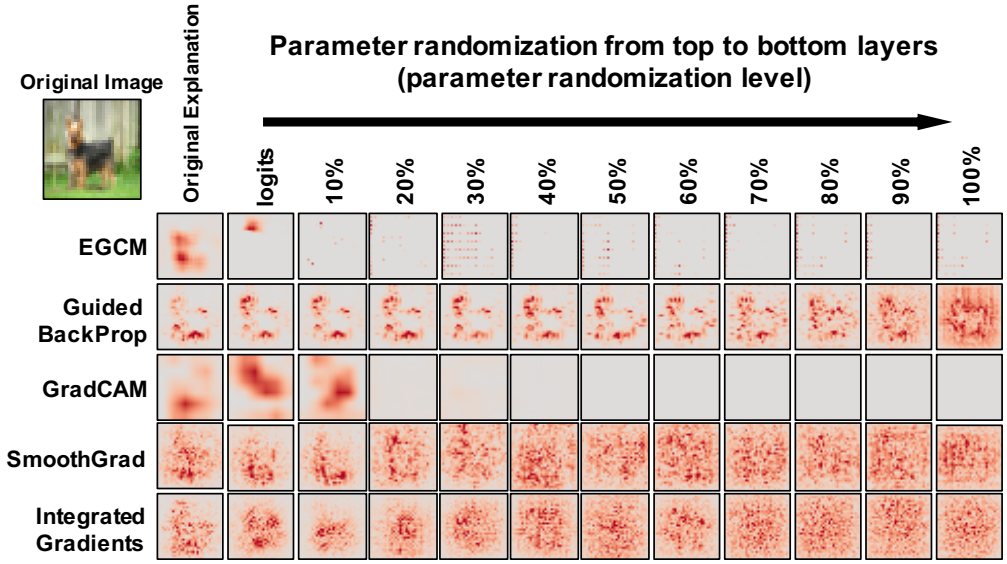
CIFAR-10 において獲得されたネットワーク構造を図 5.9 に示す．獲得されたネットワーク構造はスキップ接続や様々な大きさのフィルタをもち，人手での設計が困難な構造であることが分かる．また，獲得されたネットワークは過度なダウンサンプリングが行われない構造であることも確認された．これは過度なダウンサンプリングによる特徴量抽出は提案モデルにおいて有効でないためと考えられる．さらに実験では Chen ら [43] によって提案された Attention 機構に類似した構造が多く獲得されることが確認された．これらの結果から，提案手法は自動で適切なネットワーク構造を獲得していることが確認できる．

5.4.8　Sanity check による可視化結果の評価

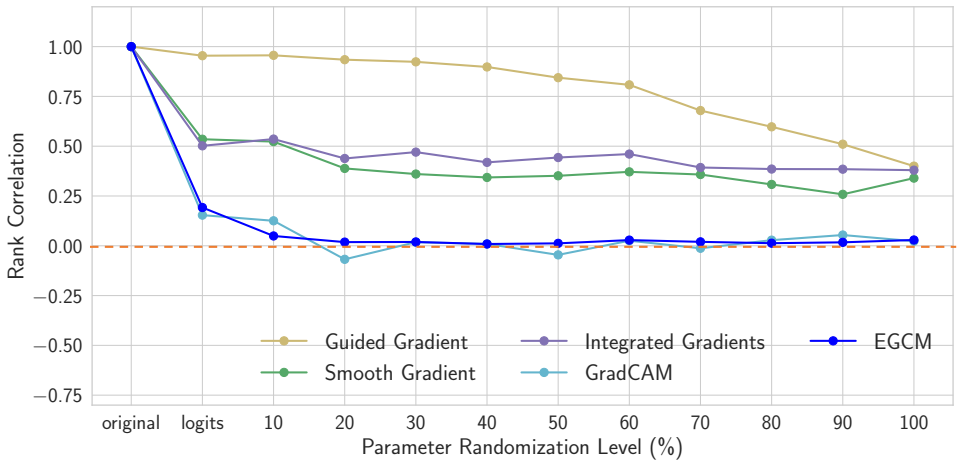
次に Sanity check [29] を用いた可視化結果の評価を行った．Sanity check は Prameter randomization テストと Data randomization テストの 2 つの評価実験から構成される．そして，これら 2 つの評価実験によって生じた変化を Spearman の順位相関係数を用いて定量的に評価する．

Parameter randomization テスト　Parameter randomization テストでは学習済みモデルの重みを出力層から順に初期化を行い，その際に生じる可視化画像の変化を比較する．Parameter randomization テストの結果を図 5.10 に示す．Adebayo ら [29] の研究では，従来手法で生成される可視化画像はネットワークの層を初期化した場合も元々の可視化から変化が生じないことが報告されている．これは生成された可視化画像が学習済みモデルの重みに依存していないことを示し，モデルの根拠説明としては不適当であることを表す．一方，EGCM で生成される可視化画像は出力層から重みを初期化した直後に可視化画像が大きく変化している．この変化は Spearman の順位相関係数にも表れており，出力層を初期化した直後に大きく値が減少している．

Data randomization テスト　Data randomization テストは正しくないラベルを用いてモデルの学習を行い，その際に生じる可視化画像の変化を比較する．Data randomization テストの結果を図 5.11 に示す．可視化結果から分かるとおり，EGCM で生成される可視化画

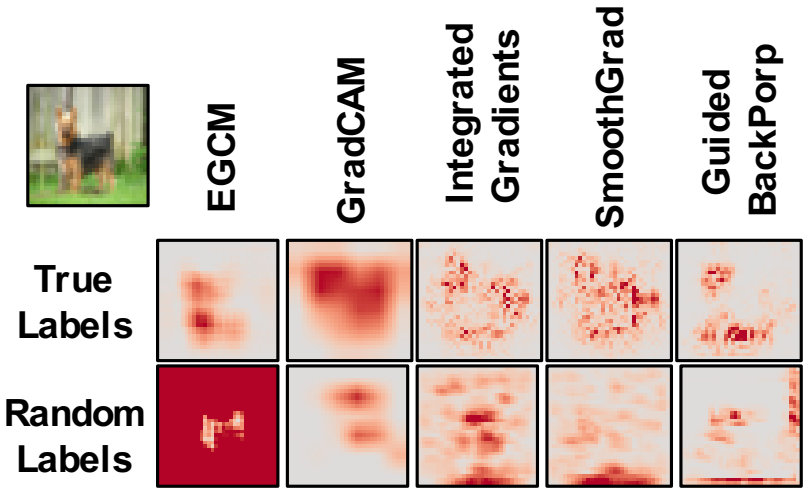


(a) Parameter randomization テストの定性評価

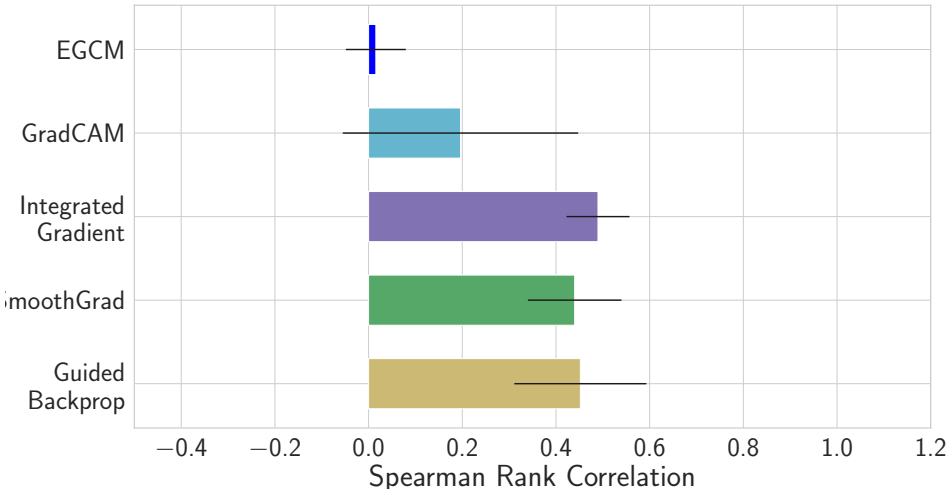


(b) Parameter randomization テストの定量評価

図 5.10: dog クラスの画像に対する Parameter randomization の結果例．可視化結果の横軸はネットワークの重みの初期化の割合を示している．



(a) Data randomization テストの定性評価



(b) Data randomization テストの定量評価

図 5.11: dog クラスの画像に対する Data randomization テストの結果例．それぞれ正しくないラベルを用いて学習したモデルで生成される可視化結果に対するテスト結果例を示す．

像は分類対象の特定に失敗している．この結果は，EGCM で生成された可視化画像は学習に用いた教師ラベルに依存していることを示している．

これら 2 つの結果より，EGCM によって生成された可視化結果はモデルの重みと学習に用いた教師ラベルに依存していることが分かる．これは EGCM が Sanity check のテストを通過したことを示し，入力画像と可視化画像の関係を正しく表現している．

5.5 まとめ

本章では，クラス分類の根拠を直感的に可視化する Evolutionary Generative Contribution Mappings (EGCM) の提案を行った．複数のクラス分類問題に適用し，提案手法によって構築されたモデルが高い精度と可読性を保つことを示した．一方，本論文では比較的に小さい個体数や子個体数，世代数を用いたことから，さらなる精度向上を見込むことができると考える．今後は，より高精度かつ可読性の高い可視化の枠組みを検討する必要がある．また，ImageNet などの他の大規模データセットに提案手法を適用し，有効性の検証を行っていく必要もある．

第6章　結論

本論文では，機械学習や進化計算法などで構築した画像分類器の分類精度と可読性の両立を目的とした手法の提案を行い，複数の画像分類問題に適用することで手法の有効性を検証した．精度と可読性のトレードオフの問題に対して，各章でそれぞれ手法を提案した．各章で得られた成果は以下の通りである．

- **If-then ルールを用いた分類器の精度と可読性の向上**
  If-then ルールを用いた分類器の精度と可読性の向上として，分類に用いられている特徴量分布を考慮したヒートマップの作成と特徴量の可視化の 2 つの可視化手法の提案を行った．提案手法を一般画像分類問題に適用し，獲得した可視化画像の有効性の検証を行った．結果として，構築した分類器の分類過程を直感的に利用者に示すことが可能となり，分類器の信頼性の向上につなげることができた．

  今後の課題として，利用者の要求度に合わせて分類クラスの違いを明確に提示することが可能な説明方法の検討が挙げられる．これは医用画像などの産業応用では判定結果に対するすべての説明を行う必要はなく，分類クラスの違いが明確になればよい場合があるためである．そのため，多クラス分類の分類過程を説明する際にクラス間の違いが明確に分るような説明方法の検討が必要であると考える．また，生成された説明の分かりやすさを数値などで評価し，提案手法の有効性を客観的に示すことも今後の課題である．

- **学習済み深層学習モデルの特徴量の可視化**
  学習済み深層学習モデルの特徴量の可視化方法として, Generative adversarial networks (GAN) の枠組みを用いた学習済み CNN の特定ユニットを活性化させる画像生成方法を提案し，その有効性を検証した．提案手法は GAN の枠組みを拡張し，学習済み CNN の特定ユニットを活性化させる自然画像を生成する．先行研究では学習済み CNN の特定ユニットの活性化と自然画像への変換を同時に満たすための制約を人手で設計した上で生成を行う．一方，提案手法は GAN の枠組みを用いることで，先行研究で提案されてきた制約を用いることなく可視化画像の生成が可能である．提案手法を学習済み AlexNet に適用して各層の特定ユニットを活性化させる画像生成を行った結果，分類に有効な特徴量がネットワークの内部で段階的に獲得されたことが確認された．また，階層的な特徴量の可視化を行うと，モデル内部でクラス分類に有効な特徴量が獲得されていることも分かった．

  今後の課題として，多様性のある可視化画像の生成が挙げられる．提案手法を用いることで学習済み CNN の任意のユニットを活性化させる自然画像を生成することができるが，Generator に入力する乱数ベクトル z を変化させた際の生成画像に多様性は見られない．一方で施行間においては多様的な画像が生成される場合もある．今

後は一施行の中で多様的な画像生成が可能なネットワークの学習方法の検討を行う必要がある．また，一般画像だけでなく，医用画像などの実画像に提案手法を適用し，有効性の検証を行っていく必要もある．

- **高精度かつ高い可読性をもつ深層学習**
  高い分類精度かつ高い可読性をもつ新しい深層学習モデルとして Evolutionary generative contribution mappings (EGCM) の提案を行い，その有効性を検証した．提案手法はクラス分類において分類に有効な領域を示す Class contribution map の作成と，進化計算法を用いたネットワーク構造の最適化によって，精度と可読性がともに高い分類器の獲得を行う．先行研究では精度と可読性のトレードオフの問題のため，可読性向上の機構を導入することで分類精度の低下が見られたが，本手法では可読性の高いモデルに適したモデル構造を自動最適化することで精度を保ったまま高い可読性を達成することが可能となった．複数のデータセットに提案手法を適用したし，データセットに対して適切なモデル構造が自動獲得され，高い精度と高い可読性を保つことを示した．また，獲得した可視化結果に対して Sanity check を行い定量的に手法の評価を行った．結果として，生成された可視化画像は入力画像を適切に説明していることを確認した．

  今後の課題として，より可読性かつ可読性の高い可視化の枠組みを検討が挙げられる．また，ImageNet などの他の大規模データセットに提案手法を適用し，有効性の検証を行っていく必要もある．

以上のように，本論文で挙げた 3 つの方策のもとでそれぞれのモデルに適切な可読性向上の手法を提案した．いずれの手法も画像分類器の可読性向上であるが，利用者の目的やニーズに合わせて適切に手法を選択する必要があると考える．具体的には，学習データが少ない場合は第 3 章で提案した If-then ルールを用いた分類器に関する手法が有効であり，すでに学習済みの深層学習の分類器が手元にある場合は第 4 章で提案した深層学習の特徴量の可視化に関する手法が有効であるといえる．そして，それ以外で学習データを十分に用意することができる場合は第 5 章で提案した精度と可読性の両立を行う手法が有効である．

また，本論文で提案した手法も精度と可読性の両立を行う手法の一つであるといえる．利用者や分類器の適用先によって求める説明やその粒度が異なることがあり，計算機が構築した処理の産業応用を考える場合はそれを考慮した上で適切な手法を選択することも重要であると考える．

謝辞

博士課程後期進学の後押しおよび研究を遂行するにあたり多大なるご指導とご助言，素晴らしい研究環境を賜りました長尾智晴先生に深く感謝いたします．また，本論文をまとめるにあたり貴重なご指導とご助言をいただきました田村直良先生，森辰則先生，原下秀士先生，富井尚志先生，白川真一先生に感謝申し上げます．

本研究に際してご支援いただきました長尾研究室の皆さま，研究内容にご助言をいただきました荒井 敏様ならびに共同研究の技術者の皆さまに感謝申し上げます．特に長尾研究室の皆さまとは飲み会や研究室合宿など，研究以外にも楽しい学生生活を送ることができました．研究室での生活は自分にとって良い思い出でです．ありがとうございました．

最後に，家族や友人をはじめ，温かく見守ってくれた方々に感謝します．ありがとうございました．今後ともどうぞよろしくお願いいたします．

参考文献

［1］ Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 427–436, 2015.

［2］ Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921–2929, 2016.

［3］ 荒井敏, 長尾智晴. 畳み込みニューラルネットワークを用いた画像分類タスクの直感的可視化方法. 情報処理学会論文誌数理モデル化と応用 (TOM), Vol. 10, No. 2, pp. 1–13, 2017.

［4］ 中山史朗, 穂積知佐, 矢田紀子, 長尾智晴. 進化的条件判断ネットワーク EDEN による画像分類. 映像情報メディア学会誌, Vol. 67, No. 7, pp. J278–J285, 2013.

［5］ J. Ross Quinlan. *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers Inc., 1993.

［6］ 崎津実穂, 菅沼雅徳, 土屋大樹, 長尾智晴. 決定木および決定ネットワークによる画像分類過程の説明文の自動生成. 情報処理学会論文誌数理モデル化と応用（TOM）, Vol. 9, No. 1, pp. 43–52, 2016.

［7］ Karen Simonyan, Andrea Vedald, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *Proceedings of the 2nd International Conference on Learning Representaions Workshop*, 2014.

［8］ Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *Proceedings of the 13th European Conference on Computer Vision*, pp. 818–833, 2014.

［9］ Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.

［10］ Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*, p. e0130140, 2015.

[11] Dumitru Erhan, Yoshua Bengio, Aaron Courville, and Pascal Vincent. Visualizing higher-layer features of a deep network. Technical report, University of Montreal, 2009.

[12] Aravindh Mahendran and Andrea Vedaldi. Visualizing deep convolutional neural networks using natural pre-images. *International Journal of Computer Vision*, Vol. 120, No. 3, pp. 233–255, 2016.

[13] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson. Understanding neural networks through deep visualization. In *Proceedings of the 32th International Conference on Machine Learning Workshop on Deep Learning*, 2015.

[14] Donglai Wei, Bolei Zhou, Antonio Torralba, and William T. Freeman. Understanding intra-class knowledge inside CNN. *arXiv preprint arXiv:1507.02379*, 2015.

[15] Dan Pelleg and Andrew W. Moore. X-means: Extending k-means with efficient estimation of the number of clusters. In *Proceedings of the 17th International Conference on Machine Learning*, pp. 727–734, 2000.

[16] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, Vol. 42, No. 3, pp. 145–175, 2001.

[17] Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *Proceedings of the Indian Conference on Computer Vision Graphics and Image Processing*, 2008.

[18] Hiroshi Satoh, Isao Ono, and Shigenobu Kobayashi. Minimal generation gap model for gas considering both exploration and exploitation. In *Proceedings of IIZUKA '96*, pp. 494–497, 1996.

[19] Zhenhua Guo, Lei Zhang, and David Zhang. A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions Image Processing*, pp. 1657–1663, 2010.

[20] Gregory Griffin, Alex Holub, and Pietro Perona. Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology, 2007.

[21] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning*, pp. 448–456, 2015.

[22] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning*, pp. 807–814, 2010.

[23] Tim Salimans and Diederik P. Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems 29 (NIPS '16)*, pp. 901–909, Barcelona, Spain, 2016.

[24] Andrew L. Maas, Awni Y. Hannun, and Andrew Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of the 30th International Conference on Machine Learning (ICML '13) Workshop on Deep Learning for Audio, Speech and Language Processing*, Gerorgia, USA, 2013.

[25] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[26] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the 22th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

[27] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representaions*, 2015.

[28] Xavier Gastaldi. Shake-shake regularization. *arXiv preprint:1705.07485*, 2017.

[29] Julius Adebayo, Justin Gilmer, Michael Muelly, Ian J. Goodfellow, Moritz Hardt, and Been Kim. Sanity checks for saliency maps. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 9505–9515, 2018.

[30] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. In *Autodiff Workshop in Neural Information Processing Systems*, 2017.

[31] Alex Krizhevsky and Geoffrey E. Hinton. Learning multiple layers of features from tiny images. *Technical report*, 2009.

[32] Bowen Baker, Otkrist Gupta, Nikhil Naik, and Ramesh Raskar. Designing neural network architectures using reinforcement learning. In *Proceedings of the 5th International Conference on Learning Representations*, 2017.

[33] Barret Zoph and Quoc V. Le. Neural architecture search with reinforcement learning. In *Proceedings of the 5th International Conference on Learning Representations*, 2017.

[34] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011*, 2011.

[35] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.

[36] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *Proceedings of the 30th International Conference on International Conference on Machine Learning*, pp. 1139–1147, 2013.

[37] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with restarts. In *Proceedings of the 5th International Conference on Learning Representations*, 2017.

[38] Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. In *Proceedings of the 30th International Conference on Machine Learning*, pp. 1319–1327, 2013.

[39] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. In *Proceedings of the 2nd International Conference on Learning Representations*, 2014.

[40] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Q. Weinberger. Deep networks with stochastic depth. In *Proceedings of the 16th European Conference on Computer Vision*, pp. 646–661, 2016.

[41] Gustav Larsson, Michael Maire, and Gregory Shakhnarovich. Fractalnet: Ultra-deep neural networks without residuals. In *Proceedings of the 5th International Conference on Learning Representations*, 2017.

[42] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *Proceedings of the British Machine Vision Conference*, pp. 87.1–87.12, 2016.

[43] Wei Chen, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3156–3164, 2017.

研究業績リスト

論文誌

1. 小林雅幸，菅沼雅徳，崎津実穂，長尾智晴：進化的条件判断ネットワークにおける画像分類過程の可視化，進化計算学会論文誌，Vol.7，No.3，pp.65-76，2017

2. Masayuki Kobayashi, Masanori Suganuma and Tomoharu Nagao: A Generative Model Approach for Visualising Convolutional Neural Networks, International Journal of Computational Intelligence Studies, Vol.7, pp.214-230, 2018

3. Masanori Suganuma, Masayuki Kobayashi, Shinichi Shirakawa and Tomoharu Nagao: Evolution of Deep Convolutional Neural Networks Using Cartesian Genetic Programming, Evolutionary Computation, MIT Press, pp.141-163, 2019

4. 藤田耕作，小林雅幸，長尾智晴：進化的画像処理を用いた指定特徴量をもつ画像の自動生成，映像情報メディア学会誌, 第 73 巻, 第 5 号，pp.987-992，2019

国際会議発表

1. Masayuki Kobayashi, Masanori Suganuma and Tomoharu Nagao: Generative Adversarial Network for Visualizing Convolutional Network, Proceedings of the 10th International Workshop on Computational Intelligence & Applications 2017 (IWCIA 2017), Hiroshima, Japan, Nov.11-12, 2017

2. Yuta Hasunuma, Chiaki Hirayama, Masayuki Kobayashi and Tomoharu Nagao: Non-parallel Voice Conversion using Generative Adversarial Networks, Proceedings of 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2020), Miyazaki, Japan, Oct. 7-10, 2018

3. Kazuki Takaishi, Masayuki Kobayashi, Miku Yanagimoto and Tomoharu Nagao: Percolative Learning: Time-Series Predictions from Future Tendencies, Proceedings of 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2020), Miyazaki, Japan, Oct. 7-10, 2018

4. Kosaku Fujita, Masayuki Kobayashi and Tomoharu Nagao: Data Augmentation using Evolutionary Image processing, Proceedings of Digital Image Computing: Techniques and Applications (DICTA 2018), Canberra, Australia, Dec. 10-13, 2018

5. Masayuki Kobayashi and Tomoharu Nagao: An Evolution-based Approach for Efficient Differentiable Architecture Search, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2020), pp.131-132, Cancun, Mexico, July 8-12, 2020

6. Masayuki Kobayashi and Tomoharu Nagao: A Multi-objective Architecture Search for Generative Adversarial Networks, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2020), pp.133-134, Cancun, Mexico, July 8-12, 2020

7. Masayuki Kobayashi, Satoshi Arai, Tomoharu Nagao: Evolutionary Generative Contribution Mappings, Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2020), Toronto, Canada, Oct. 11-14, 2020

国内学会発表

1. 小林雅幸，崎津実穂，長尾智晴：決定木および決定ネットワークの画像分類過程の可視化，2016 年電子情報通信学会総合大会，ISS 学生ポスターセッション，2016

2. 小林雅幸，菅沼雅徳，崎津実穂，長尾智晴：進化的条件判断ネットワークの画像分類過程の可視化，情報処理学会第 108 回数理モデル化と問題解決研究会，2016

3. 小林雅幸，菅沼雅徳，崎津実穂，長尾智晴：進化的条件判断ネットワークにおける画像分類過程の可視化，第 11 回進化計算学会研究会，2016 年

4. 藤田耕作，小林雅幸，長尾智晴：進化的画像処理による学習画像の自動生成，第 16 回情報科学技術フォーラム，2017 年

5. 畠　崇人，小林雅幸，長尾智晴：探索特性に可読性をもつ進化計算法，第 28 回インテリジェント・システム・シンポジウム (FAN2018)，GS2-5，2018 年

受賞

1. IEEE SMC The Brain Computer Interface Designers Hackathon 2018 受賞