

遺伝的プログラミングを用いた階層的な特徴構築による 画像分類

菅沼 雅徳^{1,2,a)} 土屋 大樹¹ 白川 真一¹ 長尾 智晴¹

受付日 2016年6月5日, 再受付日 2016年7月25日/2016年8月30日,
採録日 2016年9月7日

概要: 本論文では, 画像分類のための階層的な特徴構築手法を提案する. 提案手法では入力画像に対して, (1) 既存の画像処理フィルタの組合せ, (2) 遺伝的プログラミングで構築したフィルタ, の2層の画像変換処理を行い, 変換後の画像の各画素値を分類のための特徴量として扱う. 1層目の画像処理フィルタの組合せと2層目でのフィルタは遺伝的プログラミングを用いて段階的に構築する. 本論文では, カプセル内視鏡から撮影された小腸画像における異常画像と正常画像の分類問題に提案手法を適用し, 従来手法との比較と性能の検証を行う.

キーワード: 遺伝的プログラミング, 画像分類, 特徴構築, 多層構造

Image Classification Based on Hierarchical Feature Construction Using Genetic Programming

MASANORI SUGANUMA^{1,2,a)} DAIKI TSUCHIYA¹ SHINICHI SHIRAKAWA¹ TOMOHARU NAGAO¹

Received: June 5, 2016, Revised: July 25, 2016/August 30, 2016,
Accepted: September 7, 2016

Abstract: In this paper, we propose a hierarchical feature construction method for image classification. Our method has two feature construction stages: (1) feature construction by a combination of existing image processing filters, and (2) feature construction by evolved filters. The combination of image filters and evolved filters are constructed step-by-step using genetic programming. We verify the classification performance of the proposed method on the small bowel images taken from a capsule endoscope.

Keywords: genetic programming, image classification, feature construction, multi-layer architecture

1. はじめに

画像分類 [1], [2] や物体認識 [3], [4] などの技術は幅広い分野で必要とされているため, さかんに研究が行われている. 特に, local binary pattern (LBP) [5], histogram of oriented gradients (HOG) [6], scale-invariant feature transform (SIFT) [7] や Gabor bank [8] などの人手によ

て考案された特徴量を用いた分類手法がこれまでに高い分類精度を示してきた. しかし, これらの特徴量は特定の種類の画像分類問題では有効に働くが, ほかの種類の画像分類問題では有効に働かない可能性が考えられる. たとえば, LBP 特徴はテクスチャ画像分類では高い分類精度を示すが, シーン画像分類では有効に働かないと考えられる. 対象となる分類問題は目的によって多種多様であり, それら対象となる分類問題に応じて新たに特徴量を設計することは膨大な労力を必要とする. そこで計算機による特徴量の自動構築手法が求められている.

近年では, 画像分類のための特徴構築に深層学習 [9] が数多く適用されており, 多くの分類問題で成功を収めてい

¹ 横浜国立大学大学院環境情報学府
Graduate School of Environment and Information Sciences,
Yokohama National University, Yokohama, Kanagawa 240-
8501, Japan

² 日本学術振興会特別研究員 DC
JSPS Research Fellow

a) suganuma-masanori-hf@ynu.jp

る。深層学習では、複数の演算処理が段階的に続く多層構造によって分類に有効な特徴量を自動構築している。多層構造による特徴構築には、greedy layer-wise training [10] と呼ばれる方法が提案されており、単層構造で構築された特徴量より分類に有効な特徴量が構築されることが報告されている [11]。

また、遺伝的プログラミング (Genetic Programming; GP) [12] を用いて特徴量を自動構築する手法も数多く提案されており、画像分類問題において成功を取めている。GP を用いた手法では、(1) 事前に定義した特徴量の組合せから新たに特徴量を構築する手法、(2) 生の画素値の組合せから新たに特徴量を構築する手法が主にあげられる。

前者の手法では、GP の入力には画像全体の標準偏差や画像中心部分の平均画素値などの事前に定義した統計特徴量で構成され、それら統計特徴量の演算の組合せによって分類に有効な特徴量を構築している [13], [14]。しかし、事前に定義した統計特徴量の組合せで分類することが難しい問題では新しい統計特徴量を追加する必要がある。このため、後者の手法のように入力画像中の画素値から新たに特徴量を構築する手法の実現が望まれる。本論文では、後者の画素値から特徴量を構築する新しい手法を提案する。

後者の手法では、GP の入力は入力画像中の画素値または画像全体で構成され、それらに対する演算の組合せによって特徴量を構築する。Al-Sahaf らは、画像内から特徴量を抽出する領域と、その領域内の画素値から算出した統計特徴量の組合せによって特徴量を構築する構造を提案しており、GP を用いて構造の構築を行っている [15]。Kowaliw らは、Cartesian Genetic Programming (CGP) [16], [17] を用いて入力画像に対して画像変換を行い、変換後の画像から複数の統計特徴量を算出することで分類のための特徴量を構築している [18]。文献 [19] では入力画像からの特徴構築と分類を同時に最適化する Genetic Image Network for Image Classification (GIN-IC) を提案している。GIN-IC では、既存の画像処理フィルタによる画像変換後の画像から統計特徴量を算出し、算出した統計特徴量の演算結果を用いて分類まで行う構造を最適化する。Hirano らも同様に画像処理フィルタによる画像変換に基づく特徴構築手法を提案している [20]。これら画像変換に基づく特徴構築手法は、画像分類問題において有効であることが示されている。

一方、これら GP による特徴構築手法の多くは 1 段階の最適化によって特徴量を構築している。近年の深層学習の研究成果で示されているように、特徴量を構築する構造の多層化を行うことで単一の構造よりも分類に有効な特徴量を構築できることが GP の手法においても期待できる。実際に、Agapitos らは特徴量の最適化を行う構造の多層化を行うことで、1 段階の最適化よりも分類に有効な特徴量を構築できることを示している [21]。Agapitos らの手法は、(1) ランダムフィルタによる画像変換、(2) GP による画像

変換、(3) 平均プーリング、(4) ロジスティック回帰による分類、の 4 層構造を用いている。しかし、Agapitos らの手法では処理の初めにランダムフィルタによる画像変換を行っており、ランダムフィルタによる変換では変換後の画像から効率的に分類に有効な特徴量が抽出されることは難しいと考えられる。

そこで本論文では、(1) 既存の画像処理フィルタの組合せによる画像変換、(2) GP で構築したフィルタ処理による画像変換、の 2 種類の処理を 2 段階の進化計算によって最適化することで特徴構築を実現する手法を提案する。ランダムフィルタではなく、既存の画像処理フィルタの組合せによる画像変換処理を進化過程に組み込むことでより分類に有効な特徴量の構築が可能になると考えられる。また、近年の画像分類問題において高い分類精度を示している Convolutional neural network (CNN) [22] のように多層構造全体のパラメータを同時に最適化するのではなく、提案手法では最適化を 2 段階に分割することで解空間を小さくし最適化を容易にする。本論文では、データセット内の多様性が高いと考えられるカプセル内視鏡から撮影された小腸画像における異常画像と正常画像の分類問題に提案手法を適用し、性能の検証を行う*1。

2. 画像分類のための階層的な特徴構築

2.1 処理の流れ

提案手法の処理の概要を図 1 に示す。まず、特徴構築の第 1 段階では、既存の画像処理フィルタの組合せによって入力画像を変換し、変換後の画像に対してプーリング処理を行う。そして、プーリング処理後の各画素値を特徴量として分類器に入力し、分類を行う。このときの検証画像セットに対する分類精度が高くなるように、画像処理フィルタの組合せを GP によって構築する。提案手法における特徴構築の第 1 段階の処理の流れは次に示すとおりである。

- (1) 入力画像に対して既存の画像処理フィルタを用いて画像変換 (フィルタリング層)
- (2) プーリング処理 (プーリング層)
- (3) 分類 (分類層)

次に、特徴構築の第 2 段階では、第 1 段階で得られた変換画像をさらに GP で構築したフィルタ処理によって画像変換を行う。そして第 1 段階と同様に、変換後の画像に対してプーリング処理を行い、プーリング処理後の各画素値を特徴量として分類器に入力する。このときの検証画像セットに対する分類精度が高くなるように、フィルタを GP によって構築する。提案手法における特徴構築の第 2 段階の処理の流れは次に示すとおりである。

- (1) フィルタリング層で変換された画像に対して GP で構

*1 本論文では、著者らの国際会議論文 [23] を発展させ、文献 [23] で検証したものと異なるデータセットへの適用と考察、深層学習との比較実験などを行っている。

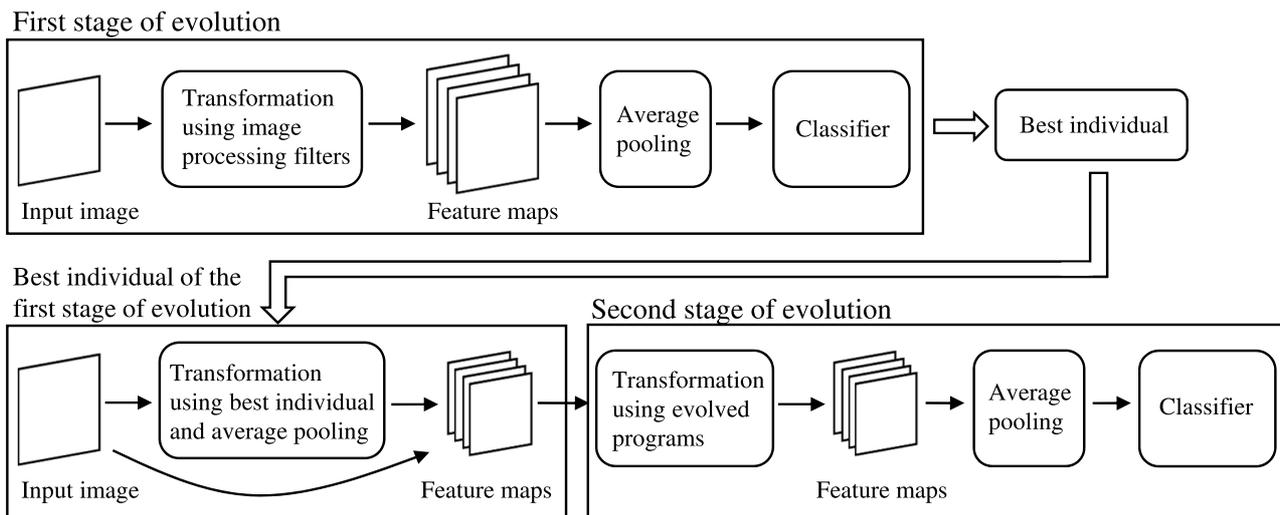


図 1 提案手法の概要

Fig. 1 Outline of the proposed method.

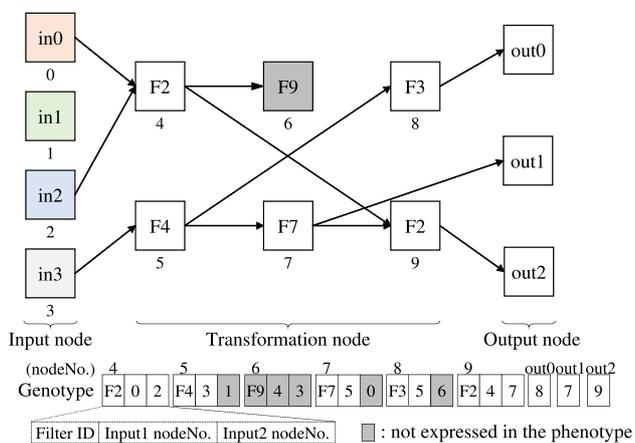


図 2 フィルタリング層の構造、表現型、遺伝子型の例

Fig. 2 Example of the filtering layer and its phenotype and genotype.

築したフィルタを用いて画像変換（変換層）

(2) プーリング処理（プーリング層）

(3) 分類（分類層）

2.2 フィルタリング層 (F)

フィルタリング層では、入力画像に対して既存の画像処理フィルタを適用することで、分類に有効な特徴構築を行う。図 2 にフィルタリング層の構造とその表現型、遺伝子型の例を示す。フィルタリング層の構造は、Cartesian Genetic Programming (CGP) と同様のフィードフォワードのグラフ構造である。各ノードは、入力ノード、変換ノード、出力ノードの 3 種類に分けられ、構造内の各ノードは入力ノードもしくは自身より入力ノードに近い 1 つまたは 2 つのノードとの接続関係を有する。入力ノードは入力画像の赤成分画像、緑成分画像、青成分画像、グレースケール画像の 4 つで構成される。変換ノードでは表 1 で

表 1 フィルタリング層で使用する画像処理フィルタ

Table 1 Image processing filters used in the filtering layer.

Function	Description
Ave3, 5	Averaging filter with 3 × 3, 5 × 5 window
Max3, 5	Maximum filter with 3 × 3, 5 × 5 window
Min3, 5	Minimum filter with 3 × 3, 5 × 5 window
Sob3, 5	Sobel filter with 3 × 3, 5 × 5 window
Lap3, 5	Laplacian filter with 3 × 3, 5 × 5 window
Gau3, 5	Gaussian smooth filter with 3 × 3, 5 × 5 window
LoG3, 5	Laplacian of gaussian filter with 3 × 3, 5 × 5 window
Exp	Expansion processing
Con	Contraction processing
Gab0	Gabor filter with 7 × 7, 11 × 11 window with orientation of 0 degree
Gab45	Gabor filter with 7 × 7, 11 × 11 window with orientation of 45 degree
Gab90	Gabor filter with 7 × 7, 11 × 11 window with orientation of 90 degree
Gab135	Gabor filter with 7 × 7, 11 × 11 window with orientation of 135 degree
Add	Add input two images pixel by pixel
Sub	Subtract input two images pixel by pixel
Mul	Multiply input two images pixel by pixel
Div	Divide input two images pixel by pixel
Abs	Absolute subtraction of input two images pixel by pixel

定義された画像処理フィルタを用いて画像変換を行う。出力ノードは入力画像に対して変換ノードを用いて画像変換を行った画像を持つ。本論文では、この画像変換を行った出力画像を特徴マップと呼び、フィルタリング層では出力ノード数分の特徴マップを出力する。ノードの種類と接続関係は進化計算を用いて最適化を行う。

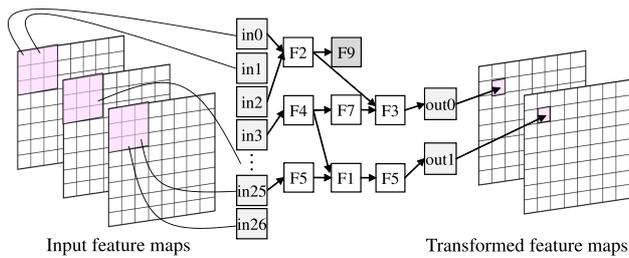


図 3 GP で構築されたフィルタを用いて 8×8 画素の 3 枚の特徴マップを 8×8 画素の 2 枚の特徴マップに変換する変換層の例。入力は注目画素の $3 \times 3 \times 3$ 近傍の画素値である

Fig. 3 Example of the transformation layer that transforms $8 \times 8 \times 3$ feature maps into $8 \times 8 \times 2$ feature maps using the evolved program. The inputs are the $3 \times 3 \times 3$ pixel intensity values surrounding a target pixel.

2.3 プーリング層 (P)

プーリング層ではフィルタリング層から出力された $w \times w \times n$ の特徴マップを入力とする。ここで、 $w \times w \times n$ の特徴マップという表現は $w \times w$ 画素の n 枚の特徴マップのことを表す。提案手法では、プーリング処理は $w \times w$ 画素の各特徴マップ上を縦横方向に p 画素ずつ $p \times p$ 画素のウィンドウをずらしながら行われる。このウィンドウの適用間隔をストライド (stride) と呼ぶ。結果として、 $w \times w \times n$ の特徴マップは $(w/p) \times (w/p) \times n$ の特徴マップに変換される。最大プーリングと平均プーリングをそれぞれ用いて 3 章の分類実験を行った結果、平均プーリングを用いた分類精度が優れていたため、本論文ではプーリング層におけるプーリング処理は平均プーリングを用いる。

2.4 変換層 (T)

変換層ではプーリング層から出力された n 枚の特徴マップと 4 枚の入力画像 (赤成分, 緑成分, 青成分, グレースケール) を $p \times p$ 画素のウィンドウでそれぞれ平均プーリングした画像を入力とする, すなわち, 変換層では $(w/p) \times (w/p) \times (n+4)$ の特徴マップを入力とする。入力画像はフィルタリング層における処理で分類に有効な情報が失われることを防ぐために用いる。変換層の構造はフィルタリング層と同様のフィードフォワードのグラフ構造であり, 入力特徴マップにおける注目画素の $s \times s \times (n+4)$ の局所領域の画素値を入力することで変換後の特徴マップの出力値が算出される。グラフ構造の出力ノード数が d の場合, $(w/p) \times (w/p) \times (n+4)$ の特徴マップは $(w/p) \times (w/p) \times d$ の特徴マップに変換される。図 3 に $8 \times 8 \times 3$ の特徴マップを $8 \times 8 \times 2$ の特徴マップに変換する例を示す。図 3 の例では, 入力は注目画素の $3 \times 3 \times 3$ 近傍の画素値である。また, 変換層の変換ノードで使用する関数は表 2 に示すとおりである。フィルタリング層と同様に, 図 2 のような構造中のノードの種類と接続関係を進化計算によって最適化する。

表 2 変換層で使用する関数セット

Table 2 Function set used in the transformation layer.

Function	# Inputs	Description
+	2	Add two inputs
-	2	Subtract two inputs
×	2	Multiply two inputs
÷	2	Divide two inputs
Max	4	The largest value of inputs
Min	4	The smallest value of inputs
Ave	4	The average value of inputs
log	1	Take the natural logarithm for an input
Sqrt	1	Extract a square root of an input
×2.0	1	Multiply an input by 2.0
×0.5	1	Multiply an input by 0.5
×0.1	1	Multiply an input by 0.1

2.5 分類層 (C)

分類層は $h \times h \times m$ の特徴マップを入力とする。 $h \times h \times m$ の特徴マップを $h \times h \times m$ 次元の特徴ベクトルに変換することで, 分類に用いる最終的な特徴量とする。本論文では, k 近傍法を分類器として使用し, 近傍数 k は $k \in \{3, 5, 7, 9, 11, 13\}$ の範囲で 3 章の分類実験を行った結果, 最も優れた精度を示した $k = 9$ に設定した。

2.6 学習手順

提案手法における特徴構築は 2 段階の最適化によって行われる。第 1 段階の最適化では, $I \rightarrow F \rightarrow P_1 \rightarrow C$ の構造でフィルタリング層 F の最適化を行う。 I は入力画像を表す。フィルタリング層は $w \times w \times n$ の特徴マップを生成し, 生成された特徴マップはプーリング層によって $(w/p_1) \times (w/p_1) \times n$ の特徴マップに変換される。 p_1 はプーリング層 P_1 におけるウィンドウサイズとストライドを表す。そして, $(w/p_1) \times (w/p_1) \times n$ の特徴マップを特徴ベクトルに変換し, k 近傍法に入力することで分類を行う。このときの検証画像セットに対する分類精度を各個体の適応度として用いる。第 1 段階の最適化によって, 最良の画像処理フィルタの組合せ F_{best} を得る。

第 2 段階の最適化では, $I \rightarrow F_{best} \rightarrow P_1 \rightarrow T \rightarrow P_2 \rightarrow C$ の構造を用いて変換層 T の最適化を行う。変換層 T では $(w/p_1) \times (w/p_1) \times (n+4)$ の特徴マップが入力され, $(w/p_1) \times (w/p_1) \times d$ の特徴マップが生成される。そして, 生成された特徴マップはプーリング層によって $(w/(p_1 p_2)) \times (w/(p_1 p_2)) \times d$ の特徴マップに変換され, k 近傍法に入力される。 p_2 はプーリング層 P_2 におけるウィンドウサイズとストライドを表す。第 2 段階目における最適化も第 1 段階目での最適化と同様に, 検証画像セットに対する分類精度を各個体の適応度とする。



図 4 カプセル内視鏡を用いて撮影された小腸画像の例

Fig. 4 Examples of small bowel images taken by a capsule endoscope.

3. 小腸画像の分類実験

3.1 実験に用いるデータセット

分類実験にはカプセル内視鏡から撮影された小腸画像を使用した。個人情報保護の観点から本論文には分類実験に使用した小腸画像を掲載することができないため、小腸画像のイメージを図 4 に示す*2。分類実験では小腸画像 220 枚の中央部 256 × 256 画素から 64 × 64 画素のサイズで切り出したブロック画像単位での異常か正常かの 2 クラス分類を行った。切り出す前の小腸画像には専門家によって 8 × 8 画素のブロックごとに異常、正常のフラグが付与されているため、64 × 64 画素のサイズで切り出したブロック画像内に占める異常フラグの割合を用いて異常ブロック画像、正常ブロック画像を作成した。本論文では、異常フラグが占める割合が 0.1 以上のブロック画像を異常ブロック画像、それ以外のブロック画像を正常ブロック画像と定義した。64 × 64 画素のサイズで切り出したブロック画像に対して、左右反転、90° ずつ回転させることで取得画像枚数を 8 倍に増やし、作成した各クラスからランダムに 1,000 枚ずつのブロック画像を選び、学習画像セット、検証画像セット、テスト画像セットを作成した。本論文で対象とした異常例は、粘膜性病変（白色を呈する）および腫瘍性病変（隆起形状を呈する）である。

3.2 実験設定

提案手法で用いた GP のパラメータを表 3 に示す。遺伝操作として遺伝子型に対する一様交叉と突然変異を用いる。本論文では処理の高速化のために GPU によって処理を行う。GPU のプログラムを記述する言語として、compute unified device architecture (CUDA)*3を用いる。GPU プログラム上でフィルタリングやプーリング、 k 近傍法 [25] に対する処理を記述することで、すべての画像に対してそれらの処理を並列に実行する。

提案手法の有効性を検証するため、次に示す 7 つの手法

*2 <http://olympusmedical.com.sg/products/all-products/endoscopes/capsule-endoscopes/endocapsule-10-system/index.html>, ENDOCAPSULE

*3 <https://developer.nvidia.com/cuda-zone>, CUDA Zone

表 3 提案手法で用いた GP のパラメータ

Table 3 Parameter settings for GP of the proposed method.

Parameter	Value
Number of generations (first evolution)	5000
Number of generations (second evolution)	5000
Generation alternation model	MGG*
Population size	50
Children size	10
Ratio for uniform crossover	0.8
Mutation rate	0.1

*Minimal Generation Gap [24]

による小腸画像の分類実験を行った。

- (1) $I \rightarrow F \rightarrow P_1 \rightarrow T \rightarrow P_2 \rightarrow C$ の構造を用いた提案手法 (Proposal 1)。フィルタリング層 F における最大変換ノード数は 40、出力ノード数は 8、変換層 T における入力ノード数は 108 ($= 3 \times 3 \times 12$)、最大変換ノード数は 60、出力ノード数は 16、プーリング層 P_1, P_2 におけるウィンドウサイズおよびストライドはそれぞれ $p_1 = 2, p_2 = 2$ に設定した。結果として、 k 近傍法で用いる特徴ベクトルの次元数は 4,096 次元である。
- (2) 全体の構造は Proposal 1 と同様であるが、プーリング層 P_1, P_2 におけるウィンドウサイズおよびストライドをそれぞれ $p_1 = 2, p_2 = 32$ と設定した構造 (Proposal 2)。結果として、 k 近傍法で用いる特徴ベクトルの次元数は 16 次元である。
- (3) Agapitos らが提案している構造を用いた分類 [21]。ここでは、 $I \rightarrow F_{\text{random}} \rightarrow P_1 \rightarrow T \rightarrow P_2 \rightarrow C$ の構造を用いる。 F_{random} はランダムフィルタリング層を表し、50 個のランダムフィルタによって構成される。各ランダムフィルタは 5×5 画素の受容野を持ち、5 つの分布 $U(-1.0, 1.0), U(-5.0, 5.0), UD(1, 5), N(1.0), N(5.0)$ からそれぞれ 10 個のランダムフィルタを生成した。 $U(a, b)$ は $[a, b]$ の範囲の実数値の一様乱数、 $UD(a, b)$ は $[a, b]$ の範囲の整数値の一様乱数、 $N(a)$ は平均 0、標準偏差 a の正規乱数を表す。結果としてランダムフィルタリング層 F_{random} は 50 枚の特徴マップを出力する。変換層 T における入力ノード数は 450 ($= 3 \times 3 \times 50$)、最大変換ノード数は 200、出力ノード数は 16 に設定した。変換層 T の構造は提案手法と同様のフィードフォワードのグラフ構造を用いた。プーリング層 P_1, P_2 におけるウィンドウサイズおよびストライドはそれぞれ $p_1 = 2, p_2 = 2$ に設定した。分類器は提案手法と同様に $k = 9$ の k 近傍法を用いた。
- (4) Convolutional neural network (CNN) [22] を用いた分類。CNN は深層学習の一手法であり、画像分類問題において高い分類性能を示している [26]。本論文で用いた CNN の構造を表 4 に示す。畳込み層 (convolu-

表 4 CNN の構造
Table 4 Structure of CNN.

# layer	type	window size	stride	output size
0	data			64 × 64 × 3
1	convolution	5 × 5	1	64 × 64 × 8
2	max pool	2 × 2	2	32 × 32 × 8
3	convolution	3 × 3	1	32 × 32 × 32
4	max pool	2 × 2	2	16 × 16 × 32
5	convolution	3 × 3	1	16 × 16 × 64
6	max pool	2 × 2	2	8 × 8 × 64
7	fully connected			1 × 1 × 256
8	softmax			1 × 1 × 2

tion)における活性化関数はReLUを用いた。epoch数は100に設定し、バッチサイズが100のミニバッチ学習を行う。学習誤差の算出には交差エントロピー誤差関数を使用し、関数の最小化にはAdam法[27]を用いた。Adam法のパラメータとして、文献[27]で用いられている $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$ を用いた。また、過学習を防ぐためにdropout[28]を全結合層に対して適用した。ユニットの選出確率 p は文献[26]で用いられている $p = 0.5$ を使用した。

- (5) Fine-tuningを行ったAlexNet[26]を用いた分類。文献[26]で提案された学習済みのCNN(AlexNet)*4のパラメータを初期値として用い、小腸画像を使用して誤差逆伝播法による学習を進める。このとき、文献[26]で学習に用いている学習画像と本論文で扱う小腸画像の画像サイズが異なることと分類するクラス数が異なることから、本論文では小腸画像を63 × 63画素にリサイズし、さらにAlexNetのプーリング層と最後の識別層の変更を行ってから小腸画像を用いた学習を行った。具体的には、1番目のプーリング処理を削除し、2番目のプーリング処理におけるウィンドウサイズを2、ストライドを1にそれぞれ変更した。また、最後の識別層の出力ユニット数を小腸画像の分類クラス数である2に変更した。epoch数は100に設定し、バッチサイズが100のミニバッチ学習を行う。学習誤差関数の最小化にはAdam法[27]を用い、Adam法のパラメータは、 $\alpha = 0.0001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$ を用いた。
- (6) 統計特徴量とuniform local binary pattern(ULBP)特徴[29]を用いた分類。ここでは画像全体の画素値の平均値、最大値、最小値、中央値、第1四分位数、第3四分位数、標準偏差、最頻値、第1 σ 内確率、第3 σ 外確率、レンジ(最大画素値と最小画素値の差分値)の11種類の統計量と10次元のULBP特徴をRGBカ

表 5 テスト画像セットに対する各手法の分類精度

Table 5 Accuracy results of each method on test images.

	Best	Average
Proposal 1	0.816	0.807
Proposal 2	0.837	0.831
Agapitos [21]	0.658	0.632
CNN	0.788	-
AlexNet+fine-tuning	0.809	-
統計特徴量+ULBP	0.759	-
原画像+プーリング	0.661	-

ラー画像の各成分からそれぞれ算出した計63次元の特徴量を k 近傍法に入力することで分類を行う。カプセル内視鏡は消化器官内を回転しながら画像を撮影するため、撮影される小腸画像は向きが一意に定まらないことから方向依存性のない特徴量を使用する必要がある。そこで本論文では、画像全体からの統計特徴量とLBP特徴を回転不変に拡張したULBP特徴を用いる。

- (7) 原画像を用いた分類。原画像(赤成分画像、緑成分画像、青成分画像、グレースケール画像)それぞれに対してウィンドウサイズとストライドが4の平均プーリングを行い、処理後の各画素値を特徴ベクトルとして k 近傍法によって分類を行う。 k 近傍法で用いる特徴ベクトルの次元数は1,024次元である。

3.3 実験結果

3.3.1 従来研究との比較

提案手法と従来手法のテスト画像セットに対する分類結果を表5に示す。提案手法およびAgapitosらの手法は5試行中の平均分類精度と最良の分類精度結果を示している。CNNおよびfine-tuningを行ったAlexNetは最大epoch数100のうち、テスト画像セットに対して最も高い分類精度を示した結果を示している。

表5の結果から、Proposal 2が平均分類精度、最良分類精度ともに最も優れた結果を示していることが分かる。Proposal 1も従来手法と比べると良好な結果を示している。Proposal 2では変換層 T の各出力特徴マップに対して特徴マップのサイズと同じウィンドウサイズで平均プーリングを行い、特徴マップ全体の平均を算出することで小腸画像分類に有効な回転不変性を有する特徴量を構築することができたため、Proposal 1に比べて高い分類精度を示していると考えられる。

Agapitosらの手法は提案手法と比べると分類精度が大きく低下してしまっている。これはフィルタリング層 F 以外にはProposal 1と同一の構造であることから、ランダムフィルタによる変換が小腸画像分類には有効に働いていないことが原因だと考えられる。小腸画像では照明条件の変化が

*4 http://caffe.berkeleyvision.org/model_zoo.html, Caffe Model Zoo

生じたり，異常部位や正常部位の種類が多様であったりすることから，非常に多様性が高い画像となっている．そのため，ランダムフィルタのみではこのような多様性が高い小腸画像から分類に有効な特徴量を構築することは困難であると考えられる．一方，提案手法におけるフィルタリング層 F では既存の画像処理フィルタを多段に適用することで，単一のランダムフィルタでは構築することが難しい特徴量を構築することを可能とし，分類精度で優れた結果を示したと考えられる．

また，提案手法は CNN と比べても優れた性能を示した．これは CNN のような深層学習の手法では一般的に大規模な学習画像を必要とするため，各クラス 2,000 枚ずつの計 4,000 枚の学習画像では十分な特徴構築を行うことができなかったと考えられる．このように学習データが少ない場合は，ほかの大規模データセットを用いた学習済みの CNN のパラメータを初期値として用いて fine-tuning を行う方法が有効であると考えられ，一般に fine-tuning による転移学習を行うことでフルスクラッチから CNN を学習するよりも良い結果を得られる場合が多い．しかし，提案手法は fine-tuning を行った AlexNet よりも優れた分類精度を示した．これは fine-tuning を行った AlexNet がフルスクラッチから学習を行った CNN の分類精度と比べて精度の向上が低いことから，本論文で対象とした小腸画像が AlexNet の事前学習に用いた ImageNet の網羅する画像との関連が低かったためであると考えられる．本論文で扱っている小腸画像のように大規模な学習画像セットの用意が難しく，かつ一般物体と性質の異なる問題では，提案手法のように少ない学習画像セットから分類に有効な特徴構築が行える手法は有望であると考えられる．

さらに，方向依存性のない特徴量である統計特徴量と ULBP 特徴を用いた分類よりも提案手法は優れた性能を示した．これは画像処理フィルタおよび GP フィルタによる画像変換とプーリング処理によって，方向依存性のない特徴量を自動で構築できたためであると考えられる．また，原画像を用いた分類方法よりも大きく分類精度で優れていることから，提案手法は原画像の情報のみから分類に有効な特徴量を構築できていることが分かる．以上のことから，提案手法は対象となる画像に応じて適切に特徴量を自動構築できることが確認できた．

表 6 に提案手法と深層学習手法の計算コストを示す．表 6 中の Training は学習に要した時間，Applying はテスト画像 2,000 枚の分類に要した時間を示している．提案手法は CNN や AlexNet と比べて学習の計算コストが高くなってしまっていることが分かる．そのため，今後はより計算コストの低い学習方法を検討する必要がある．表 6 の提案手法と CNN, AlexNet の処理時間はいずれも GPU 実装によって処理を行った結果を示しており，実験はすべて Intel core i7-5960X 3GHz CPU, 32GB RAM, NVIDIA

表 6 提案手法と深層学習手法の計算コスト

Table 6 Time costs of the proposed method and deep learning methods.

	Training	Applying
Proposal 1	4.0×10^4 [sec]	6.0×10^{-1} [sec]
Proposal 2	3.1×10^4 [sec]	4.5×10^{-1} [sec]
CNN	7.7×10^1 [sec]	1.2×10^{-1} [sec]
AlexNet+fine-tuning	5.3×10^2 [sec]	8.9×10^{-1} [sec]

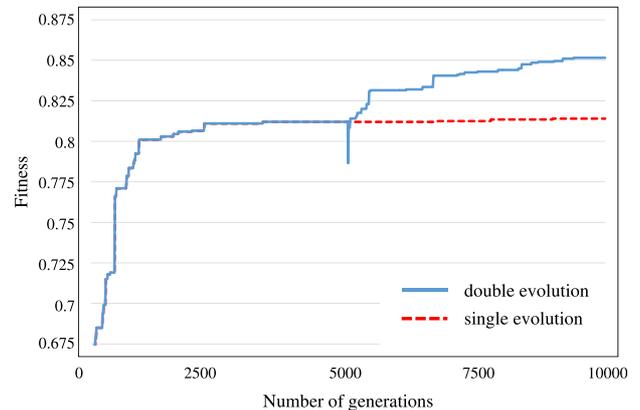


図 5 提案手法の適応度推移

Fig. 5 Transition of the fitness of the proposed method.

GeForce GTX TITAN X 上で行った．

3.3.2 提案手法に関する考察

2 種類の変換処理を行う有効性の検証

5 試行中における Proposal 2 の最良個体の適応度推移の結果を図 5 に示す．図 5 には， $I \rightarrow F \rightarrow P_1 \rightarrow C$ の構造で 10,000 世代の最適化を行った結果 (single evolution) と，先に述べた構造で 5,000 世代の時点での最良個体を固定し $I \rightarrow F_{\text{best}} \rightarrow P_1 \rightarrow T \rightarrow P_2 \rightarrow C$ の構造に変更してさらに 5,000 世代の最適化を行った結果 (double evolution) を示している． F_{best} は 5,000 世代の時点での single evolution の最良個体を表す．5,000 世代までは同一の個体を表しているため，適応度推移は一致している．図 5 で示されているように，既存の画像処理フィルタのみを用いた構造である $I \rightarrow F \rightarrow P_1 \rightarrow C$ よりも 2 種類の画像変換を用いた構造である $I \rightarrow F_{\text{best}} \rightarrow P_1 \rightarrow T \rightarrow P_2 \rightarrow C$ の方が高い適応度を示していることが分かる．特に，既存の画像処理フィルタのみを用いた構造では 5,000 世代を過ぎると進化が停滞してしまっているが，2 種類の画像変換を用いた構造では 5,000 世代を過ぎても適応度が上昇していることが分かる．これは，変換層 T では GP を用いて新たなフィルタを構築しているため，フィルタリング層 F で使用している既存のエッジ検出や平滑化処理などを行う画像処理フィルタだけでは表現できないフィルタを 2 段階目の最適化で構築できているためであると考えられる．図 6 に変換層 T で構築された GP フィルタの例を示す．図 6 から分かるように複数の変換画像と原画像の画素値を用いたフィルタが

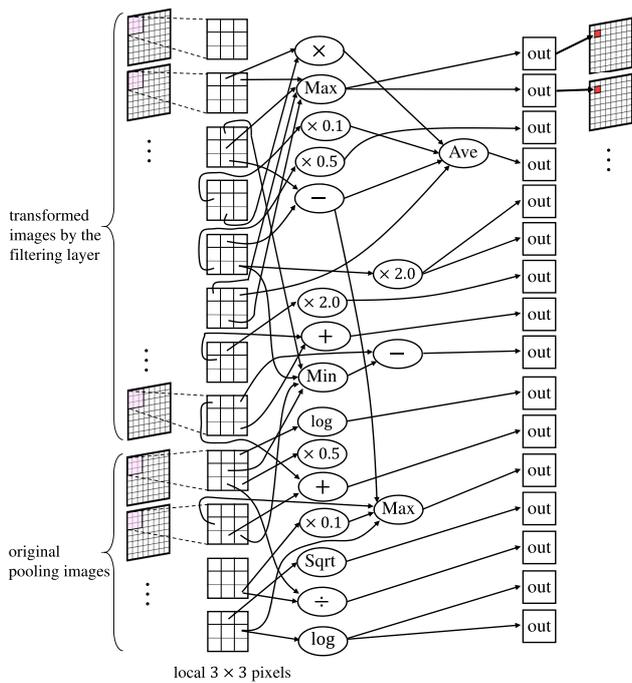


図 6 変換層 T で構築された GP フィルタの例

Fig. 6 Example of the GP filter constructed by the transformation layer T .

構築されており、既存の画像処理フィルタでは表現できない処理を実現している。このようなフィルタを人手で設計することは困難であることや図 5 で示したように変換層 T を用いることで分類精度が向上していることから、2 段階目の最適化で変換層 T を用いることに有効性があると考えられる。

変換層 T で出力する特徴マップ数に対する分類精度の検証

Proposal 2 における変換層 T で出力する特徴マップ数を変えた場合のテスト画像セットに対する分類精度を図 7 に示す。なお、図 7 は 10 試行を行った結果を示しており、output X は変換層 T で出力する特徴マップ数 X を表している。図 7 から、変換層 T で出力する特徴マップ数が 16 枚の場合に最も良好な分類精度を示しており、また特徴マップ数が 4, 8, 24, 32 枚の場合も同様に良好な結果を示していることが分かる。しかし、特徴マップ数が 2 枚の場合は、分類精度が大きく低下してしまっている。これは、特徴マップ数を極端に減らしたことで特徴量の表現力が低下してしまったためであると考えられる。したがって、単に特徴マップ数を減らしたりすればよいのではなく、適切な特徴マップ数を与えればより安定して高い性能を示すことができる可能性がある。

最適化方法に対する分類精度の検証

最後に、Proposal 1 の最適化方法を変更した場合のテスト画像セットに対する 5 試行中の平均分類精度と最良の分類精度を表 7 に示す。Proposal 1 (一括最適化) は $I \rightarrow F \rightarrow P_1 \rightarrow T \rightarrow P_2 \rightarrow C$ のフィルタリング層 F と変換層 T を同時に最適化する方法である。このとき最適

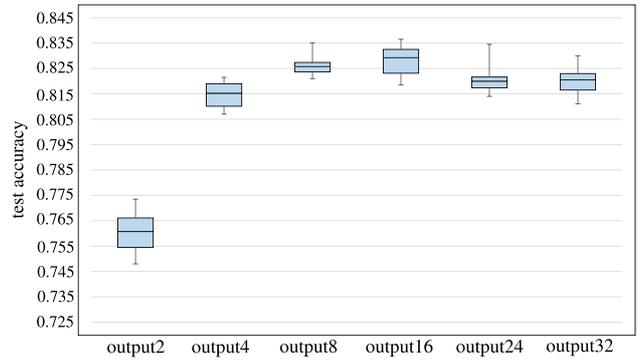


図 7 変換層 T で出力する特徴マップ数に対する分類精度の変化
Fig. 7 Test accuracy of the proposed method with changing the number of feature maps in the transformation layer T .

表 7 最適化方法に対する分類精度の変化

Table 7 Test accuracy of the proposed method with several optimization methods.

	Best	Average
Proposal 1	0.8160	0.8072
Proposal 1 (一括最適化)	0.8010	0.7974
Proposal 1 (相互最適化)	0.8010	0.7969

化の世代数は 10,000 とし、そのほかのパラメータは Proposal 1 と同様である。Proposal 1 (相互最適化) はまず、 $I \rightarrow F_1 \rightarrow P_1 \rightarrow C$ でフィルタリング層 F_1 を最適化し、次に $I \rightarrow F_1 \rightarrow P_1 \rightarrow T_1 \rightarrow P_2 \rightarrow C$ で変換層 T_1 を最適化し、さらに $I \rightarrow F_2 \rightarrow P_1 \rightarrow T_1 \rightarrow P_2 \rightarrow C$ でフィルタリング層 F_2 を最適化し、最後に $I \rightarrow F_2 \rightarrow P_1 \rightarrow T_2 \rightarrow P_2 \rightarrow C$ で変換層 T_2 を最適化する方法である。テスト画像に適用する際は、 $I \rightarrow F_2 \rightarrow P_1 \rightarrow T_2 \rightarrow P_2 \rightarrow C$ の構造を用いる。各段階の最適化の世代数は 2,500 とし、そのほかのパラメータは Proposal 1 と同様である。また、各段階の最適化を切り替えるときに最適化対象の個体集団はランダムに初期化を行っている。

表 7 の結果から Proposal 1 が最も優れた分類精度を示しており、Proposal 1 (一括最適化) と Proposal 1 (相互最適化) はほぼ同等の結果を示した。Proposal 1 (一括最適化) はフィルタリング層 F と変換層 T を同時に最適化するため、解空間が大きくなってしまい最適化が適切に行えなかったと考えられる。Proposal 1 (相互最適化) は、最適化を切り替えるときにフィルタリング層 F もしくは変換層 T のどちらかの構造が大きく変更されるため、最適化が十分に行えなかったと考えられる。最適化を切り替えるときに、前段階での最適化に用いた個体集団を複製してから次段階の最適化を開始するなどの方法で効率的な最適化を行う必要があると考えられる。

4. まとめ

本論文では、画像分類のための階層的な特徴構築手法を提案した。提案手法は (1) 既存の画像処理フィルタの組合せによる画像変換, (2) GP で構築したフィルタによる画像変換, の 2 種類の処理を 2 段階の進化計算によって最適化する。提案手法をカプセル内視鏡から撮影された小腸画像における異常画像と正常画像の分類問題に適用し, 従来手法との比較と性能の検証を行った。実験結果から, 従来手法よりも優れた分類精度を示すことが確認できた。また, 2 種類の処理を 2 段階の進化計算によって最適化することで 1 種類の処理を 1 段階の進化計算によって最適化するよりも優れた性能を示したことから, 異なる処理の構造を多層化することの有用性を示した。

今後の課題として, フィルタリング層や変換層の構造に関するパラメータも同時に最適化することで, より汎用性の高い特徴構築手法の確立があげられる。また, 多層構造の一括最適化や相互最適化などの最適化方法についてさらに検討を行うことで, より計算コストの低く分類に有効な特徴構築を行う必要がある。さらに, 提案手法の拡張として様々な画像データセットに対する分類実験や, 手法の信頼性と精度向上のために, 処理内容の解析を行う必要があると考えている。

謝辞 本研究に際して, オリパス株式会社より画像を提供していただきました。

参考文献

- [1] Chapelle, O., Haffner, P. and Vapnik, V.N.: Support vector machines for histogram-based image classification, *IEEE Trans. Neural Networks*, Vol.10, No.5, pp.1055–1064 (1999).
- [2] Nowak, E., Jurie, F. and Triggs, B.: Sampling strategies for bag-of-features image classification, *Proc. European Conference on Computer Vision*, pp.490–503 (2006).
- [3] Lowe, D.G.: Object recognition from local scale-invariant features, *Proc. IEEE International Conference on Computer Vision*, Vol.2, pp.1150–1157 (1999).
- [4] Uijlings, J.R., van de Sande, K.E., Gevers, T. and Smeulders, A.W.: Selective search for object recognition, *International journal of computer vision*, Vol.104, No.2, pp.154–171 (2013).
- [5] Ahonen, T., Hadid, A. and Pietikäinen, M.: Face recognition with local binary patterns, *Proc. European Conference on Computer Vision*, pp.469–481 (2004).
- [6] Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection, *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, Vol.1, pp.886–893 (2005).
- [7] Lowe, D.G.: Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, Vol.60, No.2, pp.91–110 (2004).
- [8] Xiao, J., Hays, J., Ehinger, K.A., Oliva, A. and Torralba, A.: Sun database: Large-scale scene recognition from abbey to zoo, *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp.3485–3492 (2010).
- [9] Bengio, Y., Courville, A. and Vincent, P.: Representation learning: A review and new perspectives, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.35, No.8, pp.1798–1828 (2013).
- [10] Hinton, G.E., Osindero, S. and Teh, Y.-W.: A fast learning algorithm for deep belief nets, *Neural computation*, Vol.18, No.7, pp.1527–1554 (2006).
- [11] Larochelle, H., Bengio, Y., Louradour, J. and Lamblin, P.: Exploring strategies for training deep neural networks, *The Journal of Machine Learning Research*, Vol.10, pp.1–40 (2009).
- [12] Koza, J.R. and Poli, R.: Genetic programming, *Search Methodologies*, pp.127–164, Springer-Verlag (2005).
- [13] Zhang, M. and Smart, W.: Multiclass object classification using genetic programming, *Applications of Evolutionary Computing*, Vol.3005 of LNCS, pp.369–378, Springer-Verlag (2004).
- [14] Zhang, M., Gao, X. and Lou, W.: A new crossover operator in genetic programming for object classification, *IEEE Trans. Systems, Man, and Cybernetics, Part B*, Vol.37, No.5, pp.1332–1343 (2007).
- [15] Al-Sahaf, H., Song, A., Neshatian, K. and Zhang, M.: Two-Tier genetic programming: towards raw pixel-based image classification, *Expert Systems with Applications*, Vol.39, No.16, pp.12291–12301, Elsevier (2012).
- [16] Miller, J.F. and Thomson, P.: Cartesian Genetic Programming, *Proc. 3rd European Conference on Genetic Programming*, Poli, R., Banzhaf, W., Langdon, W.B., Miller, J., Nordin, P. and Fogarty, T.C. (Eds.), Vol.1802 of LNCS, pp.121–132, Springer-Verlag (2000).
- [17] Harding, S.: Evolution of image filters on graphics processor units using cartesian genetic programming, *IEEE Congress on Evolutionary Computation*, pp.1921–1928 (2008).
- [18] Kowaliw, T., Banzhaf, W., Kharna, N. and Harding, S.: Evolving novel image features using genetic programming-based image transforms, *IEEE Congress on Evolutionary Computation*, pp.2502–2507 (2009).
- [19] Shirakawa, S., Nakayama, S. and Nagao, T.: Genetic image network for image classification, *Applications of Evolutionary Computing*, Vol.5484 of LNCS, pp.395–404, Springer-Verlag (2009).
- [20] Hirano, Y. and Nagao, T.: Feature transformation using filter array for automatic construction of image classification, *IEEE 7th International Workshop on Computational Intelligence and Applications*, pp.59–64 (2014).
- [21] Agapitos, A., O’Neill, M., Nicolau, M., Fagan, D., Kattan, A., Brabazon, A. and Curran, K.: Deep evolution of image representations for handwritten digit recognition, *IEEE Congress on Evolutionary Computation*, pp.2452–2459 (2015).
- [22] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-based learning applied to document recognition, *Proc. IEEE*, Vol.86, No.11, pp.2278–2324 (1998).
- [23] Suganuma, M., Tsuchiya, D., Shirakawa, S. and Nagao, T.: Hierarchical feature construction for image classification using genetic programming, *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, pp.1423–1428 (2016).
- [24] 佐藤 浩, 小野 功, 小林重信: 遺伝的アルゴリズムにおける世代交代モデルの提案と評価, *人工知能学会誌*, Vol.12, No.5, pp.734–744 (1997).
- [25] Garcia, V., Debreuve, E., Nielsen, F. and Barlaud, M.: K-nearest neighbor search: Fast GPU-based imple-

mentations and application to high-dimensional feature matching, *Proc. IEEE International Conference on Image Processing*, pp.3757-3760 (2010).

- [26] Krizhevsky, A., Sutskever, I. and Hinton, G.E.: ImageNet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp.1097-1105 (2012).
- [27] Diederik, K. and Jimmy, B.: Adam: A method for stochastic optimization, *International Conference on Learning Representations*, pp.2452-2459 arXiv: 1412.6980 (2015).
- [28] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research*, Vol.15, pp.1929-1958 (2014).
- [29] Ojala, T., Pietikäinen, M. and Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.7, pp.971-987 (2002).



長尾 智晴 (正会員)

1985年東京工業大学大学院総合理工学研究科博士課程後期中退。同年同大学助手。同大学助教授を経て、2001年横浜国立大学大学院環境情報研究院教授。工学博士。画像処理、進化計算法等の知能情報学の研究に従事。電子情報通信学会、人工知能学会、電気学会、進化計算学会、IEEE等各会員。



菅沼 雅徳 (学生会員)

2013年横浜国立大学工学部電子情報工学科を飛び級のため中退。現在、同大学大学院環境情報学府情報メディア環境学専攻博士課程後期在学中。画像処理、パターン認識を含む知能情報学の研究に従事。人工知能学会、日本医

用画像工学会各会員。



土屋 大樹

2014年横浜国立大学工学部電子情報工学科卒業。2016年同大学大学院環境情報学府情報メディア環境学専攻博士課程前期修了。進化計算の研究に従事。



白川 真一 (正会員)

2009年横浜国立大学大学院環境情報学府博士課程後期修了。2008年日本学術振興会特別研究員。2010年株式会社富士通研究所研究員。2012年青山学院大学理工学部情報テクノロジー学科助手、2013年助教。2015年筑波大学システム情報系助教。2016年より横浜国立大学大学院環境情報研究院講師。進化計算、機械学習、画像処理等の研究に従事。博士(工学)。