

注視行動によるアノテーションに基づく 曖昧さを反映した画像認識

石橋 達矢^{1,a)} 鮫島 正樹¹ 菅野 裕介¹ 松下 康之¹

概要： 深層学習により画像認識の性能が飛躍的に向上しつつある現在においても、アルゴリズムと人間の認識プロセスは大きく異なっており、人間には容易に認識できるが機械には難しい事例が存在する。その要因の一つとして、多くの場合クラスやカテゴリの定義は一意に定まるものではなく、本質的な曖昧さが存在していることが挙げられる。しかし、こうした曖昧さを反映した真値アノテーションを個別に行うのは困難であるという課題がある。本論文では、視線情報を用いた画像アノテーション手法とそれに基づく画像認識手法を提案する。提案手法では、画像分類タスクに取り組む人間の視線情報から画像の知覚の難易度を推定し、これを画像識別時の重みとして用いることで難易度を反映した画像分類を行う。評価実験では、視覚的に類似したデータセットを用いた画像分類を行い、提案手法の有効性を示す。

TATSUYA ISHIBASHI^{1,a)} MASAKI SAMEJIMA¹ YUSUKE SUGANO¹ YASUYUKI MATSUSHITA¹

1. 序論

近年、ディープニューラルネットワーク (Deep Neural Network: DNN) によって画像認識の分野は急速に発展し、クラス数が少なく学習用データが十分にあるようなタスクにおいては、機械による画像認識は人間の認識能力と同等の性能を有すようになった [9], [10]. しかし、データセットが限られている場合や複雑なタスクの場合には人間の認識能力には及ばない [1], [2]. また、現在の機械学習アルゴリズムの誤りのパターンは同じタスクに取り組む人間の誤りのパターンとは大きく異なっている [12]. そのため、人間には認識できるにもかかわらず、機械には認識できないような画像が存在している。

そのような画像を正しく認識できるようにするために、機械学習のアルゴリズムの中に人間が介入し、機械と人間が相互作用的にデータのやり取りやモデルの較正を行うことで機械学習の性能を向上させる研究が数多く行われている。人間の役割は主にデータにアノテーションを行うことで学習データの質を高めたり、機械の学習の様子や出力を確認しモデルの調整をしたりすることである。例えば、一般の画像認識タスクでは画像には単一のクラスラベルのみが与えられるが、人間によってそれ以外のラベルを与える

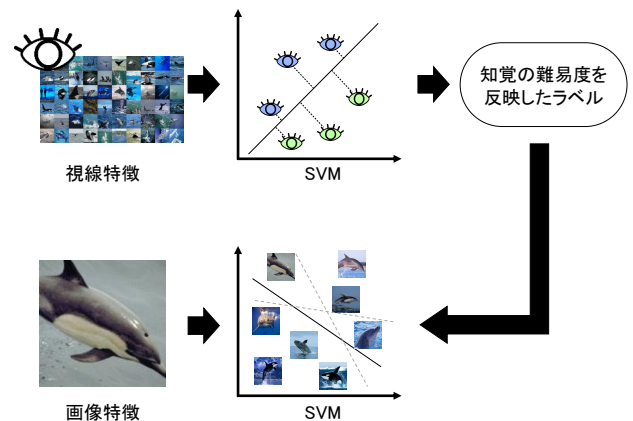


図 1: 本研究の概略。視線データを用いた分類器の学習と画像特徴を用いた分類器の学習は独立して行う。視線データによる分類器の学習によって得られた知覚の難易度を反映したラベルを用いることで、画像特徴による分類器の学習の際に知覚が容易な画像に大きな損失を与える。

ことで学習データの情報を増やすことができる。実際に人間の生体情報から得たデータを連続値のラベルに変換し、画像分類の際に使用する研究が行われている [7], [13]. これらの研究では画像を見ているときの人間の生体情報から画像の知覚の難易度を大まかに反映したラベルを生成し、画像分類の際にそのラベルを利用することで、画像以外の

¹ 大阪大学大学院情報科学研究科

^{a)} ishibashi.tatsuya@ist.osaka-u.ac.jp

データから重み付けを行う。このような研究では人間の生体情報として一般に脳活動が用いられるが、人間の脳活動を計測するためには大きなコストが必要になるという問題がある。

低コストで計測できる人間の生体情報として視線情報がある。行動認識の分野の研究では、人間の内部状態を知る手段として注視行動がしばしば用いられる。さらに、過去の研究において人間の意思決定と注視行動には関連があることが分かっている [14], [16]。また、視線推定についての研究も盛んに行われており、視線計測器を用いなくてもカメラの画像ベースで推定することが可能になりつつある [17]。しかし、注視行動を用いた画像分類の研究は未発展である。

本研究では、人間の視線データを直接画像分類に用いるのではなく、視線データを用いて画像ごとの知覚の難易度を推定する。その後、画像分類器の学習の際に知覚の難易度の情報を用いることで重み付けを行う。図 1 に本研究の概略を示す。

提案手法ではまず、画像分類タスクに取り組む人間の視線・マウスのデータを計測する。画像の分類タスクではマウスを操作し画像上でクリックするので、本研究では視線のデータとともにマウスのデータも使用する。次に、その視線・マウスのデータと画像のクラスラベルを用いて分類器の学習を行う。その分類器の決定境界と各サンプルとの差から知覚の難易度に応じた連続的なラベルを生成し、その情報を用いて画像特徴による分類の際に知覚が容易なものに大きな損失を与えるように損失の大きさを調整する。視線情報を用いるのは学習時のみであり、一度学習が終わるとその後人間の介入を一切必要としない。そのためラベルの生成に必要なコストを抑えつつ効果的な人間の生体情報を利用することができる。

本論文の構成は以下の通りである。2 節で人間の生体情報を用いた画像分類と視線を用いた研究を紹介し、3 節で提案手法について述べる。4 節で評価結果を示し、最後に 5 節で本研究のまとめを行う。

2. 関連研究

本研究では、画像分類タスクに取り組む人間の注視行動を用いて画像の知覚の難易度を推定し、その難易度を反映する画像分類を行う。本節では、2.1 節で人間の生体情報を用いた画像分類に関する研究を紹介し、2.2 節で注視行動を用いた人間の内部情報の理解に関する研究を紹介する。

2.1 人間の生体情報を用いた画像分類

画像を見ているときの人間の生体情報から知覚の難易度を推定し、その難易度を反映したラベルを画像分類モデルに組み込むことで、画像認識の精度を向上させる研究について紹介する。Scheirer ら [13] は、人間が画像を認識する

までの時間や人間の認識の精度は画像毎の知覚の難易度を反映すると考え、被験者に数枚の画像の中から顔を含む画像を選ぶタスクを与え、その時の正答率と解答速度の情報をラベルとして用いた。Fong ら [7] は、画像を見ているときの脳活動量の大きさから画像の知覚の難易度を推定できると考え、被験者が画像を見た時の脳活動量を MRI を用いて測定した。その脳活動量を用いて画像のクラスを推定する分類器を学習させ、それぞれのサンプルと決定境界との差を連続的なラベルに変換した。得られたラベルを利用して、画像特徴を用いて分類器を学習する際に画像の知覚の難易度に応じて損失の大きさを調整することで決定境界を校正した。Spampinato ら [15] は、画像を見ている被験者の EEG 信号を計測し、Long short-term memory を用いることで多次元的かつ時間的に変化する EEG 信号から特徴を抽出し画像分類器を学習させた。さらに画像の畳み込みニューラルネットワーク (Convolutional Neural Network: CNN) 特徴を EEG 特徴にマッピングし、EEG 特徴で学習させた分類器を使って画像からクラスを推定する手法を提案した。これらの研究は学習時にのみ生体情報を必要とし、一度分類器を学習させるとそれ以上のデータは不要である。しかし、これらのように脳活動を測定して利用するには大きなコストが必要となる。

2.2 注視行動を用いた人間の内部状態の理解

人間の注視行動は人間の内部状態を知る手段として研究されている。Sugano ら [16] は、2 枚の画像を見ている人間の視線のデータを視線計測器を用いて計測し、凝視の回数や合計時間などの特徴を抽出し、その特徴からどちらの画像を好むかを推定した。この研究の中で、2 枚の画像から好みの方を選択するタスクを与えた場合に、タスクを与えなかった場合より高い精度で好みの推定が可能であることと、注視行動の特徴の内、凝視の回数と合計時間が重要であることが示されている。

また、注視行動の情報を用いた行動認識に関する研究も盛んに行われている。Fathi ら [6] は、人間の手や視線の動きと日常的な行動との間には相関があると考え、凝視位置から行動の推定、行動ラベルから凝視位置の推定に取り組んだ。Bulling ら [3], [4] は、EOG 信号を計測することで眼球運動から行動予測を行った。この研究では、視線情報はまったく用いず凝視や瞬きといった簡単な眼球の動きの特徴から行動の認識を行った。

これらの研究では視線情報をそのまま特徴として学習器に入力し推定を行っているが、本研究では視線情報を入力とした学習器の出力を利用して、他の特徴を入力とする学習器の学習の際に重み付けを行う。

3. 提案手法

提案手法の概要を図 2 に示す。提案手法は 3 つのステッ

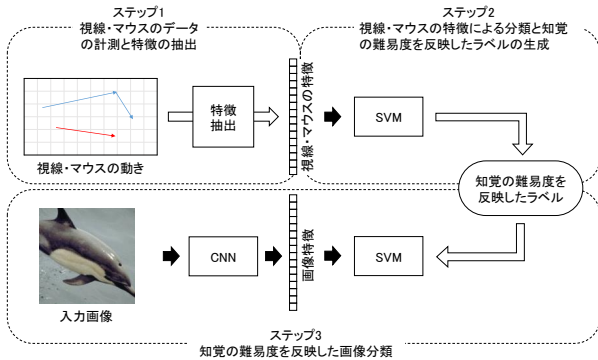


図 2: 提案手法の概要. ステップ 1 では, 画像分類タスクに取り組む人間の視線・マウスのデータを収集し, 表 1 のような 15 次元の視線・マウスの特徴を抽出する. ステップ 2 では, 視線・マウスの特徴で SVM を学習させ, 各サンプルの知覚の難易度を推定する. ステップ 3 では, CNN 特徴を用いて SVM の学習を行う際に, ステップ 2 で生成した知覚の難易度を反映したラベルを使って各サンプルに重み付けを行う.

プに分かれる. ステップ 1 では, 人間が画像の分類タスクに取り組んでいるときの視線・マウスのデータを計測し, 特徴を抽出する. 次にステップ 2 では, 計測した視線・マウスのデータと画像のクラスラベルを用いて画像のクラスを推定するサポートベクターマシン (SVM) を学習させ, それぞれのサンプルの決定境界との差をもとに知覚の難易度を反映したラベルを生成する. ステップ 3 では, 学習済みのモデルを使って画像から抽出した CNN 特徴と画像のクラスラベルを入力として, 画像を分類する SVM を学習させる. その学習の際に知覚の難易度を反映したラベルを用いることで決定境界を変化させる. 以下, 3.1 節では画像分類タスクに取り組む人間の視線・マウスのデータの計測と特徴の抽出について述べ, 3.2 節で視線・マウスのデータを用いた SVM の学習と知覚の難易度を反映したラベルへの変換について述べた後, 3.3 節で知覚の難易度を反映したラベルを用いた画像分類について述べる.

3.1 視線・マウスのデータの計測と特徴の抽出

本研究では, 画像分類タスクに取り組む人間の視線情報を利用する. まず, モニターに表示された数十枚の画像の中から指定されたクラスの画像を探し, 制限時間内に画像上でクリックするタスクを人間に与える. 人間が分類タスクに取り組んでいる間のフレーム毎の視線・マウスの座標とタイムスタンプを記録する. 視線運動は, ある一点に留まる凝視とある凝視点から次の凝視点へと移るサッケードに大別される. 本研究では, 視線の角速度が 30 [degree/second] を超えるとサッケード, 30 [degree/second] 以下のとき凝視と分類する. これらの情報から各被験者の各画像に対する凝視の合計時間, 凝視の回数, 最初の凝視までの時間,

		中央値
		分散
凝視	合計時間	中央値
	回数	分散
	最初の凝視までの時間	中央値
マウス	合計時間	中央値
	回数	分散
	最初にカーソルが入るまでの時間	中央値
マウス	最初にクリックするまでの時間	中央値
	クリックした被験者の割合	分散

表 1: 視線・マウスの特徴

マウスのカーソルの合計滞留時間, マウスのカーソルが画像内に入った回数, 最初にマウスのカーソルが画像内に入るまでの時間, マウスをクリックするまでの時間を抽出する. そして, 各値が 0 から 1 の間の値となるように正規化を行い, 画像ごとの全被験者の中央値と分散, そして画像をクリックした被験者の割合の表 1 に示す 15 次元の特徴を取得する.

3.2 視線・マウスの特徴による分類と知覚の難易度を反映したラベルの生成

視線・マウスの特徴を用いた SVM の学習とその SVM の学習結果を用いて知覚の難易度を反映したラベルを生成する手法について述べる. SVM はマージン制御に従った高い汎化性能を持つ二項分類器であり, 学習データと決定境界との間の最短距離であるマージンを最大にするようにパラメータの学習を行う. 一般に SVM のような二項分類器においては, 式 (1) に示すようなヒンジ損失関数が用いられている.

$$\phi_h(z) = \max(0, 1 - z) \quad (1)$$

$$z = yf(i)$$

ここで, $f(i)$ は画像 i の入力ベクトルと決定境界との差である.

画像を見ている人間の脳活動を扱った Fong らの研究では, 脳活動のサンプルと SVM の決定境界との差 $d(i)$ を次のような式 (2) を用いて変換することで, 画像の知覚の難易度を反映したラベル c_i を得た [7], [11].

$$c_i = \frac{1}{1 + \exp(-ad(i) + b)} \quad (2)$$

パラメータ a, b は SVM の学習の際に決定される. 式 (2) では正解のクラスラベルによらず, 推定結果が $y = 1$ で決定境界から離れるほど c_i の値が大きく, 推定結果が $y = -1$

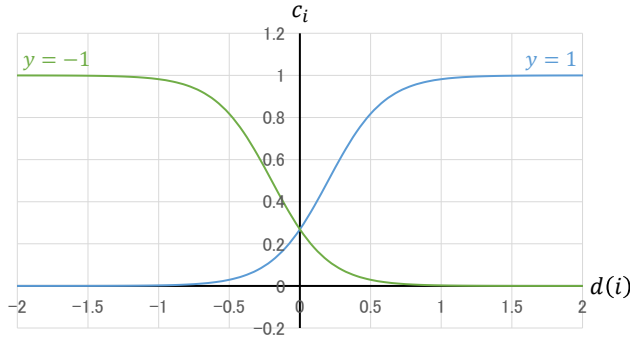


図 3: 視線・マウスの特徴によって学習した SVM の決定境界と各サンプルとの差から知覚の難易度を反映したラベルへの変換. 分類が正しくかつ決定境界から離れている場合は値が大きくなり, 分類が誤りで決定境界から離れている場合は小さくなる.

で決定境界から離れるほど c_i の値が小さくなる.

本研究では, 学習用データとして表 1 に示す 15 次元の視線・マウスの特徴ベクトルとクラスラベルを与える. このクラスラベルは, 人間に与える画像分類タスクにおいて選択されるべきクラスを 1, そうでないクラスを -1 とする. SVM の誤分類のコストを調整するパラメータ C と Radial Basis Function (RBF) カーネルのパラメータ γ については交差検証を用いて, グリッド探索で決定する. 画像 i に対する視線・マウスの特徴ベクトルのサンプルと SVM の決定境界との差 $d(i)$ を用いて知覚の難易度を反映したラベルを生成する. 視線・マウスの特徴によって学習した SVM の決定境界と各サンプルとの差と知覚の難易度との間には, 分類が正しくかつ決定境界から離れている場合は知覚が容易なもの, 分類が誤りである場合や決定境界に近い場合は知覚が困難なものであるという関係が成り立つと考える. そこで視線・マウスの特徴ベクトルと SVM の決定境界との差 $d(i)$ を次のような式を用いて変換することで, 画像の知覚の難易度を反映したラベル c_i を生成する.

$$c_i = \begin{cases} \frac{1}{1 + \exp(-ad(i) + b)}, & y = 1 \\ \frac{1}{1 + \exp(ad(i) + b)}, & y = -1 \end{cases} \quad (3)$$

$a \in \mathbb{N}, \quad b \in \mathbb{Z}$

パラメータ a, b は交差検証を用いてグリッド探索で決定する. 図 3 に $d(i)$ から c_i への変換を示す.

3.3 知覚の難易度を反映した画像分類

式 (1) で示したヒンジ損失関数では分類の誤りの度合いの大きさに比例して大きな損失を与える. 本研究では, ヒンジ損失関数の代わりに次のような重み付き損失関数を用いることを提案する.

$$\phi_w(z) = \max(0, (1 - z)M(i, z)) \quad (4)$$

$$M(i, z) = \begin{cases} \frac{1}{2} + c_i, & \text{if } z < 1 \\ 1, & \text{otherwise} \end{cases}$$

この重み付き損失関数ではラベル c_i の値の大きなもの, すなわち視線・マウスの特徴から識別されやすいような画像に対してより大きな損失を与える. この重み付き損失関数を用いることで, $z > 1$ となる場合には視線・マウスの特徴による推定の結果にかかわらず損失 $\phi_w(z)$ は 0 となり, z の値が小さくなる場合には視線・マウスの特徴による推定の正しさに比例して損失 $\phi_w(z)$ は大きくなる. 重み付き損失関数は非凸なので全体の収束は保証されないが, 重みによる調整で適当な局所解を見つけることができる.

本研究ではまず, SVM とのパラメータ C と RBF カーネルのパラメータ γ を決定するために, ヒンジ損失関数を用いて SVM を学習させる. 学習済みのモデルを用いて画像から抽出した CNN 特徴と画像のクラスラベルを入力として与える. パラメータについては交差検証を用いて決定する. 学習の際にクラス間にデータ数の偏りが生じるので学習の際のパラメータを調整することで偏りをなくした. 次に, ヒンジ損失関数を用いた学習と同様のパラメータを使用して, 式 (4) の重み付き損失関数を用いて SVM を学習させる. 決定境界の近くに存在する c_i の値の大きなサンプル, すなわち視線・マウスの特徴から正しく分類されやすいサンプルの損失が大きくなるので, そのようなサンプルが正しく分類されるように決定境界が変化する.

4. 評価実験

提案手法の有効性について, 画像特徴による分類の際に式 (1) のヒンジ損失関数を用いた場合と式 (4) の重み付き損失関数を用いた場合の SVM の分類精度を比較することで評価する. 本研究ではクラス間の違いを認識することが難しく, かつ専門家でない人間にも認識できるような例として, ImageNet [5] の dolphin, whale, killer whale, shark の画像と Places205 [18] の corn_field, golf_course, pasture, rice_paddy の画像を用いて実験を行った. ImageNet とは物体画像から構成されるデータセットで, 一般画像認識用に用いられる. Places205 とはシーン画像から構成されるデータセットで, シーン認識用に用いられる. 前者を物体データセット, 後者をシーンデータセットとする. それぞれのデータセット内の画像の例を図 4 に示す.

それぞれのデータセットに対し, 各クラス 150 枚, 計 600 枚の画像を 1 セットとし, 同じ画像を含まないように 5 セット用意した. このうち 4 セットを人間に与える分類タスクと画像特徴による分類の学習用データに使用し, 残りの 1 セットを画像特徴による分類のテスト用データとして使用する. 本研究では, オープンソースの DCNN フレー

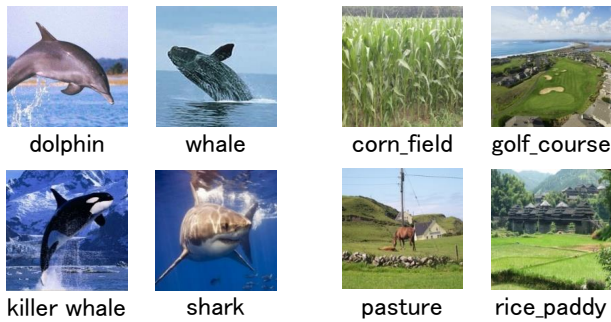


図 4: データセットの例 (左: 物体データセット, 右: シーンデータセット)

ムワークである Caffe [8] を用いる。Caffe には、物体画像から構成される ILSVRC2012 データセットを用いてあらかじめ学習された AlexNet のモデルと、シーン画像から構成される Places205 データセットを用いてあらかじめ学習された AlexNet のモデルが用意されており、これらのモデルに各画像を入力したときの中間層の出力から CNN 特徴を抽出した。SVM の学習については Python のオープンソース機械学習ライブラリである scikit-learn^{*1} を用いる。

4.1 視線・マウスのデータの収集

本節では、視線計測実験を行う際の環境や条件について述べる。視線計測器として Tobii Pro X3-120 を使用した。視線計測器を取り付けた画面サイズ 27 インチ、解像度 1920 × 1080 のモニターに画像を表示し、60 Hz で視線・マウスの動きを記録した。画像は 10 × 6 グリッドで各画像 170 × 170 pixel で表示した。画像の表示とマウスのデータの記録には C++ のオープンソースツールキットである openFrameworks^{*2} を使用した。被験者はモニターから 60 cm 離れた位置に座り、頭部を固定した状態で実験を行った。実験の様子を図 5 に示す。実験には男性 9 人、女性 1 人の計 10 人が参加し、被験者の年齢は平均 22.8 歳 (標準偏差 0.6) であった。

被験者には実験を始める前に分類タスクに用いない各クラスの画像を見てそれぞれクラス間の違いについて確認する時間を与えた。実験を始める前に視線計測器の較正を行った。一度に各クラス 15 枚、計 60 枚をランダムに並べて 45 秒間表示することを 1 クラスにつき 10 回繰り返す。被験者はそれらの画像を見ながらあらかじめ指定されたクラスの画像を探し、その画像上でクリックすることで分類を行う。それが終わると、指定するクラスを別のものに変え、同じ画像を再び見ることがないように別の画像セットを使用する。これをそれぞれのデータセットに対して 4 クラス分行う。表示する画像の例を図 6 に、実験の設定につ



図 5: 実験のセットアップ。手元のマウスを用いて画像アノテーションを行う。モニター画面の下部に取り付けた視線計測器を用いて計測を行う。



図 6: 表示する画像の例。60 枚の画像の内、各クラスの画像が 15 枚ずつ含まれている。

画像クラス数	4
一度に表示される各クラスの枚数	15
一度に表示される合計枚数	60
画面が切り替わるまでの時間 (s)	45
画面が切り替わる回数	10
1 セット当たりの各クラスの画像枚数	150
1 セット当たりの合計枚数	600
1 セット当たりの実験の時間 (s)	450

表 2: 視線計測実験の実験設定

いて表 2 に示す。

4.2 提案手法による分類精度

視線・マウスの特徴による SVM の学習では、学習用データとして 600 枚分の 15 次元の特徴ベクトルと画像のクラスラベルを使用した。学習用データ 600 枚の内、視線計測実験においてクリックするように指示されたクラスを $y = 1$ 、それ以外のクラスを $y = -1$ とする。テスト用データも学習用データで $y = 1$ のクラスを $y = 1$ 、 $y = -1$ のクラスを $y = -1$ とする。学習した SVM の決定境界と各サ

^{*1} <http://scikit-learn.org/>

^{*2} <http://openframeworks.cc/>

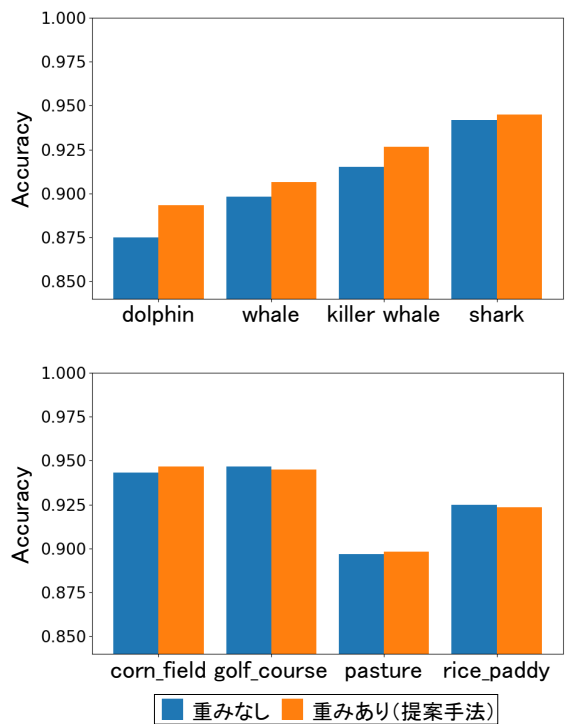


図 7: 物体データセットに対する分類精度（上）とシーンデータセットに対する分類精度（下）。青色はヒンジ損失関数を用いた場合、橙色は重み付き損失関数を用いた場合を表す。

ンプルとの差を式 (3) を使って、0 から 1 の間の値をとる知覚の難易度を反映したラベルに変換する。それぞれの値は式 (4) のように損失関数に対する重み付けに用いる。画像特徴による SVM の学習には、視線・マウスの特徴による SVM の学習で使用した 600 枚の画像の CNN 特徴とクラスラベルを使用する。この画像特徴による SVM の学習の際に、ヒンジ損失関数を用いた場合と重み付き損失関数を用いた場合についての比較を行った結果を図 7 に示す。物体データセットに対してはいくつかのクラスに対して精度の向上が見られたが、シーンデータセットに対しては精度の向上はあまり見られなかった。

人間が知覚しやすい画像の分類結果が重み付き損失関数を用いることで改善されているかどうかを確認するため、分類結果が変わった例を図 8 に示す。分類結果が悪化したの画像の中には対象物がほとんど写っていないものやシーン画像にもかかわらず画像の大部分を物体が占めているものがあり、人間にも知覚が難しいような画像が含まれていることがわかる。それに比べて分類結果が良化した画像は人間には識別できるものであるので、提案手法を用いることで画像の知覚の難易度を反映した画像分類を行えることを示唆している。



図 8: 重み付き損失関数を用いることで分類結果が変わった例。左が不正解から正解に変わった例。右が正解から不正解に変わった例。

凝視	合計時間
	回数
	最初の凝視までの時間
マウス	合計時間
	回数
	最初にカーソルが入るまでの時間
	最初にクリックするまでの時間
	クリックの有無 (0 または 1)

表 3: 個人の視線・マウスの特徴

4.3 個人差に関する考察

被験者毎に分類結果の違いがあるか、1 人分の視線・マウスのデータからも知覚の難易度の推定が行えるかを検証するため、視線・マウスの特徴として表 3 に示した 8 次元の特徴ベクトルを用いて実験を行った。被験者毎の知覚の難易度を反映したラベルによる重み付き関数を用いた場合の SVM の分類精度を図 9 に示す。多くの場合、被験者毎の知覚の難易度を反映したラベルによる重み付き関数を用いるとヒンジ損失関数を用いた場合の精度は上回るが、全被験者のデータを用いた重み付き損失関数を用いた場合の精度は下回る。さらに、同一データセット内でもクラスによって精度の差があり、個人の認識能力による差が生じたとは言えない。本研究においてはタスクに対して最適な被験者は存在せず、全体的な精度を考えると被験者毎でなく、全被験者のデータを用いた重み付き損失関数を用いる方が精度が向上すると言える。

5. 結論

本研究では、人間の注視行動を用いた画像アノテーションから画像の知覚の難易度を反映したラベルを生成し、それを用いて画像分類を行う手法について提案した。提案手法では、画像分類タスクに取り組む人間の視線・マウスのデータから特徴を抽出し、それを用いて分類器を学習させた。各サンプルと決定境界との差は知覚の難易度を反映していると捉え、その値をもとに画像の知覚の難易度を反映

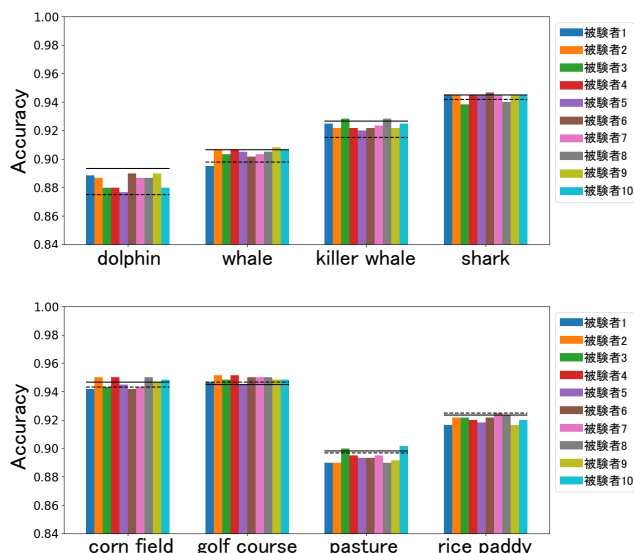


図 9: 物体データセットに対する被験者毎の分類精度 (上) とシーンデータセットに対する被験者毎の分類精度 (下). 破線はヒンジ損失関数を用いた場合の分類精度, 実線は提案手法の分類精度, 各棒グラフは被験者毎の知覚の難易度を反映したラベルを用いた場合の分類精度を表す.

した連続的なラベルを生成した. このラベルを用いて, 画像特徴による分類器の学習を行う際に各サンプルの損失の大きさを調整することで, 人間が知覚しやすい画像を正しく分類させることを可能とした.

提案手法の有効性を検証するために評価実験として物体画像とシーン画像の分類を行った. 提案手法を用いることによる著しい精度の向上は見られなかったが, 分類結果が変わった画像の例から人間が知覚しやすい画像が正しく分類され, 人間が知覚しにくいような画像の分類が誤りになることがわかる. この結果から人間の視線情報を用いることで画像の知覚の難易度を推定することができ, また, 知覚の難易度を用いることで曖昧さを反映した分類を行うことができると考えられる.

今後の課題としては, 本研究で行ったタスクは機械による分類精度が高く, 提案手法による効果があまり見られなかったが, 機械による分類精度が低いようなタスク, 例えばクラス間の違いの認識が難しいような場合に大きな効果があるかを検証することが挙げられる.

参考文献

- [1] Borji, A. and Itti, L.: Human vs. computer in scene and object recognition, *Proceedings of the IEEE Conference on CVPR*, pp. 113–120 (2014).
- [2] Borji, A., Sihite, D. N. and Itti, L.: Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study, *IEEE Transactions on Image Processing*, Vol. 22, No. 1, pp. 55–69 (2013).
- [3] Bulling, A., Ward, J. A., Gellersen, H. and Troster, G.: Eye movement analysis for activity recognition us-

- ing electrooculography, *IEEE TPAMI*, Vol. 33, No. 4, pp. 741–753 (2011).
- [4] Bulling, A., Weichel, C. and Gellersen, H.: EyeContext: recognition of high-level contextual cues from human visual behaviour, *Proceedings of the SIGCHI*, pp. 305–308 (2013).
- [5] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database, *Proceedings of the IEEE Conference on CVPR* (2009).
- [6] Fathi, A., Li, Y. and Rehg, J. M.: Learning to recognize daily actions using gaze, *Proceedings of the ECCV*, pp. 314–327 (2012).
- [7] Fong, R. C., Scheirer, W. J. and Cox, D. D.: Using human brain activity to guide machine learning, *Scientific reports*, Vol. 8, No. 1, p. 5397 (2018).
- [8] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. and Darrell, T.: Caffe: Convolutional architecture for fast feature embedding, *Proceedings of the 22nd ACM ICM*, pp. 675–678 (2014).
- [9] Karpathy, A. and Fei-Fei, L.: Deep visual-semantic alignments for generating image descriptions, *Proceedings of the IEEE Conference on CVPR*, pp. 3128–3137 (2015).
- [10] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L.: Large-scale video classification with convolutional neural networks, *Proceedings of the IEEE Conference on CVPR*, pp. 1725–1732 (2014).
- [11] Platt, J. C.: Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods, *ADVANCES IN LARGE MARGIN CLASSIFIERS*, MIT Press, pp. 61–74 (1999).
- [12] Pramod, R. and Arun, S.: Do computational models differ systematically from human object perception?, *Proceedings of the IEEE Conference on CVPR*, pp. 1601–1609 (2016).
- [13] Scheirer, W., Anthony, S., Nakayama, K. and Cox, D.: Perceptual annotation: Measuring human vision to improve computer vision, *IEEE TPAMI*, Vol. 36, No. 8, pp. 1679–1686 (2014).
- [14] Shimojo, S., Simion, C., Shimojo, E. and Scheier, C.: Gaze bias both reflects and influences preference, *Nature neuroscience*, Vol. 6, No. 12, pp. 1317–1322 (2003).
- [15] Spampinato, C., Palazzo, S., Kavasidis, I., Giordano, D., Souly, N. and Shah, M.: Deep learning human mind for automated visual classification, *Proceedings of the IEEE Conference on CVPR*, pp. 6809–6817 (2017).
- [16] Sugano, Y., Ozaki, Y., Kasai, H., Ogaki, K. and Sato, Y.: Image preference estimation with a data-driven approach: A comparative study between gaze and image features, *Journal of Eye Movement Research*, Vol. 7, No. 3 (2014).
- [17] Zhang, X., Sugano, Y., Fritz, M. and Bulling, A.: Appearance-Based Gaze Estimation in the Wild, *Proceedings of the IEEE Conference on CVPR* (2015).
- [18] Zhou, B., Lapedriza, A., Xiao, J., Torralba, A. and Oliva, A.: Learning deep features for scene recognition using places database, *Advances in NIPS*, pp. 487–495 (2014).