教師なし学習による物体認識器の構成法に向けて
Toward Unsupervised Construction of Object Identification Model

河井翔太郎 [1)]，日高昇平 [2)]
KAWAI Shotaro [1)]，HIDAKA Shohei [2)]
s1810042@jaist.ac.jp, shhidaka@jaist.ac.jp

1) 北陸先端科学技術大学院大学
1) Japan Advanced Institute of Science and Technology

【キーワード】教師なし学習，オブジェクト同定

1. 背景

人は視覚的タスクを行うとき，入力された情報から，視覚的パタンの背後に潜在する何らかの法則を見つけ出す(Marr, 1982). この意味で視知覚処理は Helmholtz の無意識的推論とも呼ばれる.

一方，近年の機械学習分野の研究では，大量のデータの学習によって視覚的処理を行う認識器を構成し，特定の課題では人と同等以上の成績を収めることが報告されている(Park et al., 2017). しかし，こうしたモデルは学習データの十分でない領域において，人の情報処理からは想定しえない出力がされる場合が多々ある(Goodfellow et al., 2015).

こうした人と機械学習モデルの違いは，逆に人が視覚的情報処理をする際に，視覚データの個々の要素を独立にではなく，何らかの統合的な制約の下で解釈していることを示唆する．言い換えれば，人と同等の制約を機械学習モデルに与えることで，より人として自然な情報処理が可能になると考えられる.

2. 研究目的

機械学習モデルのパラメタに関する制約の例として，なるべく実効的なパラメタ数を減らす制約を導入するスパースモデリングなどがある．一般に，このようなモデルに関する制約は，課題の種類などに応じてモデルのユーザーが選択し与えている．こうした制約を適切に選択すれば，特定の課題で求められる“正解”の存在する領域を狭めるのに有効であるが，その選択は課題領域に関する知識を要求する.

人はこうした知識を自律的に獲得すると考えられるため，外的に与えられた制約に依存せずに，データに潜在する何らかの規則性を見つける方法論が模索されている．日高(2018)は，一切のラベル付き学習データなしに，与えられる 2 時点の視覚像だけで，ある線形写像を共有する点集合からそれを共有しない点集合を認識可能であることを示している.このことは抽象化した制約を与えることで，そこから適切な制約を決定していることを示している．しかし，先行研究の方法論は点集合を対象としており，標準的に視覚的物体認識で想定される画像を想定した方法論の開発には至っていない.

本研究の目的は，コンピュータが人のように，視覚的タスクを行えるようにすることを目指し，先行研究を基に視覚的タスク実行における制約を自ら決めることができるモデルの構築を行うことである.

まず，予備実験で示した課題を重回帰分析を用いて行うことで，学習モデルがどのような推論を行うか確認を行った.

3. 関連研究

Agrawel et al.(2015)では 2 つのペア画像を入力として，2 つの画像を撮影時のカメラ姿勢の変換を出力(教師信号)として与えたとき，中間層に有用な視覚的表現を学習できることを示している．この研究では，カメラ姿勢の変換を明示的に教師信号として与えるが，本研究では，その変換が適切に表現されるモデルの制約を模索する.

4. 予備実験

4.1 予備実験概要

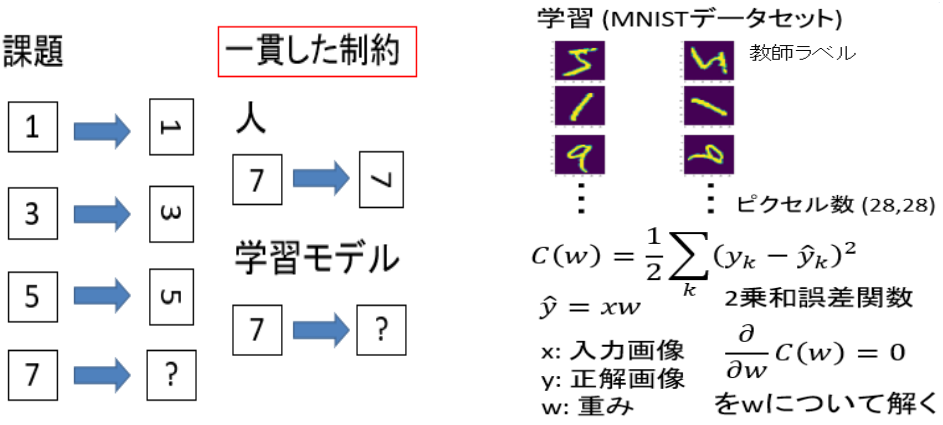

図 1： 予備実験の概要図

予備実験として考えたタスクの概要を図 1 に示す．学習データとして MNIST データセット，学習モデルとして 2 乗和誤差関数を用い，予備実験の再現を行った．

4.2 推論結果

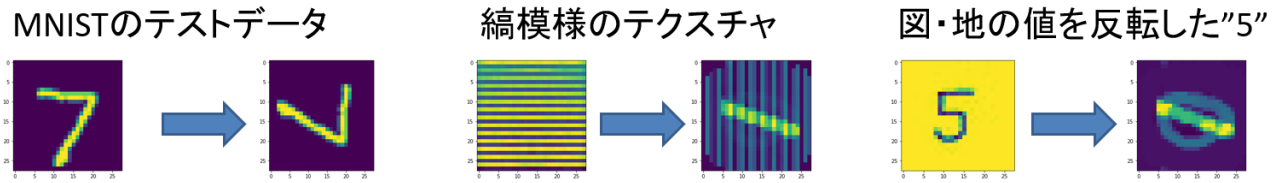

図 2： 予備実験の推論結果

推論結果を確認したところ，MNIST のテストデータ(訓練データと類似のデータ)に関しては教師信号(人であれば, 90 度回転という規則性が見出されると想定される)と類似の出力を得た．

人の見出すであろう変換(すべての画素の 90 度回転に相当する置換行列)と同じ重みを学習したか確認するため, MNIST データ以外のデータを入力とした．その結果，縞模様のテクスチャ，図・地の値を判定した“5”に関しては，想定しえない出力が得られた．これは MNIST のデータは数字(図)以外の背景(地)部分は 0 であるという特殊性を過学習したためだと考えられる．

これを解決するため，0～0.01 の一様な乱数を全ての画像の各ピクセルに足した後，最初の工程からやり直し，出力を確認した．

4.3 推論結果(条件変更後)

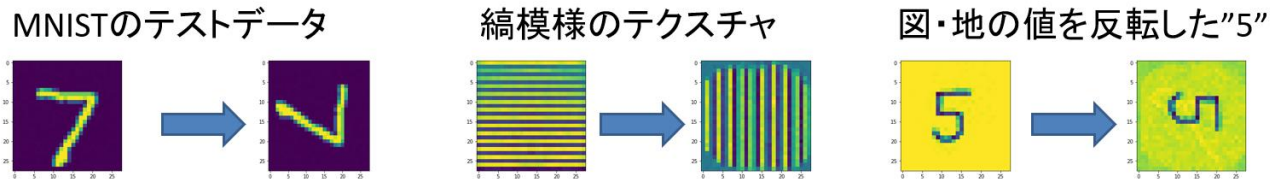

図 3： 予備実験の推論結果(条件変更後)

条件変更後の推論結果を確認したところ，MNIST のテストデータに加えて，縞模様のテクスチャ，図・地の値を反転した“5”の入力に対しても 90 度回転した画像に類似の結果が得られた.

しかし，縞模様のテクスチャ図・地の値を反転した“5”の出力は画像の周辺がぼやけている．これは数字(図)の提示頻度が低く，情報価の低い周辺部分ではデータのパタンでは無く，モデルの制約(この場合は，重みベクトルノルム最小化)が反映されたと思われる.

5. 今後の方向性

重回帰分析において，人のような出力(画像を 90 度回転させる置換行列)は，回帰係数行列が直交かつスパースという条件を満たす．よって，そのような制約を学習時に与えると人と同じような出力となる可能性がある.

今後の方向性として，先行研究の理論を基にモデルを作成し，この課題において，自ら画像を 90 度回転させる重み，つまり置換行列となる重みを学習できるようにすることを目指す.

6. 謝辞

本研究は科学研究費補助金　若手研究 A(16H05860)の補助を受けて行われた.

参考文献

Marr, D. (1982). Vision, MIT press.
Park et al. (2017). IMAGENET Large Scale Visual Recognition Challenge (ILSVRC) 2017 Overview.
Goodfellow et al. (2015). Explaining and Harnessing Adversarial Examples, ICLR.
日髙昇平 (2018) 「記号接地問題における地とは何か：視覚的物体の同一性の分析」
日本認知科学会第 35 回大会論文集, pp. 306-310.
Agrawal et al. (2015) Learning to See by Moving, ICCV.

連絡先
住所：〒923-1292　石川県能美市旭台 1-1 北陸先端科学技術大学院大学
名前：河井翔太郎
E-mail：s1810042@jaist.ac.jp