

CNN型AEとLSTMを用いた，転移学習モデルD-CAFLの提案

D-CAFL: transfer learning model using Convolutional Auto-Encoder and LSTM

芦原 佑太^{*1} 佐藤 聡^{*2} 栗原 聡^{*3}
Ashihara Yuta Sato Akira Kurihara Satoshi

^{*1*3}電気通信大学 大学院情報システム学研究所 社会知能情報学専攻

The University of Electro-Communications, Graduate School of Information Systems, Department of Social Intelligence and Informatics

^{*2}株式会社クロスコンパス
XCompass Ltd.

DeepLearning has built a great trajectory in the field of machine learning. However, while the success of many tasks, it is less likely to be discussed for the reuse of trained middle layers. Then, we propose transfer learning model named D-CAFL(Deep-CNN-AE-FV-LSTM), which can reuse trained middle layers. We adopt this model to predict the rotation of the object. A final surprising result is that rotation can be estimated in the object that has not been learned by our model.

1. はじめに

本稿は，Convolutional Neural Network(CNN)型のAutoencoder(AE)の中間層を入力層にFlag Vectorを付加したLong Short Term Memory(FV-LSTM)で書き換えることで，物体の回転画像を想起させる転移学習モデル，Deep-CNN-AE-FV-LSTM(D-CAFL)を提案する。

DeepLearningは入力層，中間層，出力層にそれぞれ配置されたニューロンが値を受け渡すことで，様々な問題を解決する。データを受け付ける入力層と，結果の出力を行う出力層の間にある中間層は，問題を解決するための情報表現を学習によって獲得する。特に，CNNにおいては，中間層が多層である場合，入力層に近いニューロンは抽象的な特徴を，出力層に近いニューロンは具体的な特徴を得ていることが知られている[1]。AEは入力されたデータを復元するように学習する過程で，情報を圧縮する能力を中間層で得ている。しかし，これら中間層の表現は，問題を解く上で副産物のように扱われることが多い。そこで，本稿では，中間層の表現を再利用するモデルであるD-CAFLを提案し，物体の回転画像の想起問題に適用する。

2. 関連研究

Yoshinkiら[3]では，ある画像データで学習させたDeepLearningの中間層を他の画像データに適用する転移学習の実験を行い，抽象的な特徴量を学習した中間層の再利用が，有用であることを示した。Kirosら[5, 6]では，画像の特徴を抽出した中間層にテキスト情報をベクトル化して付加することで，画像データを説明する英文を生成するDeepLearningのモデルを提案している。[5]では，中間層に情報を付加する工夫がなされていることによって，画像データに対して次元数の少ないテキスト情報のベクトルが無視されることを防ぐこともできる。

これらの研究からDeepLearningの中間層が抽象的な特徴量を学習することで，他のデータに対しても適用可能であること，あるいは，他のタスクに適用することが可能であることを示している。本研究では，[3]のように事前に別のデータで学

連絡先: 芦原佑太，電気通信大学大学院情報システム学研究所
社会知能情報学専攻，y.ashi@ni.is.ucc.ac.jp

習したDeepLearningの中間層を再利用し，DeepLearningが学習できるような転移学習モデルを提案する。

2.1 提案モデル

本節では，画像の特徴量を抽出するためのCNN型自己符号化器としてのCNN-AEと，外部からのベクトル情報を付加した入力进行学习するFV-LSTM，及びそれらを組み合わせたD-CAFLについて述べる。

2.2 D-CAFL

D-CAFLのモデル図を図1に示す。

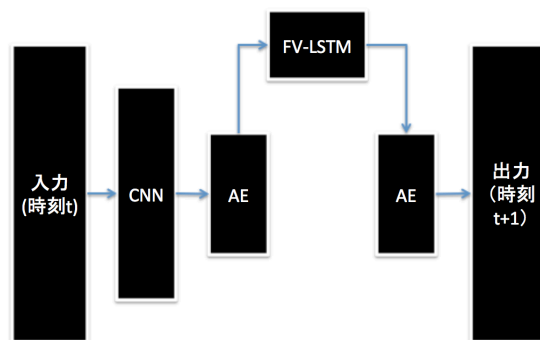


図1: CAFLの概略図

入力部から出力部まで，CNNで抽象的な特徴量に分解され，AEで情報を圧縮し，FV-LSTMで時系列処理を行った後，AEによって画像が復元される流れになっている。次節以降で各部分の処理について述べる。

2.3 CNN-AE

CNNは，畳み込み層，プーリング層を数層重ねることにより，入力される画像にある特徴を抽出することができる。AEは入力された画像を復元するように学習し，中間層の数を入力次元より少なくすることで，次元圧縮を実現することができる。本研究では，あらかじめ別のデータセットで事前に学習した状態のCNNとAEをそれぞれ用いる。

2.4 FV-LSTM

Pascanu ら [7] によれば、ネットワーク内に再帰構造があることが、時系列情報を扱う上で有用であるということが示されている。特に、Long Short-Term Memory(LSTM) において、入力されるデータの制御は4つのゲートによって行われる [8]。本研究では、[8] に基づいた LSTM を採用し、そのデータ入力部に、物体の回転角を表現した Flag Vector(FV) を追加した FV-LSTM を構築した。FV-LSTM では、入力されるデータの次の時刻情報を予測して出力するように学習する。

3. 実験と考察

本節では、D-CAFL を用いた物体の回転画像の想起について行った実験結果を示す。

3.1 データセット

今回用意したデータセットは、中心に物体が写っている。その物体が1枚毎に左回りに18度回転する画像を2物体準備した。画像サイズは50*50ピクセルで準備し、各ピクセルの値について最大値が1となるように、画像内のピクセルを255で割った値を使用する(図2.参照)。図2の右側2枚(物体B)については1周分の20枚全てを学習に用い、左側2枚(物体A)については後ろ姿と正面画像のみを学習に用いる。



図 2: 使用するデータセットの一部

3.2 回転予測実験

D-CAFL で実際に2つの物体の回転画像の想起について実験を行った。まず、物体Aについての想起実験を行ったところ、事前学習を行うことによって、モデルが上手く回転を認識していることがわかった(図3.参照)。

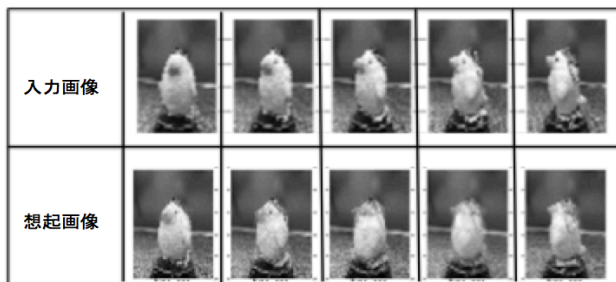


図 3: 物体 A の回転画像想起実験結果

次に、物体Bについても同様の実験を行った。物体Bは後ろ姿と正面の画像のみを学習したが、学習したことのない横向きなどの様子を含め、回転する様子が想起できた。

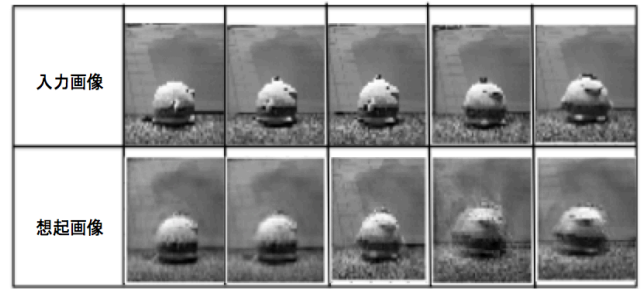


図 4: 物体 B の回転画像想起実験結果

4. おわりに

本稿では、DeepLearning の中間層を再利用し、物体の回転画像を想起するモデル D-CAFL を提案した。D-CAFL は、事前学習された CNN と AE, AE で圧縮された情報と回転情報を与える Flag Vector が付加された FV-LSTM の組み合わせによって実現される。D-CAFL を用いた回転画像の予測実験では、物体が回転する様子を出力することができたため、転移学習を行うことができたと考えられる。今後については、物体の種類を増やして実験を行うことや、回転以外の時系列処理について行いたい。

参考文献

- [1] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, Aude Oliva, "Learning Deep Features for Scene Recognition using Places Database", Neural Information Processing Systems, 2014
- [2] Koray Kavukcuoglu, Pierre Sermanet, Y-lan Boureau, Karol Gregor, Michael Mathieu and Yann Lecun, "Learning Convolutional Feature Hierarchies for Visual Recognition", Advances in Neural Information Processing Systems 23, pp.1090-1098, 2010.
- [3] Jason Yosinski, Jeff Clune, Yoshua Bengio, Hod Lipson, "How transferable are features in deep neural networks?", Advances in Neural Information Processing Systems 27, pp.3320-3328, 2014.
- [4] 野田 邦昭, 有江 浩明, 菅 佑樹, 尾形 哲也, "Deep neural network を用いたヒューマノイドロボットによる物体操作行動の記憶学習と行動生成", The 27th Annual Conference of the Japanese Society for Artificial Intelligence, 2013
- [5] Ryan Kiros, Richard S. Zemel, Ruslan Salakhutdinov, "Multimodal Neural Language Models", Proceedings of The 31st International Conference on Machine Learning, pp.595-603, 2014
- [6] Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, Sanja Fidler, "Skip-Thought Vectors", Advances in Neural Information Processing Systems 28, 2015
- [7] K.N. Ramamurthy, J.J. Thiagarajan, R. Sridhar, K. Premnisanth and N. Ramanathan, "Consensus inference with multilayer graphs for multi-modal data". IEEE Asilomar SSC, 2014.
- [8] Gers FA, Schmidhuber J, Cummins F, "Learning to forget: continual prediction with LSTM", Artificial Neural Networks, pp.850 - 855 vol.2, 1999.