CNN型AEとベクトル付加LSTMを用いた物体の回転画像の想起モデル

Recall of Rotated Image Model using Autoencoder with Convolutional Neural Network and Long Short-Term Memory

芦原佑太[1*]　佐藤 聡[2]　栗原 聡[1]

Yuta Ashihara[1]　Akira Sato[2]　Satoshi Kurihara[1]

[1] 電気通信大学　大学院情報システム学研究科　社会知能情報学専攻

[1] 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

[2] XCompass Ltd.

Abstract: 近年，ニューラルネットワークを多層化したDeepLearningという手法が注目を集めている．DeepLearningを使ったネットワークの中間層は入力されるデータに内在する特徴量を抽出することができるが，多くの研究は，分類問題，回帰問題の精度向上という目標に関して扱っており，抽出された特徴量が利用できるか，といったアプローチを取っている研究は少ない．そこで，本研究は，中間層で得られた特徴量を利用することのできるモデルとして，Convolutional Neural Network型Auto Encoder(CNN-AE)と，Flag Vectorを付加したLong Short-Term Memory (FV-LSTM)を用いたCNN-AE-FV-LSTM(CAFL)を提案する，

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モデル，CNN-AE-FV-LSTM(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によって問題を解決する．つまり，中間層のニューロンが問題を解決するための情報表現を，入力されるデータから得ており，特にCNNにおいては，中間層を深くするにつれ，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ことで，情報を圧縮する表現を中間層で得ることができる．しかし，中間層の表現は，回帰問題，分類問題を解く上で得られた副産物のように扱われることが多い．そこで，本稿では，物体の回転画像の想起する問題について，中間層の表現を利用して解くモデルであるCAFLを提案する．2節では関連研究を紹介し，3節では提案するモデルについて述べる．4節ではモデルを使った実験について述べ，最後に5節でまとめを述べる．

*連絡先：電気通信大学大学院情報システム学研究科
〒182-8585 東京都調布市 調布ケ丘1丁目5－1
E-mail: y.ashi@ni.is.uec.ac.jp

2　関連研究

DeepLearningのモデルにおいて，中間層の表現を利用して別のタスクに応用する研究は，野田らによる[2]があげられる．[2]では，感覚運動統合学習システムとして，900次元の画像情報を多層AEによって30次元に圧縮し，その圧縮された中間表現とロボットの関節情報を統合した情報がTime-delay型Neural Networkに入力されることで，環境に応じたロボットの行動選択を可能としている．そこで用いられているAEは，入力される高次元データを低次元データに圧縮しながら，特徴を上手く取り出すように学習している．

また，中間層にベクトルを付加する研究として，Kirosらの[3]では，モデルの中間層に新たな情報を付加することで，入力された画像を説明するテキストを生成するモデルを提案している．このモデルの中間層では，抽象化された画像情報とテキストを一致させる問題が解かれている．

上記で挙げた二つの研究は，いずれも中間層が入力されるデータを抽象化する空間として利用しており，中

間層の表現を利用することで，解くタスクの幅を広げることができることを示唆している．

3 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を学習する FV-LSTM，及びそれらを組み合わせた CAFL を提案する．

3.1 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．また，Yoshinki らの [4] によれば，CNN の入力層に近い層は，general な特徴を抽出しており，別のタスクに対しても使うことが可能であるとしている．本研究では，AE の入力部に事前学習された CNN の 3 層を加え，入力画像の再現精度を上げた CNN-AE を用いて，入力された画像を再現するネットワークを構築した．(図 1. 参照)

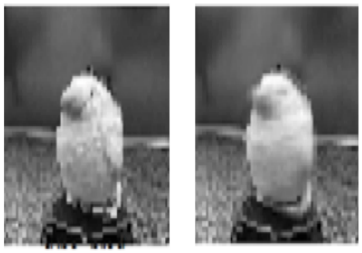

図 1: 左:入力された画像 右:CNN-AE によって復元された画像

3.2 FV-LSTM

Pascanu らの [5] 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) においては，入力されるデータの制御を 4 つのゲートによって行われる [6]．本研究では，[6] に基づいた LSTM を採用し，そのデータ入力部に，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．

3.3 CAFL

これまで述べた CNN-AE，FV-LSTM を組み合わせた CAFL(CNN-AE-FV-LSTM) は，CNN-AE のエンコード部分で圧縮された次元を FV-LSTM の入力とし，FV-LSTM の出力を CNN-AE のデコード部分につなげることで，特徴量の変換によって次の時系列の画像を予測するモデルとなる．(図 2. 参照)

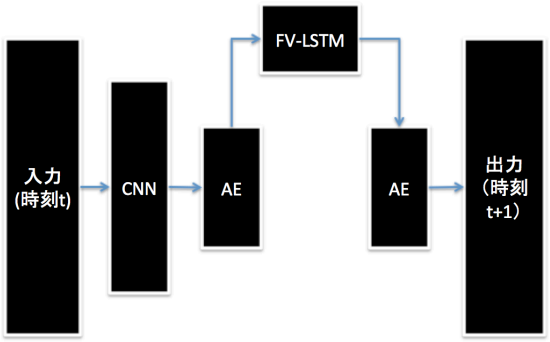

図 2: CAFL の概略図

4 実験と考察

本節では，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．

4.1 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おり，その物体が 1 枚毎に左回りに 20 °回転する画像を準備した．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 3. 参照)

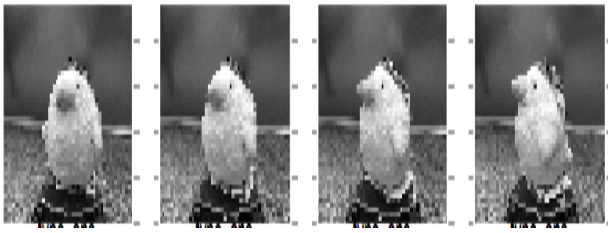

図 3: 使用するデータセットの一部

4.2 事前学習

CAFL ではデータが入力されると，はじめに CNN のネットワークによって特徴分解される．ここでの CNN は，今回の想起に使う画像だけでは十分なデータ数に

ならず，過学習に陥る可能性があるため，事前に [7] で公開されている CIFAR-10 の 10 クラス分類問題を解いて事前学習を行い．その時に得た学習済み CNN の最初の 3 層を用いた．

また，AE についても同様に CIFAR-10 の画像を自己符号化に使い，事前学習を行った．

4.3 FV-LSTM の学習

図 2. より，FV-LSTM は AE から情報を受け取るが，本モデルでは AE によって 25 次元まで圧縮された中間層を FV-LSTM の入力データとする．Flag Vector は物体の回転状態に応じた 11 次元のベクトルを付加し，合計 36 次元の入力を 1 時刻当たりの入力とする．

4.4 回転予測実験

CAFL で実際に回転画像の想起について実験を行った．まず，入力された画像の左回りに 20 °回転した後の画像を想起する実験を行った．実験の結果，想起された画像は入力された画像に対して，左回りに回転していることが伺えた．(図 4．参照)

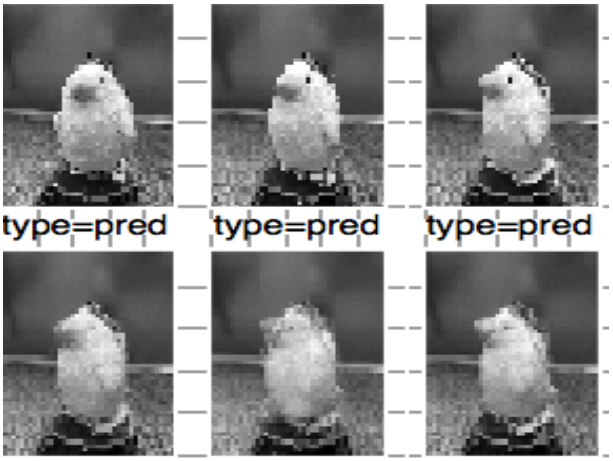

図 4: 上段:入力した画像データ，下段:想起された画像データ

次に，学習した物体以外についても同様の実験を行った．実験に用いた物体は学習に用いた物体に比べて，ある程度見た目の似たものを用意し，正面を向いた画像 (0 °) と，後ろを向いた画像 (180 °) の二つの画像のみを学習させ，横向の画像を学習させることなく，物体が回転する様子を想起できるかどうかについて行った．実験の結果，目や口の位置を把握し，回転するという動作に応じて位置を変化させていることがわかった．(図 5. 参照)

これらの実験により，中間層の表現を利用することで，物体を回転させる結果が示された．実際に利用した中間層と，FV-LSTM によって変換された中間層部分を可視化してみると，FV-LSTM が予測した結果が，理想的な中間層と似た表現を得ていることがわかった．(図 6. 参照)

しかし，想起された画像は，理想とする画像に比べて，全体的にぼやけた画像となっている．図 6 にある白い丸で囲った部分を比較すると，FV-LSTM によって変換された中間層は理想的な中間層に比べて，白い部分のベクトルの値がやや小さく，灰色に近い色となり，近隣のベクトルとの値の差が曖昧になっている．想起画像が理想的な画像と同じように，口や目といった特徴を鮮明に想起するためには，FV-LSTM の学習について工夫することが課題として挙げられる．

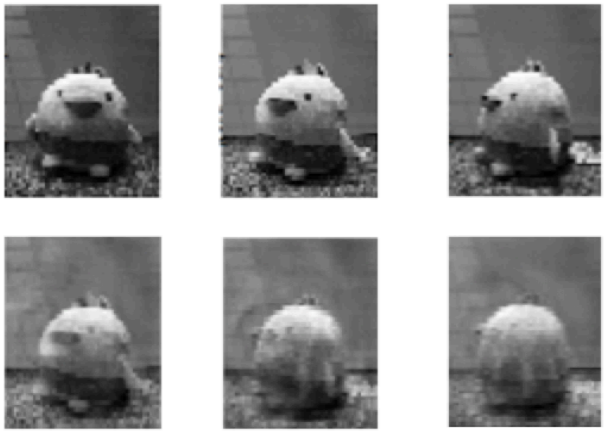

図 5: 上段:入力した画像データ，下段:想起された画像データ

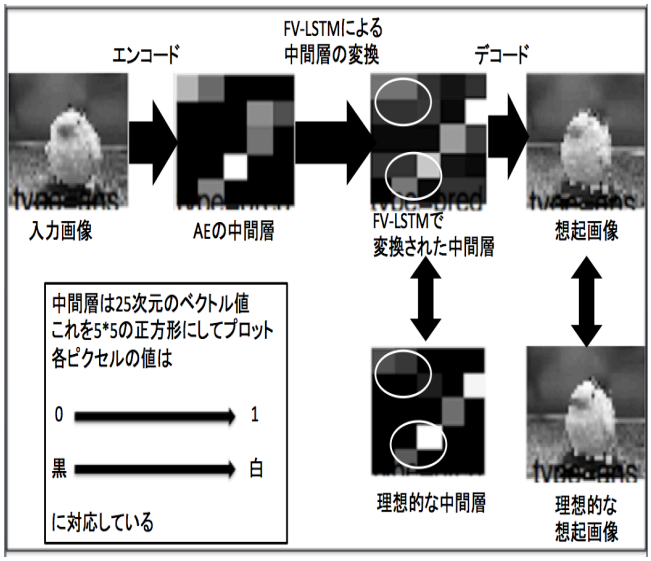

図 6: 画像を想起するまでの中間層を可視化と，理想的な中間層との比較

5　終わりに

本稿では，物体の回転画像を想起する CAFL を提案した．提案したモデルは，近年注目を集めている DeepLearning の計算論を取り入れており，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．実験では学習した物体と正面と後ろの画像のみ学習した物体のそれぞれの回転画像の想起について行い，物体が回転する様子を想起することができた．しかし，問題として，想起に失敗しやすい角度が物体の後ろ姿に近い角度で存在しており，後ろ姿からの特徴獲得による想起は課題となる (図 7. 参照)．

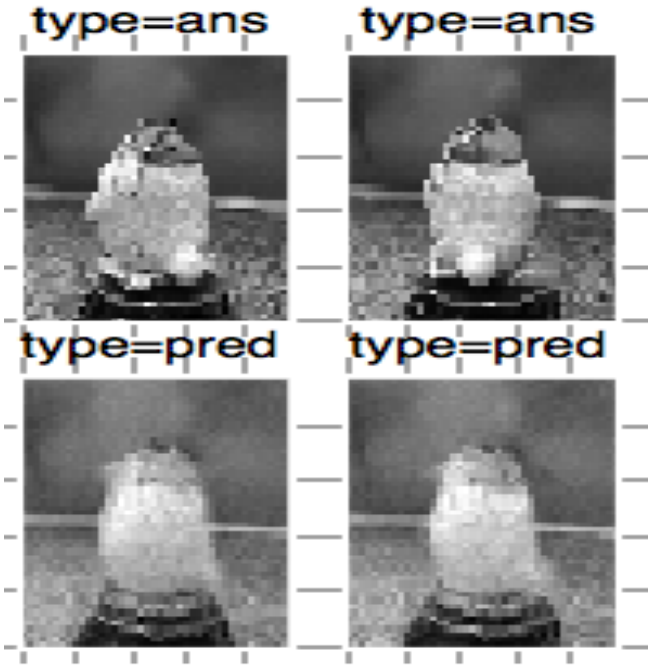

図 7: 上段:入力した画像データ，下段:想起された画像データ

また，今後の課題として，学習モデルとなる物体を増やして，より多くの物体の回転画像が想起できること，Flag Vector の中間層への付加の仕方を工夫することがあげられる．前者の課題は，CAFL が物体によらず回転の概念を獲得したモデルであることを主張するために必要な条件であり，後者の課題は，この CAFL を他のタスクに応用するために，Flag Vector がタスクに応じて回転以外の情報としても使えるようにすることが必要となるからである．

今後は，以上の課題を踏まえ，CAFL をより広範なタスクを扱えるモデルとして発展させていきたいと考えている．

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] 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)

[3] Ryan Kiros,Richard S. Zemel,Ruslan Salakhutdinov.: Multimodal Neural Language Models, *Proceedings of The 31st International Conference on Machine Learning*, pp. 595–603 (2014)

[4] Jason Yosinski,Jeff Clune,Yoshua Bengio,Hod Lipson.: How transferable are features in deep neural networks?, *Neural Information Processing Systems*, pp. 3320–3328 (2014)

[5] Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho,Yoshua Bengio.: How to Construct Deep Recurrent Neural Networks, *International Conference on Learning Representations* ,(2014)

[6] Alex Graves.: Generating Sequences With Recurrent Neural Networks, *The Centre for Computational Statistics and Machine Learning*(2014)

[7] CIFAR-10 and CIFAR-100 datasets `https://www.cs.toronto.edu/~kriz/cifar.html` (アクセス日: 2015/9/6)