第 27 回　Weekly Report

深層学習を用いた動画像からの危険認知手法のための基礎的研究

12 月　15 日　(火)

小松　大起

1　本実験の目的

人間の認知は時間経過による視覚世界の変化の予測が可能である. 従来の深層学習においては静止画の画像処理が中心であったが, 本研究では動画像から未来フレームの画像生成へと拡張を行い, 危険認知や行動予測などへの予測画像応用を行うための予測画像生成及び, その生成画像の評価方法を明らかにすることである.

2　PredNet

PredNet は Deep Recurrent Convolutional Neural Network の 1 種で 神経科学の概念である Predictive Coding を組み込んで作られたモデルである. 2016 年に William Lotter, Gabriel Kreiman, David Cox の 3 氏によって公開された.

2.1　PredNet の層構造

PredNet の各層には 4 つの素子が存在しておりそれぞれ、Target, Representation, Prediction, Error と呼ぶ. Target は下層からの出力である誤差信号をエンコード, 符号化する. Representation は Recurrent unit で, 上層からの出力, 側方からの誤差信号, 1 ステップ前の自分の出力を受け取る. Representation unit は Target の予測をする Prediction unit に投射し, 入力の予測が出力される. Error は Prediction と Target の誤差であり, Error は上層に送られる. この Error が小さくなるように学習を進めていく. また, 層構造の例を図 1 に示す.

3　使用した動画像

3.1　KITTI

ドイツの都市環境を運転している車の屋根に取り付けられたカメラによってキャプチャされたデータセットである.City, Residential, Road のカテゴリに分かれており, それぞれ都市街, 住宅地, 高速道路というようなシーン分けがされている. それぞれのカテゴリには City には 28 シーン, Residential には 21 シーン, Road には 12 シーン存在している. 合計 61 のシーンがあるが, それぞれのカテゴリから 10 フレームがテストデータとしてサンプリングされ予測画像生成に用いられ, 57 シーンが訓練に用いられ, 4 シーンが検証に用いられる. また, 用いられる画像は中央でトリミングされて,128 × 160 ピクセルの画像となっている. 1 フレームは 0.1 秒である. 最大 5 フレーム先の画像を生成することが可能.

4　オートエンコーダ

2006 年に hinton らによって提案された, 自己符号化器とも呼ばれるオートエンコーダはニューラルネットワークのモデルの一つである. 画像や音声データなどを符号化と呼ばれる圧縮作業を行い, 意味のある特徴量を残した後にそれを用いて復号化と呼ばれる次元の復元を行うことを目的として用いられることが多い.

4.1　本実験で用いるオートエンコーダの構造

使用する画像のサイズは 100 * 100 を RGB の 3 次元で入力とした. 中間層での素子の値 z_i 及び, 出力層の素子の値 y_k は, 入力のフィドフォワードで求められる. それぞれ, z_i の値と y_k の値は以下の式で与えられる.

$$z_j = \sum_i f_{ji} x_i \tag{1}$$

$$y_k = \sum_j w_{ji} z_i \tag{2}$$

また、中間層の素子での値を $\vec{z}$, 出力層の素子での値を $\vec{y}$ のベクトルで表す. f_{ji} を (j, i) 成分に持つ行列 F とした時, 以下のように表すこともできる.

$$\vec{z} = F\vec{x} \tag{3}$$

$$\vec{y} = W\vec{z} \tag{4}$$

また、オートエンコーダの概略図を図 2 に示す. 最適化関数には Adam を用いて, 損失関数には binary crossentropy を用いた. 中間層は 32 次元にして圧縮を行っている. 現段階での出力の結果を図 3 に示す.

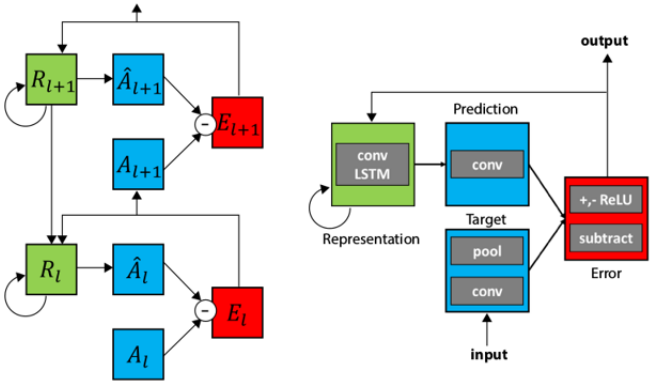

Fig. 1: PredNet の層構造の例

5　今週の作業

- 活性化関数の決め方を全部 relu かシグモイド関数でためしたがどちらの場合も再生成ができなかった. （勾配消失が起きて損失関数が減らなくなった）もともと中間層 3 層で最後の元の次元に戻す層のみをシグモイド関数にするとうまく再生成できていたので、特徴量として抽出したい層もシグモイド関数にして実行してみたがうまくできなかった。
- 物体認識系の論文探して現時点でどれくらいできているのか調べる（主に車載映像系？）
- 物体認識を軽く試してみる
- ドコモインターンの選考は通ったので GD の対策をしておく

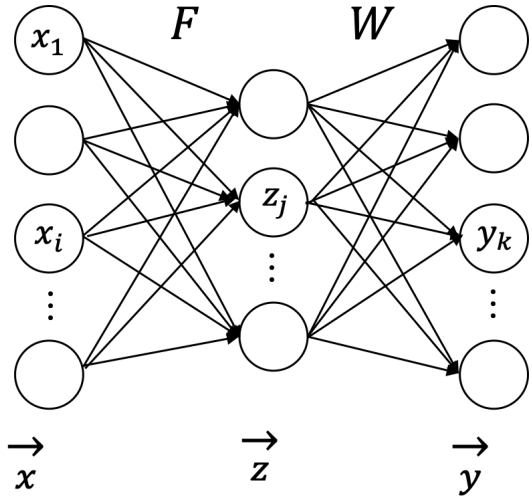

Fig. 2: autoencoder の概略図

Fig. 3: autoencoder の出力結果（カラー）