第 7 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

5 月　19 日　(火)

小松　大起

1　本実験の目的

危険認知手法実現のために動画像を用いた深層学習によって次フレームの画像を用いた深層学習によって次フレームの画像を予測し生成を行い, その画像と実際の画像を比べる.

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.

2.1　Predictive Coding

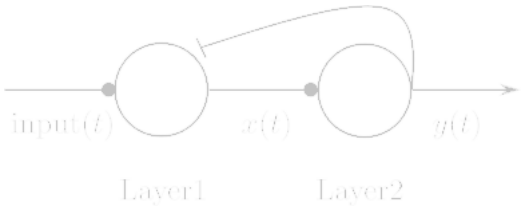

Fig. 1: 2 つのニューロンの模式図

2 種類のニューロンが存在するとし, それを図 1 で示す. 図の上では 2 つのニューロンであるが, Layer1 と Layer2 という 2 層について, それぞれを代表するニューロンを 1 つずつ描いている. まず, Layer1 のニューロンは下層から $input(t)$ を受け取る. これは入力刺激と呼ばれる. 次に Layer1 のニューロンは $x(t)$ を出力する. これは誤差信号または残差と呼ばれ, 以下の式が成り立つ.

$$x(t) = input(t) - y(t) \tag{1}$$

また, $y(t)$ は Layer2 のニューロンからの出力で, Layer1 のニューロンへの入力を予測する. ここで

$$dy/dt = x(t) = input(t) - y(t) \tag{2}$$

であり,$dy/dt = 0$ すなわち $y(t) = input(t)$ となるように学習が進行するモデルである.

2.2　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 2 に示す.

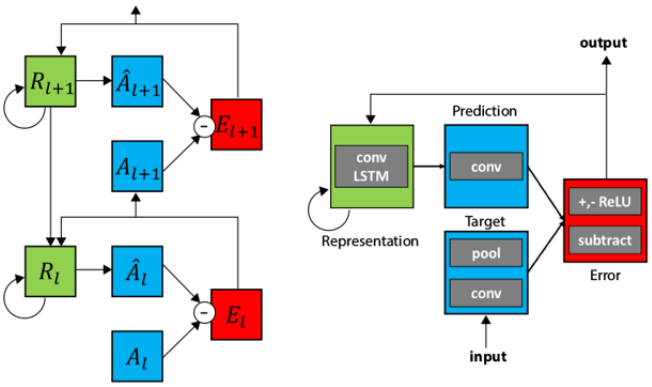

Fig. 2: PredNet の層構造の例

更新式は (1) から (4) のように表される.

3　使用した動画像

3.1　lyft level5

lyft という団体が公開しているデータセット．lyft level5 が 2 年間にわたって収集してきた．約 13 万枚のフレームごとの画像があるが, テキストデータもありファイルの形式は json．シーンだけでなく車や人といったカテゴリやその状態, その物体のどれだけの部分が写っているかという意味での可視率などもテキストファイルとして扱われている. 画像の例として以下をあげる．

4　先週までの作業

- PredNet について
-

5　今週の作業

- 授業課題
- ゼミの予習

6　来週以降の作業

- lyft level5 でサンプルコードを動かせるようにする．

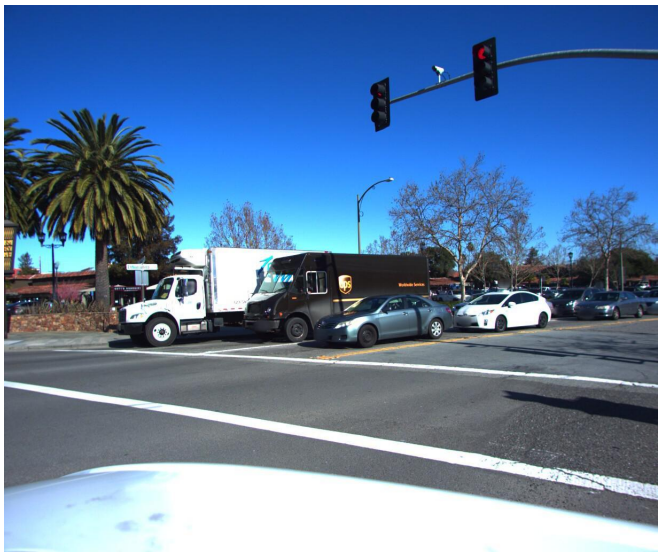

Fig. 4: lyft の写真の例

$$A_l^t = \begin{cases} x_t & \text{if } l = 0 \\ \text{MaxPool}(\text{ReLU}(\text{Conv}(E_{l-1}^t))) & l > 0 \end{cases} \quad (1)$$

$$\hat{A}_l^t = \text{ReLU}(\text{Conv}(R_l^t)) \quad (2)$$

$$E_l^t = [\text{ReLU}(A_l^t - \hat{A}_l^t); \text{ReLU}(\hat{A}_l^t - A_l^t)] \quad (3)$$

$$R_l^t = \text{ConvLSTM}(E_l^{t-1}, R_l^{t-1}, R_{l+1}^t) \quad (4)$$

Fig. 3: PredNet の更新式