

# デジタルツイン構築のための 脳の認知機構を用いた オブジェクト認識手法の 実装および評価

大阪大学基礎工学部情報科学科 村田研究室  
久保 快斗

CQ研究会 (3/10)

0

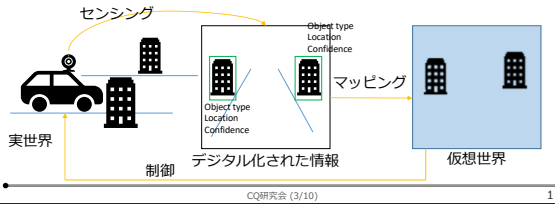
## 研究背景

### ・高まるデジタルツインへの期待

- ・ 実世界のセンシング結果を仮想世界にマッピングして表現
- ・ 仮想世界上のシミュレーション結果を実世界にフィードバック
- ・ 物流管理や自動運転技術などへの応用が期待されている

### ・リアルタイム性と仮想世界上の情報の精度が重要

- ・ 自動制御などでは移り変わる現実情報に即座に対処することが必須
- ・ センシングしたデータには一般に誤差が含まれており、そこから可能な限り正確な情報を取り出すことが必須

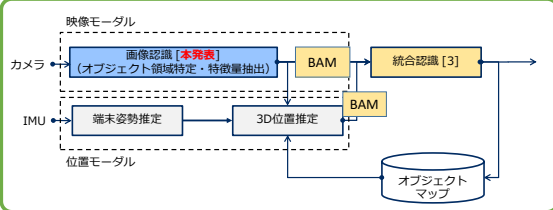


CQ研究会 (3/10)

1

1

## オブジェクト認識システムの全体構成



- ・ **先行研究**：人の脳の情報処理に着想を得た物体認識手法<sup>[3]</sup>
  - ・ 不確実な観測結果を用いて意思決定を行う仕組みを応用
  - ・ Bayesian Attractor Model (BAM) を用いたユニモーダル認識
- ・ **本研究**：先行研究の統合認識手法に適した形式の特徴量をカメラ映像から抽出する方式を提案・実装

3. 間良我, 小南大樹, 下西英之, 村田正幸, 藤若雅也, 野上耕介, "脳のマルチモーダルな情報処理に着想を得た物体推定手法の提案と評価," 電子情報通信学会技術研究報告 (CQ2021-14) 2024年5月号 (3/10)

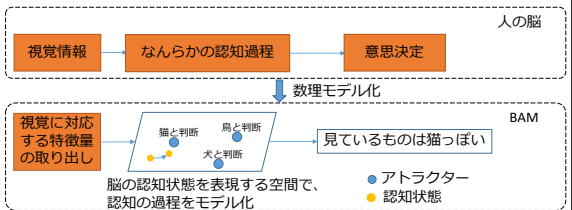
2

2

## 先行研究での認識手法<sup>[3]</sup>

### ・ Bayesian Attractor Model を用いた物体認識

- ・ 観測したものが記憶にある選択肢(アトラクター)のいずれであるのかを脳が判断する認識過程をモデル化したもの



- ・ [3] の認識手法では、あらかじめ識別する対象となる物体の特徴量を記憶しておき、観測した物体が記憶したもののうちどれに該当するのかを確信度とともに出力する
- ・ 特徴量の取り出し方は十分検討していなかった

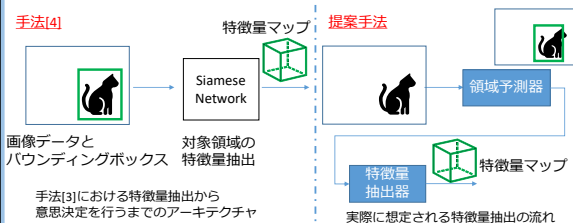
CQ研究会 (3/10)

3

3

## 手法<sup>[3]</sup>での特徴量の取り出し手法

- ・ **オブジェクト認識を行うための観測特徴量を取得する状況として、理想的な状況を想定していた**
  - ・ オブジェクトを予測したバウンディングボックスから特徴量を抽出
  - ・ 動画中の認識対象物体の検出が**成功したとき**のバウンディングボックス内の画像から、Siamese Network における軽量CNNを用いて特徴量を取得しBAMへ入力
  - ・ 実際にはバウンディングボックスは**誤差を含んで推定**される



手法[3]における特徴量抽出から意思決定を行うまでのアーキテクチャ

CQ研究会 (3/10)

4

4

## 研究目的とアプローチ

### ・研究目的

- ・ BAM の処理に適した形式での特徴量抽出アーキテクチャを実装
  - ・ 領域推定と特徴量抽出の両方が必要
- ・ 実際の特徴量を BAM に入力したときの精度や計算時間を評価

### ・アプローチ

- ・ Siamese RPN (Region Proposal Network) を使用
  - ・ 「Siamese Network + RPN」アーキテクチャにより認識対象物体の存在する領域位置を推定し特徴量抽出
- ・ 取り出した特徴量が有効か BAM に入力し評価

CQ研究会 (3/10)

5

5

## 構築するアーキテクチャ

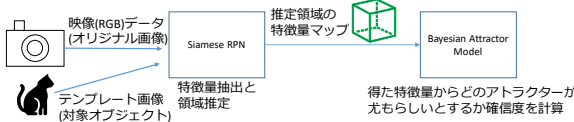
- Siamese RPN と BAM を用いたユニモーダル認識アーキテクチャ

### 1. 学習モデルの作成

- Siamese RPN
  - 特徴量抽出手法と領域推定手法の学習

### 2. 学習モデルを用いた特徴量抽出

- Siamese RPN
  - 映像データとテンプレートを受け取り領域推定したのち特徴量抽出
- Bayesian Attractor Model
  - 取得特徴量とこれまでの推論結果でアトラクターに対する確信度計算



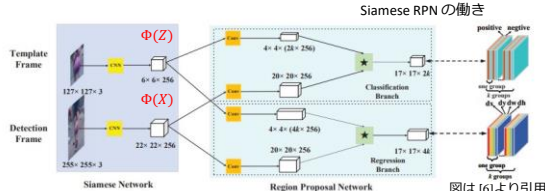
CQ研究会 (3/10)

6

6

## Siamese RPN による領域推定・特徴量抽出

- テンプレート画像特徴量マップ  $\Phi(Z)$  と オリジナル画像特徴マップ  $\Phi(X)$  を用いて RPN が領域を推定
- 推定領域の映像特徴量  $\Phi(X)$  から取り出して物体の特徴量を抽出



図は [6] より引用

6. L. Bo, Y. Junjie, W. Wei, Z. Zheng, and H. XiaoLin, "High performance visual tracking with siamese region proposal network," in Proc. CVPR 2018.

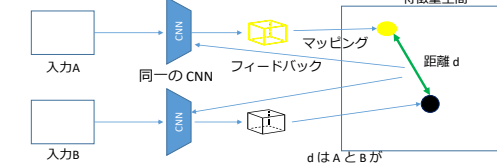
CQ研究会 (3/10)

7

7

## Siamese Network における学習 ([9]使用)

- 入力 A と 入力 B から同じ CNN で特徴量マップを取得
- 特徴量空間上にマッピングした時の距離  $d$  が
  - A と B が同じクラスであれば近くなるように CNN をチューニング
  - A と B が違うクラスであれば遠くなるように CNN をチューニング
- CNN 構造として 3 種類の構造を実装した
  - 3 層、4 層の CNN / 特徴量の次元が小さく、計算時間小
  - AlexNet / 特徴量の次元が大きく、計算時間大



[9] "Youtube-boundingboxes dataset," available at <https://research.google.com/youtube-bb>

CQ研究会 (3/10)

8

8

## RPN での学習 ([9]使用)

- 領域対象選定基準となるアンカー、アンカーボックス設定
  - $\Phi(X)$  の  $(x-2)$  と  $(y-2)$  の積の数だけアンカーが画像領域中に生成
  - アンカーを基準とした SCALES \* RATIOS サイズのアンカーボックスが ANCHOR\_NUM だけ設定
- $\Phi(X)$  と  $\Phi(Z)$  の畳み込みによる物体分類、領域推定
  - Classification ブランチ
    - 各アンカーボックスが物体か背景を囲っているかの確認
    - IoU 0.7 以上で物体、0.3 以下で背景
      - IoU:  $(B_{ans} \cap B_{box}) / (B_{ans} \cup B_{box})$
      - 正解 ( $B_{ans}$ )、アンカーボックス ( $B_{box}$ )
  - Regression ブランチ
    - IoU  $\geq 0.7 \vee$  IoU  $\leq 0.3$  のアンカーボックスの正解とのずれを回帰
    - 0.3 < IoU < 0.7 のアンカーボックスは学習に用いない



設定変数	使用目的
RATIOS	縦横比
SCALES	縦の長さ
ANCHOR_NUM	アンカーボックス数

CQ研究会 (3/10)

9

9

## BAM での推定

- 時刻  $t$  における意思決定状態  $z_t$  を保持し、観測特徴量  $x_t$  を受け状態  $z_t$  を更新

- $z_t - z_{t-\Delta t} = \Delta t f(z_{t-\Delta t}) + \sqrt{\Delta t} w_t$ 
  - $f$  は勝者総取りのホップフィールドダイナミクス
  - 意思決定状態の選択肢  $i$  に対応するアトラクター  $\phi_i$  が  $z$  空間に設定
  - $w_t$  はノイズ項
- $x_t = M\sigma(z) + v_t$ 
  - $\sigma$  は値域を 0~1 にするシグモイド関数
  - $M$  は各選択肢のノイズなしの特徴量を並べたもの
  - $v_t$  はノイズ項

- $i, ii$  を逆に推定し事後推定確率  $P(z_t | x_t)$  を計算

- $P(z_t = \phi_i | X_{0:t}) \geq \lambda$  を満たす  $\phi_i$  の対応選択肢を意思決定

- これまでの観測値から、意思決定の確信度を計算
- $X_{0:t}$  は時刻 0 から  $t$  までの観測値集合
- $\lambda$  は閾値

CQ研究会 (3/10)

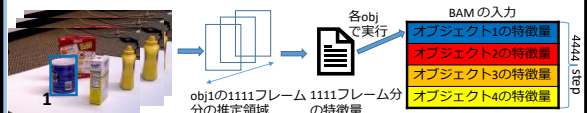
10

10

## 性能評価

- 先行研究で使用していた公開データセット<sup>[12]</sup>を用いた評価 (4 オブジェクトが写る 1111 フレームの連続した画像)

- Siamese RPN を用いた領域推定と特徴量抽出
  - 特定オブジェクトに対する推定を 1111 フレーム実行
    - 各オブジェクトに対して実行
  - IoU (Intersection over Union) による推定精度の評価
- BAM は推定領域の抽出特徴量を受け取り確信度計算
  - タイムステップを 4444 として設定
  - ステップ 1, 1112, 2223, 3334 においてそれぞれオブジェクト 1, 2, 3, 4 を認識し始めるものとして 1 で得た推定領域からの特徴量を入力とする
  - 各アトラクターについての確信度を計算し、その中で最大のものと対応している選択肢を意思決定結果として選定



[12] "YCB benchmarks-object and model set," available at <http://www.ycbenchmarks.com/>

CQ研究会 (3/10)

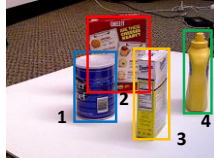
11

11

## アーキテクチャ構成要素の事前設定

### Siamese RPN

- Siamese Network の特徴量抽出のための CNN
  - 3層[24,32,16]で構成
  - 4層[16,16,16,16]で構成
  - AlexNet<sup>[10]</sup>で構成
- アンカーボックスの設定
  - RATIOS[0.33, 0.5, 1, 2, 3]
  - SCALES[2, 4, 8, 16, 64]
- 与えるテンプレート画像
  - 612フレーム目の正解領域から切り出した各オブジェクト



テンプレートとして用いる  
612フレーム目の画像

### BAM

- 記憶させるアトラクター
  - 各オブジェクトで、IoUが一番大きいフレームから取り出した特徴量
- ノイズ項
  - $w_t \sim N(0, \frac{\sigma^2}{\Delta t})$ ,  $v_t \sim N(0, r^2 I)$ , において  $q = 1.5$ ,  $r = 3.0$

10. K. Alex, S. Ilya, and E. H. Geoffrey, "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, 2012.  
CQ研究会 (3/10) 12

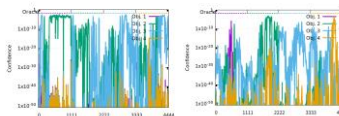
## 評価結果：軽量 CNN を用いた場合

3層 CNN [24,32,16]

オブジェクト	1	2	3	4
平均 IoU	0.484	0.452	0.241	0.178
RPN の平均計算時間 (s/f)	0.0293	0.0297	0.0304	0.0294

4層 CNN [16, 16, 16, 16]

オブジェクト	1	2	3	4
平均 IoU	0.279	0.502	0.196	0.131
RPN の平均計算時間 (s/f)	0.0243	0.0238	0.0233	0.0236



BAM の推定結果

	精度[%]	計算時間[f/s]
3層	32.6	0.0539
4層	51.4	0.0393

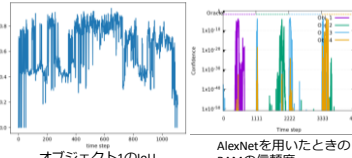
3層構造における BAM の信頼度 4層構造における BAM の信頼度

CQ研究会 (3/10)

## 評価結果：AlexNet を用いた場合

オブジェクト	1	2	3	4
平均 IoU	0.626	0.461	0.306	0.382
RPN の平均計算時間 (s/f)	0.0509	0.0488	0.0484	0.0490

- ほとんどの場合において、軽量 CNN の IoU を上回る
  - 一方で計算時間は4層 CNN を用いた際の2倍程度
- どの時間においても比較して、高い IoU が得られている



オブジェクト1のIoU AlexNetを用いたときの BAMの信頼度

- BAM における認識精度 54.36%
- BAM における計算時間 0.0657s/f

CQ研究会 (3/10)

## まとめと今後の課題

### まとめ

- デジタルツインの実現を目指し、脳の情報処理にならう物体認識手法を実装
- BAM への入力として与えるための特徴量抽出器を実装
  - 領域推定と特徴量抽出を同時に実施
- Siamese RPN の精度は特徴量マップの取り出し方に依存
  - IoU の精度が高いことが必要で、特徴量次元の大きな AlexNet が有効
    - 領域内に含まれる物体の特徴を抽出する可能性が高くなる
- BAM のアトラクター設計が認識精度にとって重要

### 今後の課題

- マルチモーダル認識への拡張
  - 映像モーダルだけで得られる認識精度には限界がある
  - 映像モーダルのみならず他のモーダルの認識による精度向上を図る

CQ研究会 (3/10)