

知能情報論 画像認識とは

2016年4月27日

東京大学 大学院情報理工学系研究科

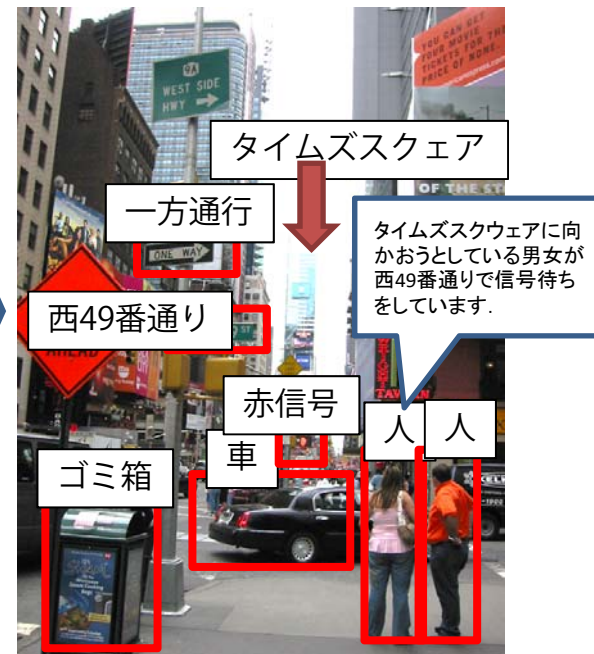
原田達也

Schedule

- 4/13: Introduction
- 4/20: Short history
- 4/27: What's image recognition? Detectors and descriptors
- 5/04: Golden week holidays!
- 5/11: Statistical feature extraction
- 5/18: Image coding and pooling
- 5/25: Classifiers
- 6/01: 講義は実施しない
- 6/08: Deep learning and convolutional neural networks
- 6/15: Object detection
- 6/22: Caption generation (Prof. Ushiku)
- 6/29: Cancel (CVPR)
- 7/06: Image and language (Prof. Ushiku)
- 7/13: Instance recognition and search

画像認識とは、画像に写る内容を理解すること

出力結果



物体認識



1. neck brace
2. bullet train
3. potter's wheel
4. seat belt
5. barbell



1. mountain bike
2. hartebeest
3. yurt
4. bighorn
5. coho



1. mask
2. ski mask
3. jack-o'-lantern
4. jellyfish
5. teddy bear



1. volleyball
2. bittern
3. shower curtain
4. crane
5. suspension bridge



1. toilet seat
2. scanner
3. hard disc
4. scale
5. backpack



1. aircraft carrier
2. paddle
3. bullfrog
4. water ouzel
5. mantis

- 画像をシステムに入力し，画像に写る物体を理解し適切なラベルを付与する過程のこと

シーン認識

街中



浜辺

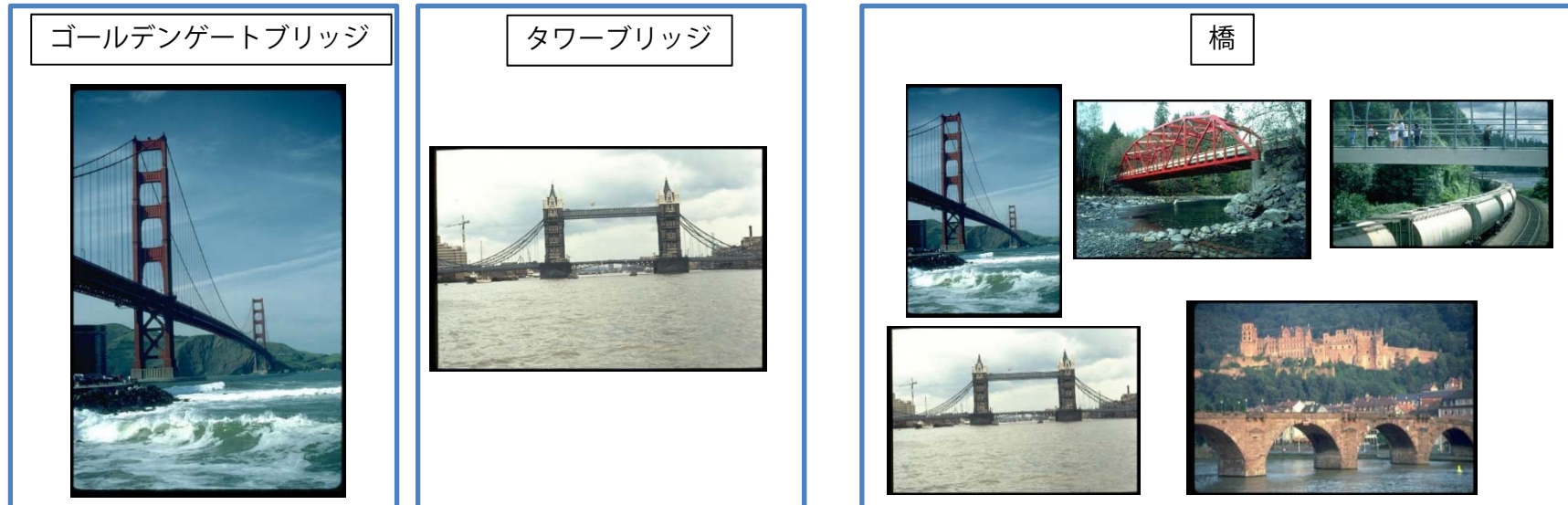


屋内



- シーンとは実世界の環境であり，複数の物体などが存在し何らかの意味を成す状況を表現したものと捉えることができ，画像が表現する包括的な状態を理解する過程

インスタンス認識とカテゴリ認識



(a) インスタンス認識

(b) カテゴリ認識

- インスタンス認識
 - ゴールデンゲートブリッジの画像を入力として, 「橋」と出力するのではなく, 「ゴールデンゲートブリッジ」として出力する過程
 - インスタンス認識は膨大なデータベースから内容もしくは見た目の近い画像を高速に探し出す類似画像検索と関連が深い
- カテゴリ認識
 - 入力画像内の物体の属する上位概念を予測する過程
 - 例えば, 橋には様々な形や色のものが存在するが, システムはどの橋を見ても「橋」として出力

インスタンス認識の難しさ



- 同じ建造物であっても撮影条件により見た目が大きく異なる

カテゴリ認識 の難しさ

例) 椅子



同じカテゴリ内の見た目が多様であり、単純な局所特徴のマッチングでは認識は難しい。

詳細物体認識



English setter



Siberian husky



Australian terrier



English springer



malamute



Great Dane



Walker hound



Welsh springer spaniel



whippet



Scottish deerhound



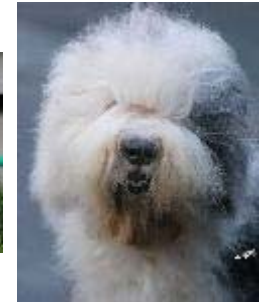
Weimaraner



soft-coated wheaten terrier



Dandie Dinmont



Old English sheepdog



otterhound



bloodhound



Airedale



giant schnauzer



black-and-tan coonhound



papillon



Staffordshire bullterrier



Mexican hairless



Bouvier des Flandres



miniature poodle



Cardigan



malinois

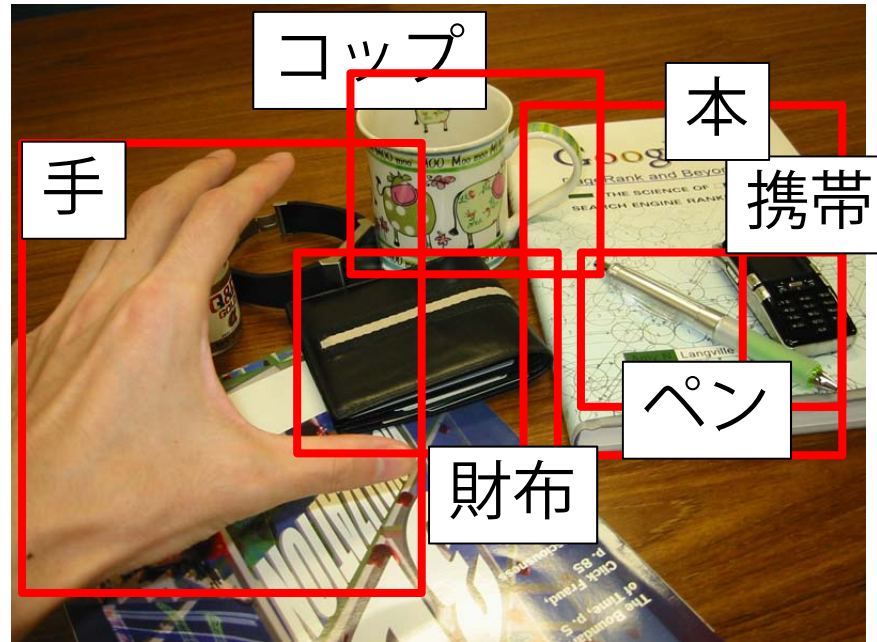
画像アノテーション



認識結果：
財布, コップ, 本, 携帯電話, ペン, 手

- 狭義としては複数ラベルが付与されたデータセットを用いて, 入力画像に複数のラベルを付与すること
- 広義としては物体認識, シーン認識を含む広い概念

物体検出



- 画像から物体のインスタンスやカテゴリだけを予測するのではなく、物体の画像における領域まで推定する課題

画像セグメンテーション



- 物体の詳細な領域
まで切り出す課題



http://image-net.org/challenges/talks/ILSVRC+MSCOCO_12_17_15_detection.pdf

キャプション生成



A giraffe standing next to a tree in a fence.



A yellow train on the tracks near a train station.



A dog laying on the side of a pool.

- 物体間の関係性を考慮し，理解する課題

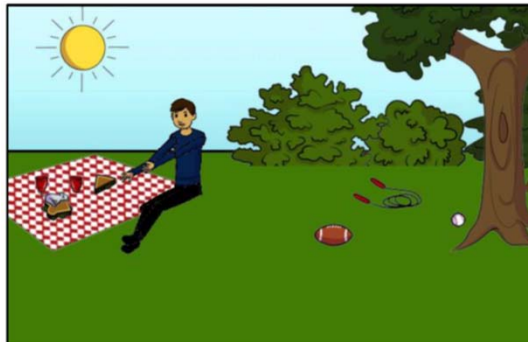
VQA



What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?



Does it appear to be rainy?
Does this person have 20/20 vision?

Aishwarya Agrawal, Jiasen Lu, Stanislaw Antol, Margaret Mitchell, C. Lawrence Zitnick, Dhruv Batra, Devi Parikh. VQA: Visual Question Answering. arXiv:1505.00468.

- Visual Question Answering (VQA)
- 自然文の画像に関する質問に対して、自然文で回答する.

動画像認識



ApplyEyeMakeup



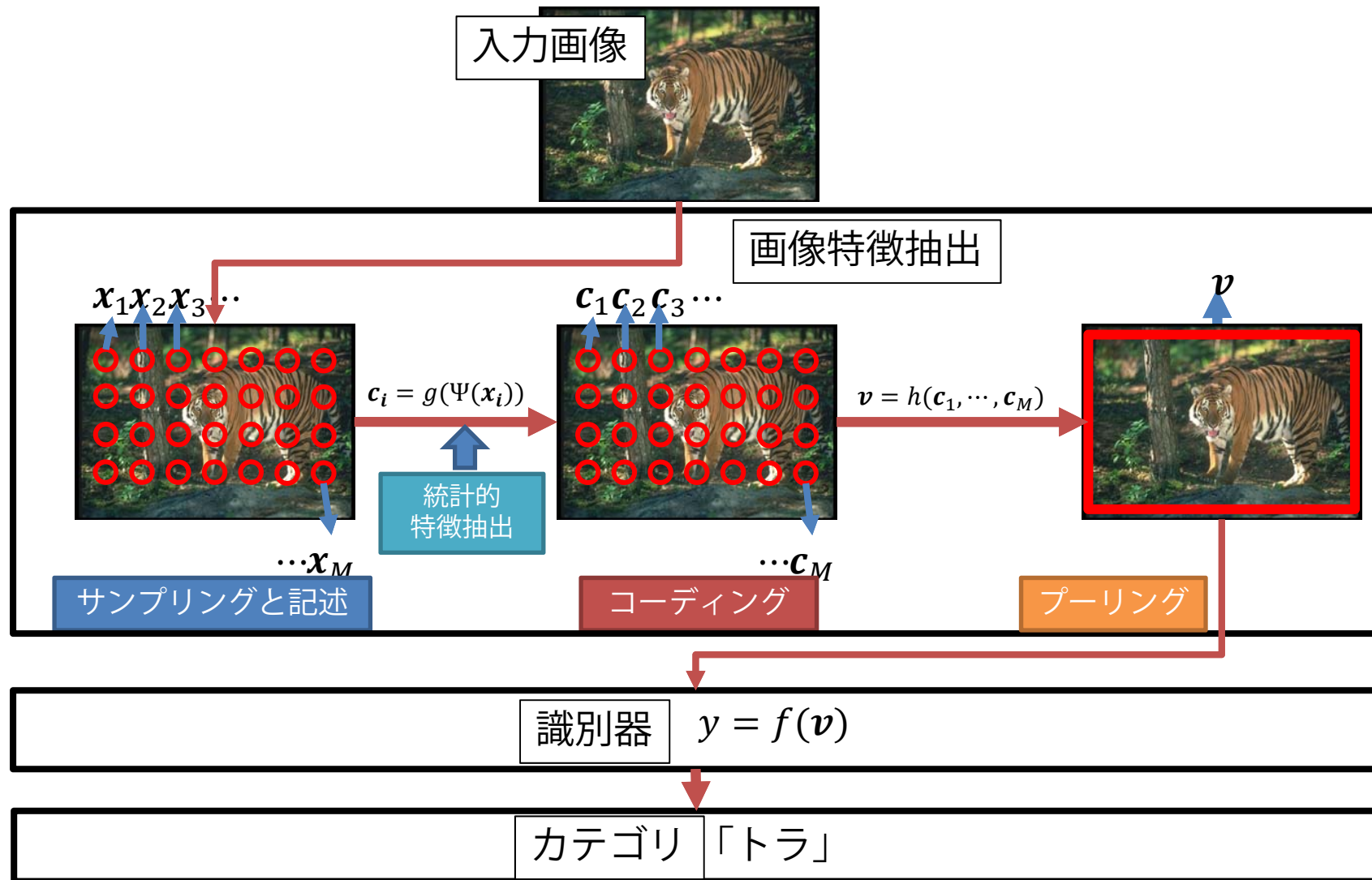
BabyCrawling



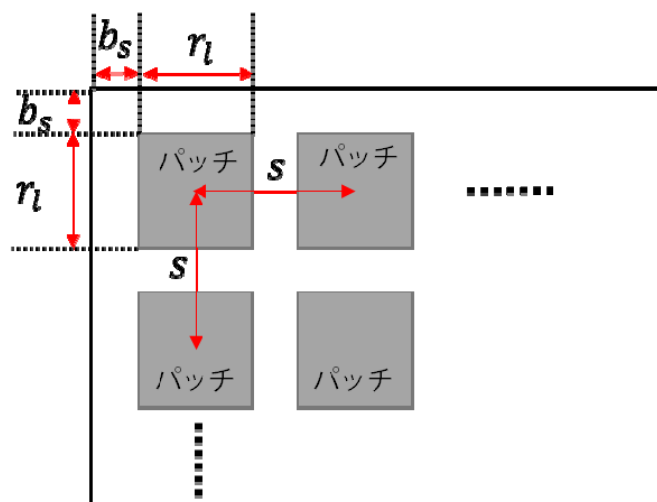
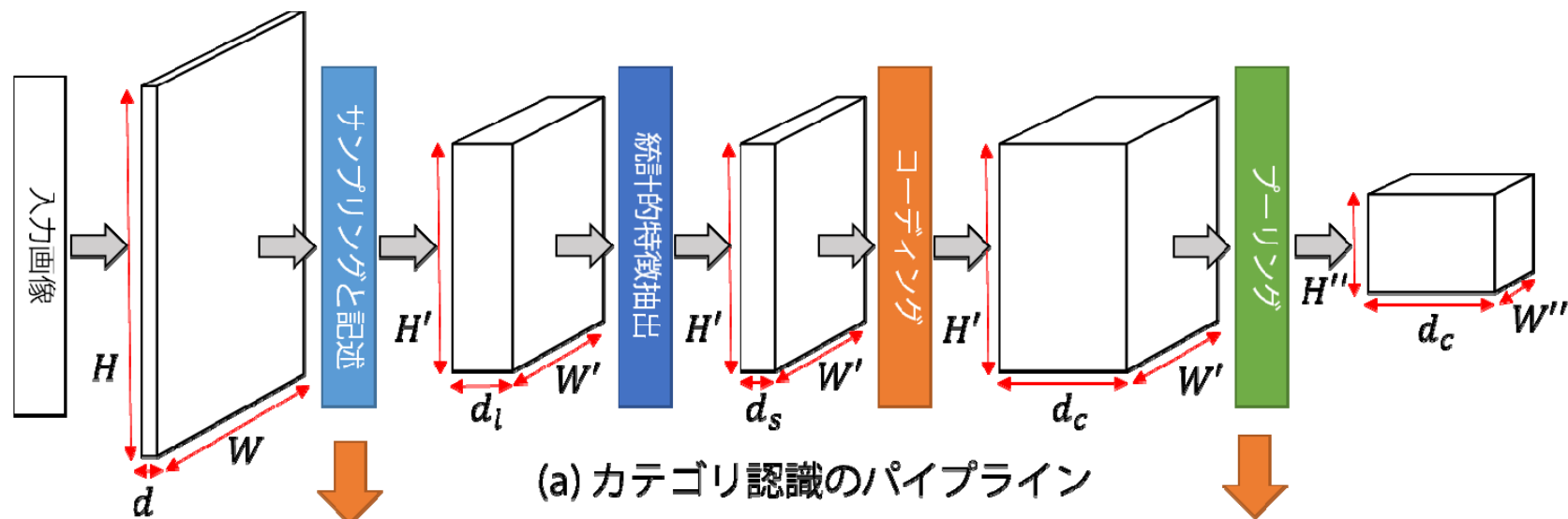
BasketBallDunk

- 「歩く」と「走る」動作の識別を考えます. これらの動作は静止画だけでは分類が困難
- 時系列情報が必要不可欠

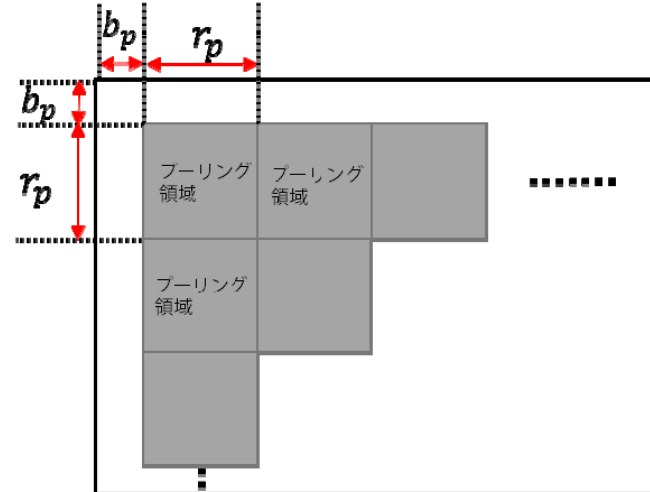
カテゴリ認識の流れ



カテゴリ認識のデータ構造

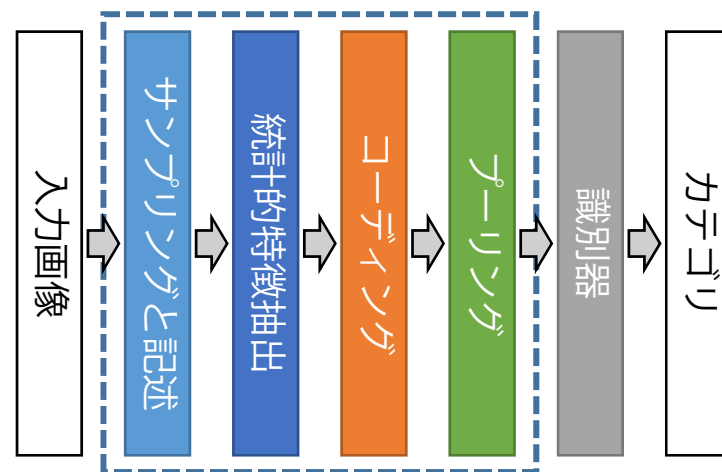


(b) 局所特徴のサンプリング

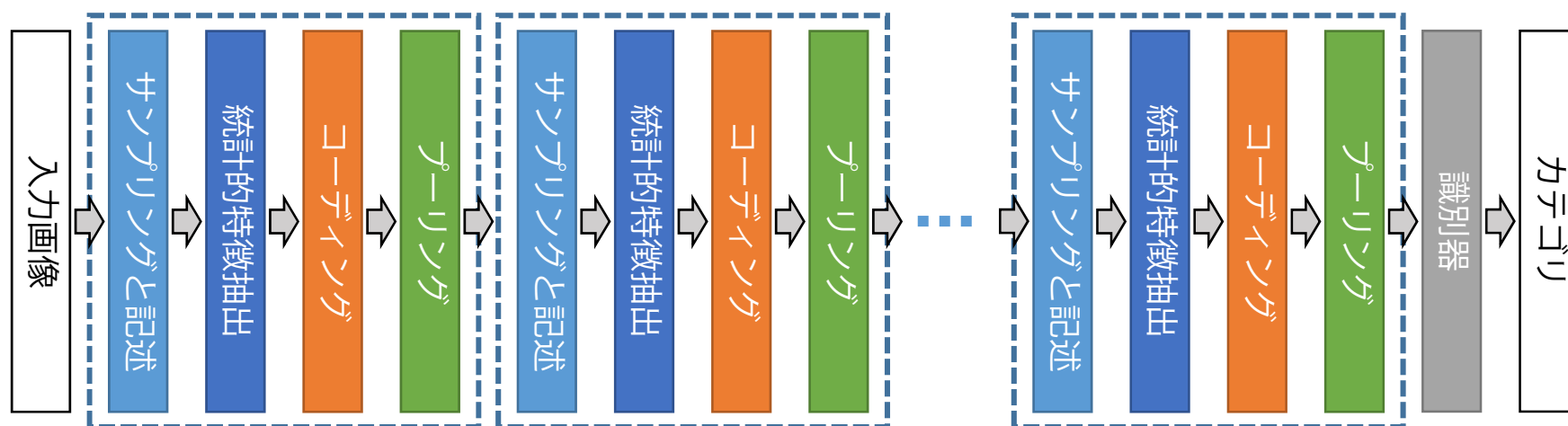


(c) プーリング

浅い構造と深い構造

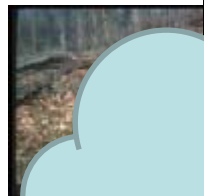


(a) 浅い構造

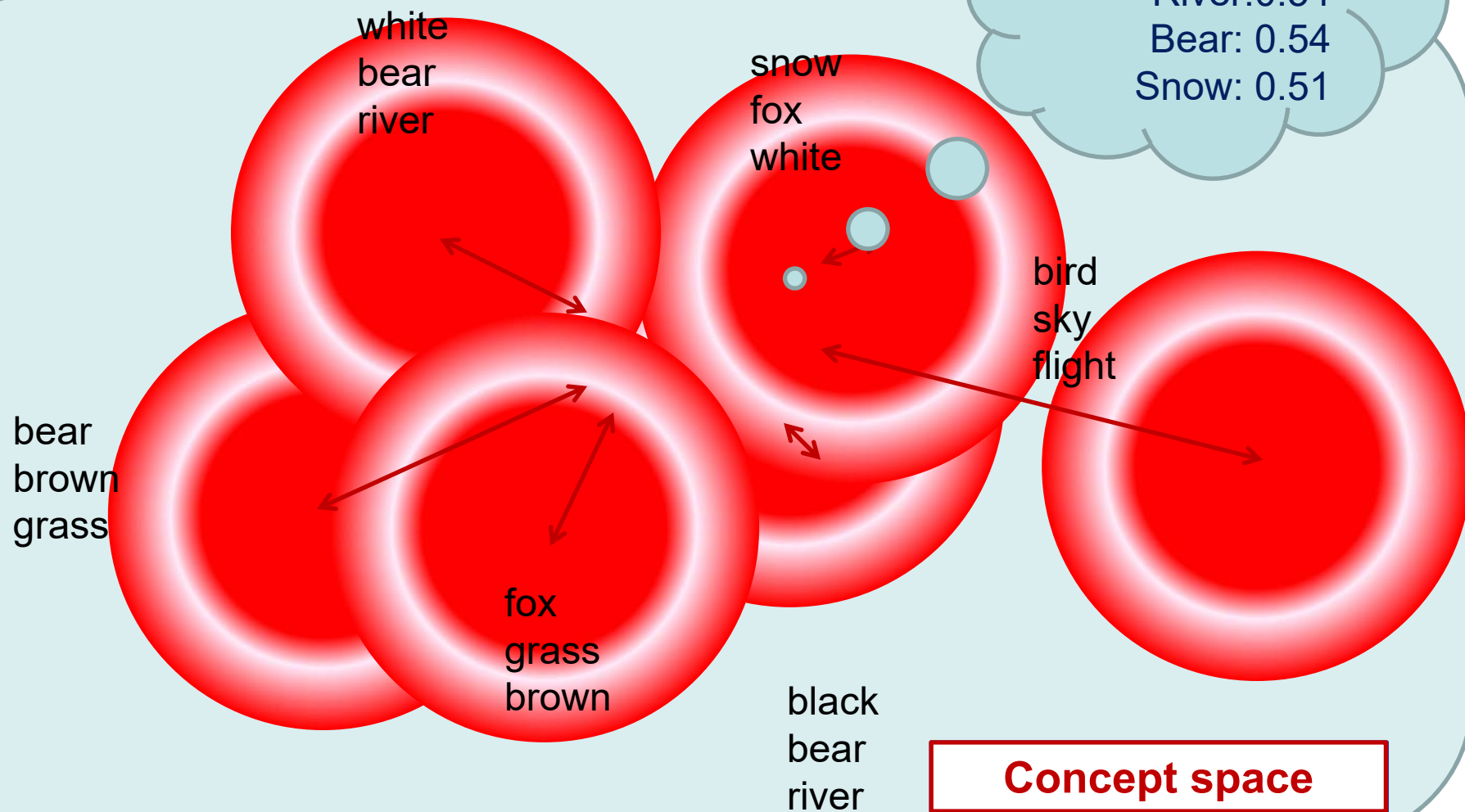


(b) 深い構造

画像アノテーション



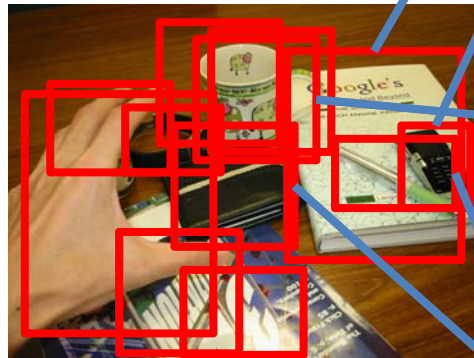
Fox: 0.90
White: 0.83
River: 0.54
Bear: 0.54
Snow: 0.51



物体検出



(a) 入力画像



カップらしさのスコア

カップ
識別器

s_1

カップ
識別器

s_2

カップ
識別器

s_i

カップ
識別器

s_{N_r-1}

カップ
識別器

s_{N_r}



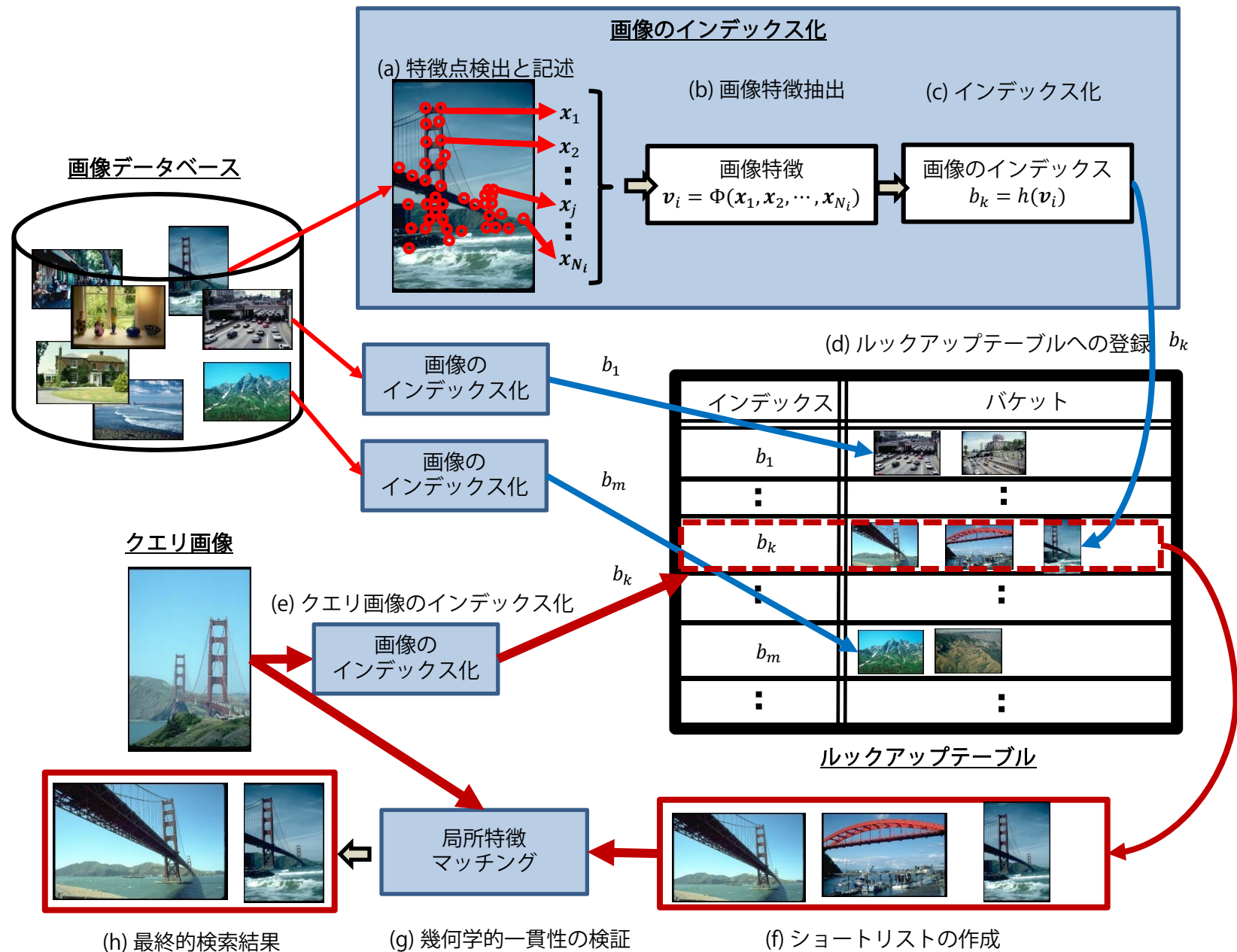
(e) 最終結果



$$s_j > \theta$$

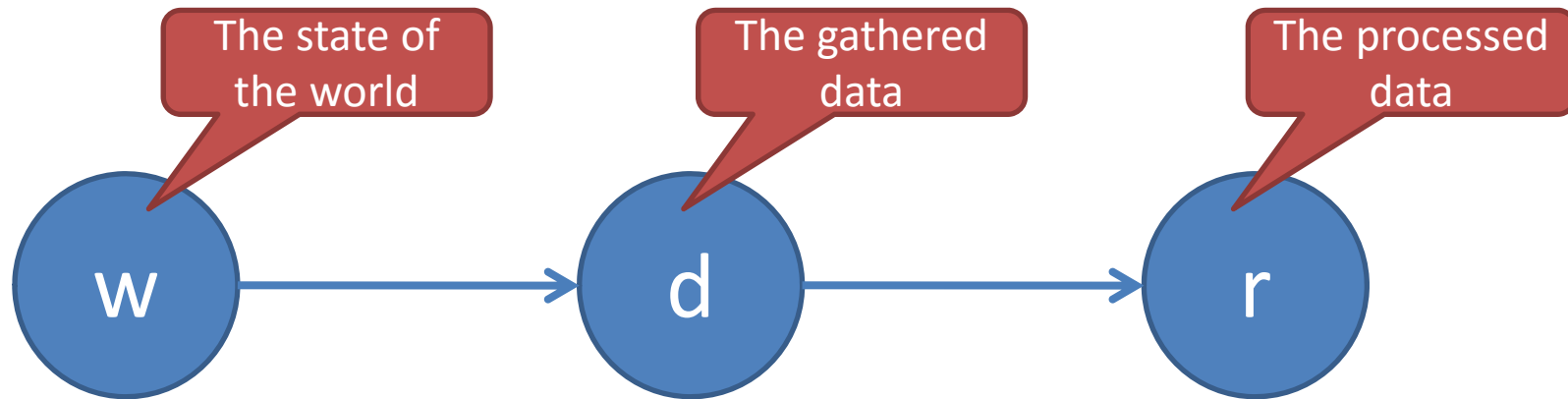
(b) 物体候補領域の抽出 (c) 物体候補領域の物体認識 (d) 非最大値出力の削除

インスタンス認識と検索



認識性能を向上させるには？

The data processing theorem



Markov chain

$$P(w, d, r) = P(w)P(d|w)P(r|d)$$

The average information

$$I(W; D) \geq I(W; R)$$

The data processing theorem states that data processing can only destroy information.

画像認識のプロセスと必要機能

訓練時



識別時



- 処理を重ねる毎にデータの持つ情報は減少
 - データ，特徴抽出，モデルの順に高い質が必要
 - データの大規模化，リッチな画像表現

Data matter!
Features matter!