

タカとハヤブサはどこが違う？

～新たな認識領域「詳細画像識別」の展開と応用～

東京大学 大学院情報理工学系研究科

創造情報学専攻 中山研究室

中山 英樹

クイズ1

- ▶ どちらがハヤブサでしょう？



<http://plaza.rakuten.co.jp>

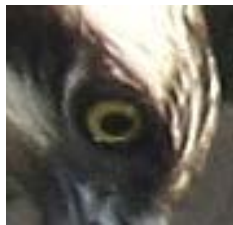


<http://birds.mints.ne.jp>

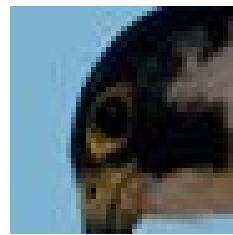
クイズ1

- ▶ どちらがハヤブサでしょう？

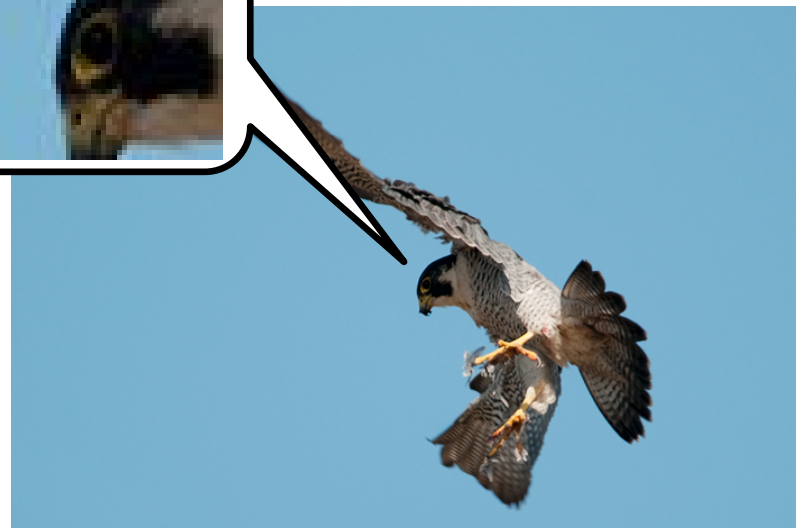
タカ



ハヤブサ



<http://plaza.rakuten.co.jp>



<http://birds.mints.ne.jp>

クイズ2

▶ 車種はなんでしょう？



クイズ2

▶ 車種はなんでしょう？



識別結果

Mitsubishi
Lancer

Acura RL

Toyota Camry



識別結果

Audi S4

Honda Accord

Mercedes-Benz C-Class

詳細画像識別

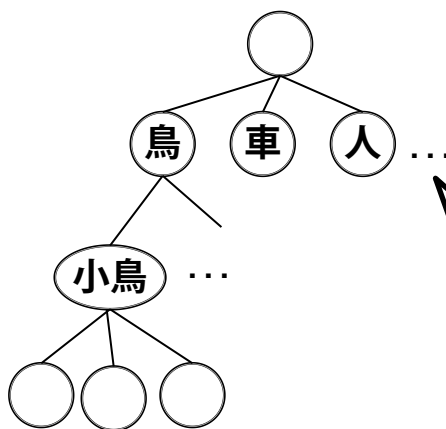
Fine-grained Visual Categorization (FGVC)
Subordinate-level Categorization

- ▶ **ある特定の対象領域における高粒度の多クラス画像識別**
 - 意味的・視覚的に非常に類似したクラスを扱う
 - ・ クラス間の違いは極めて細かく、局所的であることが多い
 - 一般的な人間を遥かに超えるレベルの詳細な識別能力を目指す

詳細画像識別



Pine Warbler Prairie Warbler Yellow Warbler



従来的一般物体認識



Car Flower Dog

何に使える？

▶ 研究分野でよく扱われる題材

○ 動植物の種類識別



○ 人の動作識別

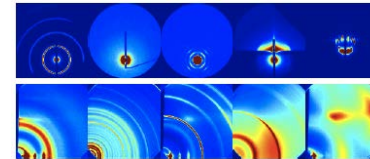


○ 顔の属性識別



▶ いろいろな応用可能性

- 官能検査, 目視検査
- 物性予測 (X線拡散)



[Kiapour et al.,
WACV'14]

○ 道路交通状況認識



[Stark et al.,
BMVC'12]

○ 商品画像認識・検索



○ 医療画像診断

具体的なアプリケーションに特化した対象に絞り、
人間以上の識別能力の実現を目指す！

アプリケーション例

Leaf snap : 植物画像識別



<http://leafsnap.com/>

Neeraj Kumar et al., Leafsnap: A Computer Vision System for Automatic Plant Species Identification, In Proc. ECCV, 2012.

Bird snap : 鳥画像識別

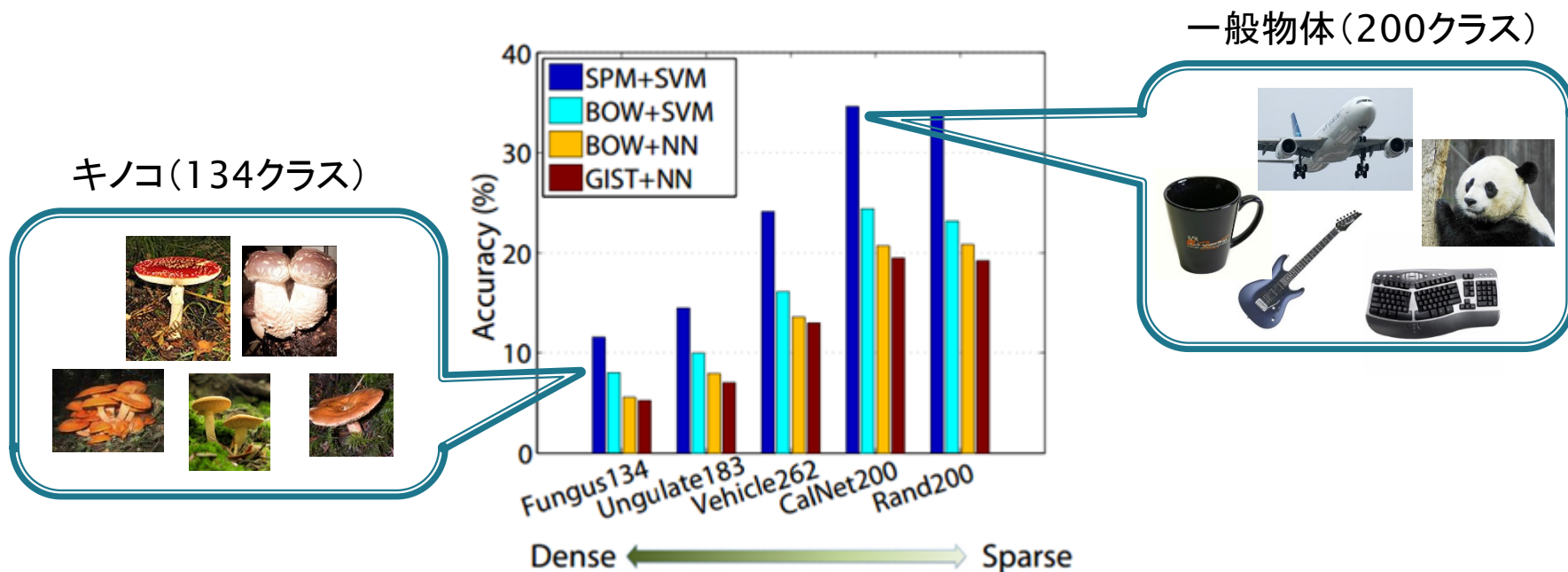


<http://birdsnap.com/>

Thomas Berg et al., Birdsnap: Large-scale Fine-grained Visual Categorization of Birds, In Proc. CVPR, 2014.

詳細画像識別のむずかしさ

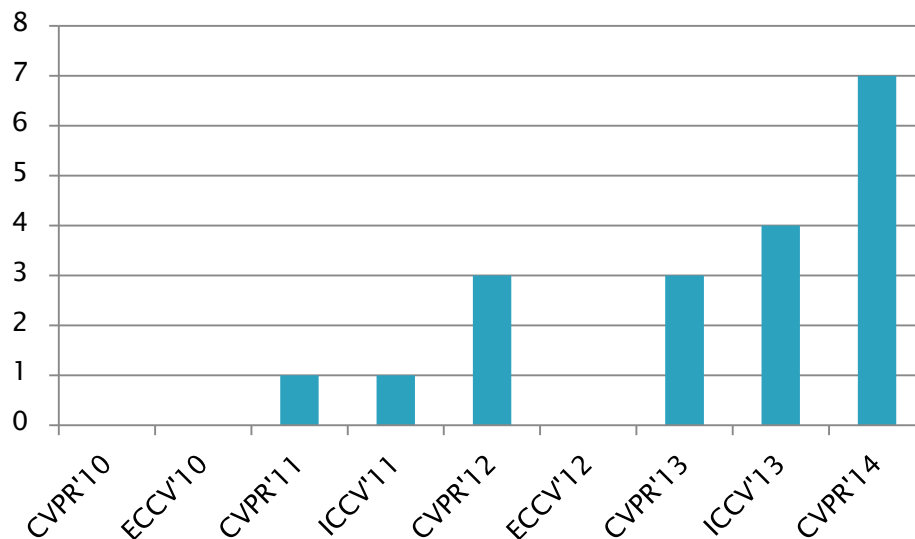
- ▶ 意味的・視覚的に類似したクラスは識別困難
 - 同じ程度のクラス数の識別問題でも、**一般物体認識に比べて著しく低い識別精度**しか得られない



詳細画像識別の歴史

- ▶ 大規模一般物体認識からスピンオフする形で誕生
- ▶ ECCV 2010: 詳細画像識別の難しさが指摘される [Deng et al.]
- ▶ CVPR 2011: 1st workshop on Fine-grained visual categorization
- ▶ CVPR 2013: 2nd workshop on Fine-grained visual categorization
- ▶ ICCV 2013: FGVC competition
 - ImageNet large-scale visual recognition challenge と同時開催

タイトルに
“Fine-grained”
が入った論文数



現在の研究動向

- ▶ **1. 完全自動の識別における精度向上**
 - 基本的には従来のパラダイムの延長
 - 対象が“物体”であることを前提とするものが多い

- ▶ **2. 識別過程の意味づけ・可視化**
 - 最終的には人間が納得できるかどうか重要
 - 人間とのハイブリッドシステム

詳細画像識別の手法

- ▶ 明確な方向性は必ずしも定まっていない
- ▶ 既存の大規模一般物体認識・検出技術を総動員

技術要素①：領域を記述する画像特徴量

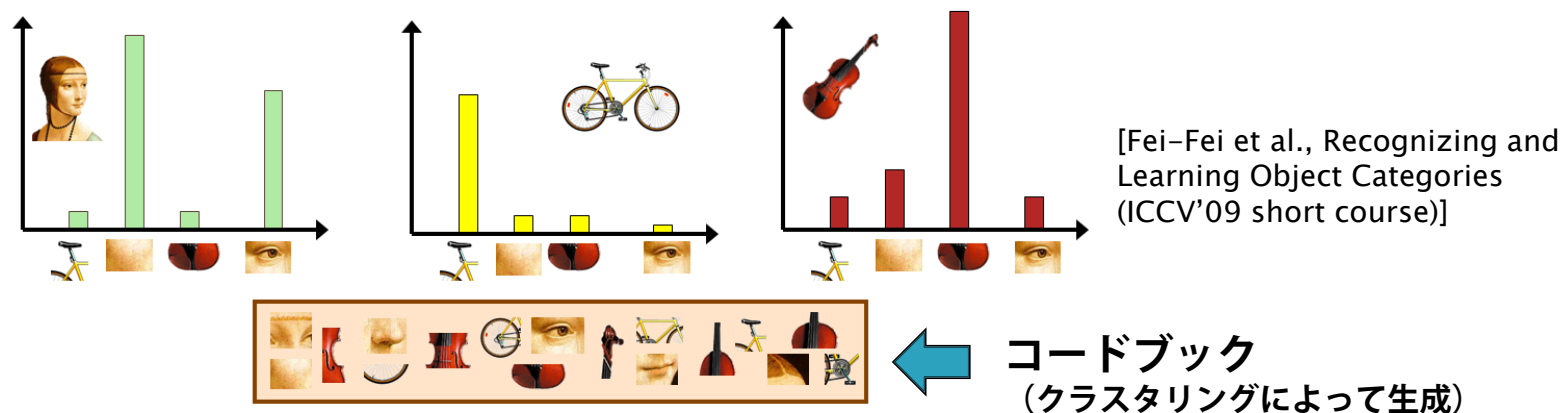
技術要素②：物体領域のセグメンテーション

技術要素③：部位モデルに基づく正規化

これらの組み合わせを軸として手法開発が進められている

技術要素①：画像特徴量 (1/3)

▶ Bag-of-Features が基本



- **Locality-constrained linear coding** [Wang et al., 2010], **Improved Fisher vector** [Perronnin et al., 2010] 等の最新的手法を利用
- 一般物体認識よりも**コードブックを大きくする必要がある**ことが指摘されている
 - 識別に有用なパッチはごく少数であり、大部分は似通っているため量子化誤差に埋もれやすい

J. Lin et al., Is Fine Grained Classification Different?, In Proc. CVPR workshop on FGVC, 2013.

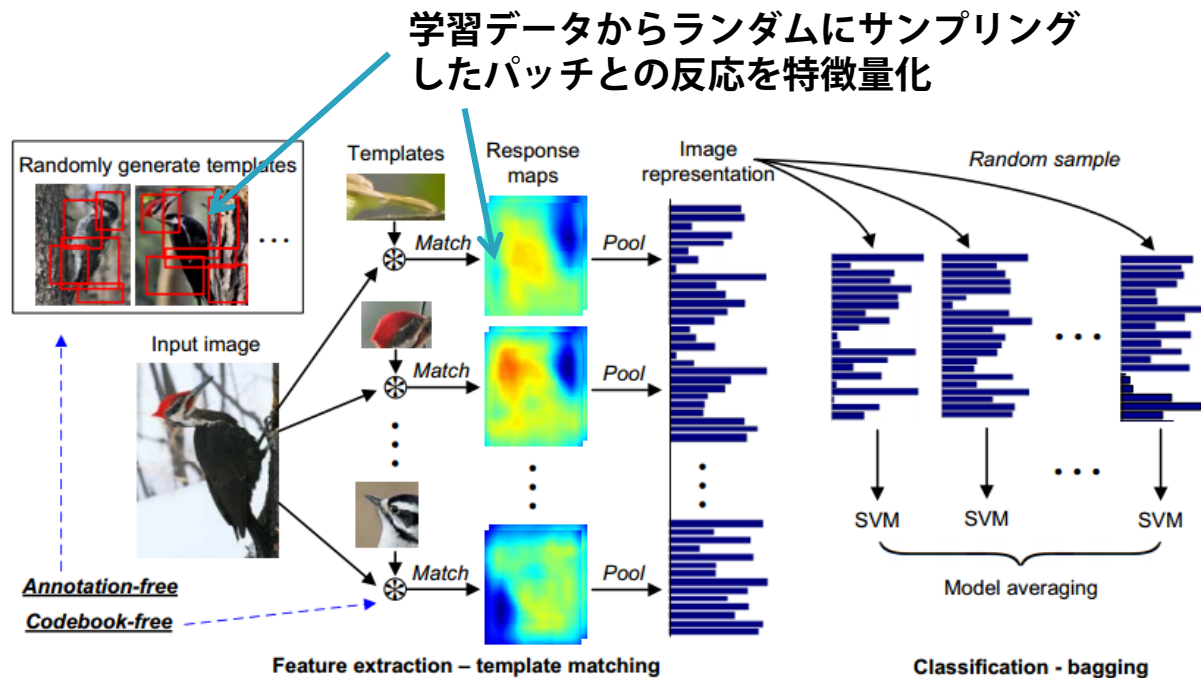
P.-H. Gosselin et al., Inria+Xerox@FGcomp: Boosting the Fisher vector for fine-grained classification, INRIA Tech. Rep. 2013.

技術要素①：画像特徴量 (2/3)

▶ テンプレートベース

- 量子化を行わず、生の画像パッチをコードブック的に用いる
- 量子化誤差の影響を緩和

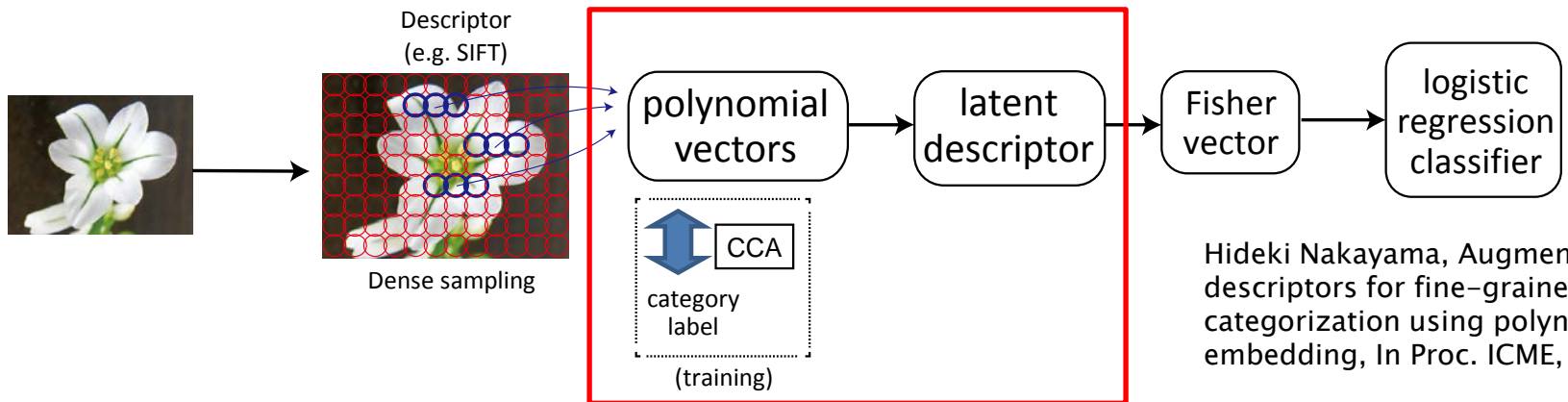
[B. Yao et al., CVPR'12], [S. Yang et al., NIPS'12]



Bangpeng Yao, Gary Bradski, Li Fei-Fei, A Codebook-Free and Annotation-Free Approach for Fine-Grained Image Categorization, In Proc. CVPR, 2012.

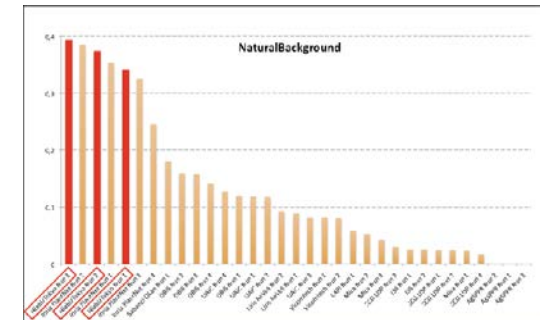
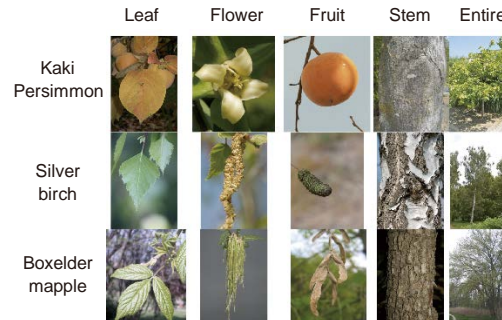
技術要素①：画像特徴量 (3/3)

- ▶ 局所特徴量の教師付次元圧縮 [中山, 2013]
 - 識別的な情報を局所特徴量レベルで抽出した後、bag-of-featuresへ利用
 - 隣接する局所特徴の共起関係を埋め込み
 - 一般的な方法 (PCAによる圧縮) から大きく識別精度向上



Hideki Nakayama, Augmenting descriptors for fine-grained visual categorization using polynomial embedding, In Proc. ICME, 2013.

ImageCLEF 2013の
植物画像識別タスクで第一位



技術要素②：セグメンテーション

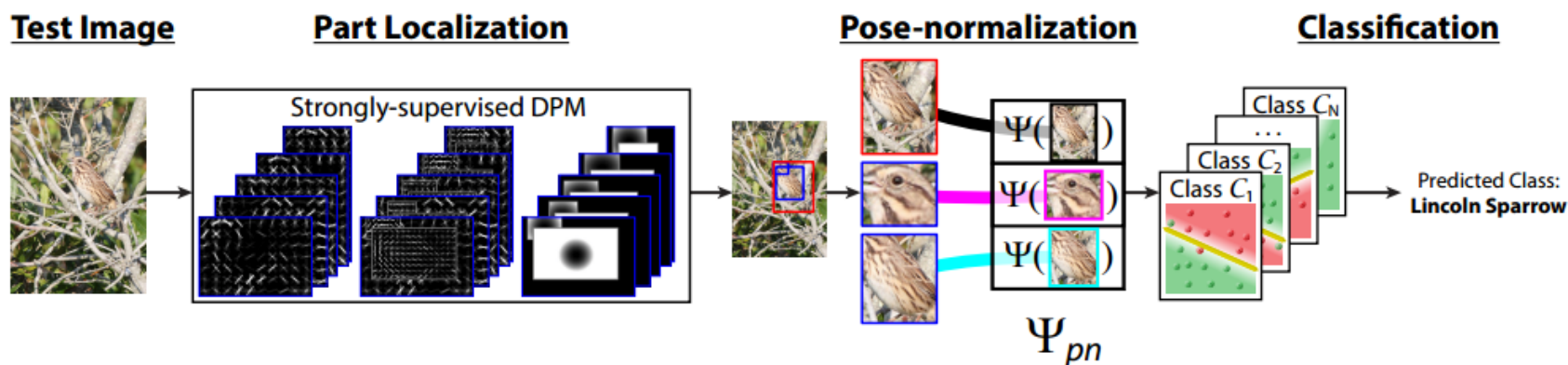
- ▶ 特徴抽出の前処理として、物体領域のみを抽出
 - GrabCut等の標準的な手法がよく用いられる
 - [A. Angelova et al., CVPR'13], [Chai et al., ECCV'12]



- (1) 画像中の各スーパーピクセルが物体領域に属するかを判定する識別器を学習
- (2) スーパーピクセルの識別結果を初期入力とし、label propagationにより物体全体の詳細なセグメンテーションを行う
- (3) 物体領域のみからbag-of-features (LLC) を求め、カテゴリ識別を行う

技術要素③：部位モデル (1/4)

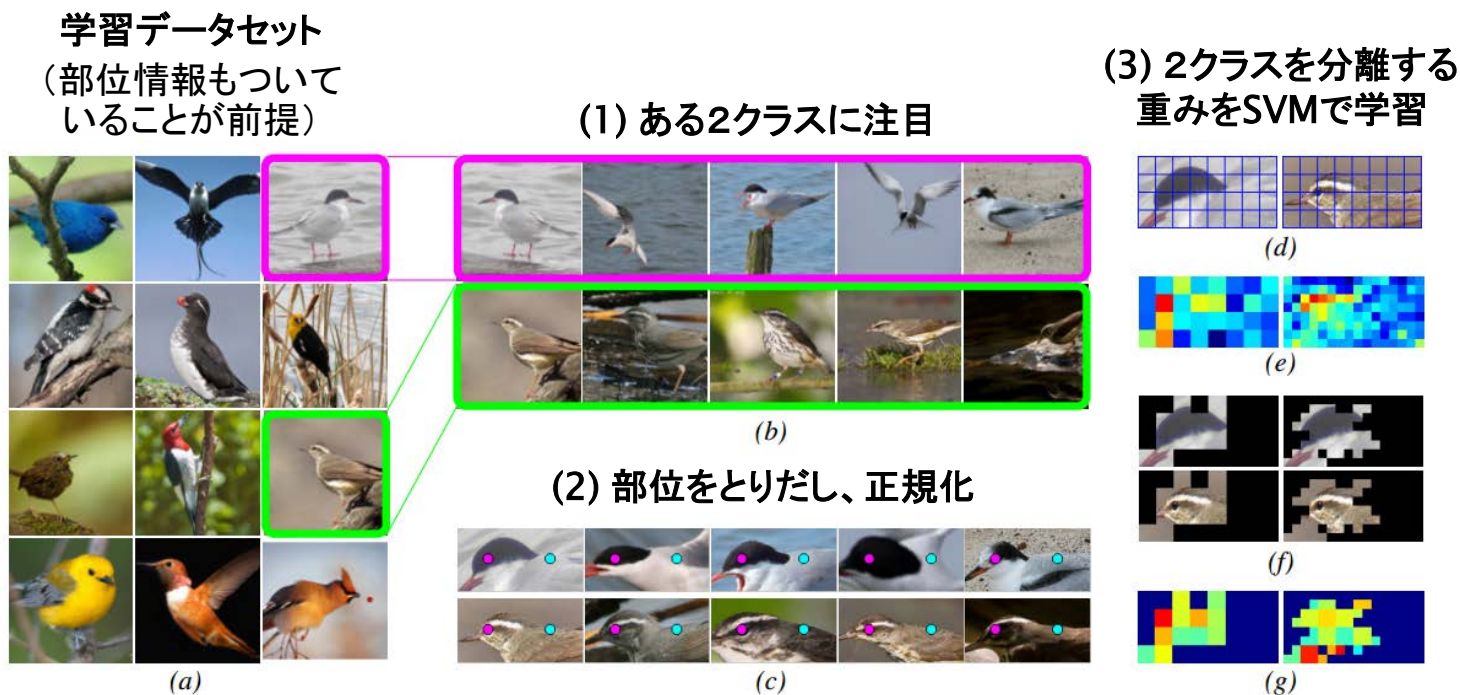
- ▶ 物体の各部位（頭、羽、etc.）を検出し、位置や姿勢の違いを吸収
 - セグメンテーションベースよりも頑健である場合が多い（対象による）
- ▶ 検出手法はDeformable Part Model [Felzenszwalb et al., 2008] が定番
- ▶ Deformable Part Descriptor [Zhang et al., ICCV'13]
 - DPMで検出された部位の領域内の特徴を抽出し、識別に利用



Ning Zhang, Ryan Farrell, Forrest Iandola, Trevor Darrell, Deformable Part Descriptors for Fine-grained Recognition and Attribute Prediction, In Proc. ICCV, 2013.

技術要素③：部位モデル (2/4)

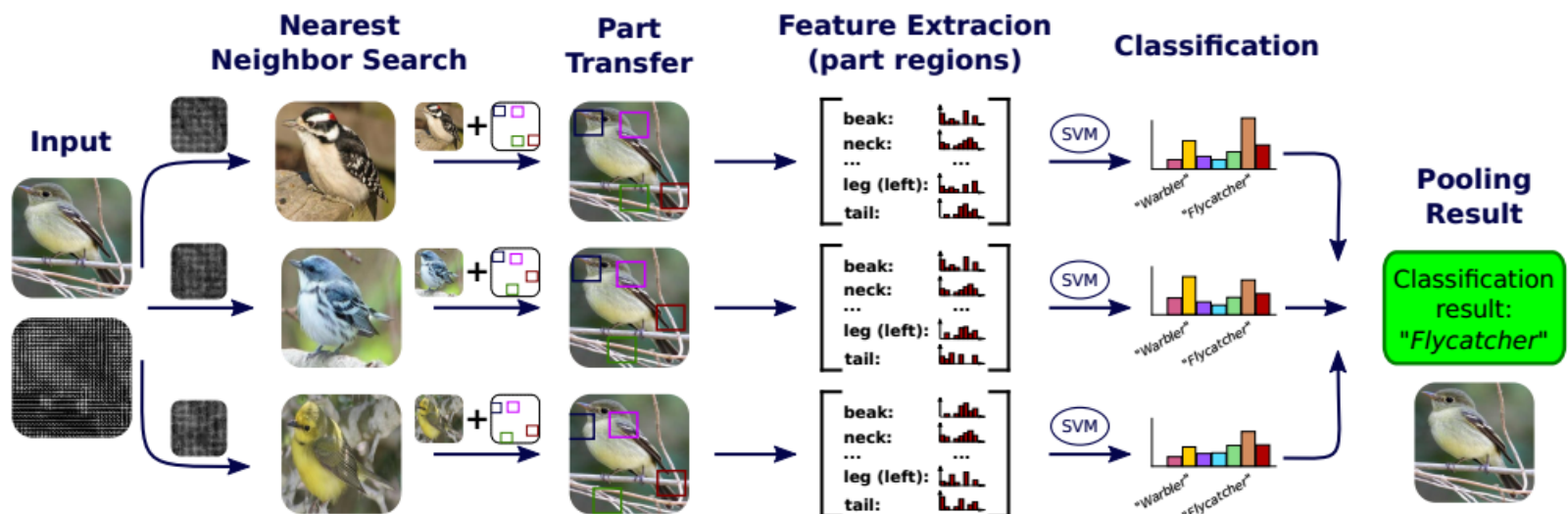
- ▶ **POOF: Part-Based One-vs-One Features** [Berg et al., CVPR'13]
 - 部位を正規化した上で、任意の2クラスをできるだけ分離する重みを学習
 - 最終的に、重みに応じた特徴量を各部位領域から抽出



Thomas Berg, Peter N. Belhumeur, POOF: Part-Based One-vs-One Features for Fine-Grained Categorization, Face Verification, and Attribute Estimation, In Proc. CVPR, 2013.

技術要素③：部位モデル (3/4)

- ▶ **Part transfer** [Göring et al., CVPR'14], [Gavves et al., ICCV'13]
 - 類似する学習サンプルの部位の位置をそのまま利用
 - 学習サンプルに部位情報がアノテーションされていることが前提
 - **Exemplar-SVM** [Malisiewicz, ICCV'11] の考え方に基づく



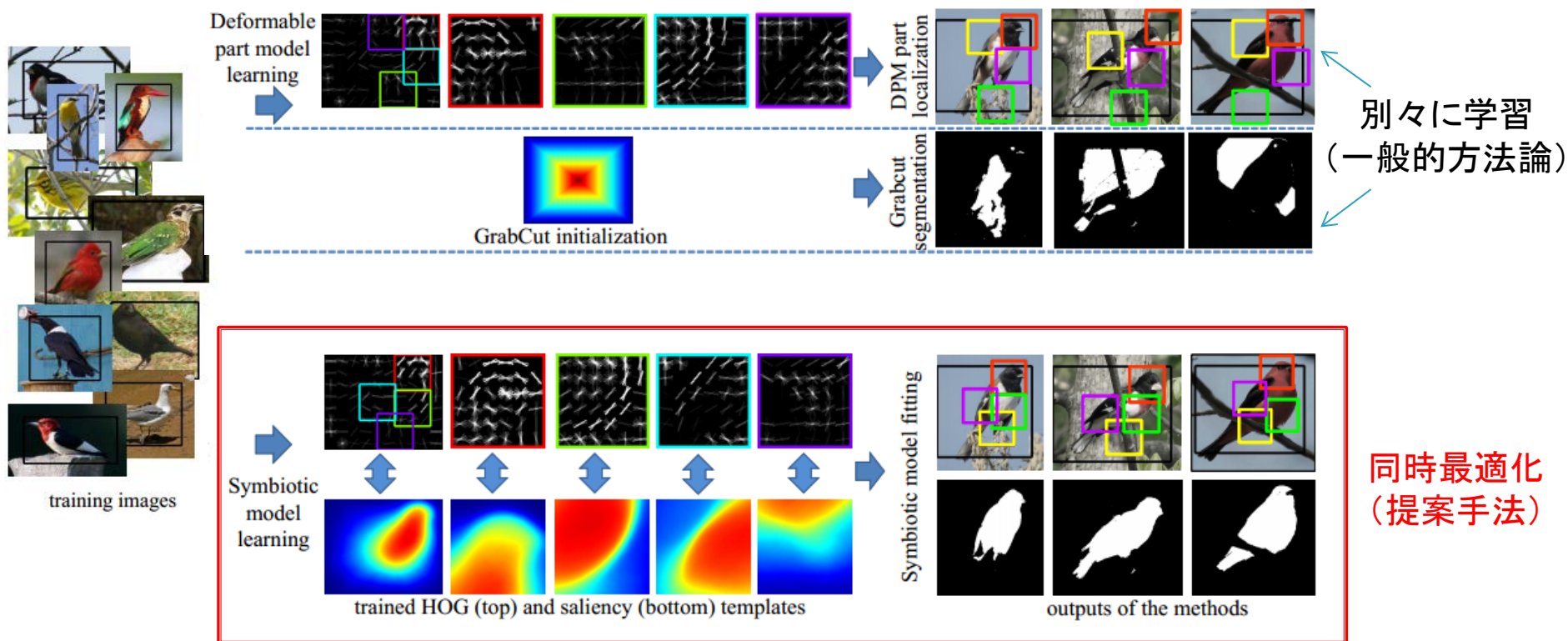
(1) HOG特徴で類似サンプルを検索

(2) 類似サンプルの部位の位置から特徴抽出

(3) それぞれ識別したあと結果を統合

技術要素③：部位モデル (4/4)

- ▶ セグメンテーション手法との同時最適化 [Chai et al., ICCV'13]
 - Saliencyモデルを介してDPMによる部位検出とGrabCutを統合



Fine-grained competition 2013

<https://sites.google.com/site/fgcomp2013/>



Fisher vector
をチューニング

1.

Team	Aircraft	Birds	Cars	Dogs	Shoes	Overall
Inria-Xerox	81.46	71.69	87.79	52.90	91.52	77.07
CafeNet	78.85	73.01	79.58	57.53	90.12	75.82
Inria-Xerox	75.88	66.29	84.70	50.42	88.62	73.18
VisionMetric	75.49	63.90	74.33	55.87	89.02	71.72
2. Symbiotic	75.85	69.06	81.03	44.89	87.33	71.63
Inria-Xerox	80.59	58.54	84.67	35.62	90.92	70.07
CognitiveVision	67.42	72.79	64.40	60.56	84.83	70.00
DPD_Berkeley	68.47	69.57	67.40	50.84	89.52	69.16
VisionMetric	73.93	51.35	69.31	38.63	87.33	64.11

Deep
learning

※網掛けのエント리는訓練セット外のデータを学習に使用 (ランキング対象外)

▶ 優勝チームはFisher vectorのチューニングのみに注力

- 詳細画像識別においても特徴量は最も大事 (ノウハウが必要)
- 識別精度を上げるだけなら、従来の枠組でもまだまだ工夫の余地がある

セグメンテーション
+
パートディテクション
+
Fisher vector

P.-H. Gosselin et al., Inria+Xerox@FGcomp: Boosting the Fisher vector for fine-grained classification, INRIA Research Reports, 2013.

現在の研究動向







- ▶ **1. 完全自動の識別における精度向上**
 - 基本的には従来のパラダイムの延長
 - 対象が“物体”であることを前提とするものが多い
- ▶ **2. 識別過程の意味づけ・可視化**
 - 最終的には人間が納得できるかどうか重要
 - 人間とのハイブリッドシステム

識別過程の意味づけ・可視化

- ▶ 詳細画像識別は非常に困難なタスク
 - 識別の根拠や過程が人間にとって納得できる形であることが重要
 - 完全自動の識別手法は基本的にブラックボックス
- ▶ キーワード：**Human in the loop**
 - 人間とのハイブリッドシステム
 - データ作成、学習時、識別時のさまざまなタイミングで人間の手助け(知見)をインタラクティブに取り入れる
 - 識別の根拠を“人間の言葉”で説明できるように

基本的な道具

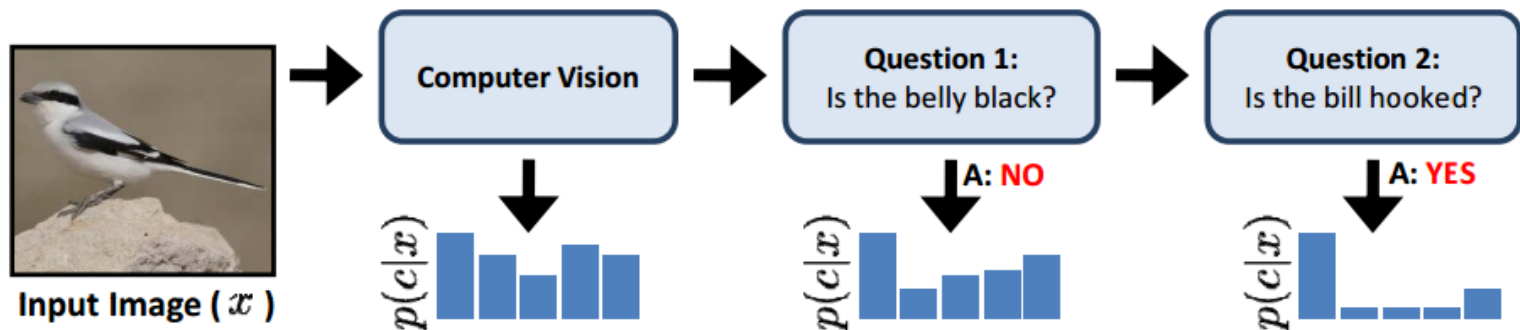
- ▶ **属性 (Attribute)** [Lampert et al., CVPR'09], [Farhadi et al., CVPR'09]
 - 異なるクラス間で潜在的に共有される (と思われる)
基本的な視覚プリミティブ
 - 人間にも理解できる中間表現として、識別結果の意味づけに利用可能
 - (例) “しましま” 属性識別器が強く反応した結果、最終的にシマウマと判断
 - 多くの場合、属性の辞書は人間が天下りの的に準備
 - 明確にルールが決まっているわけではない
 - コンピュータが本当に抽出可能であるかは保証されていない

<u>otter</u>			
black:	yes		
white:	no		
brown:	yes		
stripes:	no		
water:	yes		
eats fish:	yes		
<u>polar bear</u>			
black:	no		
white:	yes		
brown:	no		
stripes:	no		
water:	yes		
eats fish:	yes		
<u>zebra</u>			
black:	yes		
white:	yes		
brown:	no		
stripes:	yes		
water:	no		
eats fish:	no		

Christoph H. Lampert Hannes Nickisch Stefan Harmeling, Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer, In Proc. CVPR, 2009.

人間が識別時にヘルプ (1/2)

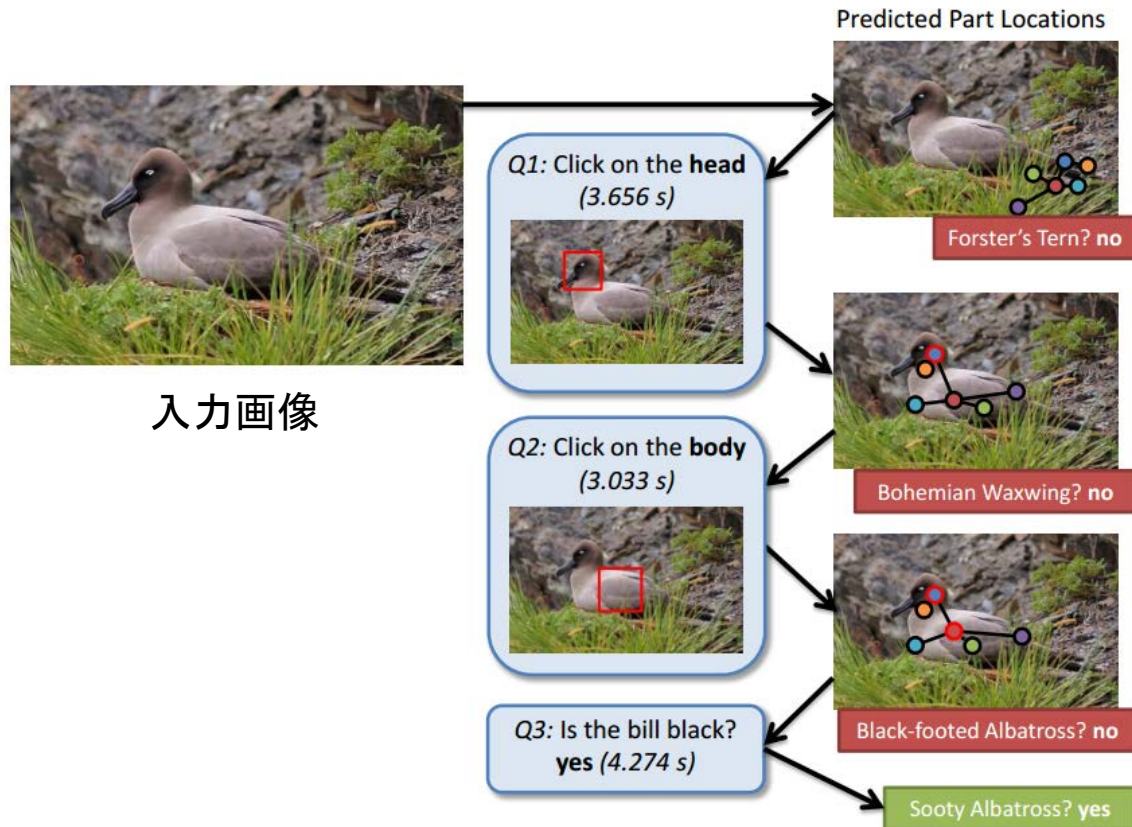
- ▶ コンピュータは可能な範囲で解の候補を生成
- ▶ **難しいところは人間に問い合わせる**
- ▶ インタラクティブに処理を進め、徐々に解答を絞りこむ (20の扉)
 - できるだけ情報量が得られる質問を生成
 - 識別の精度を上げると同時に、納得感も得られる
- ▶ 属性 (attribute) を問い合わせ [Branson et al., ECCV'10]



Steve Branson, Catherine Wah, Florian Schro, Boris Babenko, Peter Welinder, Pietro Perona, and Serge Belongie, Visual Recognition with Humans in the Loop, In Proc. ECCV, 2010.

人間が識別時にヘルプ (2/2)

- ▶ 属性に加え、部位の位置も問い合わせ [Wah et al., ICCV'11]



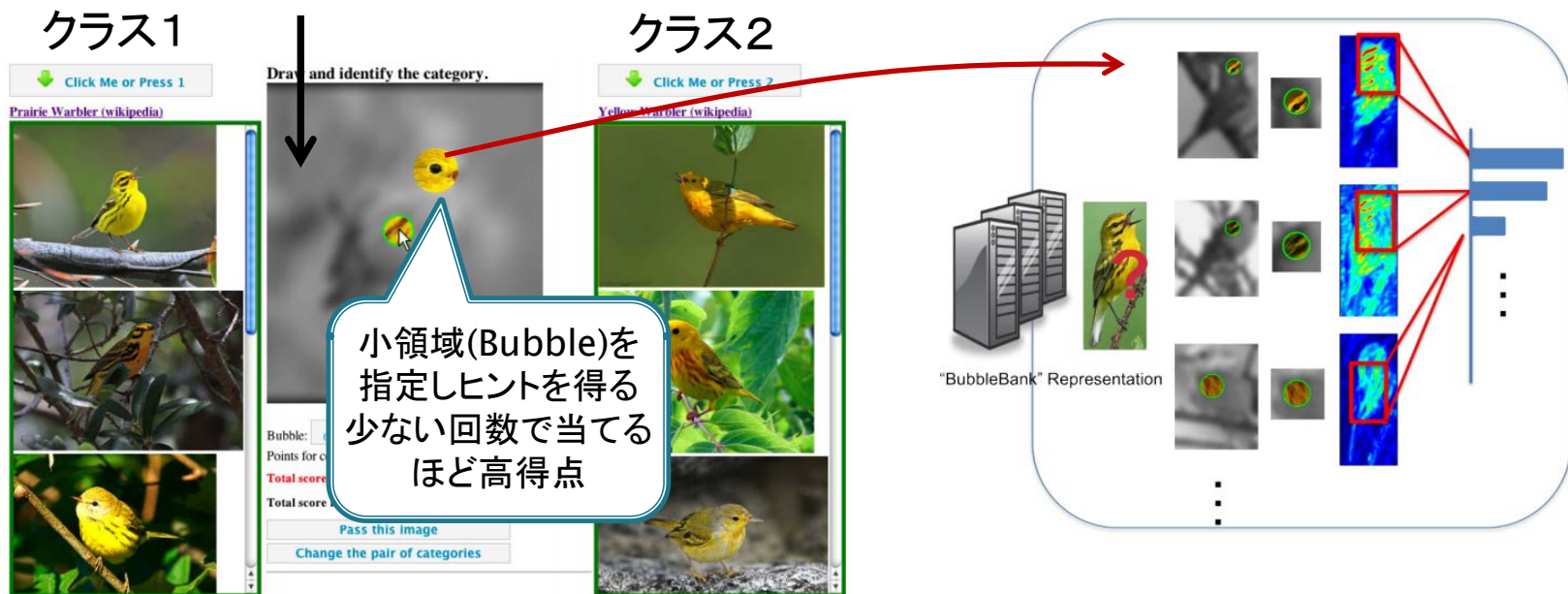
Catherine Wah, Steve Branson, Pietro Perona, Serge Belongie, Multiclass Recognition and Part Localization with Humans in the Loop, In Proc. ICCV, 2011.

人間が学習時にヘルプ (1/2)

- ▶ 識別に有効なテンプレートパッチをクラウドソーシングで獲得
 - **ゲーム形式**で人間にBubble (パッチ)を選択させる [Deng et al., CVPR'13]
 - 学習時に集合知を活用することで、完全自動のシステムを大きく上回る識別精度を達成

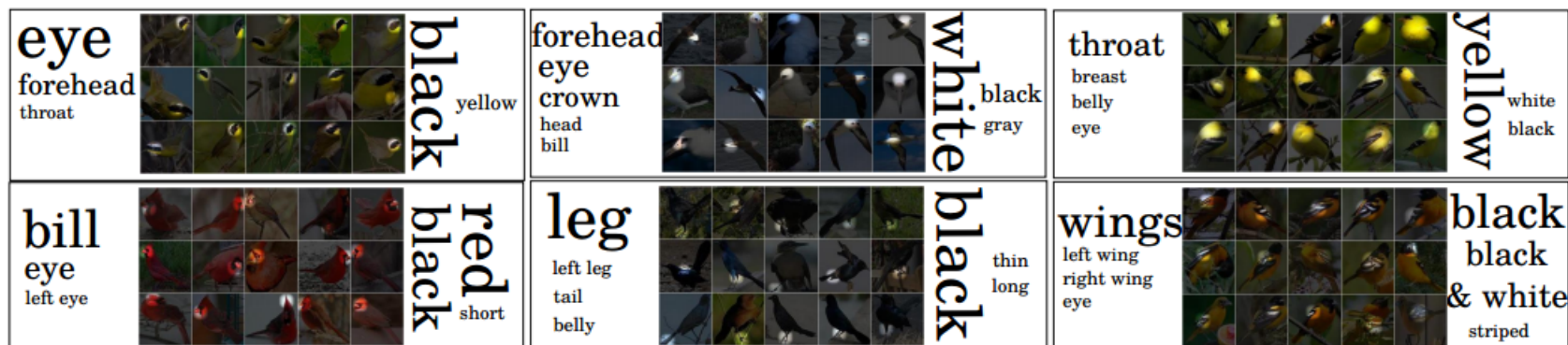
与えられた低品質画像がクラス1・2
のどちらであるかを答えるゲーム

得られた多数のBubbleをテンプレート
として利用し特徴ベクトルを生成



人間が学習時にヘルプ (2/2)

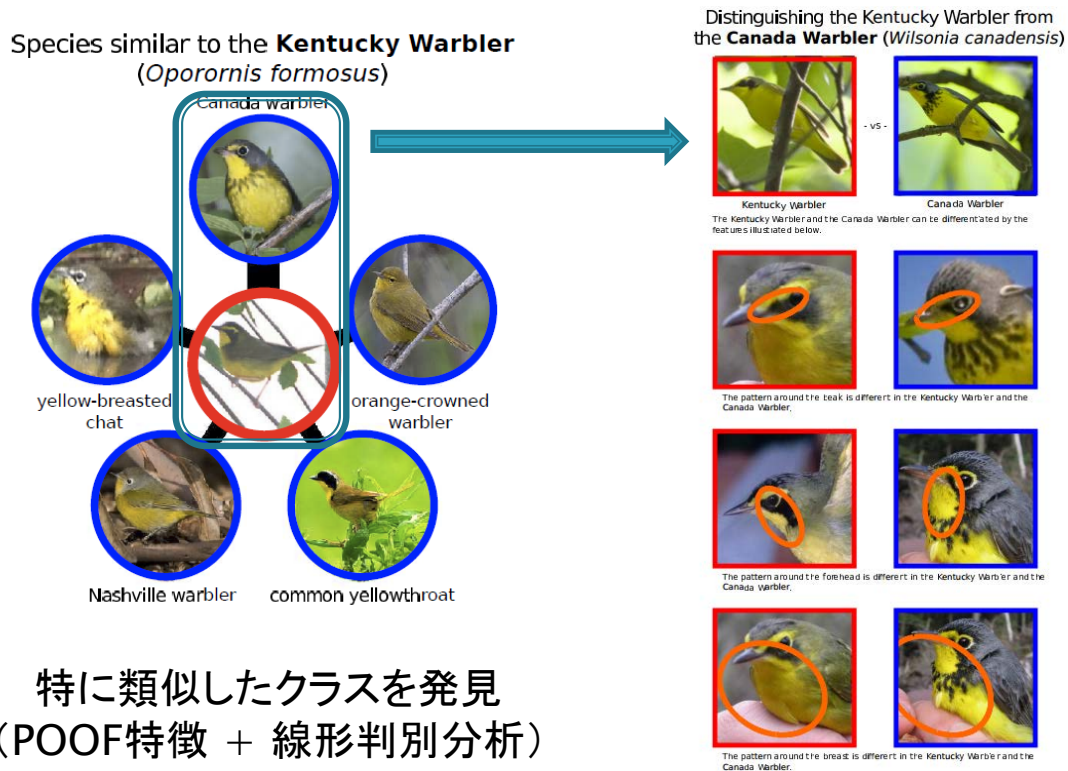
- ▶ 属性 (attribute) の対話的な発見 [Duan et al., CVPR'12]
 - クラスを弁別するために有効と思われる局所パターンをlatent CRFで自動的に抽出し、妥当性を人間へ問い合わせ
 - **コンピュータにとっても人間にとっても理解できる中間表現**
 - 天降りで定義されたattributeと異なり、システムが検出できることが保証されている



システムが自動的に発見したパターンと、ユーザが与えた名称
(フォントの大きさはそれを与えた人数に対応)

人間に教える

- ▶ 似通ったクラスの識別に有効な特徴を発見し、提示する
 - 図示に最も適したサンプルを自動的に選択 [Bert et al., ICCV'13]



最も識別的な特徴を図示
この際に、出来る限り以下を
満たすサンプルを選ぶ。

- (1) 図示したい特徴がはっきり表れている
- (2) その他の特徴が(たまたま)差として表れていない
- (3) 似たような構図を持つ

まとめ

- ▶ **詳細画像識別 (Fine-grained visual categorization, FGVC)**
 - 意味的・視覚的に非常に類似した対象の多クラス識別
 - 特定のドメインで、非専門家を大きく超える識別能力の実現
 - 応用先は広い
 - 官能検査、概観検査、医療画像識別、商品識別...
- ▶ **識別手法の発展**
 - 大規模一般物体認識の技術をもとに、独自の進化を遂げつつある
- ▶ **識別に“意味”を与える**
 - 人とのハイブリッドで成長するシステム
 - 数年後には、人を超え、人に教えるシステムの登場を期待