どこまでできる画像検索？！
深層学習技術で目指す画像を見つけ出せ

国立情報学研究所
佐藤真一

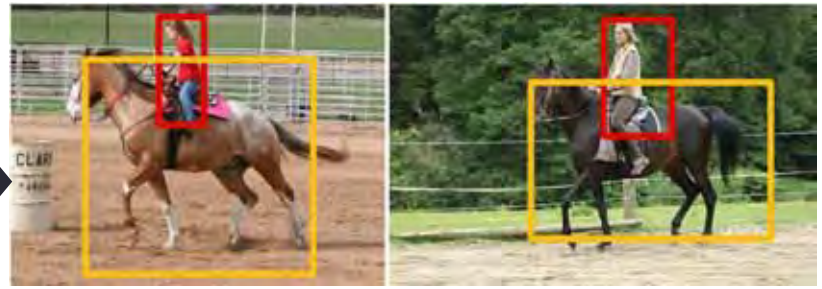
物体検索技術を用いた大規模画像検索

入力 (問い合わせ)

‘ジャンプしている
スノーボーダー’

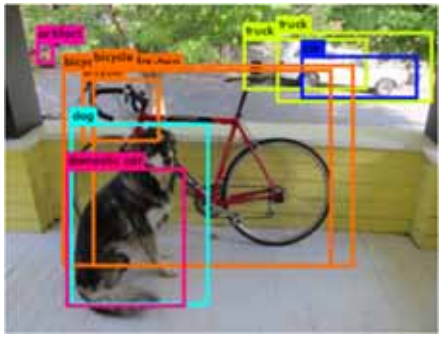
“白馬にまたがる人”

出力



深層学習による物体検出技術

物体検出



[Redmon16]

フレーズによる検出

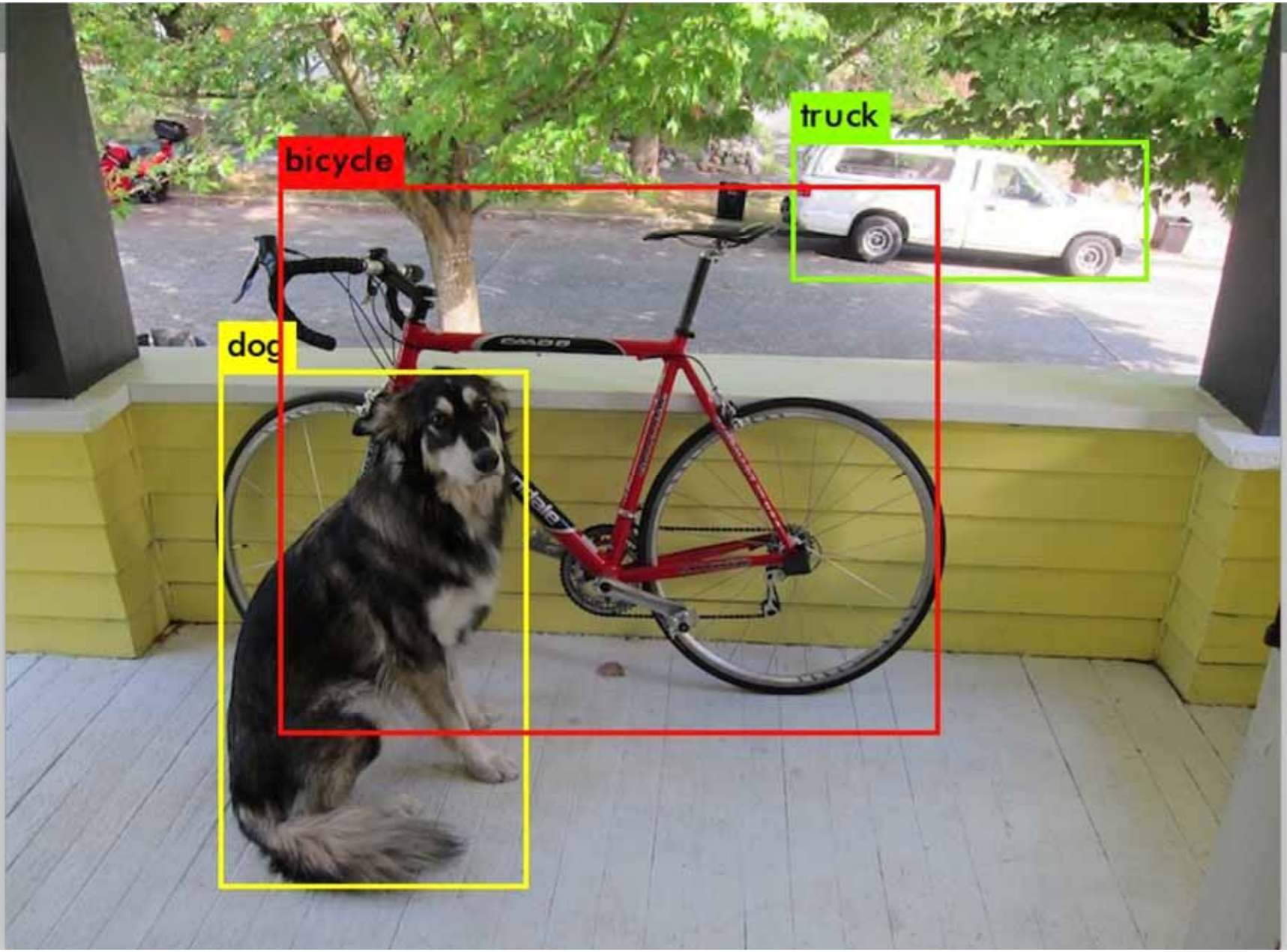
[Plummer17]

関連性検出

[Lu16]

シーングラフ

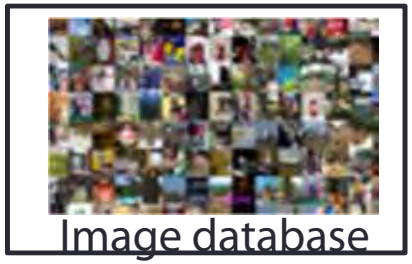
[Xu17]


truck
bicycle
dog

目標: 大規模画像データベースを対象とした物体検出

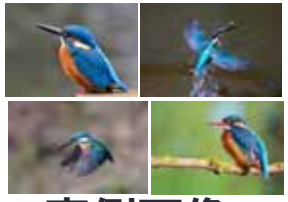
もし物体カテゴリが固定であれば…

'イヌ'
'ネコ'



- 事前にすべて検出しておける
- しかし未知カテゴリに対応できない

未知カテゴリの物体検出



事例画像

物体検出システム

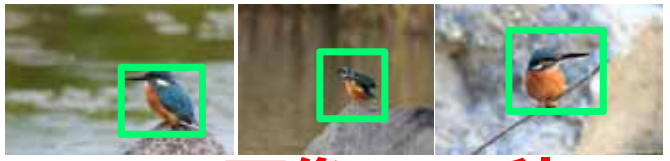
1画像当たり0.1秒ほどかかる

- 100万画像だと1日!

提案手法: 事前に大規模物体検出向け索引を作っておく

物体検出システム

索引構造

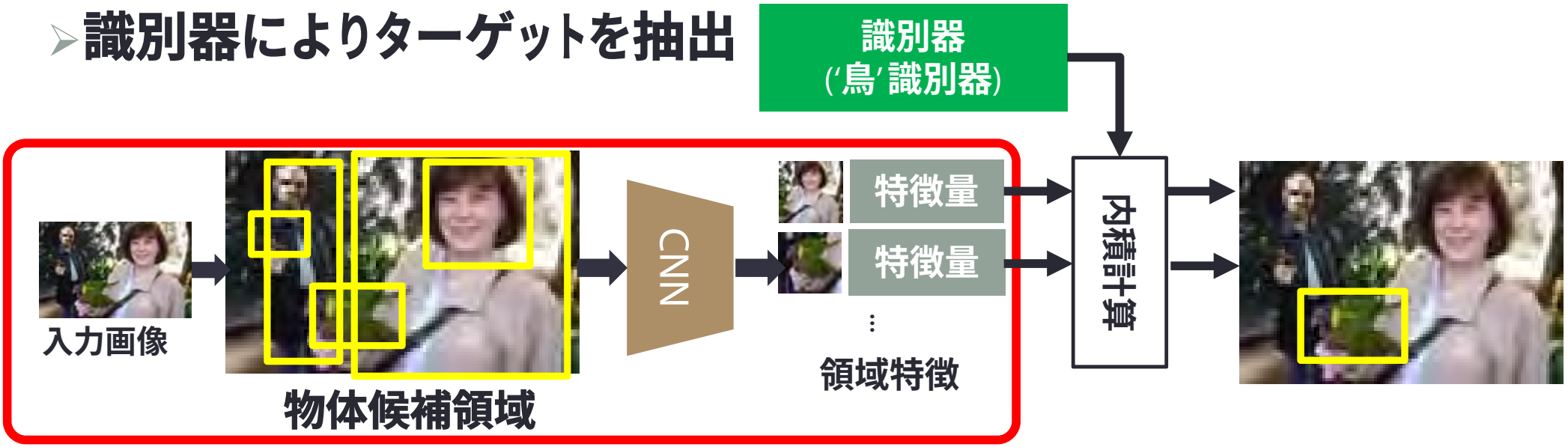


…

- 10万画像で0.1秒

R-CNN [Girshick+14]

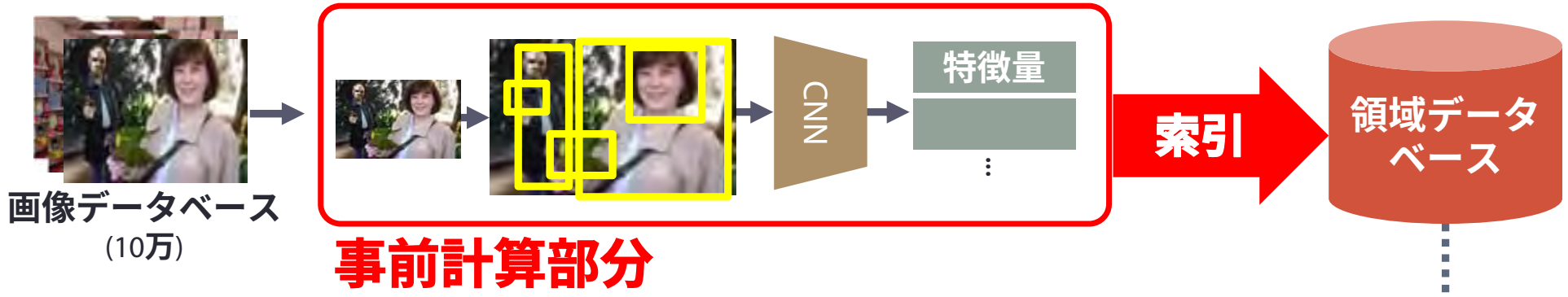
- 深層学習に基づく物体検出手法
- 物体「候補」領域を検出: 画像当たり100-2000 候補領域
- 深層学習により各候補から「特徴量」を抽出
  → 物体単位の意味情報として利用
- 識別器によりターゲットを抽出



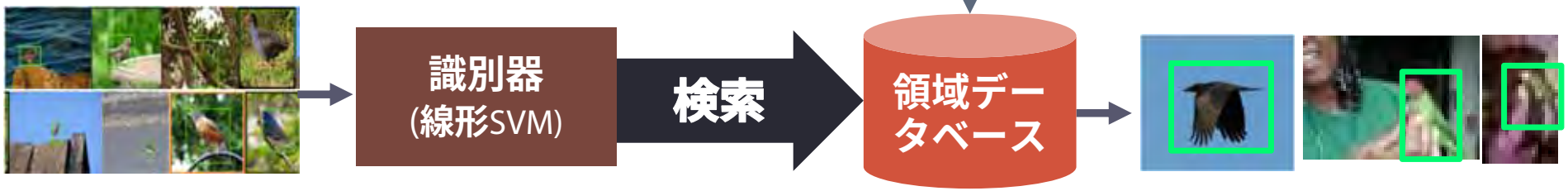
物体カテゴリ(問い合わせ) と無関係→ 事前に計算しておける

提案手法: 大規模R-CNN

索引作成 (事前計算)

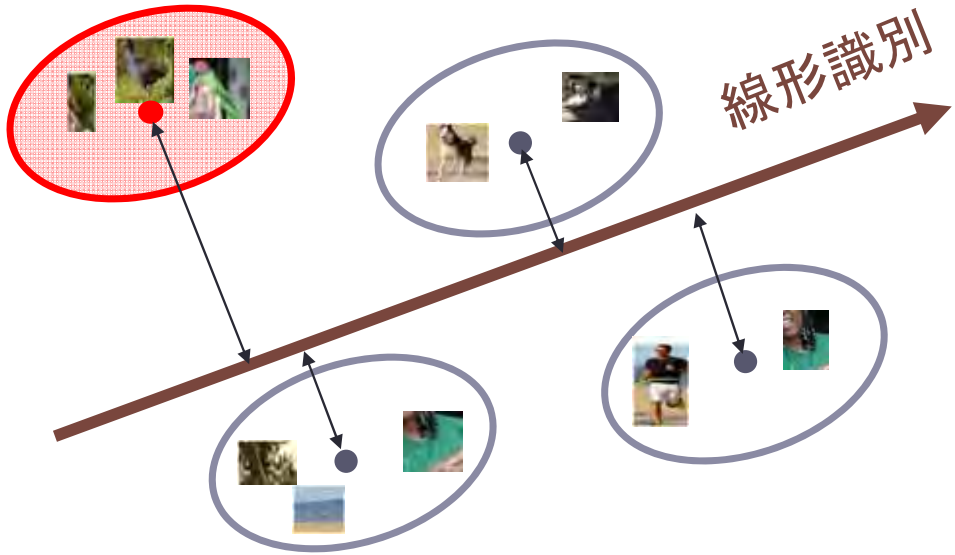


検索時



大規模R-CNNの索引の作成 (概要)

転置索引



- 探索すべき領域の削減(250×)
- 線形識別に適した「見出し語」作成

ベクトル量子化

深層学習特徴

$x \in \mathbb{R}^{4096}$

直積量子化

(PQ) [Jegou10]

3.2TB
10万画像

25GB
10万画像

- メモリ使用量削減(128×)
- 距離計算の高速化

通常のR-CNN と大規模R-CNNの比較

R-CNN: 全特徴量に線形SVMを適用した場合

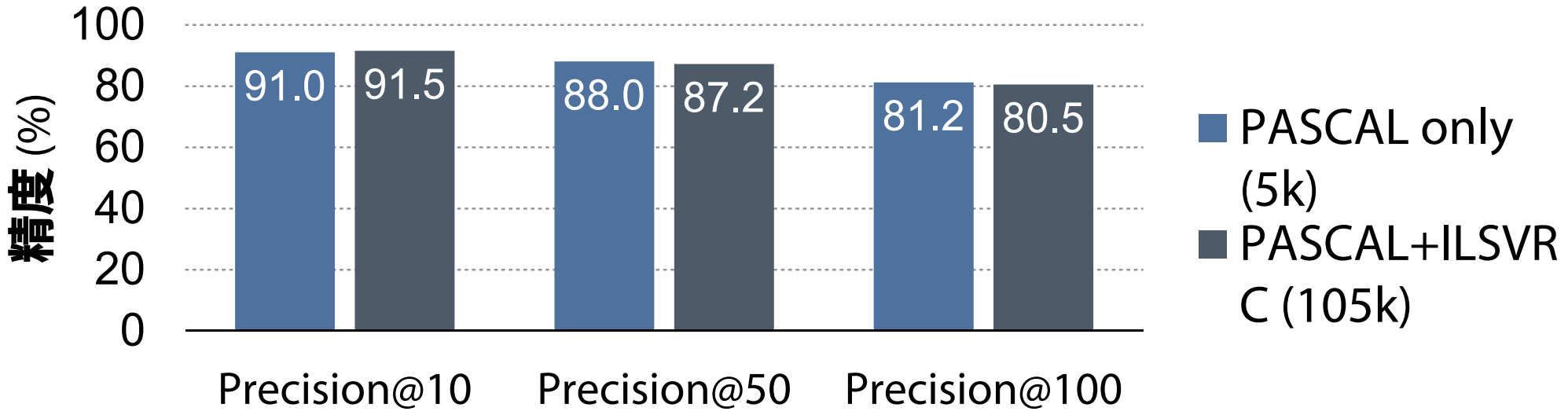
大規模R-CNN: 転置索引(IVF) とベクトル量子化(VQ)を使用

Dataset: PASCAL VOC	IVF	VQ	mAP	Time	Memory
R-CNN			**54.2%**	6258 ms	163 GB
w/ Vector quantization		✓	52.4%	69.5 ms	1.27 GB
w/ Inverted index	✓		52.4%	518.0 ms	163 GB
w/ VQ, IVF	✓	✓	50.7 %	**24.5 ms**	**1.54 GB**

250倍高速×106メモリ効率

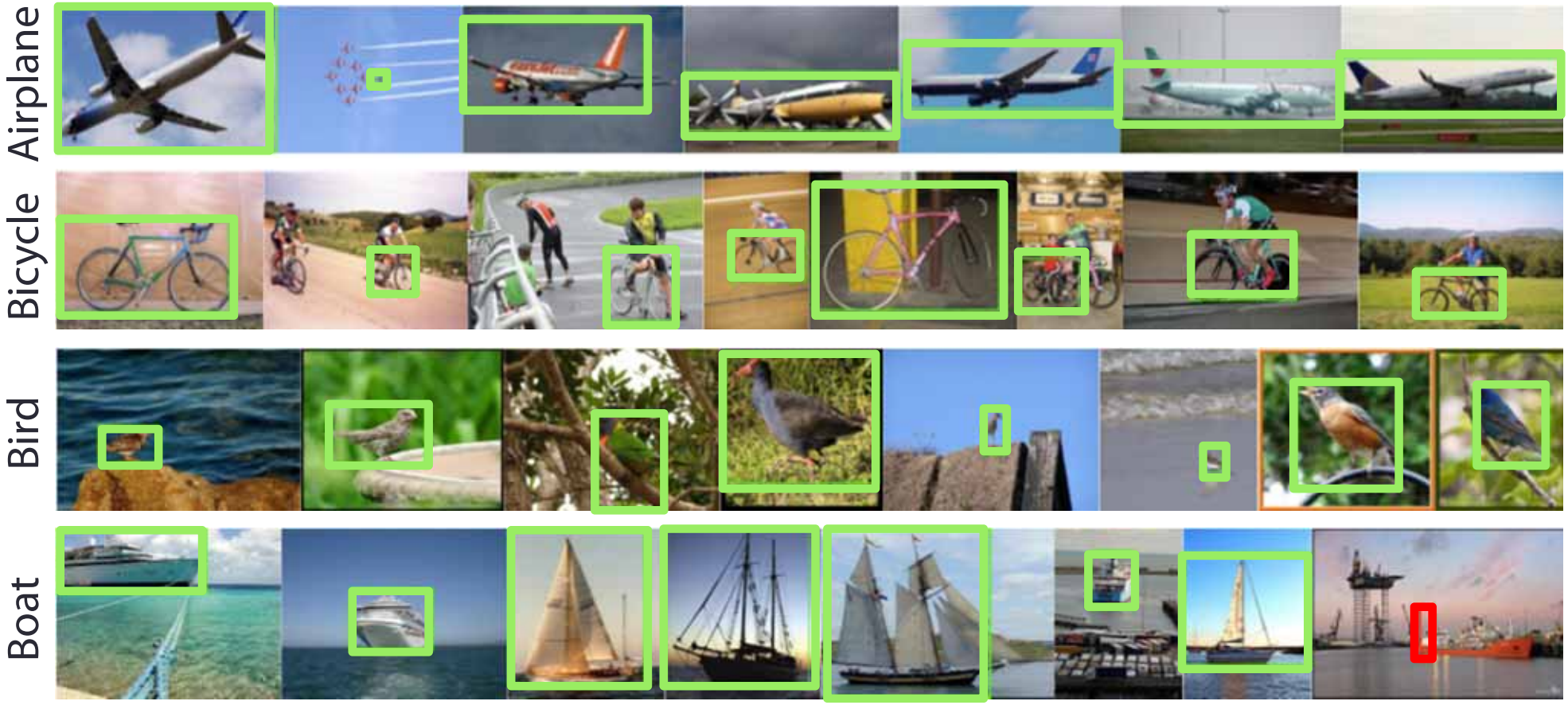
大規模実験

10万画像に対する性能評価(PASCAL5K+ILSVRC100K)

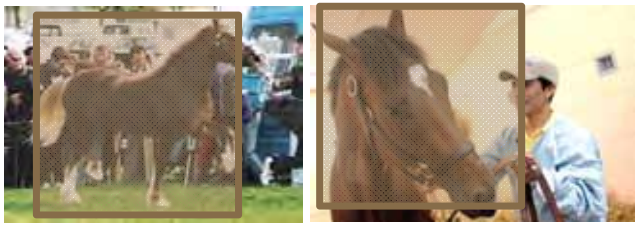


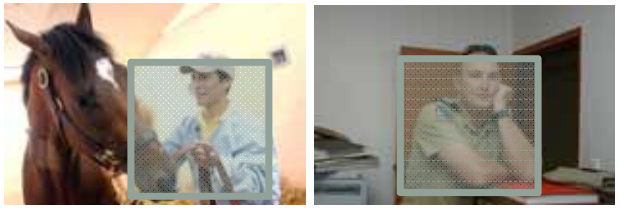
- 10万画像にしても精度劣化は見られず
- 10万画像に対して100ms

10万画像に対する検索結果



複数物体への対応

‘ウマ’→  …

‘ヒト’→  …

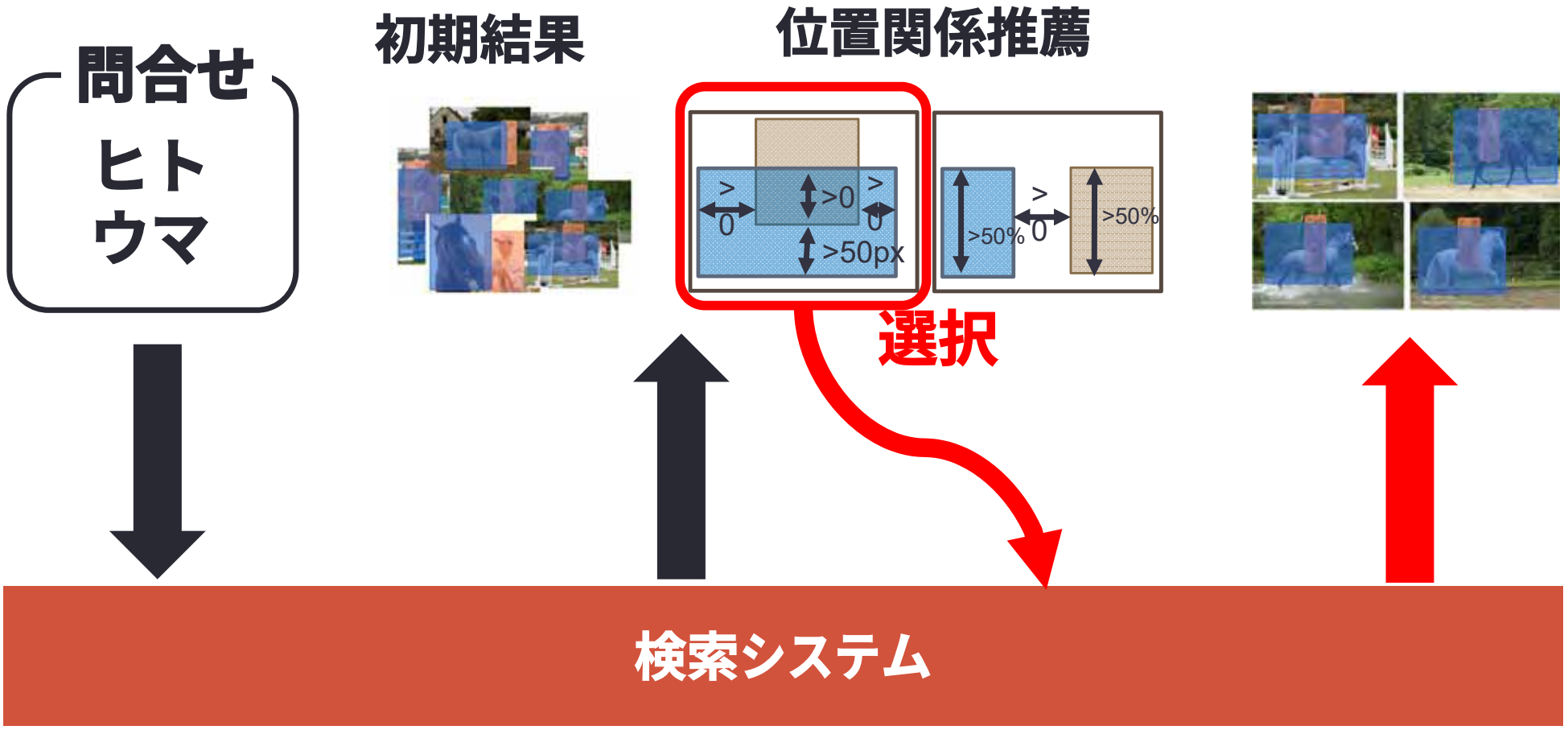
&

物体間の関連性を指定できるか?

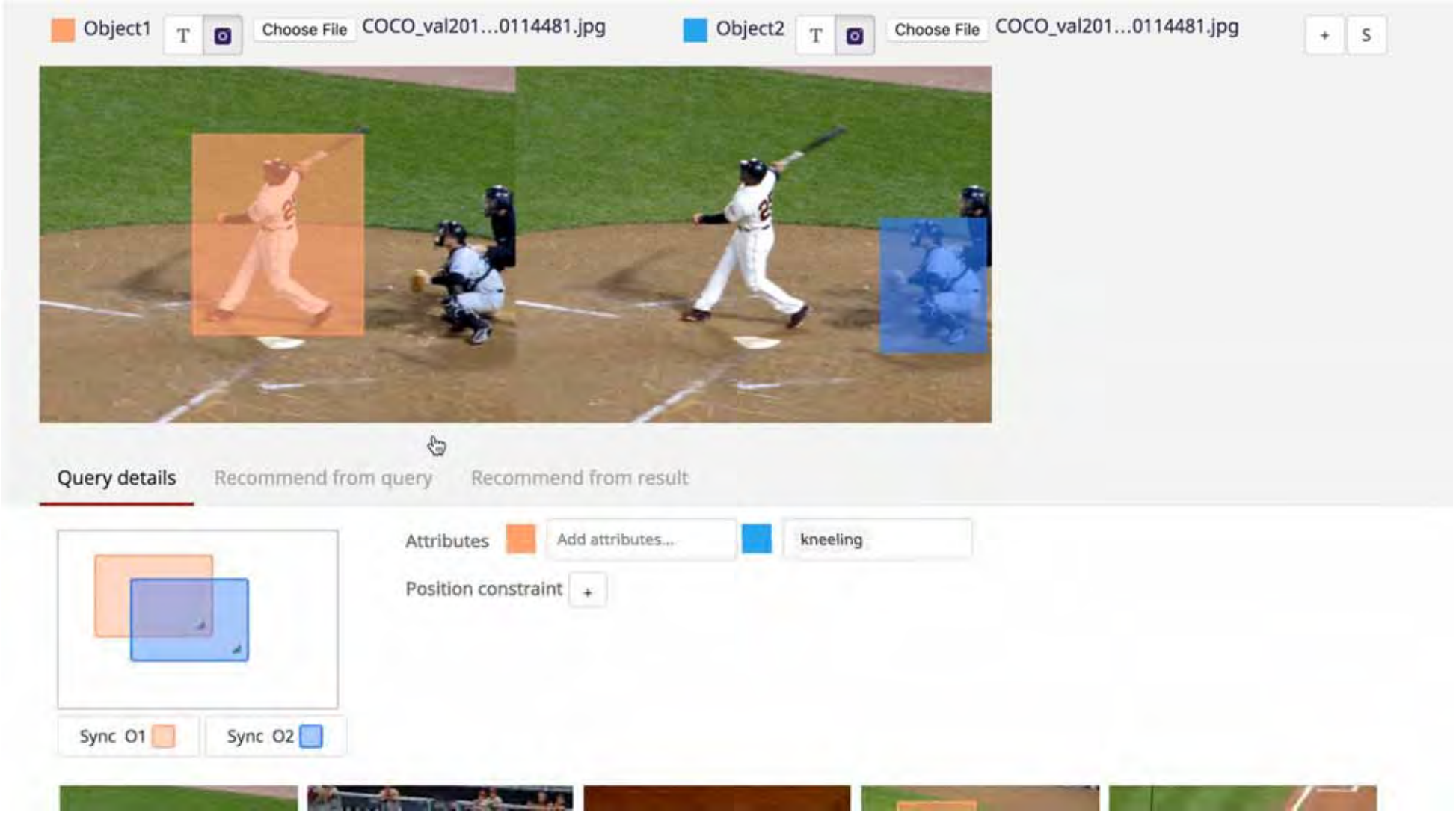
<man, ride, horse>

もっと込み入った関連性

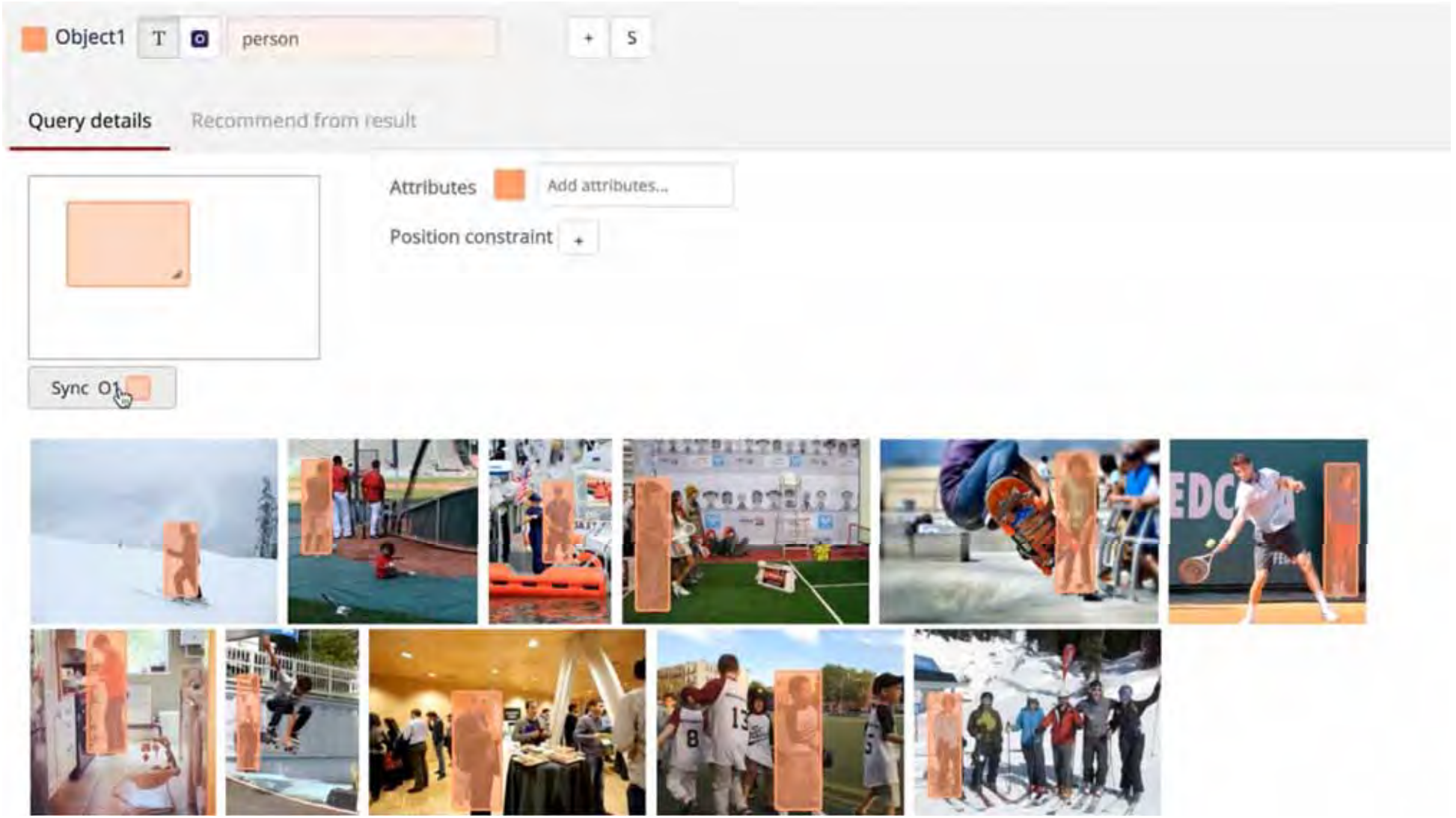
適切な位置関係の推薦



初期結果に基づく推薦


Object1
T
Choose File
COCO_val201…0114481.jpg
Object2
T
Choose File
COCO_val201…0114481.jpg
+
S
Query details
Recommend from query
Recommend from result
Attributes
Add attributes...
kneeling
Position constraint
+
Sync O1
Sync O2

位置関係の直接指定


Object1
T
person
+
S
Query details
Recommend from result
Attributes
Add attributes...
Position constraint
+
Sync O1

問い合わせ=‘自転車を運んでいる車’

O_1 自転車
O_2 車
‘上に’

問い合わせ

結果

問い合わせ=‘イヌとスキー’

O_1 スキーをする人
O_2 イヌ
‘横に’

問い合わせ

結果

問い合わせ=‘馬車に乗っている’

O_1 座っている人

O_2 馬

O_3 

関連性推薦

結果

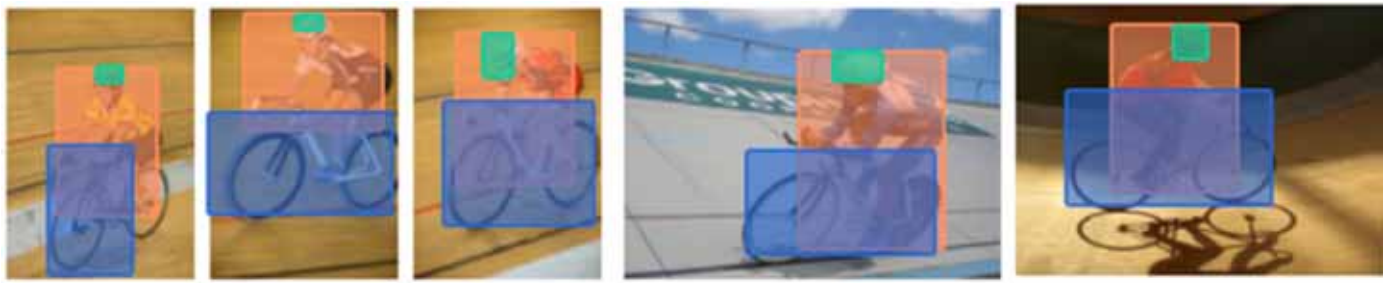
問い合わせ=‘ヘルメットをかぶって自転車に乗る人’

O_1 乗っている人

O_2 自転車

O_3 

関連性推薦



結果

おわりに

- 深層学習(AI技術)により、画像中の事物の意味内容まで深く解析することが可能になった
- 大規模処理により画像検索まで実現可能であることを紹介した
- 今後は、画像中の事物の動作の解析や検索、さらには映像も用いた動作の解析や検索について検討していきたい