LLINOS

MOTIVATION

Task: Phrase localization (*i.e.*, given a phrase and image find the corresponding bounding box described in the phrase)

- Challenging multimodal task
- Joint understanding of language and image
- Variety of objects and instance level descriptions

Contributions:

- State-of-the-art unsupervised textual grounding approach
- Utilize co-occurrence of words and image concepts
- Hypothesis testing formulation for unsupervised learning.

INTRODUCTION

Problem Formulation:

- Input: x = (Q, I)
- Output: $y = (y_1, ..., y_4)$

Intuition of our Approach:

- Learning from word-concept co-occurrences
- Identify relevant "image concepts" given query

Bounding Box Prediction:

- $t_s(Q) \in \{0, 1\}$: existence of the token $s \in S$ in a query Q
- $a_c(I) \in \{0, 1\}$: existence of the concept $c \in C$ in an image I
- Find the least independent activated image concept given query

$$c^* = \underset{c \in \mathcal{C}: a_c(I)=1}{\operatorname{argmin}} \min_{s \in \mathcal{S}: t_s(Q)=1} E(s)$$

- Image concept c is attached to a score map $\phi_c(I) \in \mathbb{R}^{W \times H}$
- Estimated bounding box from concept ϕ_{c^*} via efficient subwindow search

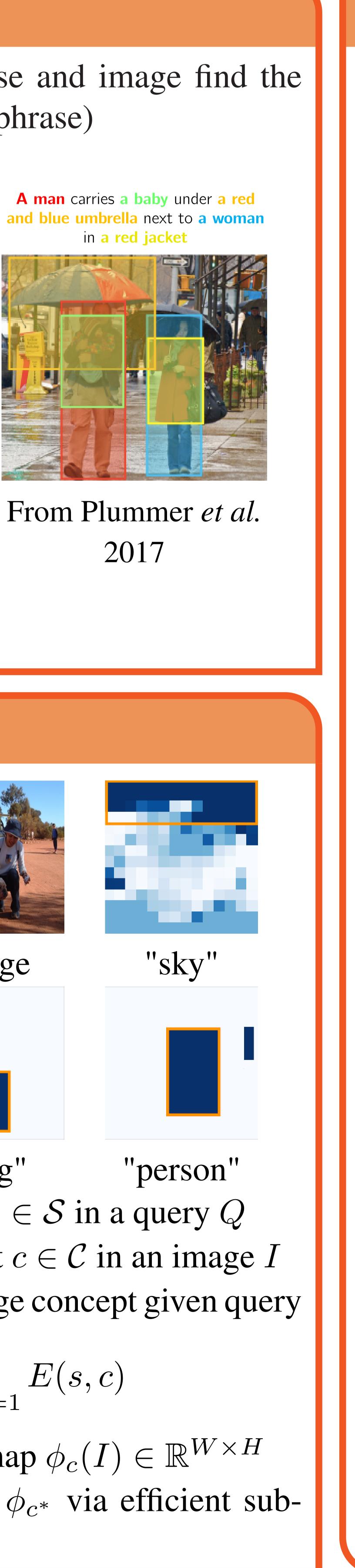


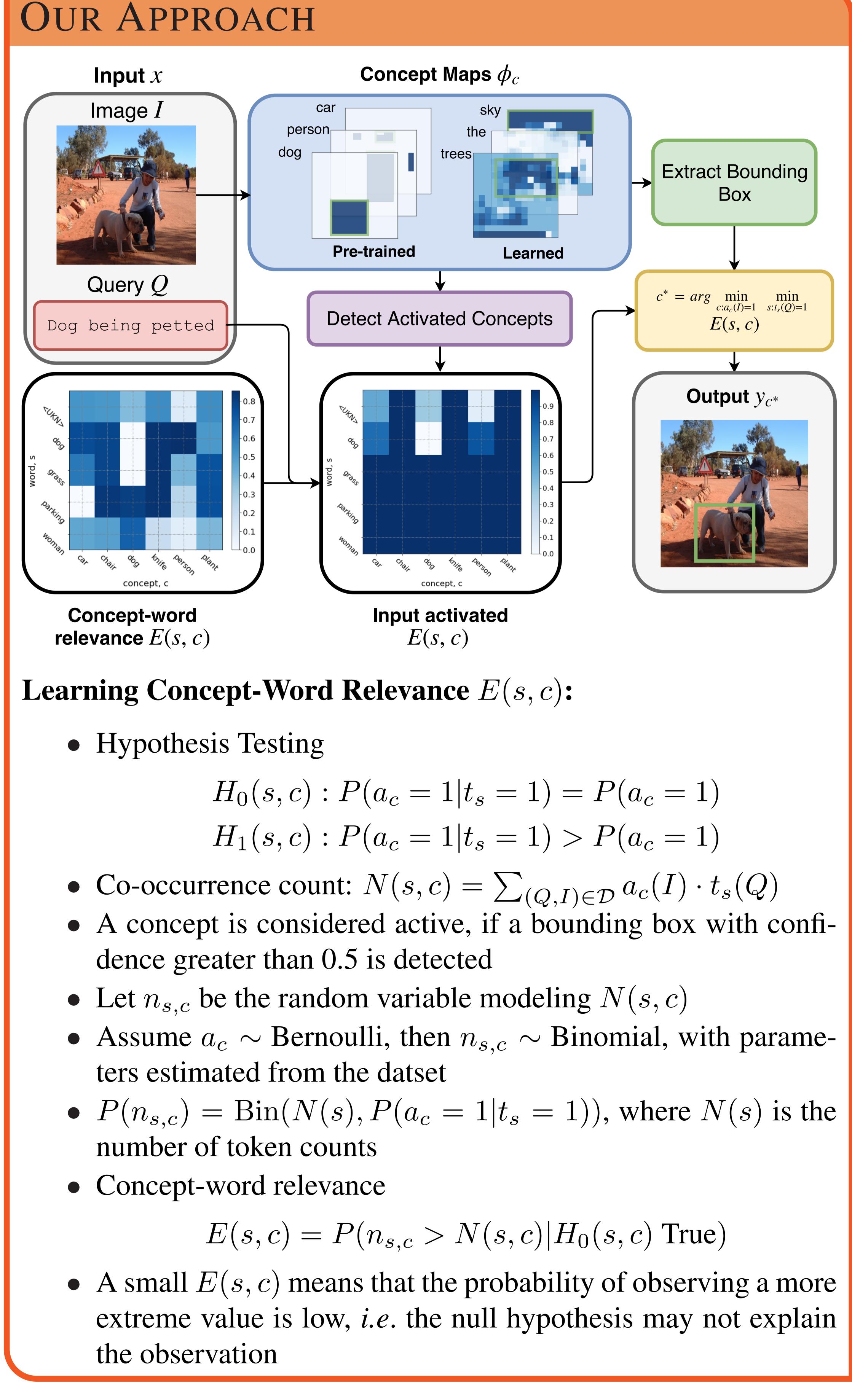




"dog"

UNSUPERVISED TEXTUAL GROUNDING: LINKING WORDS Alexander G. Schwing Raymond A. Yeh Minh N. Do University of Illinois at Urbana-Champaign





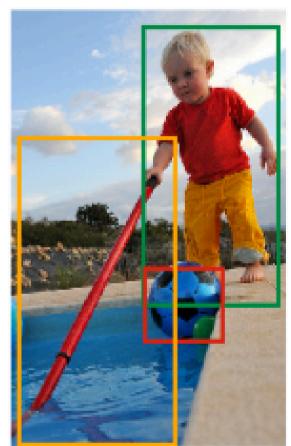
$$t = 1|t_s = 1)$$
, where $N(s)$ is the

RESULTS

Quantitative Results:

Approach	Features	Acc (%)	Approach	Features	Acc (%)
GroundeR	VGG-cls	10.70	GroundeR	VGG-cls	24.66
GroundeR	VGG-det	_	GroundeR	VGG-det	28.94
Mutual Info.	VGG-det	16.00	Mutual Info.	VGG-det	31.19
Ours	VGG-cls	18.68	GloVe	YOLO-det	(35.40)
Ours	VGG-det	17.88	Ours	VGG-det	35.90
Ours	VGG-cls+det	20.10	Ours	YOLO-det	36.93

Success Cases:

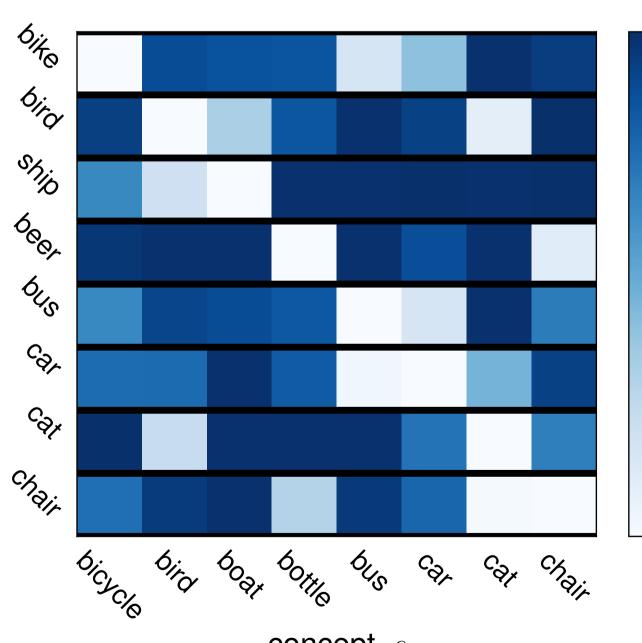


The baby boy is playing with a red stick with other balls in

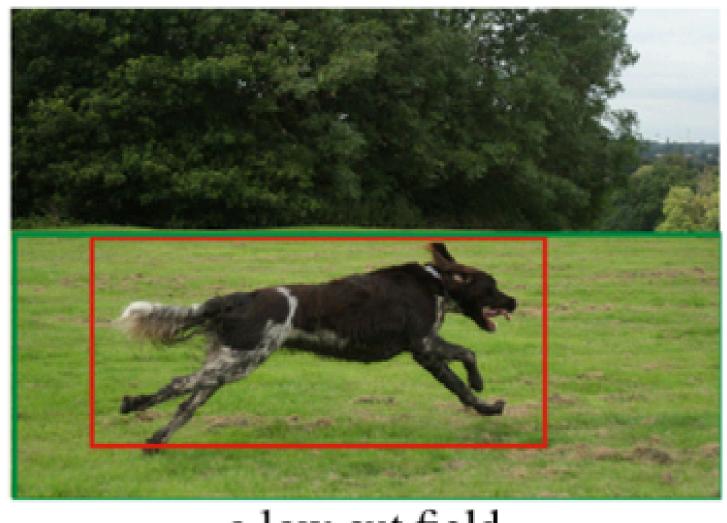


A woman tries to volley a a tennis

Interpretability of E(s, c):



Failure Cases:



a low cut field

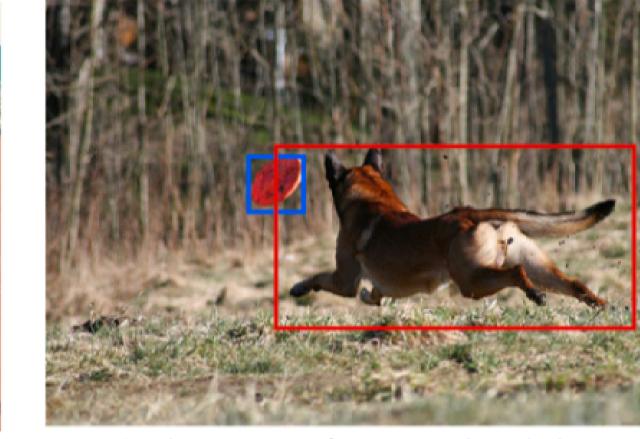




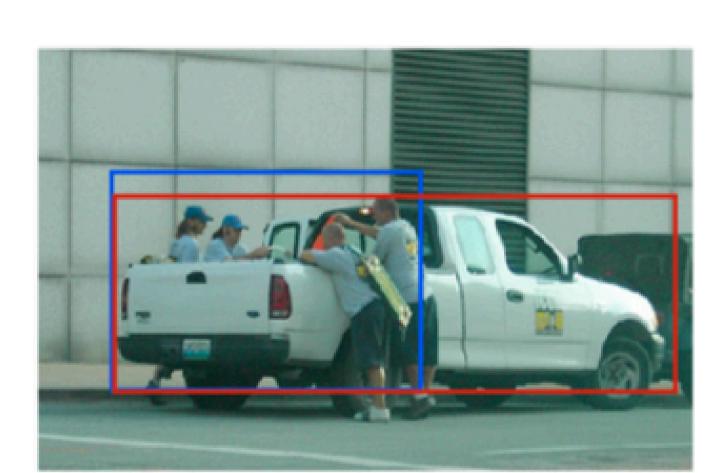




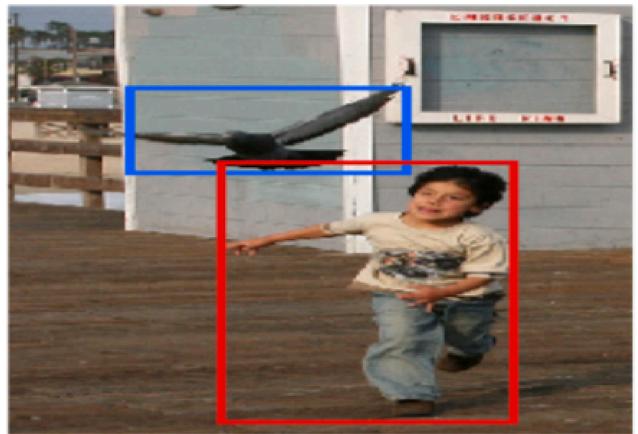
front of a British Airways airplane.



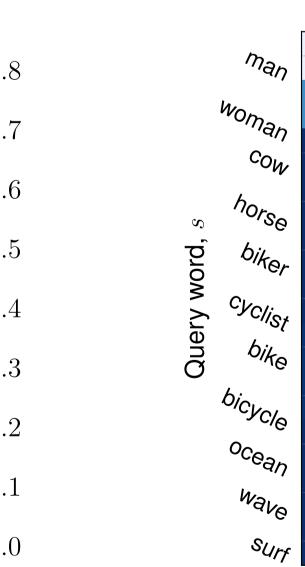
A dog runs for a red Frisbee.

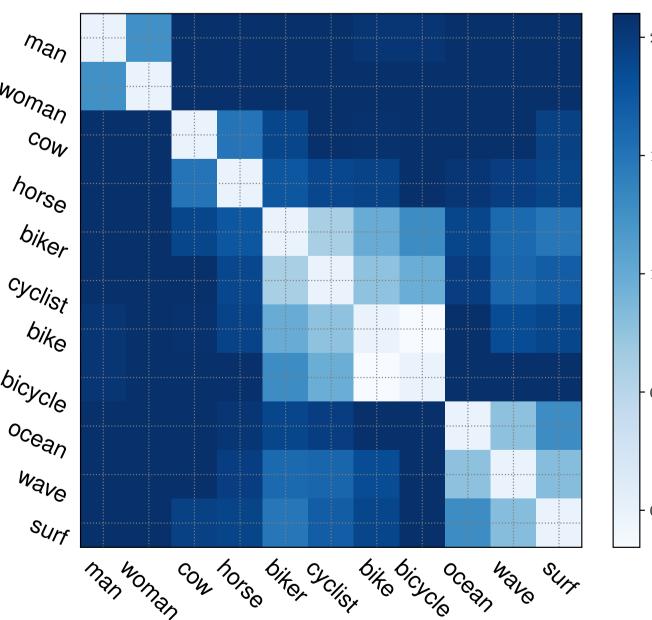


Two pilots are standing and talking in Four guys standing around the bed of a white pickup truck.



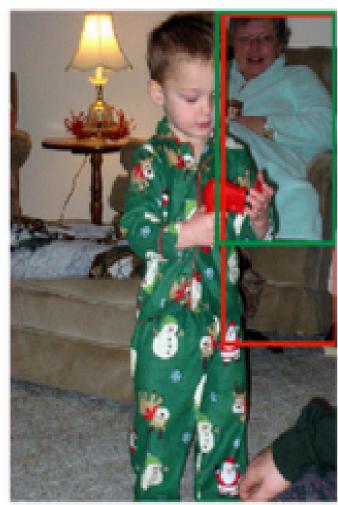
Little boy trying to chase a black







a rough stretch of river



a robe