



University of Pittsburgh

Detecting Sexually Provocative Images

Debashis Ganguly,
Mohammad H. Mofrad,
Adriana Kovashka

*Department of Computer Science,
University of Pittsburgh*





Real Life Challenges









- Overwhelming amount of visual data on the Internet
- Parents may want to restrict the visual contents which their children can see.
- Lots of manual effort is invested by digital content administrators to classify images in age restricted categories.



Limitations of Existing Approaches

- Existing approaches detect pornographic contents based on percentage of skin area exposed by the subjects in such images.

Input Image	HSV threshold
	
	
	

Jiao et. al., "Detecting adult image using multiple features", Info-tech and Info-net 2001

Duan et. al., "Adult image detection method based on skin color model and support vector machine", Asian Conference on Computer Vision 2002

Zheng et. al., "Shape based adult image detection", International Journal on Image and Graphics 2006

Lee et. al., "Naked image detection based on adaptive and extensible skin color model", Pattern recognition 2007



Limitations of Existing Approaches (contd.)

- Current methods can not differentiate between pornographic content, portrait or harmless body shot like below.



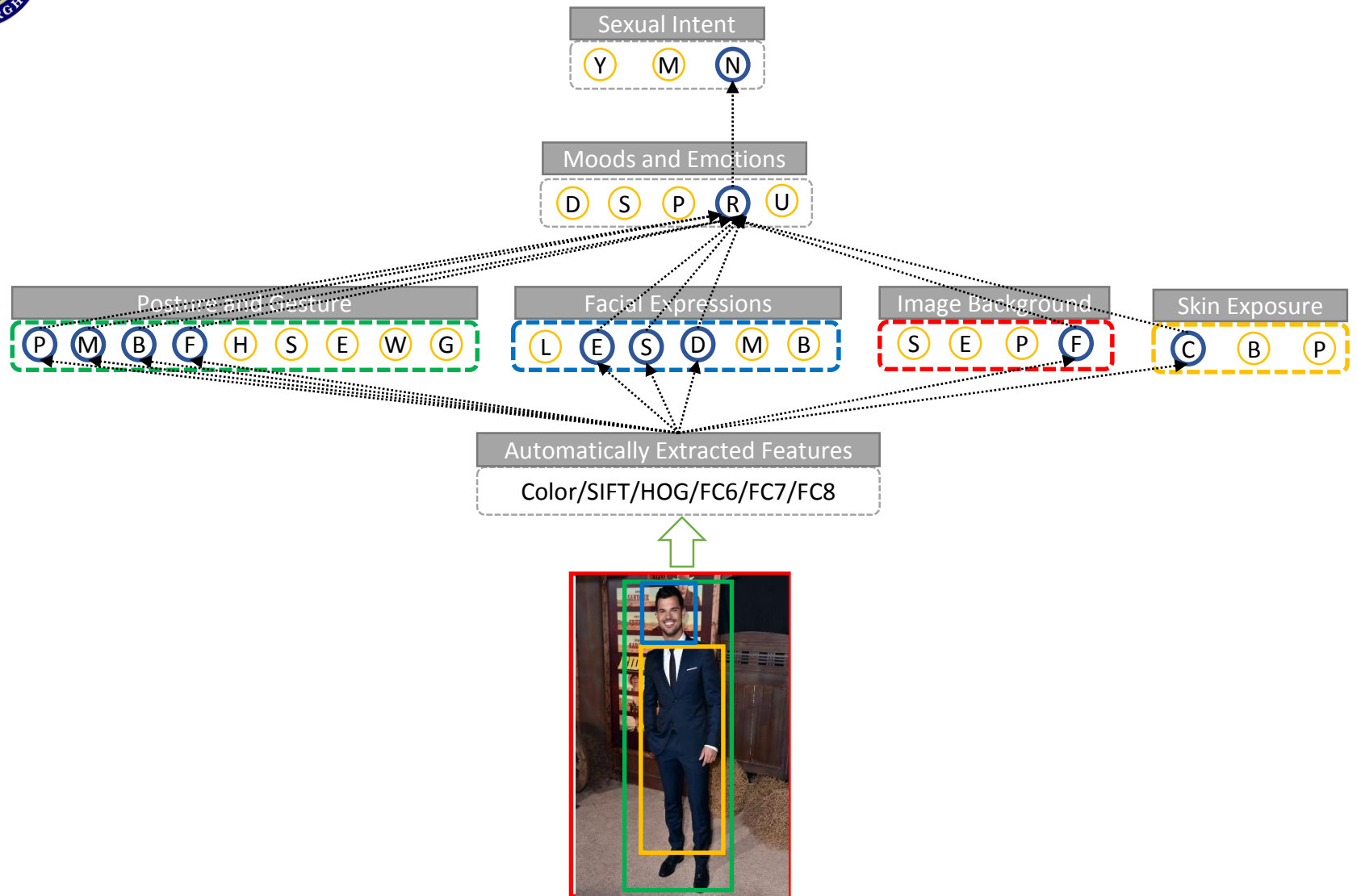


Approach: Identifying Features

- 17 types of Attributes composed from:
 - Posture and gesture
 - Posture, gesture with fingers, movement, head position, direction of body and face relative to camera, etc.
 - Facial expression
 - Mouth open or closed, type of smile, biting lips, eyebrows, eyelids, looking direction
 - Scene context
 - Outdoor scene, outdoor events, indoor scenes with props or with flat background
 - Skin exposure
 - Fully clothed, bare bodied, private body parts exposed
- 5 types of Moods and Emotions:
 - Defensive, suggestive, playful, relaxed, upset
- 3 Sexual Intentions
 - Yes, maybe, no



Hierarchical Framework





Experiments: Dataset

- 1,146 celebrity images
 - 203 Hollywood celebrities from *people.com*
 - 892 and 254 images of female and male candidates respectively
 - 5.6 images per person ratio
- 19 questions per image for annotations
 - Amazon Mechanical Turk by majority voting of 3 annotators per image
 - 70.5% annotator consensus



Experiments: Baseline

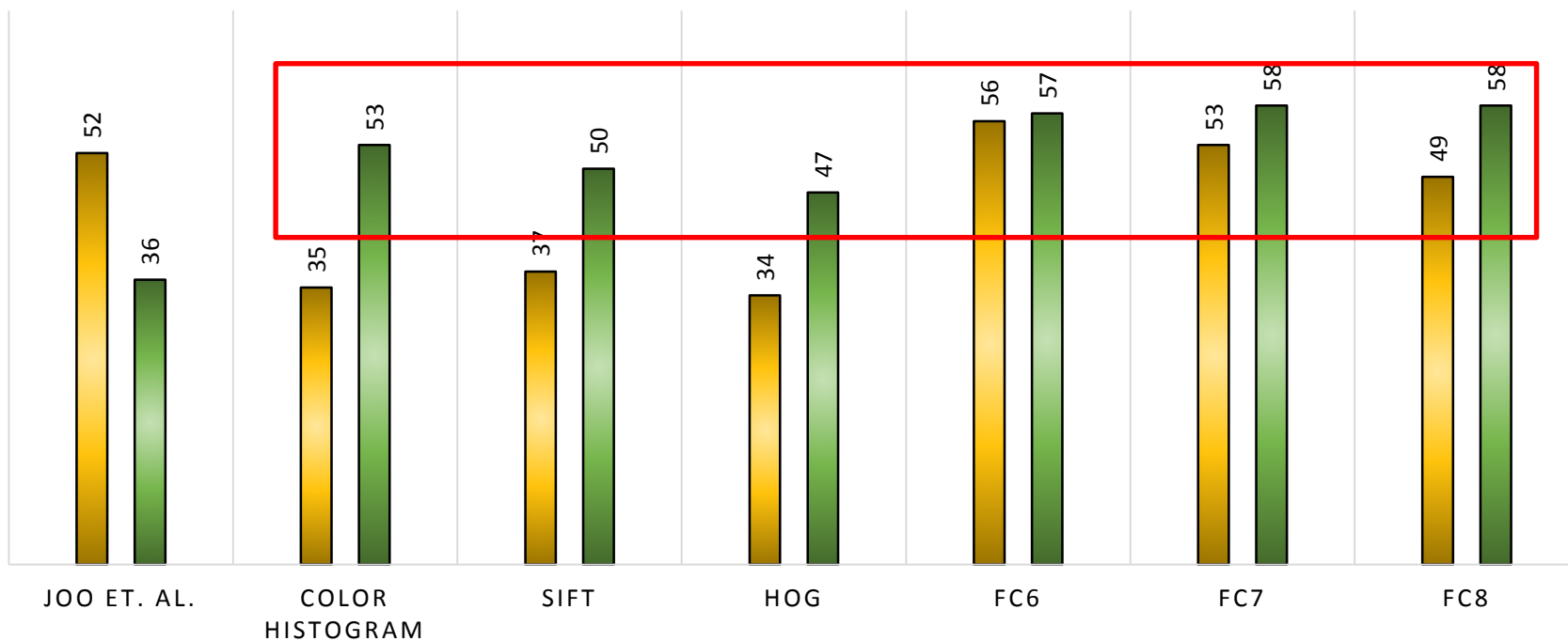
- Automatically extracted features
 - Low level features: Color histogram, SIFT, HOG using VLFeat
 - CaffeNet Features: FC6, FC7, FC8 using Caffe
- Direct model
 - Single level of classification hierarchy trained from automatically extracted features to predict sexual intent
- Joo et. al., “Visual persuasion: inferring communicative intents of images”, CVPR 2014
 - Subset of features mapped based on relevance to our problem domain



Results: Overview

F-MEASURE

Direct Hierarchical

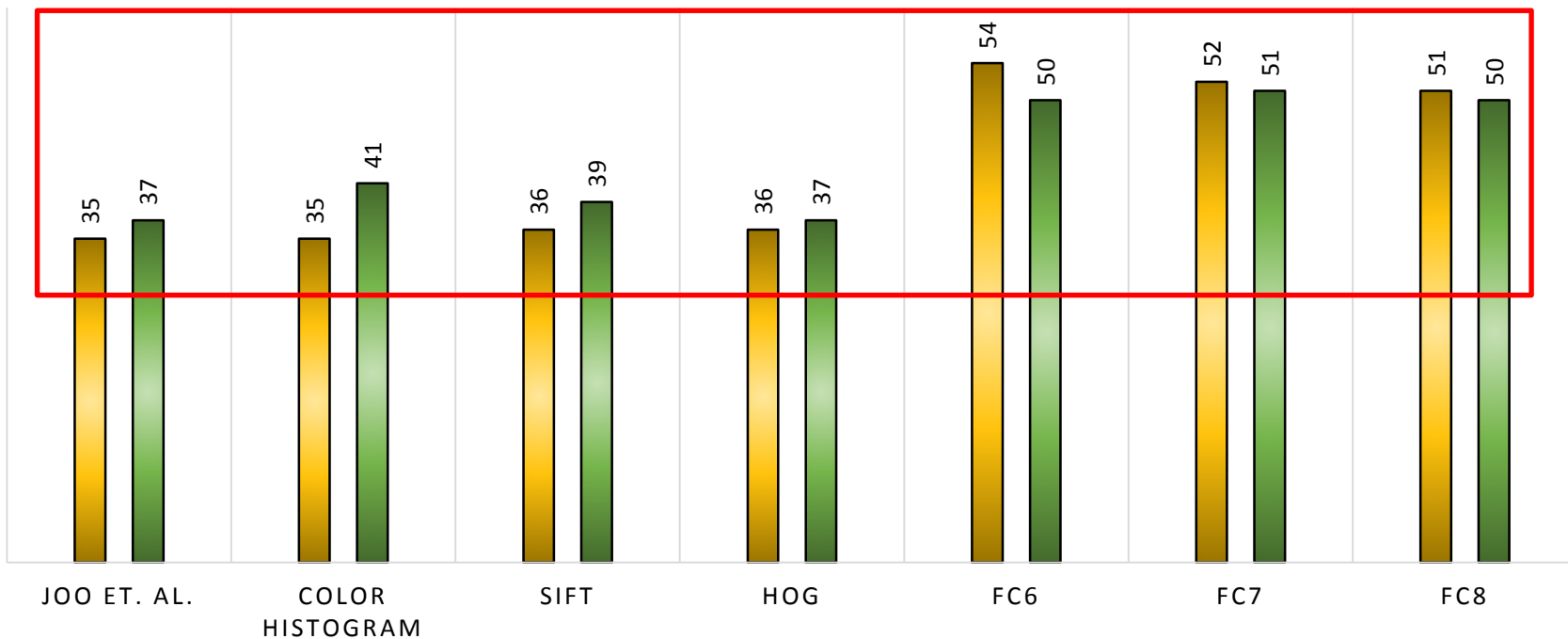




Results: Overview

ACCURACY

Direct Hierarchical

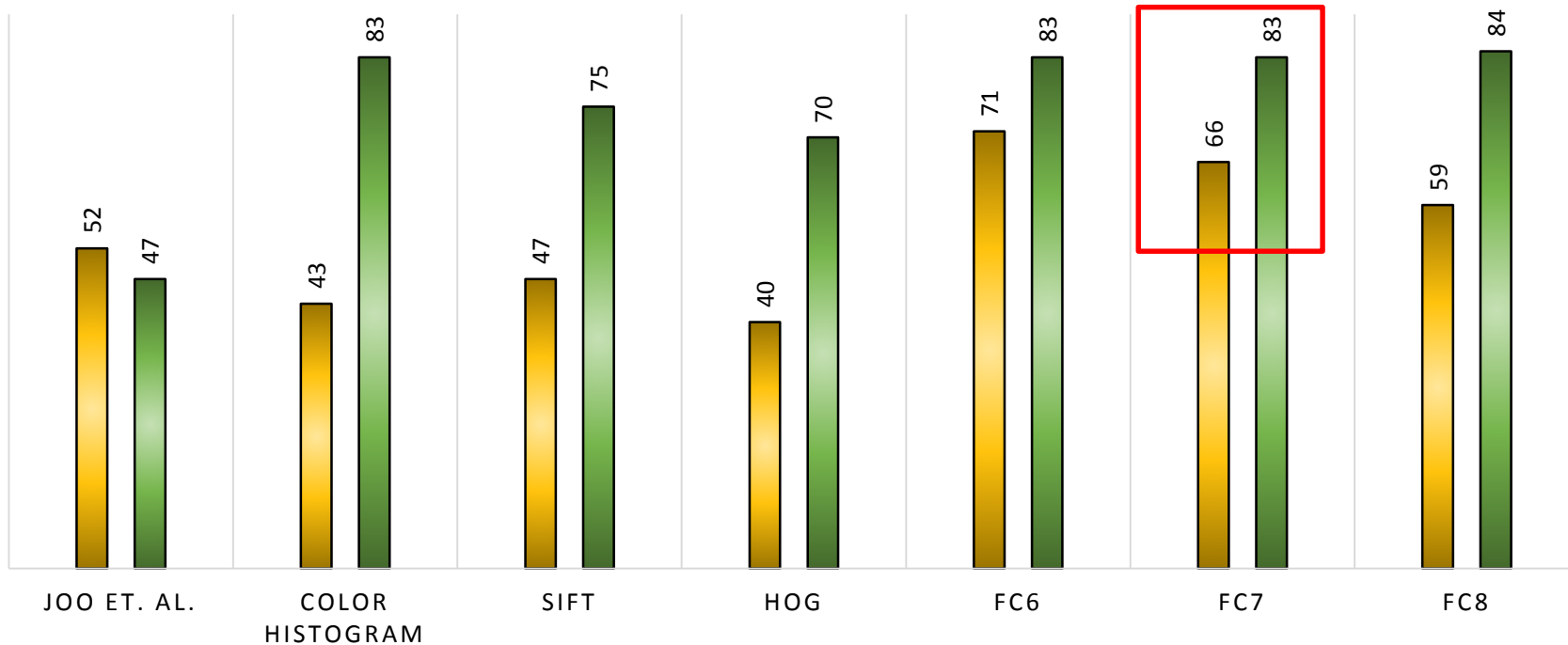




Results: Overview

SENSITIVITY

Direct Hierarchical

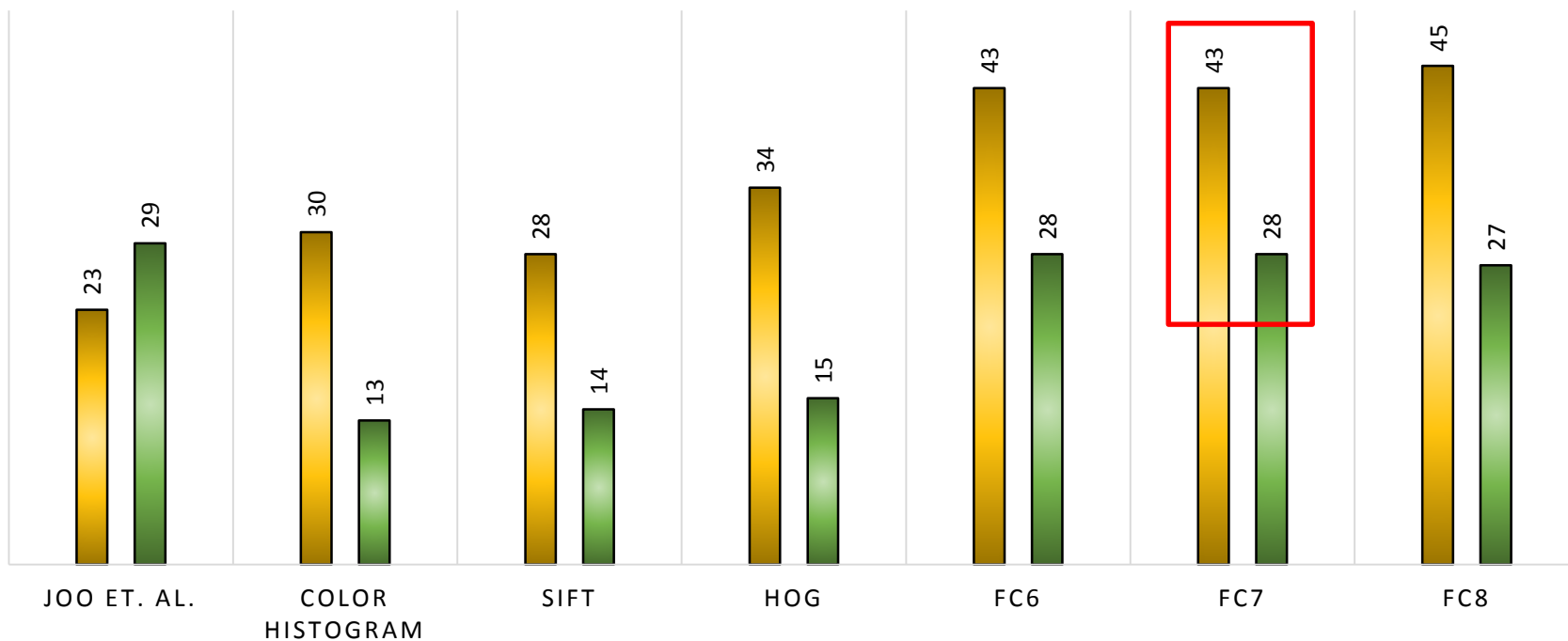




Results: Overview

SPECIFICITY

Direct Hierarchical





Conclusion

- Our method enables automated contents classification based on behaviors and intents of the portrayed subjects.
- It allows prompt intervention of human experts upon integrating the proposed methodology with mobile apps, social media websites, and media streaming websites.