

Spring 2021

ADVANCED TOPICS IN COMPUTER VISION

Atlas Wang

Assistant Professor, The University of Texas at Austin

A photograph of a young girl with blonde hair tied back in a ponytail, wearing a blue school uniform, sitting on a dark wooden bench and hugging a white and orange humanoid robot. The robot has a white head with an orange visor, orange shoulder panels, and orange knee pads. The girl is smiling and looking at the robot.

Computer Vision: Ethics and Privacy

A close-up photograph of a person's eye, showing the iris and pupil. The eye is brown with some redness or shadowing around the eyelid. The background is dark and out of focus.

Let's see an example: “Predicting Criminality from Facial Images”

Israeli startup, [Faception](#)

“Faception is first-to-technology and first-to-market with proprietary computer vision and machine learning technology for profiling people and revealing their personality based only on their facial image

Offering specialized engines for recognizing “High IQ”, “White-Collar Offender”, “Pedophile”, and “Terrorist” from a face image.

Main clients are in homeland security and public safety.

Predicting Criminality

[“Automated Inference on Criminality using Face Images” Wu and Zhang, 2016.](#)
arXiv

1,856 closely cropped images of faces;
Includes “wanted suspect” ID pictures
from specific regions.

*[...] angle θ from nose tip to two
mouth corners is on average 19.6%
smaller for criminals than for
non-criminals ...”*



[Physiognomy’s New Clothes](#)

Predicting Criminality - The Media Blitz...

arXiv Paper Spotlight: Automated Inference on Criminality Using Face ...

www.kdnuggets.com/.../arxiv-spotlight-automated-inference-criminality-face-images.... ▾

A recent paper by Xiaolin Wu (McMaster University, Shanghai Jiao Tong University) and Xi Zhang (Shanghai Jiao Tong University), titled "Automated Inference ...

Automated Inference on Criminality Using Face Images | Hacker News

<https://news.ycombinator.com/item?id=12983827> ▾

Nov 18, 2016 - The automated inference on criminality eliminates the variable of meta-accuracy (the competence of the human judge/examiner) all together.

A New Program Judges If You're a Criminal From Your Facial Features ...

<https://motherboard.vice.com/.../new-program-decides-criminality-from-facial-feature...> ▾

Nov 18, 2016 - In their paper 'Automated Inference on Criminality using Face Images', published on the arXiv pre-print server, Xiaolin Wu and Xi Zhang from ...

Can face classifiers make a reliable inference on criminality?

<https://techxplore.com/.../computer-sciences> ▾

Nov 23, 2016 - Their paper is titled "Automated Inference on Criminality using Face Images ... face classifiers are able to make reliable inference on criminality.

Troubling Study Says Artificial Intelligence Can Predict Who Will Be ...

<https://theintercept.com/.../troubling-study-says-artificial-intelligence-can-predict-who...> ▾

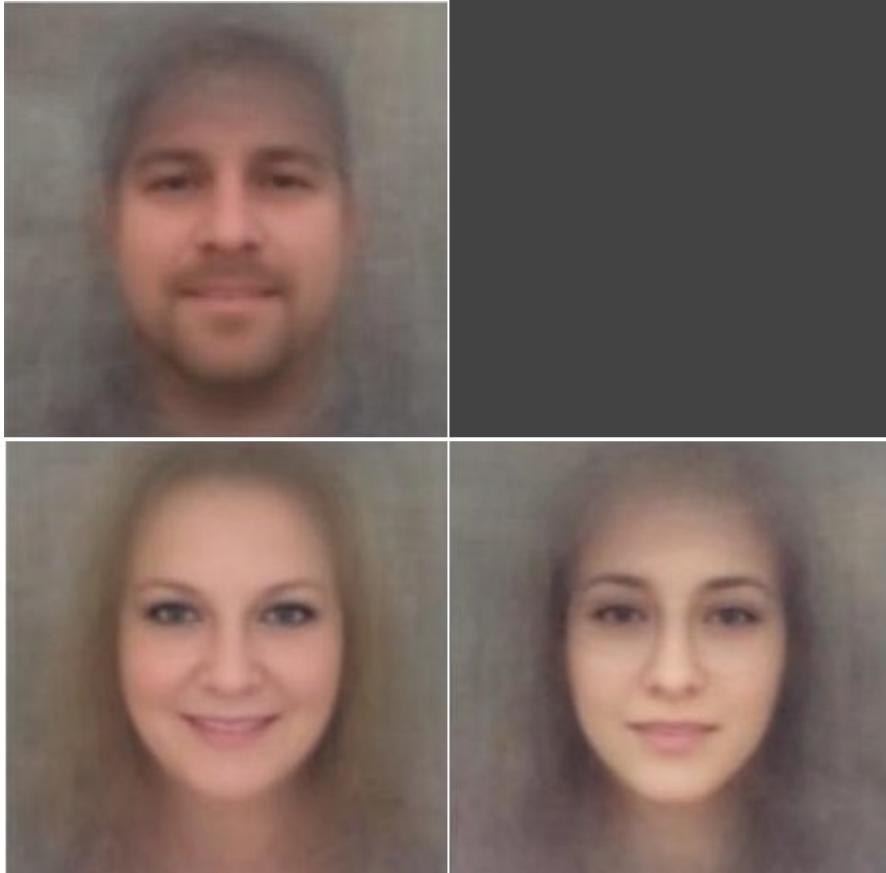
Nov 18, 2016 - Not so in the modern age of Artificial Intelligence, apparently: In a paper titled "Automated Inference on Criminality using Face Images," two ...

Automated Inference on Criminality using Face Images (via arXiv ...)

<https://computationallegalstudies.com/.../automated-inference-on-criminality-using-fa...> ▾

Dec 6, 2016 - Next Next post: A General Approach for Predicting the Behavior of the Supreme Court of the United States (Paper Version 2.01) (Katz, ...

Let's see another example: “Predicting Homosexuality”



Composite Straight Faces

Composite Gay Faces

- Wang and Kosinski, [Deep neural networks are more accurate than humans at detecting sexual orientation from facial images](#)
- “Sexual orientation detector” using 35,326 images from public profiles on a US dating website.
- “Consistent with the prenatal hormone theory [PHT] of sexual orientation, gay men and women tended to have gender-atypical facial morphology.”

Predicting Homosexuality

Differences between lesbian or gay and straight faces in selfies relate to grooming, presentation, and lifestyle — that is, **differences in culture, not in facial structure**

See more on Medium: [“Do Algorithms Reveal Sexual Orientation or Just Expose our Stereotypes?”](#)



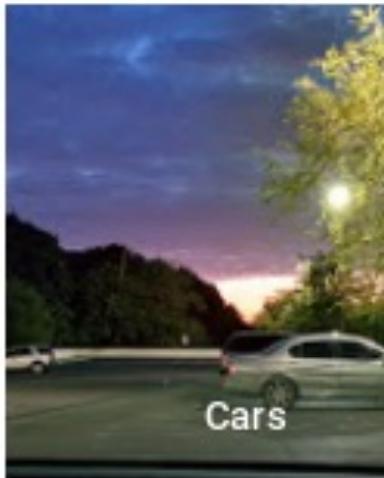
Bias and fairness



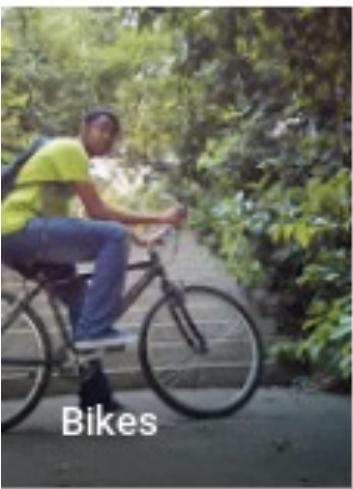
Skyscrapers



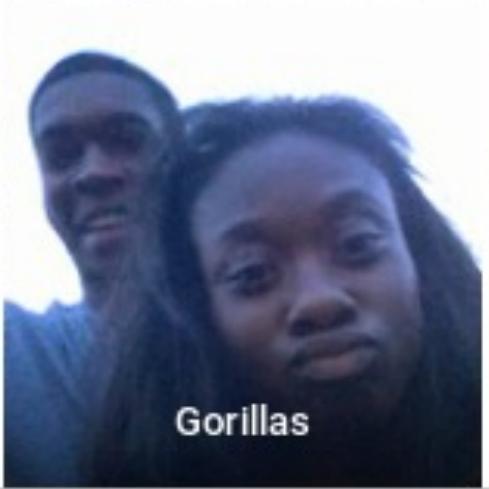
Airplanes



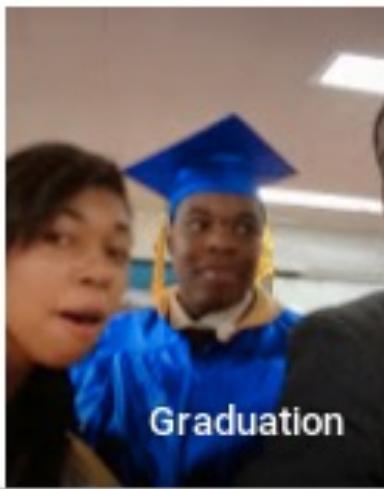
Cars



Bikes



Gorillas



Graduation

<https://bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas/>

Bias and fairness

- Concerns
 - AI will inadvertently absorb biases from data
 - Making important decisions based on biased data will exacerbate bias: especially for law enforcement, employment, loans, health insurance, etc.
 - Even well-intentioned applications can create negative side effects: filter bubbles, targeted advertising
 - Outcomes cannot be appealed because AI systems are opaque and proprietary
- Potential solutions
 - Regulation and transparency: e.g., [right to explanation](#)
 - More inclusivity among AI technologists: [AI4ALL](#)

Human Biases in Data

Reporting bias	Stereotypical bias	Group attribution error
Selection bias	Historical unfairness	Halo effect
Overgeneralization	Implicit associations	
Out-group homogeneity bias	Implicit stereotypes	
	Prejudice	

Training data are collected and annotated

Human Biases in Collection and Annotation

Sampling error	Bias blind spot	Neglect of probability
Non-sampling error	Confirmation bias	Anecdotal fallacy
Insensitivity to sample size	Subjective validation	Illusion of validity
Correspondence bias	Experimenter's bias	
In-group bias	Choice-supportive bias	

Evaluate for Fairness & Inclusion: Confusion Matrix

		Model Predictions		
		Positive	Negative	
References	Positive	<ul style="list-style-type: none">• Exists• Predicted True Positives	<ul style="list-style-type: none">• Exists• Not predicted False Negatives	Recall, False Negative Rate
	Negative	<ul style="list-style-type: none">• Doesn't exist• Predicted False Positives	<ul style="list-style-type: none">• Doesn't exist• Not predicted True Negatives	
		Precision, False Discovery Rate	Negative Predictive Value, False Omission Rate	LR+, LR-

Evaluate for Fairness & Inclusion

Female Patient Results

True Positives (TP) = 10	False Positives (FP) = 1
False Negatives (FN) = 1	True Negatives (TN) = 488

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{10}{10 + 1} = 0.909$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{10}{10 + 1} = 0.909$$

Male Patient Results

True Positives (TP) = 6	False Positives (FP) = 3
False Negatives (FN) = 5	True Negatives (TN) = 48

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{6}{6 + 3} = 0.667$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{6}{6 + 5} = 0.545$$

Evaluate for Fairness & Inclusion

Female Patient Results

True Positives (TP) = 10	False Positives (FP) = 1
False Negatives (FN) = 1	True Negatives (TN) = 488

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{10}{10 + 1} = 0.909$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{10}{10 + 1} = 0.909$$

Male Patient Results

True Positives (TP) = 6	False Positives (FP) = 3
False Negatives (FN) = 5	True Negatives (TN) = 48

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{6}{6 + 3} = 0.667$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{6}{6 + 5} = 0.545$$

“Equality of Opportunity” fairness criterion:
Recall is equal across subgroups

Evaluate for Fairness & Inclusion

Female Patient Results

True Positives (TP) = 10	False Positives (FP) = 1
False Negatives (FN) = 1	True Negatives (TN) = 488

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{10}{10 + 1} = 0.909$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{10}{10 + 1} = 0.909$$

Male Patient Results

True Positives (TP) = 6	False Positives (FP) = 3
False Negatives (FN) = 5	True Negatives (TN) = 48

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{6}{6 + 3} = 0.667$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{6}{6 + 5} = 0.545$$

“Predictive Parity” fairness criterion:
Precision is equal across subgroups

Towards Fairness in Visual Recognition (CVPR'20)

MODEL NAME	MODEL	TEST INFERENCE	BIAS (\downarrow)	ACCURACY (% , \uparrow)		
				COLOR	GRAY	MEAN
BASELINE OVERSAMPLING	N-way softmax	$\arg \max_y P(y x)$	0.074	89.0	88.0	88.5 ± 0.3
	N-way softmax, resampled	$\arg \max_y P(y x)$	0.066	89.2	89.1	89.1 ± 0.4
ADVERSARIAL	w/ uniform confusion [1, 46]	$\arg \max_y P(y x)$	0.101	83.8	83.9	83.8 ± 1.1
	w/ ∇ reversal, proj. [51]	$\arg \max_y P(y x)$	0.094	84.6	83.5	84.1 ± 1.0
DOMAINDISCRIM	joint ND-way softmax	$\arg \max_y \sum_d P_{\text{tr}}(y, d x)$	0.844	88.3	86.4	87.3 ± 0.3
		$\arg \max_y \max_d P_{\text{te}}(y, d x)$	0.040	91.3	89.3	90.3 ± 0.5
		$\arg \max_y \sum_d P_{\text{te}}(y, d x)$	0.040	91.2	89.4	90.3 ± 0.5
	RBA [52]	$y = \mathcal{L}(\sum_d P_{\text{tr}}(y, d x))$	0.054	89.2	88.0	88.6 ± 0.4
DOMAININDEPEND	N-way classifier per domain	$\arg \max_y P_{\text{te}}(y d^*, x)$	0.069	89.2	88.7	88.9 ± 0.4
		$\arg \max_y \sum_d s(y, d, x)$	0.004	92.4	91.7	92.0 ± 0.1

Table 1. Performance comparison of algorithms on CIFAR-10S. All architectures are based on ResNet-18 [20]. We investigate multiple bias mitigation strategies, and demonstrate that a domain-independent classifier outperforms all baselines on this benchmark.

Computer Vision Everywhere = Privacy Intrusion?



Smart home, Smart hospitals, Behavior study and data sharing ...



**MIT
Technology
Review**

Facial Recognition Technology Raises Privacy Concerns

Amazon's camera-equipped Echo Look raises new questions about smart home privacy

Facial recognition has to be regulated to protect the public, says AI report

The research institute AI Now has identified facial recognition as a key challenge for society and policymakers —but is it too late?



In the Practice ...

The Dilemma

- We would like a camera system to recognize important events and assist human daily life by understanding its videos
- ... while preventing it from obtaining “too sensitive” visual information that can intrude people's privacy.
- Would classical cryptographic solutions suffice?
 - They secure the communication against unauthorized access from attackers
 - But not applicable to preventing authorized agents (such as the backend analytics) from the unauthorized abuse of information

Existing Solutions

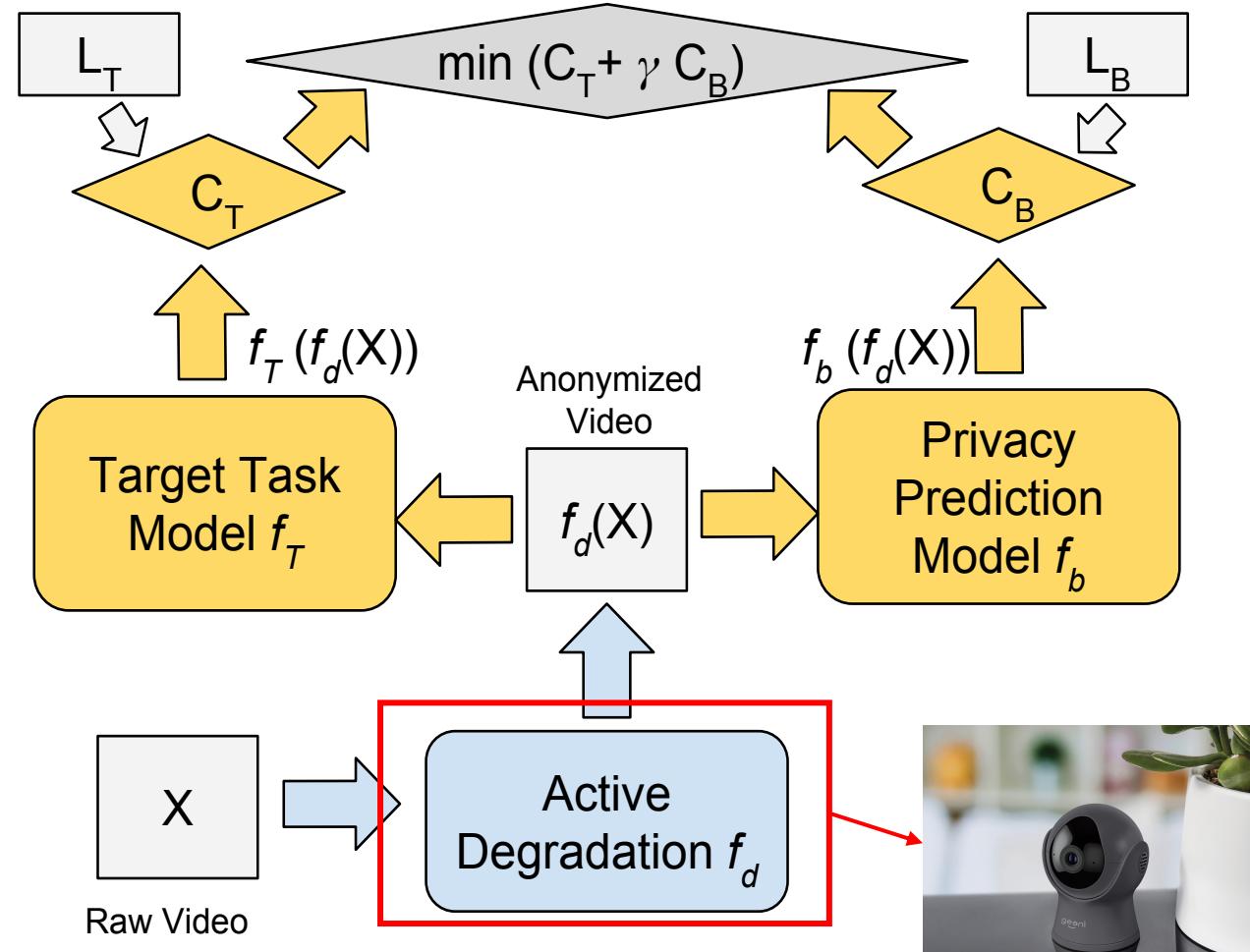
- Privacy Protection in Computer Vision Systems
 - Transmit feature descriptors to the cloud? **Not safe**
 - Homomorphic cryptographic solution? **Expensive, working on only simple classifiers**
 - Downsample the video aggressively, and strategically? **Cheap, works empirically, but usually no competitive trade-off**
 - A few game-theoretic or learning-based recent solutions ... **IMPORTANT to distinguish between model-specific and model-agnostic privacy!**
- Privacy Protection in Social Media and Photo Sharing
 - Add empirical obfuscations? **Not safe, sometimes sacrificing utility**
 - Deep learning-based adversarial perturbations? **model-specific privacy, and may no longer generalize when the computer vision models are upgraded ...**

Privacy Protection via Adversarial Training (ECCV'18, IEEE TPAMI'2-)

Our goal is to seek such a transform for the original data, such that on the transformed data:

- The achievable **target task performance** is minimally affected compared to using raw data
- The **privacy leak risk** is greatly suppressed compared to raw data
 - Can be defined by the predictive performance of the privacy attributes

It can be formulated via an adversarial deep learning framework.



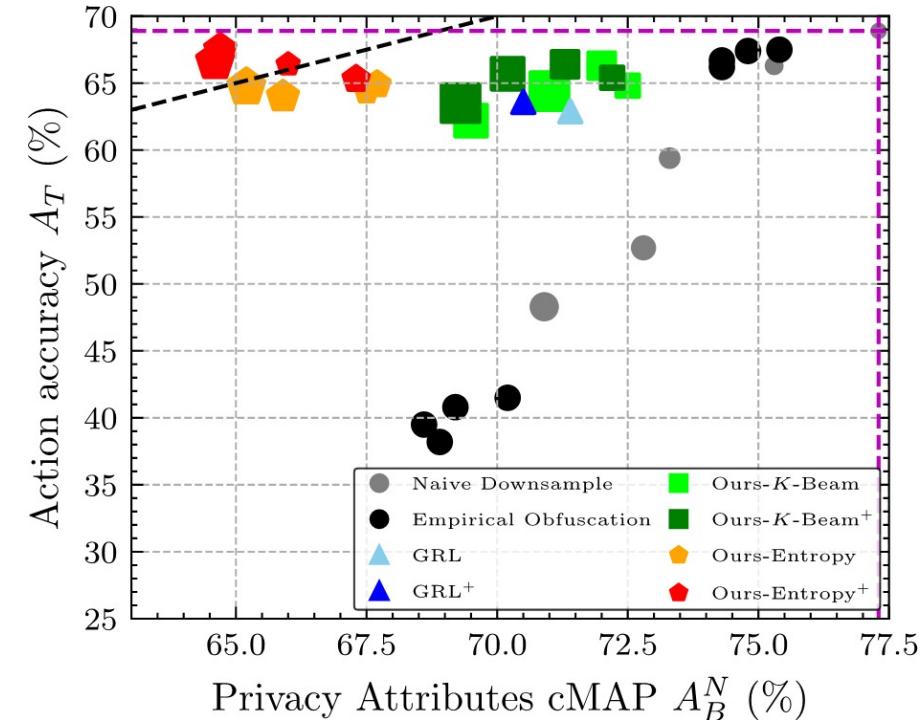
Result Visualization



A New Privacy CV Benchmark, and more

Frame	Action	Privacy Attributes
	brush hair	<ul style="list-style-type: none">• skin color: white• face: invisible• gender: female• nudity: semi-nudity• relationship: unidentifiable
	situp	<ul style="list-style-type: none">• skin color: black• face: completely visible• gender: male• nudity: semi-nudity• relationship: unidentifiable

Privacy Annotated HMDB51 (PA-HMDB51)



GOOD SYSTEMS
A UT Grand Challenge



Summary

- We should be aware of all these issues when developing computer vision technologies!
 - Privacy violations
 - Potential for deception, misuse and manipulation
 - Exacerbating bias and unfair outcomes
 - Lack of transparency and due process
 - Threats to human rights and dignity
 - Weaponization
 - Unintended consequences



The University of Texas at Austin
**Electrical and Computer
Engineering**
Cockrell School of Engineering