Vulnerability of Face Age Verification to Replay Attacks

Pavel Korshunov (www.idiap.ch/∼pkorshunov)

April 18, 2024

Children are exposed to harmful content online

Children are harmed online[1]

The New York Times | U.S.

Video Games and Online Chats Are 'Hunting Grounds' for Sexual Predators

Criminals are making virtual connections with children through gaming and social media platforms. One popular site warns visitors, "Please be careful."

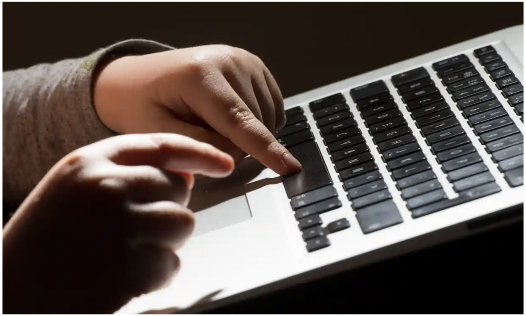
By NELLIE BOWLES and MICHAEL H. KELLER DEC. 7, 2019

[1] www.nytimes.com/interactive/2019/12/07/us/video-games-child-sex-abuse.html

Children are harmed online[2]

Almost half of children in England have seen harmful content online – survey

Children's commissioner raises fears of another tragedy like that of Molly Russell after poll findings



Children are being exposed to pornography, sexualised and violent imagery and anonymous trolling, as well as content featuring self-harm and suicide, according to the survey. Photograph: Dominic Lipinski/PA

[2] www.theguardian.com/society/2022/sep/29/almost-half-of-children-in-england-have-seen-harmful-content-online-survey

The legislation in UK[3]

GOV.UK

Topics

Home > Society and culture > Online safety

Press release

Landmark laws to protect children and stop abuse online published

The draft Online Safety Bill will help safeguard young people and clamp down on racist abuse, while upholding democratic debate online.

[3] www.gov.uk/government/news/landmark-laws-to-protect-children-and-stop-abuse-online-published

The legislation in EU[4]

REUTERS® World Business Legal Markets Breakingviews Technology Investigations More

Summary Law firms

- EU law cracks down on targeted ads, promoted content
- Tech firms that fail to remove illegal content face fines
- DSA could set global example, say child rights campaigners

Aa

BRUSSELS, July 12 (Thomson Reuters Foundation) - A new EU law to rein in tech giants could serve as a benchmark for worldwide legislation to protect children online, as concern grows globally about the impact of social media on young people, children's rights campaigners say.

The bloc's **Digital Services Act (DSA)** includes a ban on targeted advertising aimed at children and prohibits the algorithmic promotion of content that could be harmful for minors such as videos related to eating disorders or self-harm.

[4] www.reuters.com/legal/litigation/can-an-eu-law-save-children-harmful-content-online-2022-07-12/

The legislation in US[5]

The New York Times

Sweeping Children's Online Safety Bill Is Passed in California

The new rules, which would require many online services to increase protections for children, could change how popular social media and game platforms treat minors.

[5] www.nytimes.com/2022/08/30/business/california-children-online-safety.html

PAD and age verification

Age verification

- Rapidly expanding field
- Often used when recognition is not desirable
- Under-studied

Presentation attacks in age verification

- Easier to perform than for biometric systems
- No datasets available
- No research related to PAD for age verification

UTKPAD — dataset of face presentation attacks

Built specifically for age verification

- Original UTKFace dataset (20K images) of faces from 1 to 110 years old
- Replay attacks to iPhone 12, Galaxy S9, and Huawei Mate 30

Evaluation of the dataset

- Vulnerability of state of the art age verification systems
- Assess the existing attack detection systems developed for biometrics

Creating the dataset

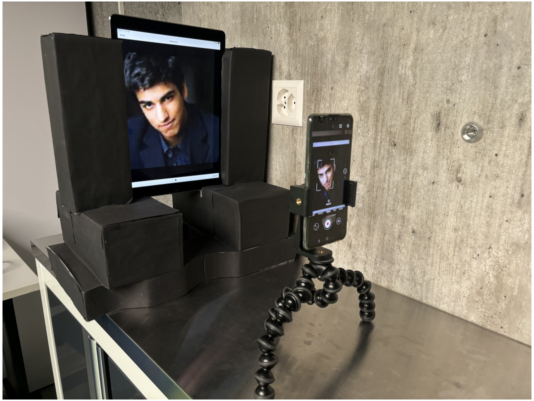
Process original images from UTKFace

- Upsample images (face region) with CodeFormer[6]
- Arrange all images into videos (display each image for 2 seconds)

[6] https://github.com/sczhou/CodeFormer

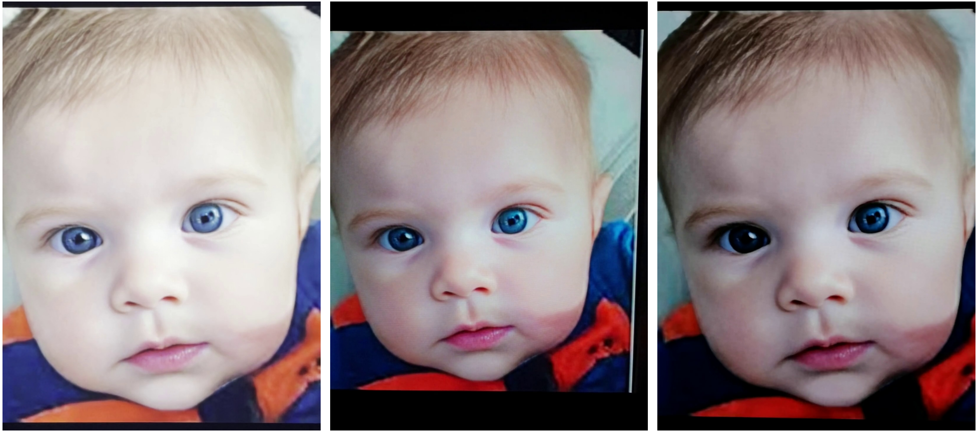
Replay the images

- Display videos on iPad Pro (2018)
- Recording room with controlled and consistent lighting
- Record videos with iPhone 12, Galaxy S9, and Huawei Mate 30



Extract final images

- Split recorded videos using the original order
- Save middle frame from each segment as an image
- UTKPAD: three sets of the same faces as in UTKFace



Vulnerability of age verification

Image-based age verification methods[7]

Classification

Classify into predefined classes

Regression

Regress to a true age or class

Regression via classification (RVC)

Many classifiers with average regressing to a true age

Distribution

Classification with Gaussian 'fuzzy' labels

Adaptive distribution

Gaussian sigma depends on aging characteristics

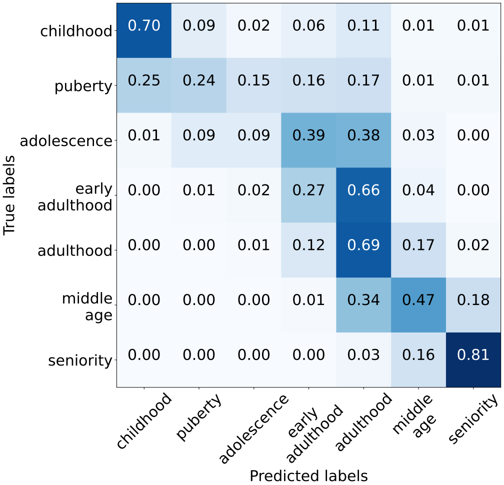
[7] P. Korshunov and S. Marcel, "Face Anthropometry Aware Audio-Visual Age Verification", ACM

Age detection on bona fide and replay attacks

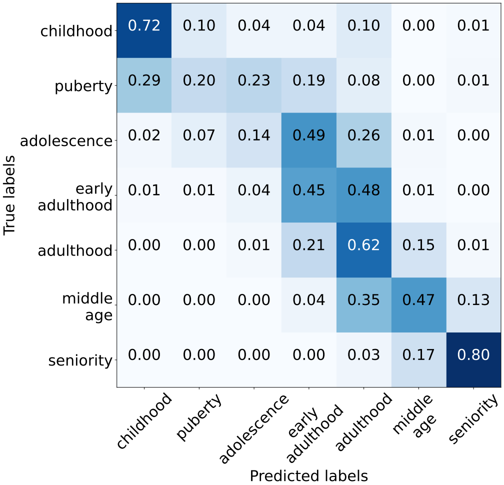
Classification scenario (predicting one age category out of seven)

Train DB	Method	Bona fide	iPhone 12	Galaxy S9	Huawei Mate 30
UTK	*adaptive*	0.599	0.566	0.567	0.586
Several	*rvc*	0.596	0.571	0.573	0.583
Several	*adaptive*	0.591	0.587	0.584	0.595
UTK	*class.*, ResNet50	0.591	0.540	0.561	0.561
Several	*distribution*	0.589	0.574	0.585	0.597
UTK	*rvc*	0.581	0.534	0.554	0.573
UTK	*classification*	0.574	0.529	0.543	0.560
Several	*classification*	0.567	0.516	0.510	0.536

Confusion matrices on bona fide and replay attacks



(a) *adaptive* on original UTKFace



(b) *adaptive* on attacks with Huawei

Assess existing PAD systems

Presentation attack detection for face biometrics[10]

DeepPixBiS

Fully connected network trained on image patches

CDCN++

Central difference convolution network with multiscale attention fusion module

Trained on protocol 1 of the OULU-NPU dataset

Well-known dataset of photo and replay attacks

[8] A. George and S. Marcel, "Deep pixel-wise binary supervision for face presentation attack detection", ICB 2019

[9] Z. Yu *et al.*, "Searching central difference convolutional networks for face anti-spoofing", CVPR 2020

[10] Z. Boulkenafet *et al.*, "OULU-NPU: A mobile face presentation attack database with real-world variations", IEEE FG 2017

Performance on PAD for biometrics

Accuracy of attack detection following OULU-NPU Protocol 1

	ACER $\downarrow$ (EER)	ACER $\downarrow$ (BPCER5)	ACER $\downarrow$ (BPCER20)
DeepPixBiS	2.1	20.0	9.6
CDCN++	7.5	8.3	6.2

Performance on UTKPAD

The threshold is set on OULU-NPU *dev* set

Model	Replay attacks	ACER $\downarrow$ (EER)	ACER $\downarrow$ (BPCER5)	ACER $\downarrow$ (BPCER20)
DeepPixBiS	iPhone 12	38.9	40.1	37.3
	Galaxy S9	48.7	52.4	50.2
	Huawei Mate 30	57.7	59.9	59.4
CDCN++	iPhone 12	45.4	34.8	42.9
	Galaxy S9	52.9	51.9	53.3
	Huawei Mate 30	61.2	61.3	63.3

Impact of replay attacks on age verification

- Face age verification is vulnerable to replay attacks
- Existing presentation attack detection is not generalizing
- We need more dataset with children and corresponding attacks
- The critical issue of children safety drives this type of work

Scan QR for source code

Related papers

- P. Korshunov and S. Marcel **"Face Anthropometry Aware Audio-visual Age Verification"**, ACM Multimedia 2022.
- A. George and S. Marcel, **"Deep pixel-wise binary supervision for face presentation attack detection"**, ICB 2019

Thank you for your attention!

Pavel Korshunov (www.idiap.ch/∼pkorshunov)

Idiap Research Institute, Martigny, Switzerland