HUMAN DISTINGUISHABLE VISUAL KEY FINGERPRINTS

Mozhgan Azimpourkivi[1], Umut Topkara[1], and Bogdan Carbunar[2]

[1] Bloomberg LP
[2] Florida International University

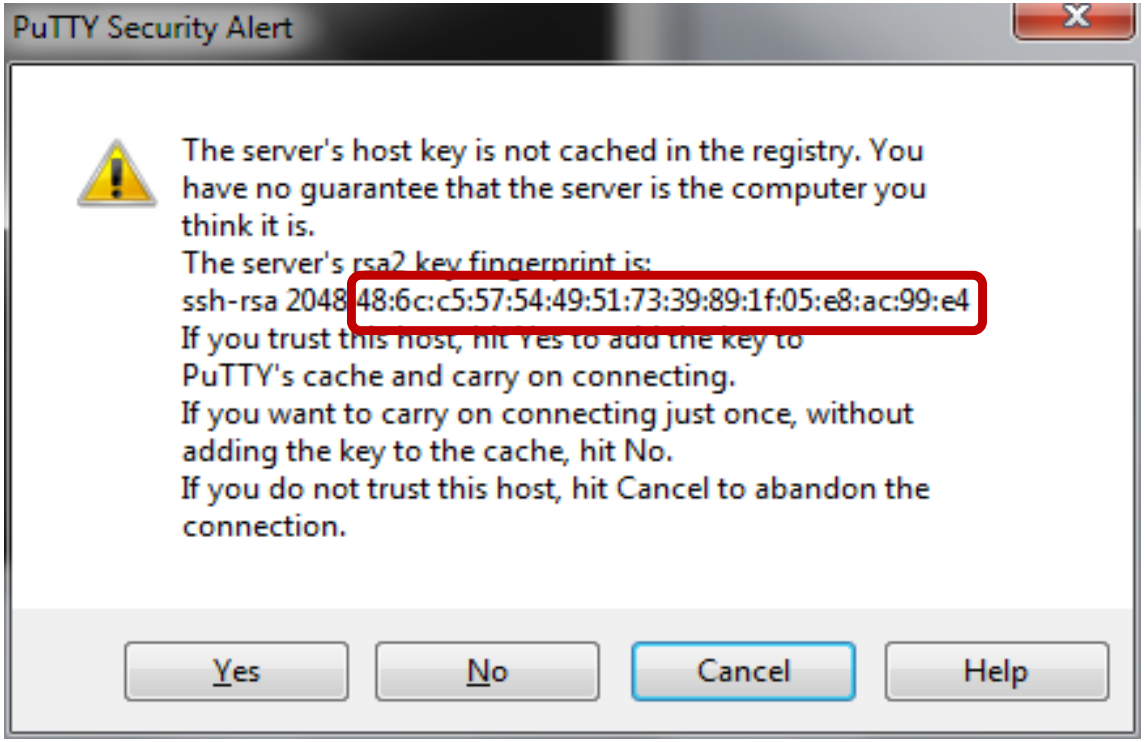
CaSPRLab
Cyber Security and Privacy Research

August 2020

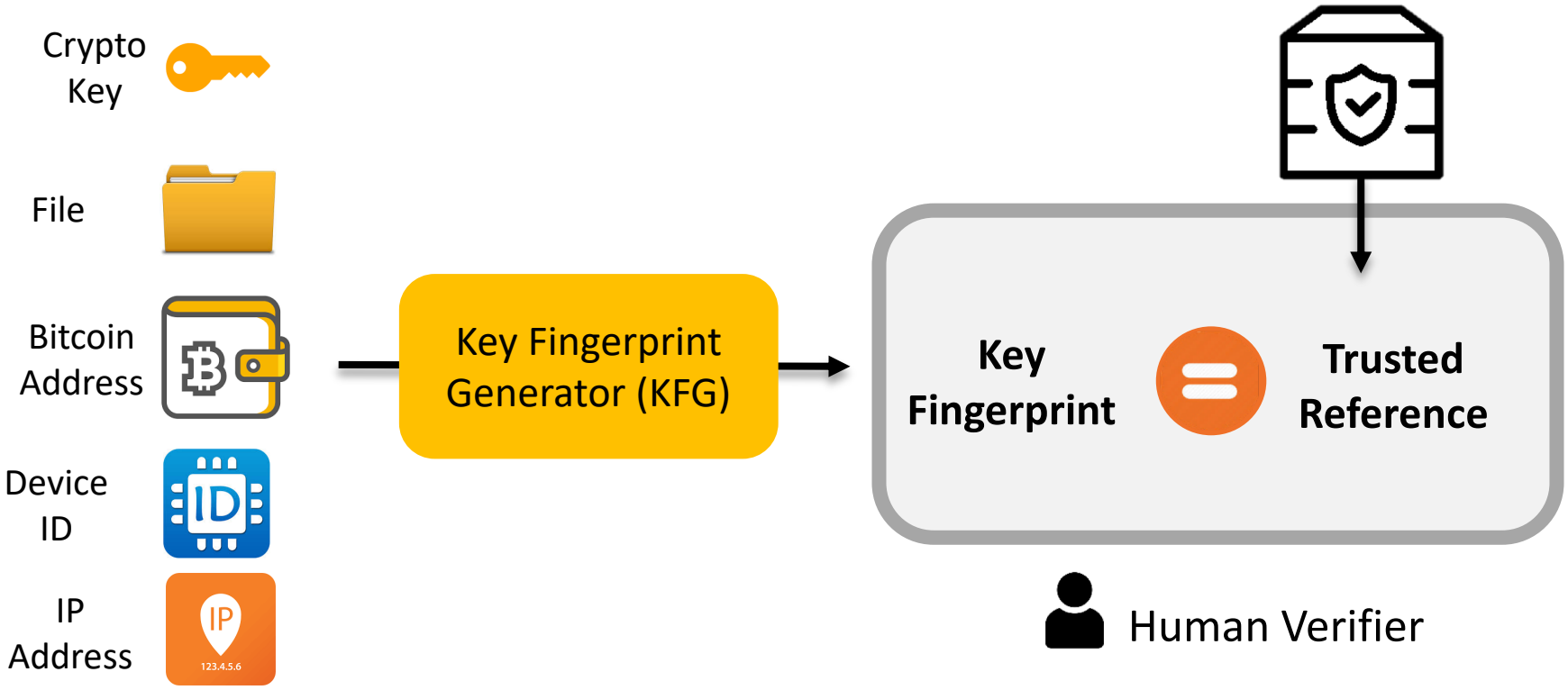
USENIX Security '20

Key Fingerprints (KF)

- Compact version of a crypto key
- Used for authentication
  - Easier to compare by humans against reference value

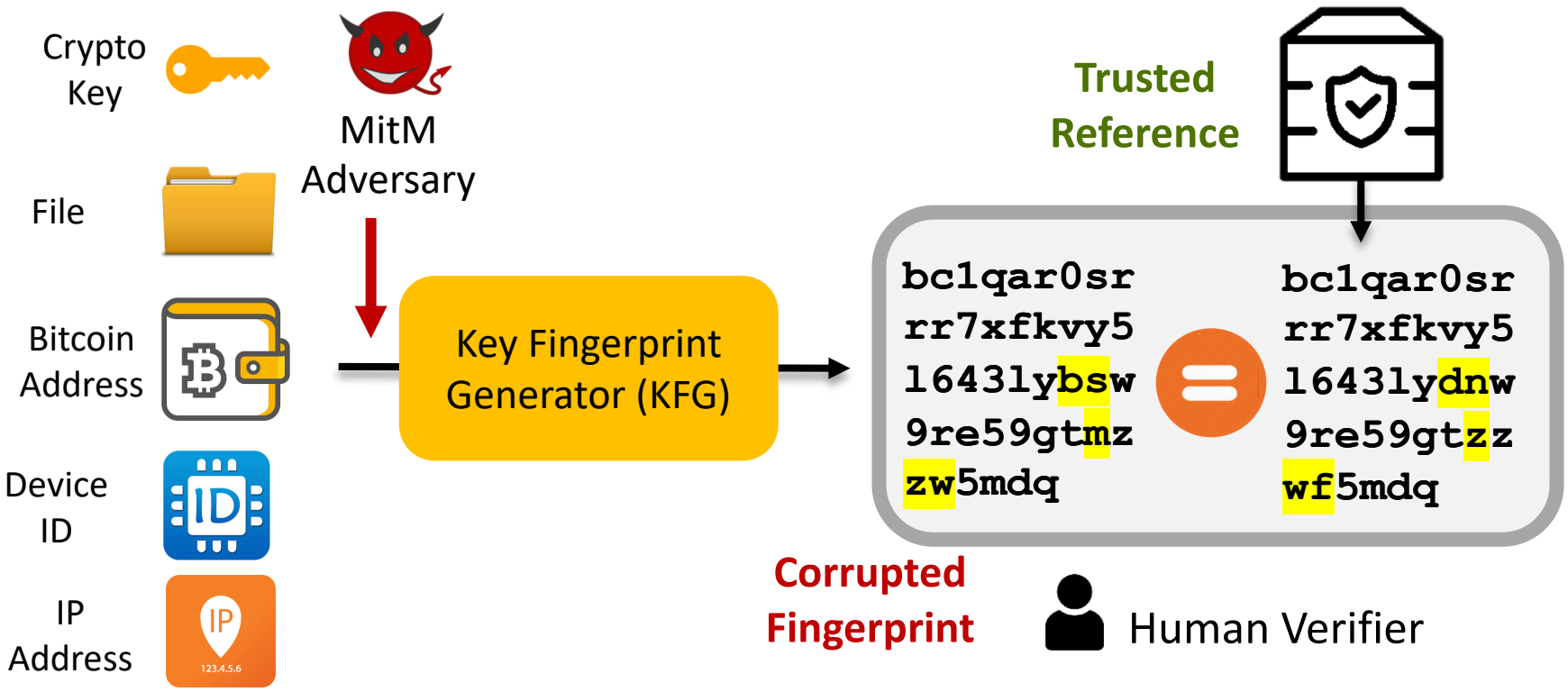


Key Fingerprint Authentication



Adversary Model

Generate and inject string whose key representation is *human-indistinguishable* from expected value



Applications

1. Remote authentication (SSH, OpenPGP/GnuPG)
   - ❑ Encode pub key hash into human readable format
2. End-to-End Encrypted (E2EE) messaging applications
   - ❑ WhatsApp, Viber, Facebook messenger
3. Device pairing
   - ❑ Bluetooth Secure Simple Pairing using ECDH
4. Prevent phishing & Bitcoin clipboard hijacking attacks
5. File checksums

Example Key Fingerprints (KF)

3A70 F9A0 4ECD B5D7 8A89 D32C EDA0 A352 66E2 C53D	learning equal education bent collar religion new shelf angle table train sad keep meal	The basket ends your right cat on his linen. Her range repeats her nerve.
Alphanumeric	Pronounceable words	Sentences

OpenSSH Visual Host Key

Vash

Unicorn

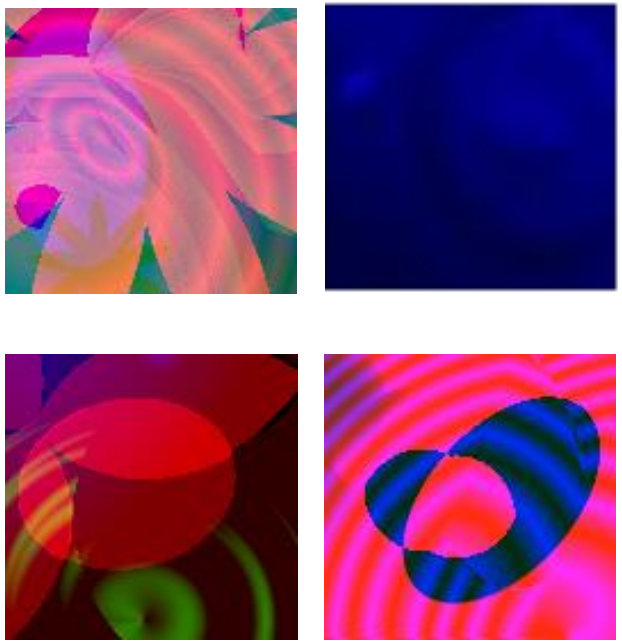
Textual representation

Visual representation

Tan, Joshua, et al. "Can unicorns help users compare crypto key fingerprints?." *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017.

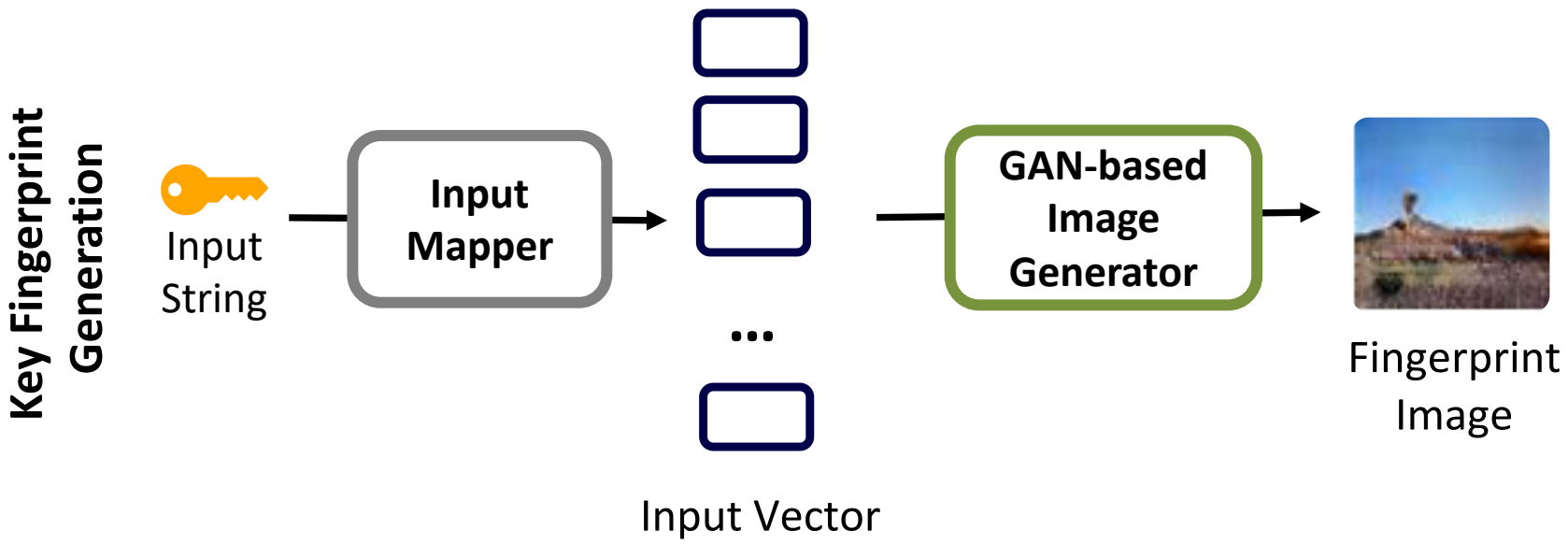
Vash: Visual KFG (VKFG)

- Tan et al., CHI'17:
  - Visual representations verified faster and easier than text-based
- Generate images using
  - Set of rules
  - Hand curated functions
- Human visual system limitations
  - Human error rate > 10%

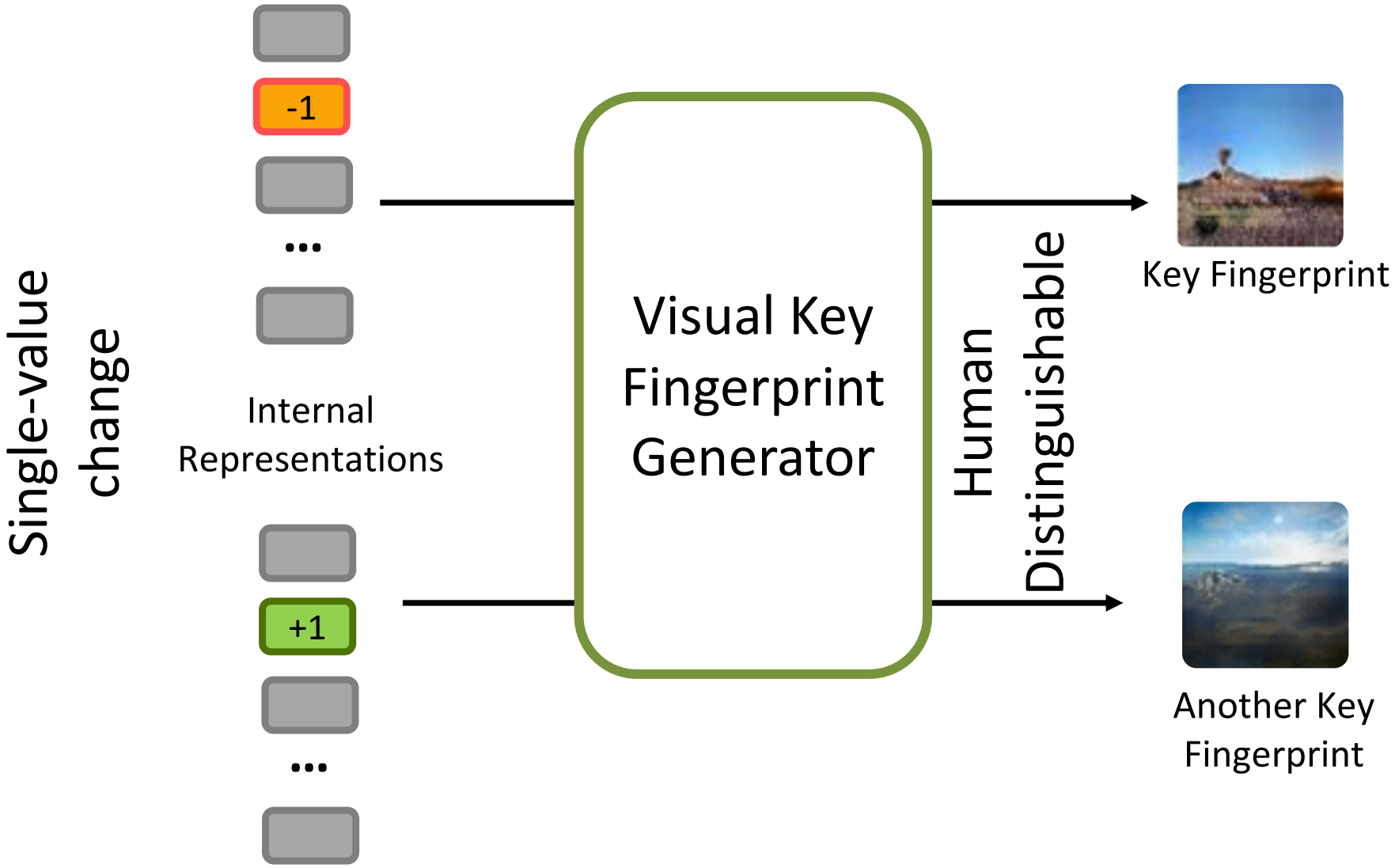


CEAL: DNN for KFG

Generate *realistic* images to improve usability

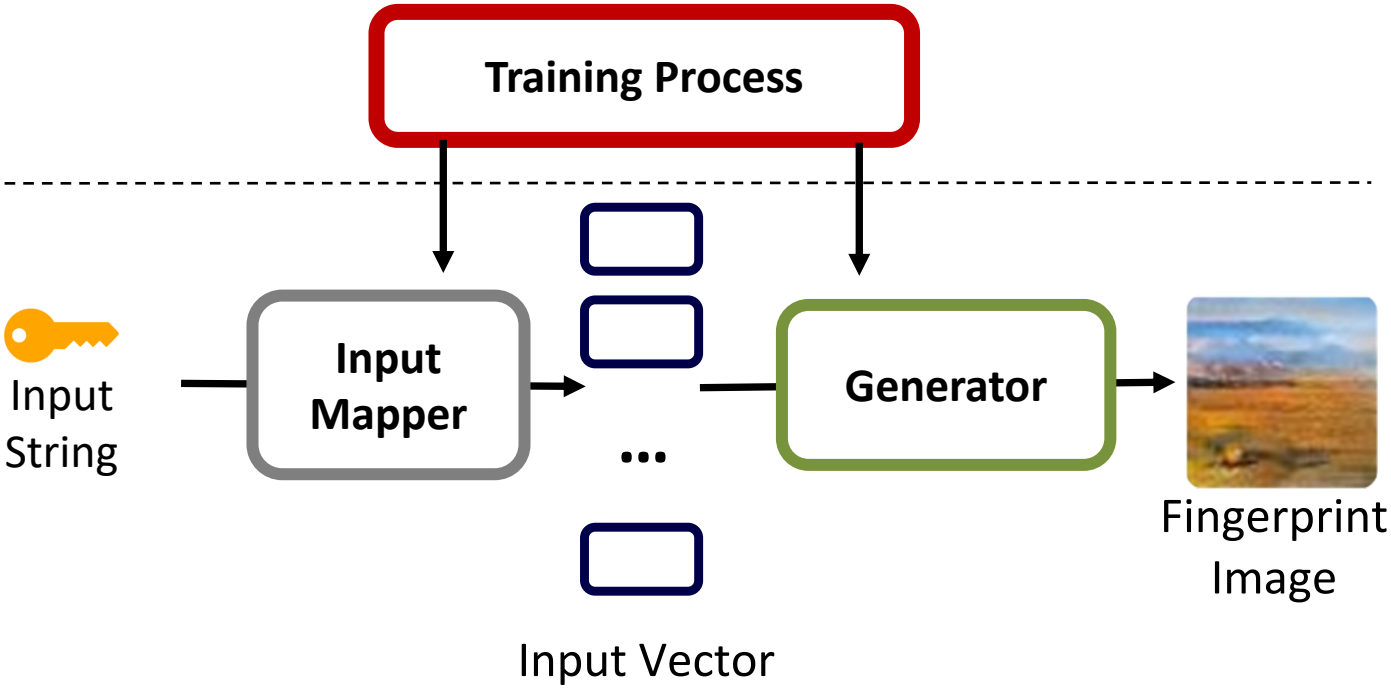


Visual Key Fingerprint Generator

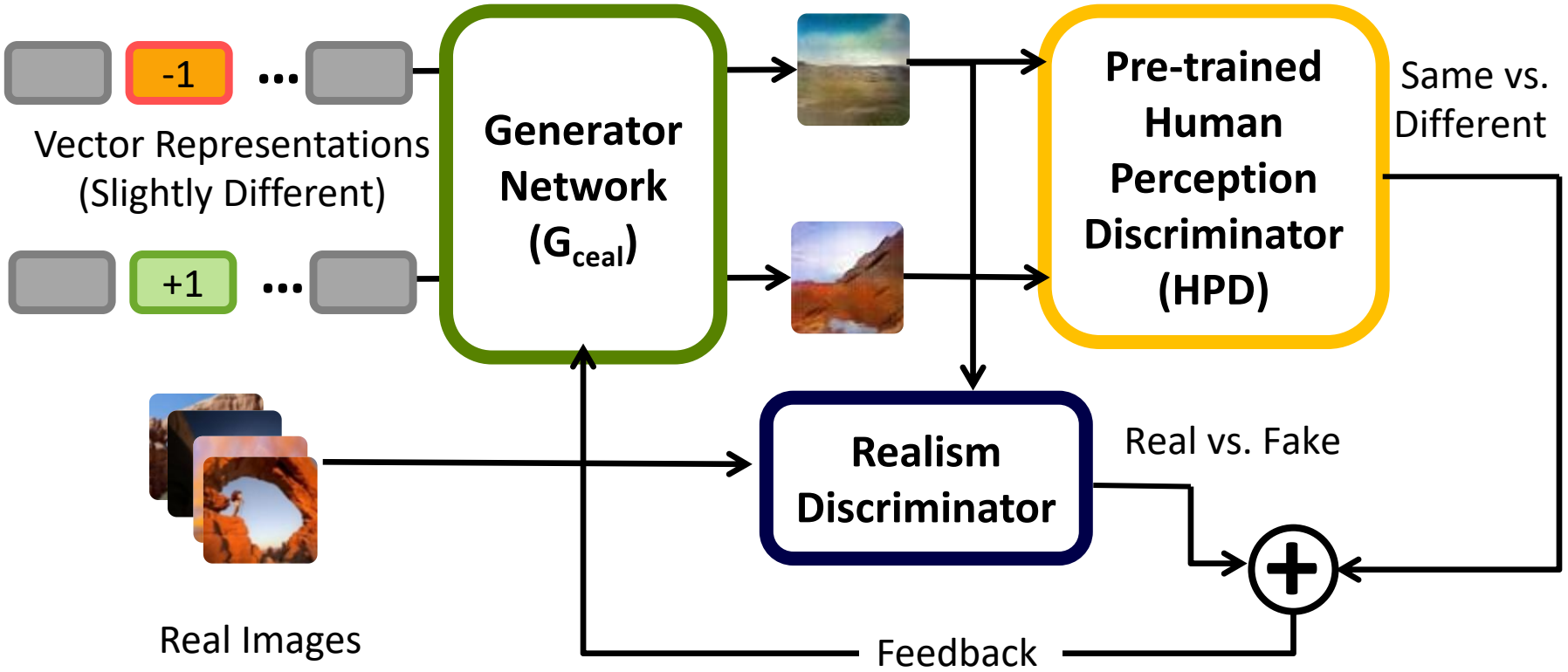


CEAL (CrEdential Assurance Label)

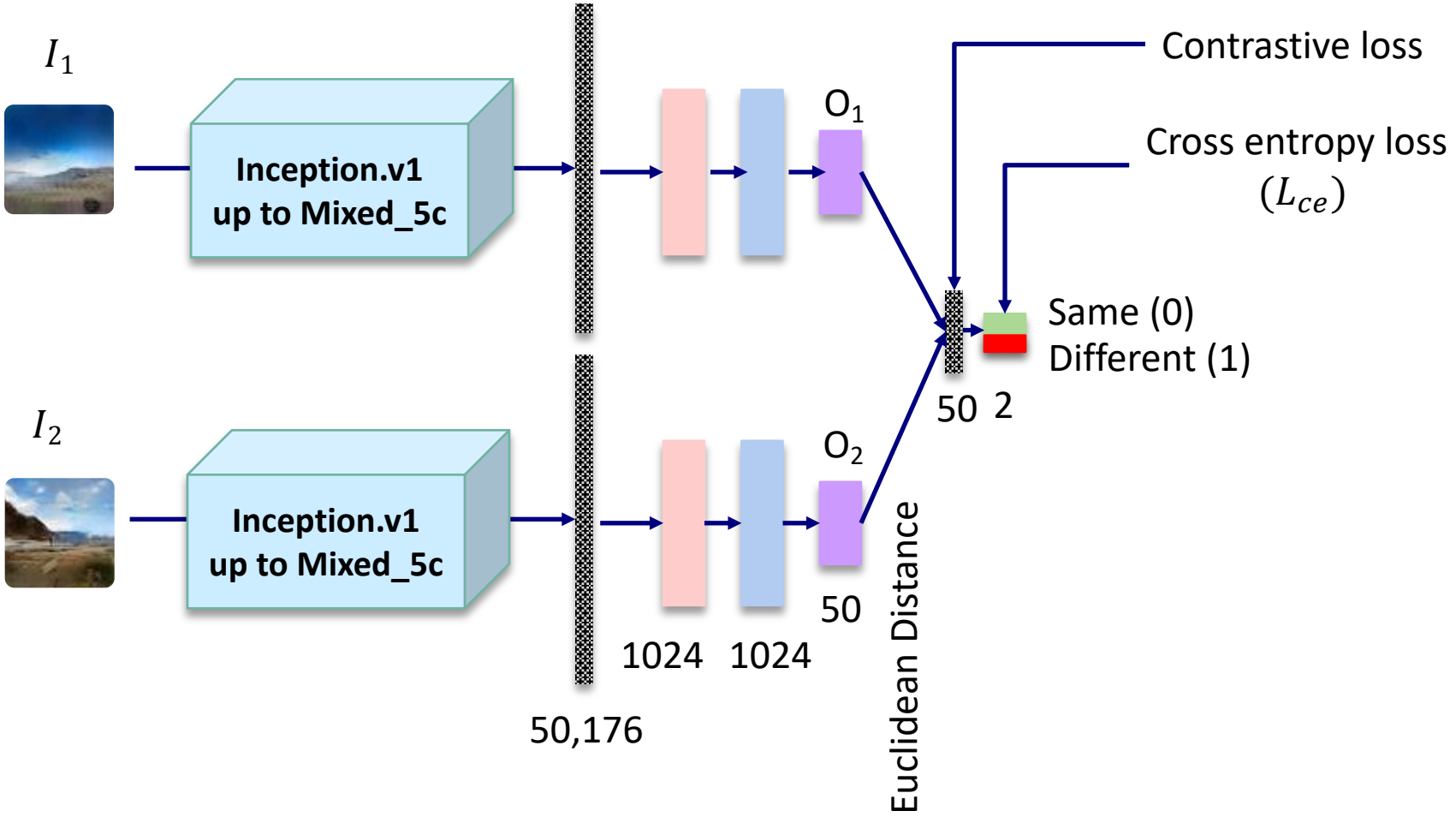
- Fingerprints should be *realistic* and *human-distinguishable*
- Remove humans from evaluation process



CL-GAN


-1
+1
Vector Representations
(Slightly Different)
Generator Network
(G_{ceal})
Pre-trained Human Perception Discriminator (HPD)
Same vs. Different
Realism Discriminator
Real vs. Fake
Real Images
Feedback

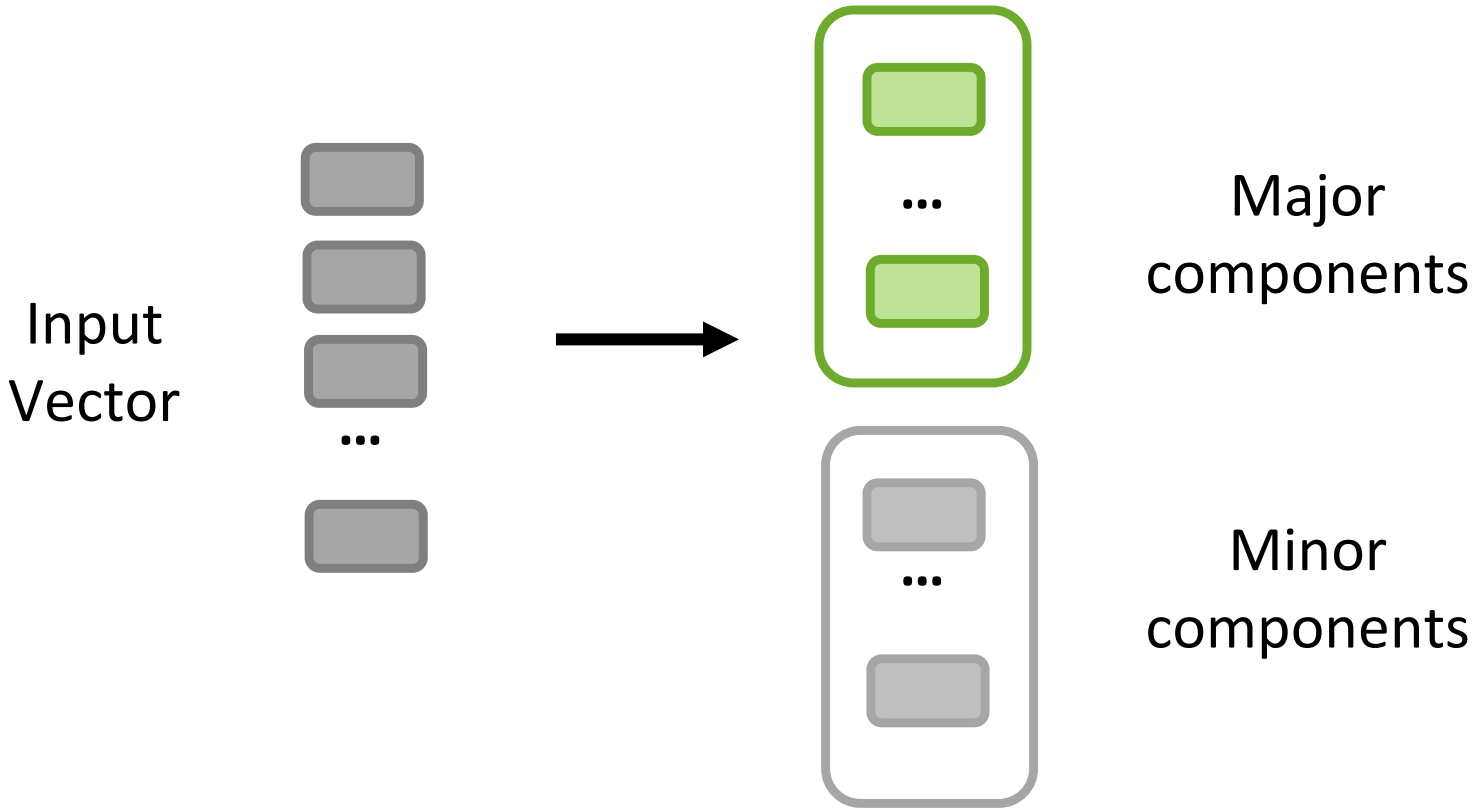
Human Perception Discriminator (HPD)



HPD Evaluation

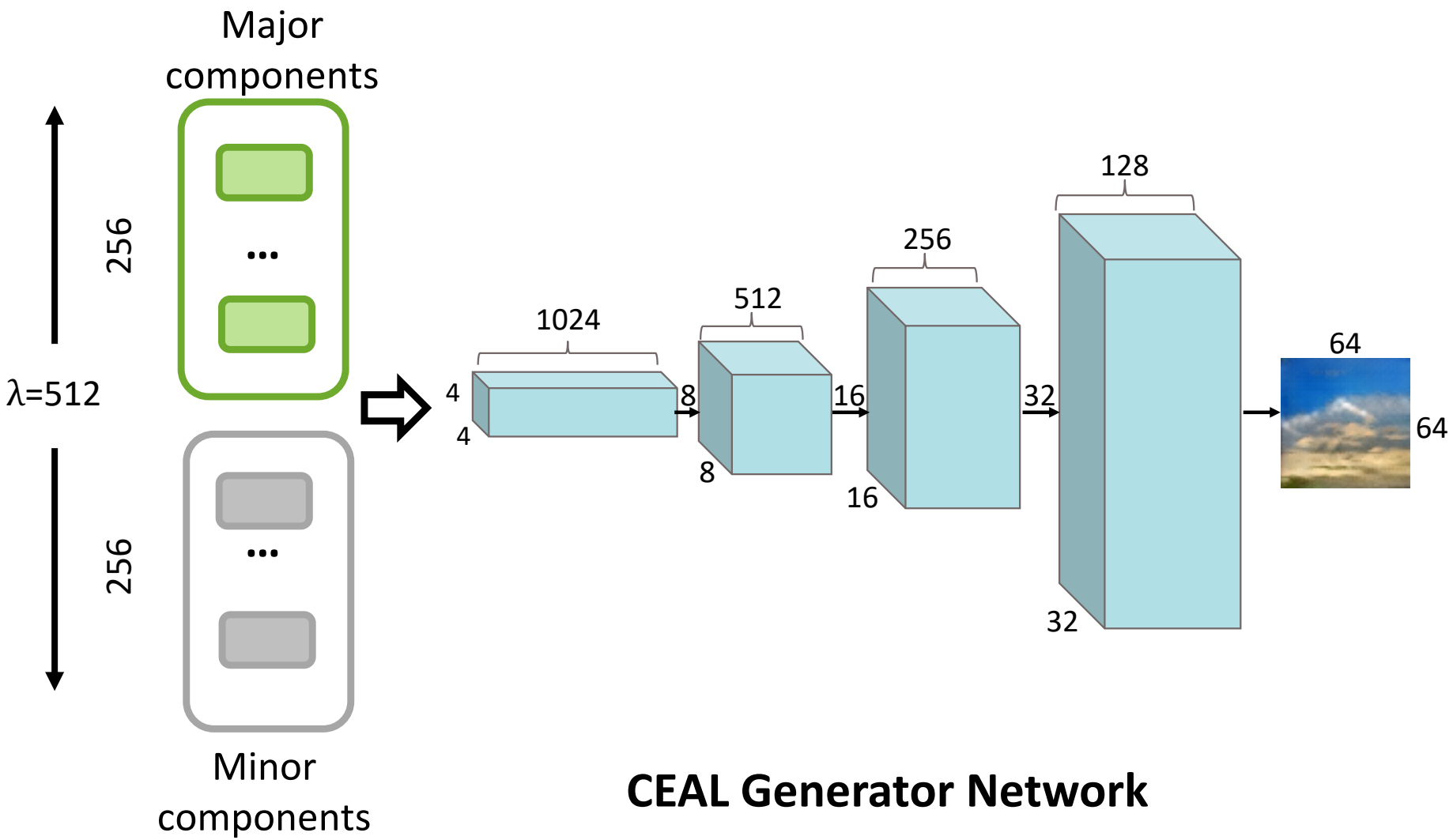
- Training: > 26,000 image pairs
  - 558 labeled by Mechanical Turk (MTurk) workers
    - Each image labeled by up to 100 workers
  - 26,244 synthetically generated images
- 84% Precision, 82% F1-score
  - Holdout subset of 112 image pairs

Major vs. Minor Components



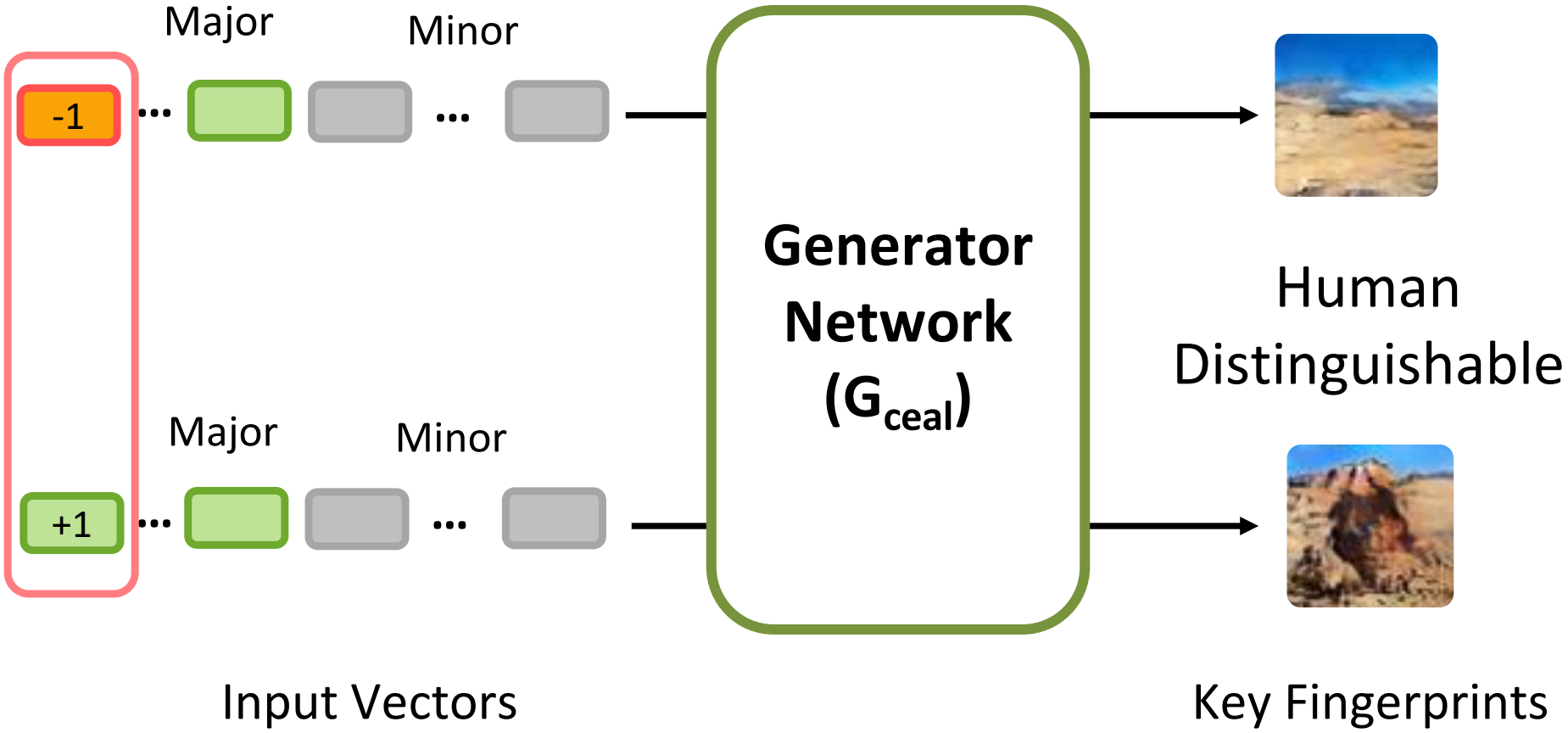
- Some components are equivalent of others
- We can *train* some components to be major

CL-GAN generator

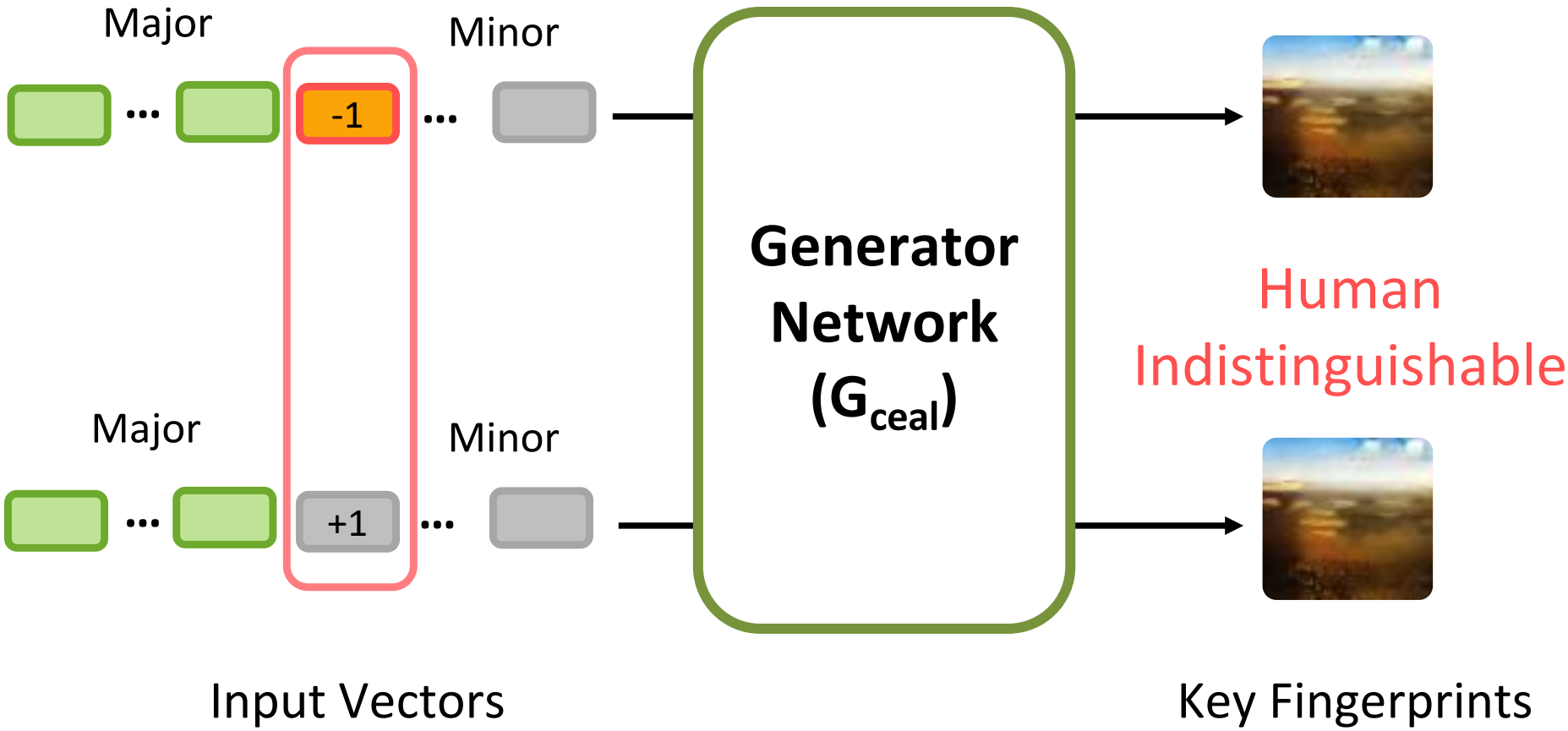


CEAL Generator Network

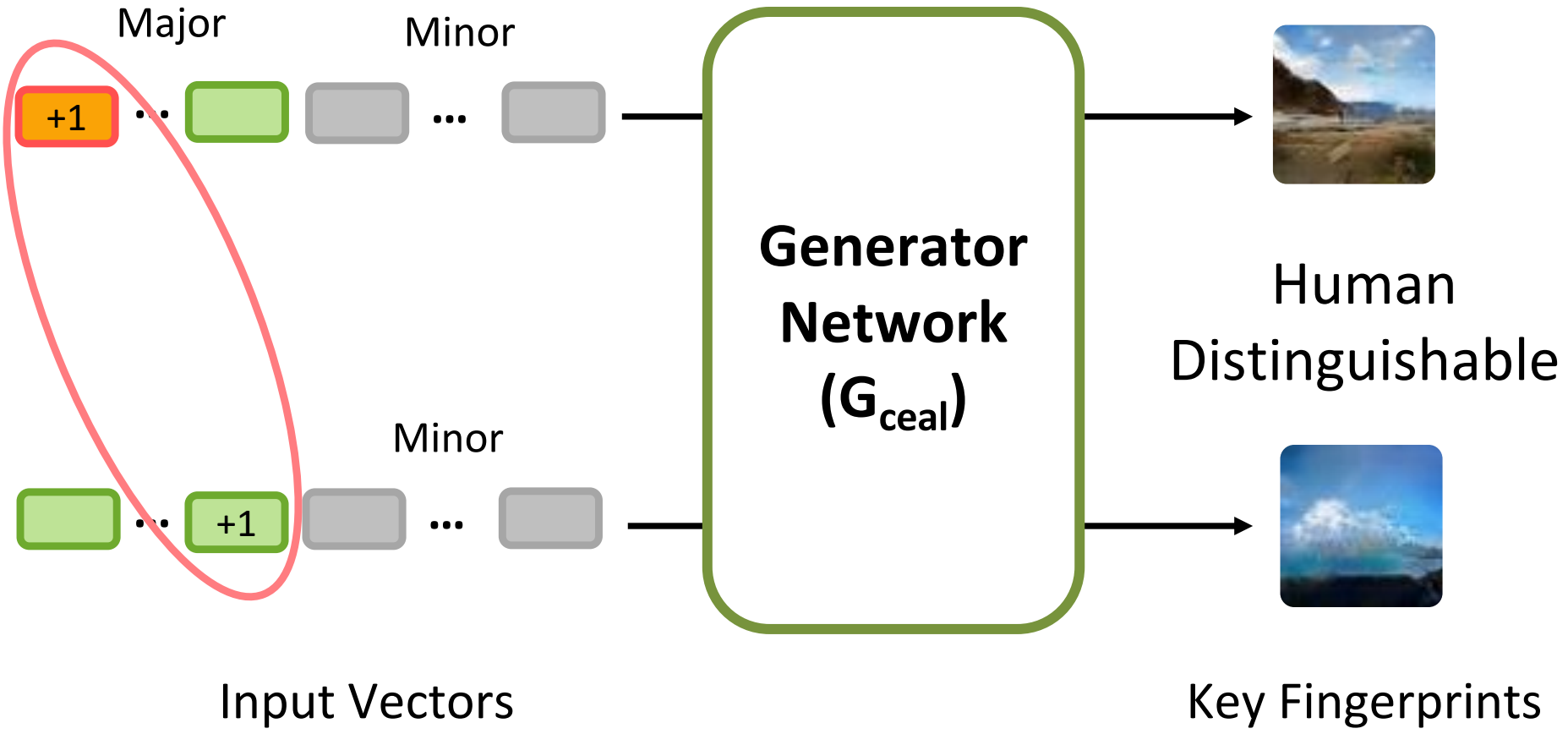
Train Majors for Distinguishability



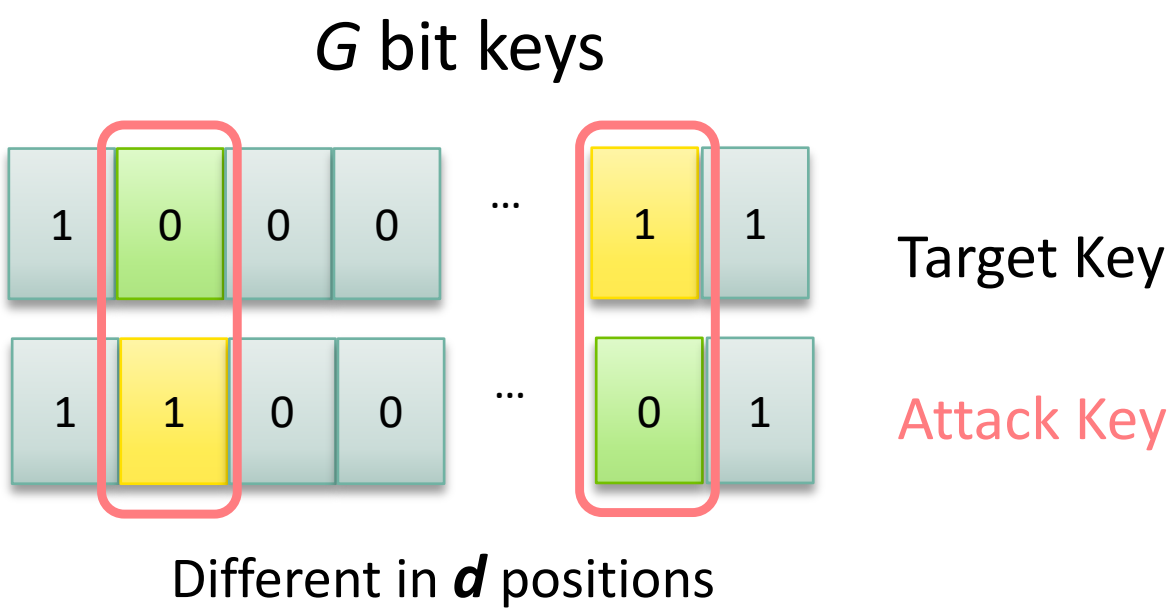
Train Minors



Train Majors for Diversity



(G, d)-adversary [Dechand et al. Usenix '16]



- ❑ Generate target keys (G bits)
- ❑ Generate attack keys different in d bits from target
- ❑ Generate corresponding visual key fingerprints
- ❑ Use a HPD to filter similar fingerprints to target

CEAL Under (G, d)-attack

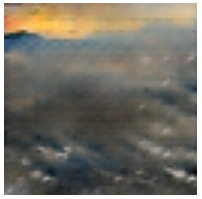
Attack Dataset	Dataset Size	# Attacks Identified by HPD-Attacker	Human Verified Attacks
(123,1)-adversary	123M	121	2 (1.7%)
(123,d)-adversary	123M	1,473	23 (1.6%)

Evaluate potential attack images using 374 MTurk workers

(G, d) Attack Examples

Targets







Attacks







Humans labeled as different

Humans labeled as same

CEAL vs. Vash

- Generate 10,000 random Vash and CEAL images
- Compare all key fingerprint pairs using HPD
  - Approx. 50 million image pair comparisons

VKFG	Attack Dataset Size	# Attacks Identified by HPD	Human Verified Attacks
CEAL	~50M	1	0 (0%)
Vash	~50M	150	24 (16%)

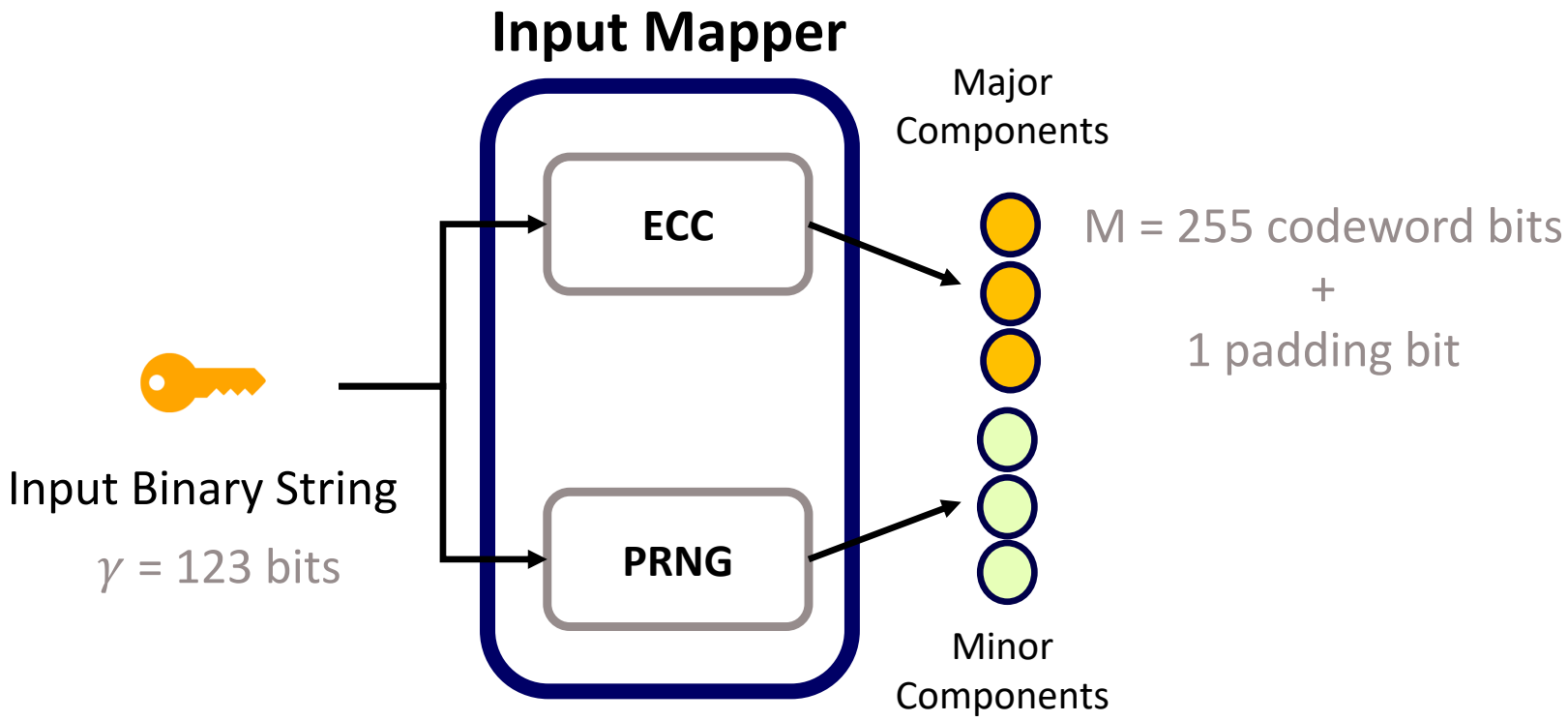
Attack datasets of 10,000 random images

Conclusions

- CEAL: Visual key fingerprint generation solution
  - Human-distinguishable fingerprints
  - Resilient to powerful adversaries
- CEAL improves on state-of-the-art Vash
  - Resilient to attack
  - Fast to compare: 2.73s for CEAL vs. 3.03s for Vash
- Incentive to adversaries to improve HPD
- Applications to CAPTCHA

Backup Slides

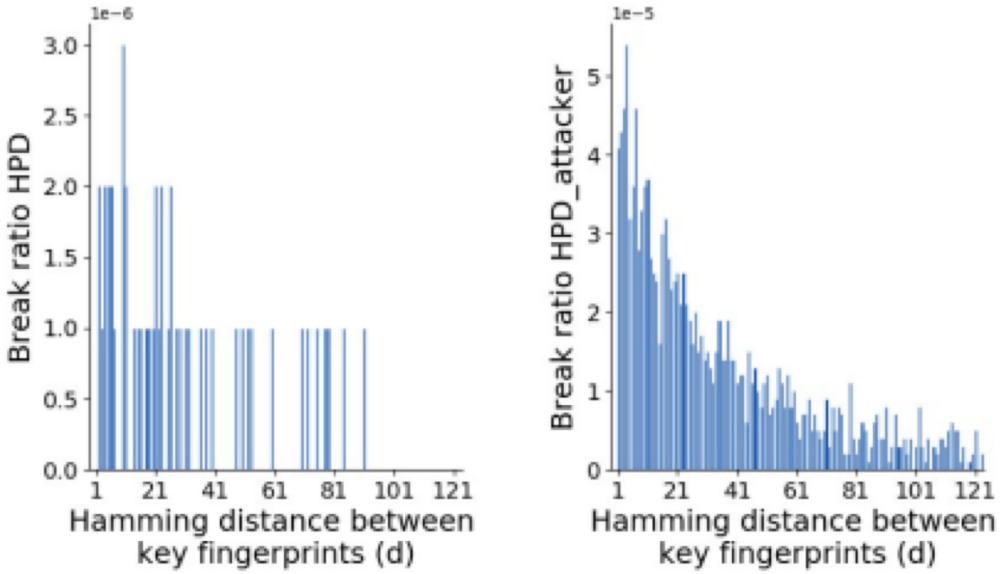
Input Mapper



BCH(n=255, k=123, d_{min}=19) for CL-GAN

d_{min} = min Hamming distance between codewords

Attack Success Relation to *d*



The break ratio of 1 million target CEAL images for each value of *d*, the Hamming distance between the attack and the target binary fingerprints, according to (left) HPD_model_1 and (right) HPD_attacker.

Datasets for Training HPD

Dataset Name	# Image Pairs	Labels
Labeled Synthetic Image Pairs	558	Mixed
Unrealistic DCGAN Image Pairs	11,072	Same
Minor Change Image Pairs Dataset	7,040	Same
Blob Image Pairs Dataset	2,108	Different
10%-different Image Pairs Dataset	1,024	Different
Enhanced Synthetic Image Pairs Dataset	5000	Different
Total	***26,802***	***Mixed***

Ground Truth Human Perception and Synthetic Image Pair Datasets we used to train HPD

HPD Performance on Vash Images

Model	F1	FPR	FNR	Recall	Precision
HPD_model_1	0.76	0.21	0.14	0.86	0.69

Performance of HPD over 120 labeled Vash images

CEAL vs. Vash: Time to Verify

Vash: 3.03s (SD=5.42s) avg over 150 attacks
CEAL: 2.73s (SD=2.33s) avg over 48 attacks