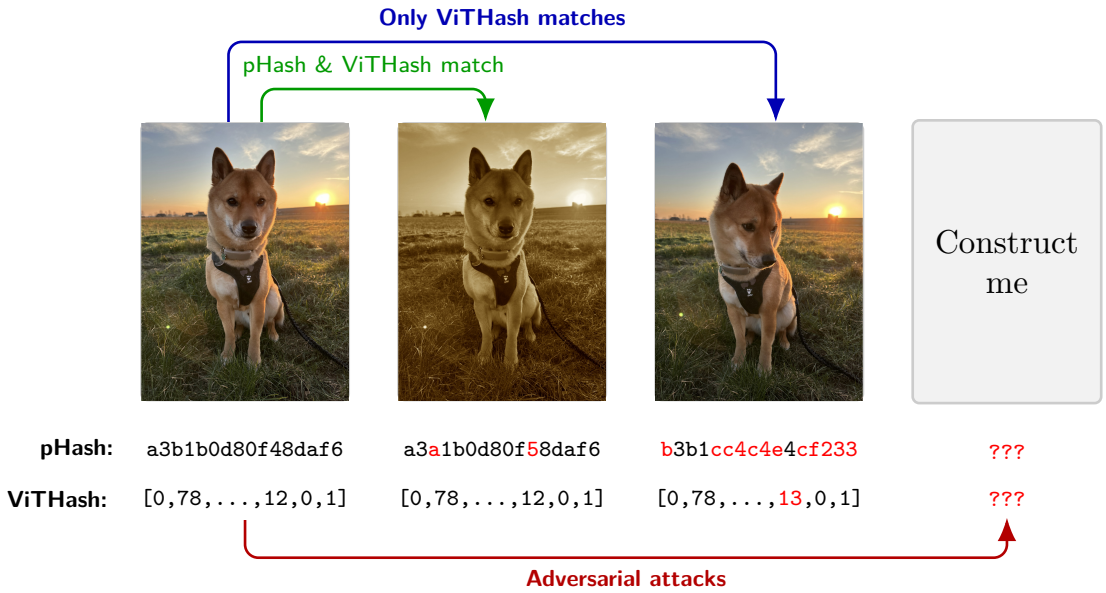
Cryptanalysis of ViTHash: Can You Fool an AI System?



Perceptual hash functions generate similar outputs for similar inputs, making them ideal for identifying near-duplicate images. However, as they are designed to find images derived from the same original image, they typically cannot be used to recognise visually similar images that originate from a different acquisition origin.

In recent work [1], a new type of perceptual hash function has been proposed to tackle this exact challenge. The authors introduced ViTHash, using a Vision Transformer (ViT) for this purpose. They compared it to existing perceptual hash functions (such as pHash) and observed both advantages and disadvantages, but an adversarial context was not considered.

The main objectives of this project include:

- Understanding the mathematical properties of ViTHash and relate these to cryptographic strengths or vulnerabilities.
- Developing adversarial attacks, such as preimage, second preimage, collision and evasion attacks.
- Improving the robustness of the system against benign image transformations.

After discussion with the student, several of these research directions can be explored.

More information on perceptual hashing can be found in this article, with recent adversarial attacks to be found here, here and here.

[1] J. De Geest et al., "Exploring Human Perception-Aligned Perceptual Hashing," *IEEE Consumer Electronics Magazine*, vol. 15, no. 1, pp. 41–54, Jan. 2026. doi: 10.1109/MCE.2025.3551813.

Practicalities

Promotor:	Bart Preneel		
Daily supervisor(s):	Maxime Deryck	maxime.deryck@esat.kuleuven.be	office B01.19
	Diane Leblanc-Albarel	diane.leblanc-albarel@esat.kuleuven.be	office B01.23

Nature of the work:	20% literature, 30% theoretical work, 50% software
Number of students:	1 or 2