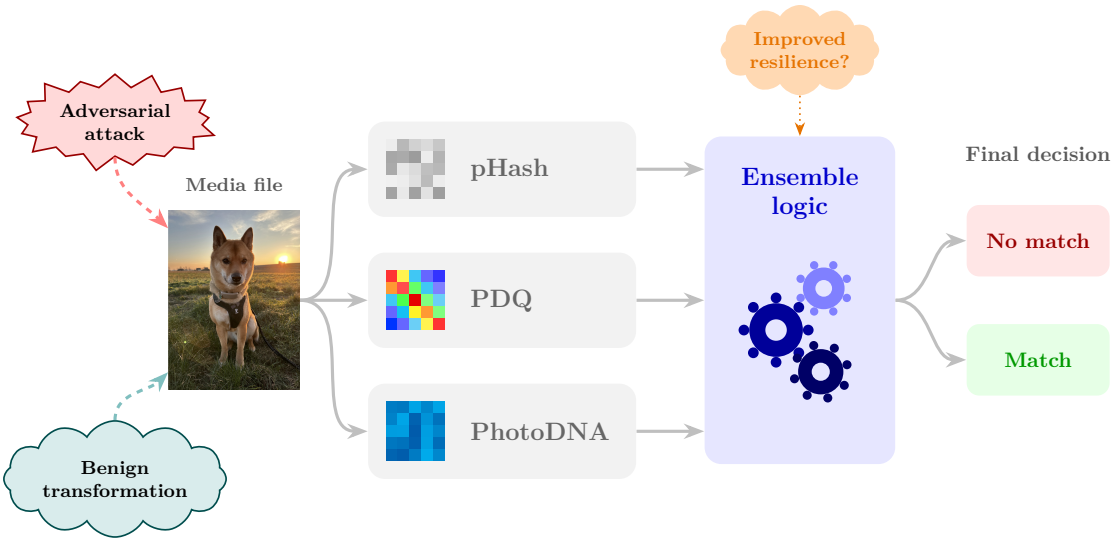
Stronger Together? Exploring Ensemble Learning for Perceptual Hashing



Perceptual hash functions generate similar outputs for similar inputs, making them ideal for matching media content. However, recent work [1] shows that false positive and false negative rates are much higher than claimed. In addition, several adversarial attacks against widely deployed algorithms have been constructed, see, for example, here, here and here.

As several designs, such as pHash, PDQ, and (Alleged) PhotoDNA, are publicly available and show different strengths and weaknesses, the question arises whether a combination of them can improve performance.

Ensemble learning has long been used in machine learning to improve the performance of weak learners, but whether this also improves decision making, and more interestingly, resistance against adversarial attacks, for the use case of perceptual hash function, remains an open problem.

The main objectives of this project include:

- Understanding how ensemble learning techniques can be used for perceptual hash functions;
- Evaluating the robustness (efficiency) of these systems against benign transformations; and
- Studying the effects of these techniques on adversarial attacks.

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

More information on perceptual hashing can be found in this article.

[1] D. Leblanc-Albarel and B. Preneel, "Breaking Widely Deployed Perceptual Hash Functions: Black-Box Collisions in Apple NeuralHash and Microsoft PhotoDNA" *35th USENIX Security Symposium (USENIX Security 26)*, pp. 1907–1926, Aug. 2026.

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, 20% theoretical work, 60% software		
Number of students:	1 or 2		