

## Ẩn dữ liệu và chia sẻ thông tin

### Đồ án môn học

# ALASKA2 Image Steganalysis

*Detect secret data hidden within digital images*

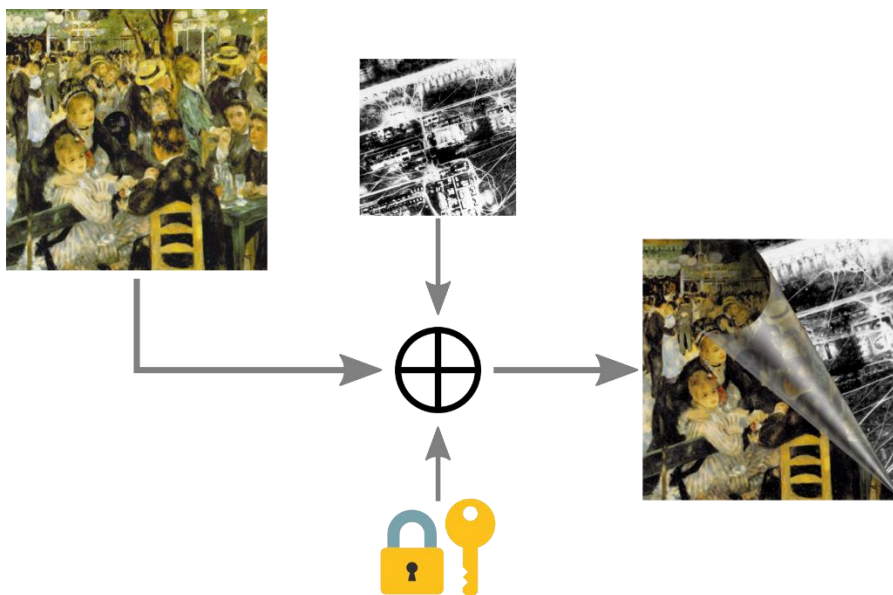
## 1. Giới thiệu

Đồ án này sẽ dựa trên cuộc thi [ALASKA2 Image Steganalysis](#) trên Kaggle

### Mô tả:

Các quan chức thực thi pháp luật cần những phương pháp tốt hơn để chống lại tội phạm sử dụng tin nhắn ẩn. Cộng đồng khoa học dữ liệu và các nhà nghiên cứu khác có thể giúp phát hiện một cách tự động các thông điệp ẩn tốt hơn. Các phương pháp chính xác hơn có thể giúp truy bắt những tên tội phạm với thông điệp bị ẩn đi.

Trong cuộc thi này, SV sẽ cần tạo ra một phương pháp hiệu quả và đáng tin cậy để phát hiện dữ liệu bí mật ẩn trong các hình ảnh kỹ thuật số tưởng chừng như vô hại. Thay vì giới hạn nguồn dữ liệu, những hình ảnh này đã được thu thập với 50 máy ảnh khác nhau (từ điện thoại thông minh đến máy ảnh cao cấp có định dạng đầy đủ) và được xử lý theo nhiều kiểu khác nhau. Phương pháp thành công sẽ bao gồm các thuật toán phát hiện mạnh mẽ với tỷ lệ sai sót tối thiểu.



## 2. Yêu cầu của đề án

Các nhóm sẽ tham dự cuộc thi trên. Download bộ dữ liệu, xây dựng mô hình dự đoán, submit lên Kaggle, xem kết quả.

- a) Tổ chức thư mục cho đề án: các nhóm có thể dùng một hay nhiều file notebooks nhưng phải được tách biệt rõ ràng cho từng giai đoạn: giới thiệu bài toán, tiền xử lý dữ liệu, rút trích đặc trưng, xây dựng mô hình, đánh giá và phân tích kết quả.
- b) Nên có nhiều hình vẽ biểu đồ, hình ảnh trực quan để giải thích và trình bày trong quá trình làm.
- c) Phải có giải thích rõ ràng cho mọi cell code trong file jupyter notebook. Tức là, mỗi cell code nên có một cell markdown kèm theo để giải thích.

## 3. Quy định nộp bài:

Mỗi nhóm cần nộp:

- File jupyter notebook. File này sẽ bao gồm source code và báo cáo. Nội dung cần trình bày bao gồm
  - Giới thiệu thành viên và phân công công việc
  - Giới thiệu chung về đề án
  - Giới thiệu và trình bày chi tiết bài toán
  - Quá trình tiền xử lý dữ liệu
  - Quá trình tạo đặc trưng (feature engineering)
  - Xây dựng mô hình (Có thể dùng nhiều mô hình)
  - Thử nghiệm:
    - Training
    - Testing
    - Kết quả đạt được, nhận xét và phân tích
  - Kết luận, những khó khăn gặp phải, những gì đã học được qua quá trình làm đề án này, và hướng phát triển. (Cần đưa kết quả trên leaderboard của Kaggle)
- Các mã nguồn python khác (nếu có)
- Một đoạn video ngắn (15-20p) trong đó nhóm trình bày đề án của mình.

#### **4. Thang điểm:**

- a. Kết quả trên Leaderboard (2đ). (Dự kiến)
- b. Nội dung file notebook + code (5đ)

Theo yêu cầu như phần 3. Các nhóm cần trình bày rõ ràng, chi tiết về mô hình mình đã dùng để đạt được kết quả như phần 4a.

- c. Trình bày seminar (3đ)

#### **5. Tham khảo**

[1] <https://www.kaggle.com/code/prashant111/alaska2-image-steganalysis-all-you-need-to-know/notebook>

[2] <https://www.kaggle.com/competitions/alaska2-image-steganalysis/discussion/146622>