TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN - ĐHQG TPHCM
KHOA TOÁN - TIN HỌC

THỰC HÀNH XỬ LÝ DỮ LIỆU ĐA CHIỀU

THÔNG TIN LIÊN HỆ

GMAIL: ntktrang@hcmus.edu.vn

GHI CHÚ:

Tiêu đề mail (bắt buộc):

[XLDLDC2023] [Tiêu đề thư]

VD: [CSLT2023] HỎI BÀI

Gửi email kèm giới thiệu họ tên, MSSV và tên ca học.

CLUSTERING

CLUSTERING

UNSUPERVISED LEARNING

K-means

CLUSTERING

- Hai thành phần thiết yếu của phân tích cụm:
  - Đo khoảng cách: Là khái niệm về khoảng cách hoặc độ giống nhau của hai vật: Hai vật gần nhau khi nào?
    - Ví dụ: chuẩn Euclide, chuẩn Manhattan
  - Thuật toán cụm: Một quy trình để giảm thiểu khoảng cách của các đối tượng trong nhóm và/hoặc tối đa hóa khoảng cách giữa các nhóm
    - Ví dụ: Hard clustering, Hierarchical clustering, K-mean, SOMs (Self-Organizing Maps), Autoclass, mixture models

THUẬT TOÁN

Mô tả bằng ngôn ngữ tự nhiên:

Cho biết K cụm cần phân tách, và k điểm center đầu (có thể chọn random)

Tính toán khoảng cách của các điểm với k centers. Điểm đó gần với center nào thì vô nhóm của center đó.

Cập nhật center bằng cách lấy trung bình cộng của các điểm.

Lặp lại

HẠN CHẾ

- Bị phụ thuộc vào điểm đầu tiên
- Cần biết số lượng cụm cần phân ra.
- Các cụm cần có lượng điểm tương đương nhau, không quá chênh lệch.

NÉN ẢNH - TÁCH VẬT THỂ

Các thuật toán nén hình ảnh tận dụng nhận thức trực quan và các thuộc tính thống kê của dữ liệu hình ảnh để mang lại kết quả vượt trội.

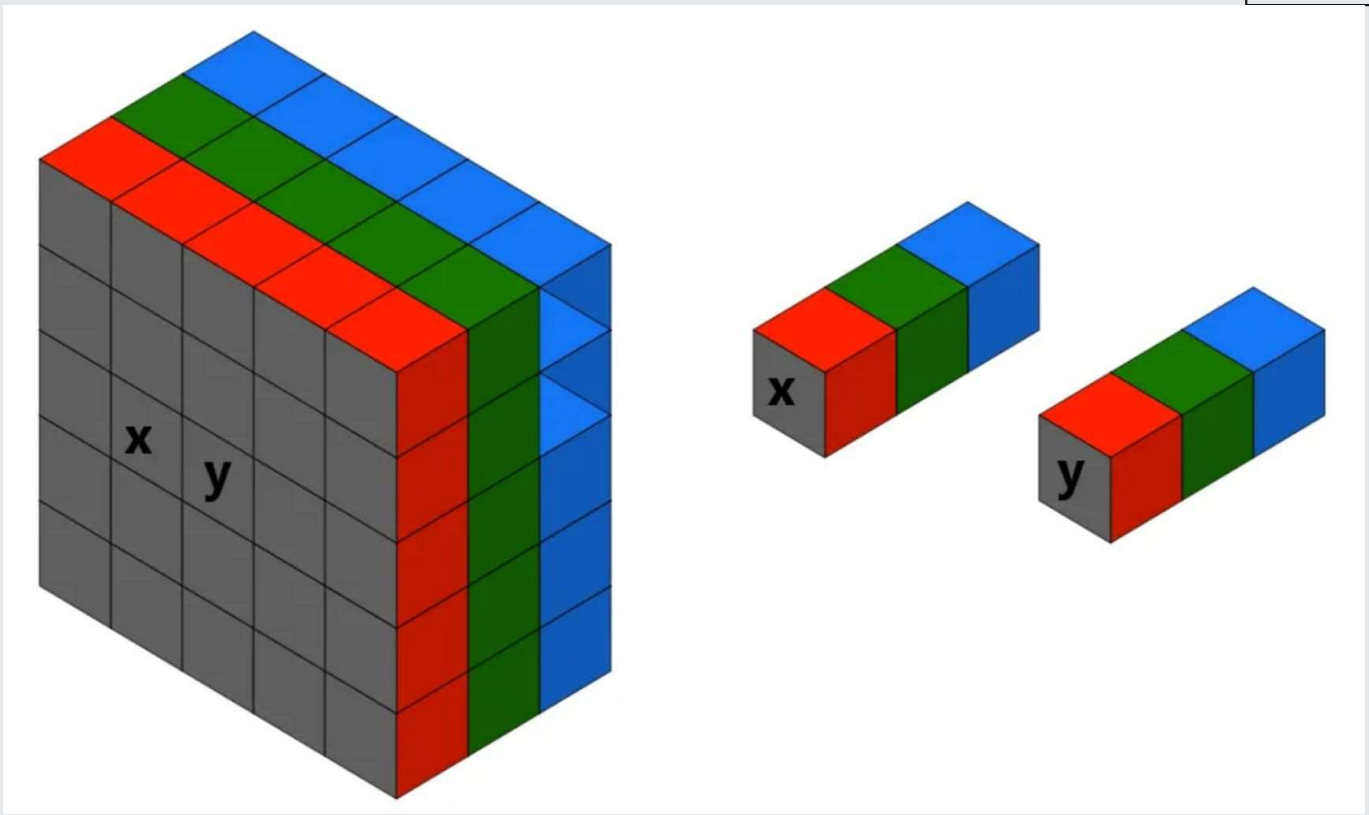
Có hai loại phương pháp nén hình ảnh - lossless và lossy.

+ Lossless: có thể khôi phục lại hoàn toàn chính xác

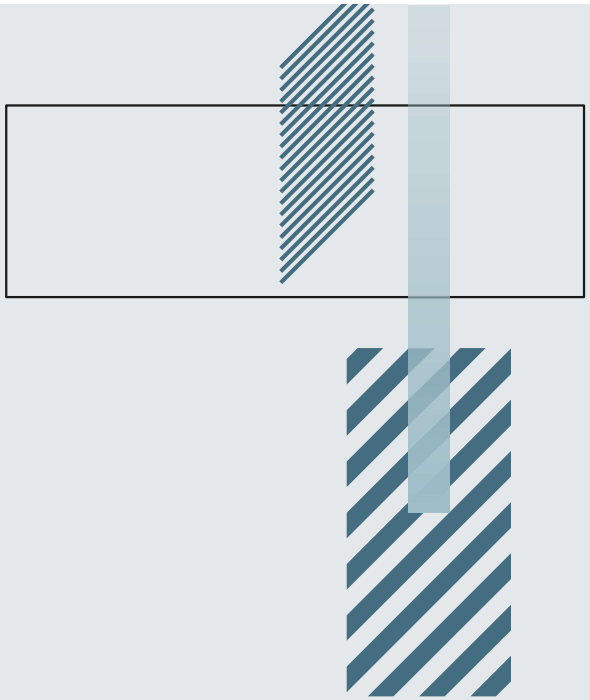
+ Lossy: chấp nhận một số lỗi hoặc sai số trong quá trình tái tạo.

NÉN ẢNH - TÁCH VẬT THỂ

NÉN ẢNH - TÁCH VẬT THỂ



NÉN ẢNH - TÁCH VẬT THỂ



Giả sử

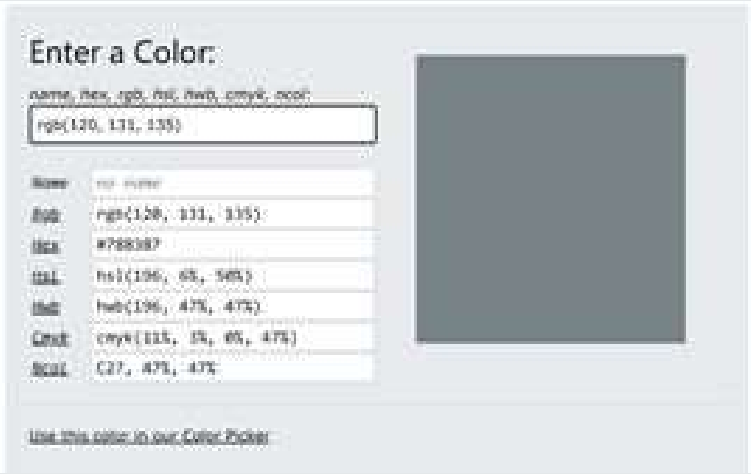
360x640= 230.400

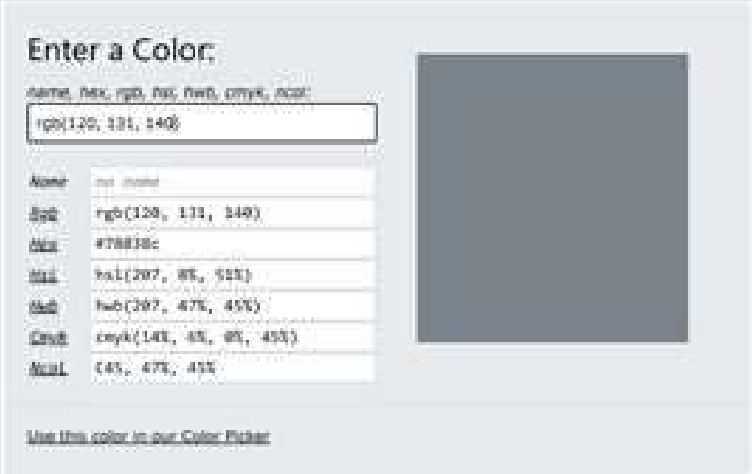
255*255*255 = 16,581,375

x in RGB(120, 131, 135)

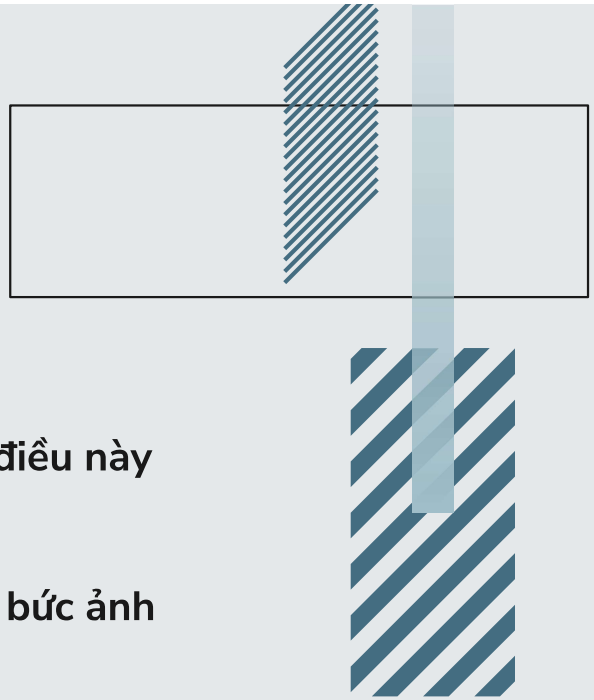
y in RGB (120, 131, 140)

NÉN ẢNH - TÁCH VẬT THỂ





NÉN ẢNH - TÁCH VẬT THỂ



Sau đó, ta sẽ thay thế màu trong 1 cluster bằng center của nó, và điều này sẽ dẫn đến mất dữ liệu.

Khi lưu, ta chỉ cần lưu centers của mỗi điểm ảnh là đã có được một bức ảnh nén.

Tương tự với bài toán tách vật thể.

BÀI TẬP

1. Làm tiếp các bước còn lại của bài ví dụ
2. Hãy thực hiện nén ảnh sau:
   a. Đọc file
   b. Flatten bằng reshape
   c. Thực hiện K-means (gợi ý: sklearn.cluster import KMeans)
   d. Thay thế các điểm ảnh trong cluster bằng center của nó. (gợi ý: cluster_centers_ và np.clip)
   e. Vẽ và lưu hình sau nén

BÀI TẬP

Exercise XV. K-means Clustering

1. Suppose we have some data points (the black dots) as in Figure 2. We want to apply K-means to this data, using $K = 2$ and centers initialized to the two circled data points.

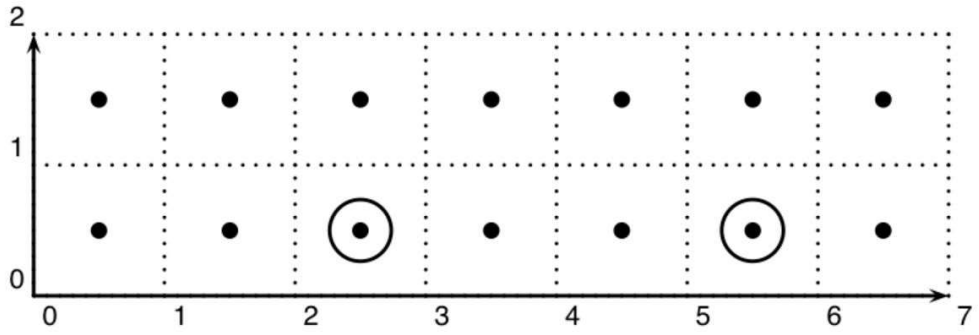


Figure 2: Cluster the black dots with K-means. The two circles denote the initial guess for the two cluster centres, respectively.

THE END