Image Processing with Salient Object Detection

Trần Gia Phát, Lâm Hoàng Khánh

# Abstract

Vấn đề phân đoạn foreground và background hiệu quả trong ảnh tĩnh (still image) có tầm quan trọng rất lớn trong lĩnh vực chỉnh sửa ảnh (image editing). Có thể kể đến các công cụ phân đoạn hình ảnh cổ điển sử dụng thông tin về kết cấu (texture information), ví dụ: Magic Wand, hoặc thông tin về cạnh (edge information). Thông thường, hầu hết các kỹ thuật phân đoạn ảnh sử dụng một trong hai phương pháp thông tin về vùng (region information) hoặc thông tin về cạnh (edge information). Trong khi đó, “GrabCut” là một kỹ thuật phân đoạn tiên tiến sử dụng kết hợp cả hai phương pháp thông tin về vùng và cạnh [1]. “GrabCut” phân đoạn bằng cách sử dụng các đồ thị để lưu trữ thông tin về vùng và cạnh. Trong đó thuật toán Min-Cut/Max-Flow, là một kĩ thuật cắt đồ thị được sử dụng trong trường hợp này để phân đoạn đồ thị và cũng bằng cách đó phân đoạn được hình ảnh. Phương pháp phát hiện vùng và cạnh là đủ để phân đoạn hình ảnh, tuy nhiên nếu cạnh của vật thể được phân đoạn bị mờ đi thì đây chưa được xem là một phương pháp hiệu quả. Vì thế để giải quyết vấn đề trên, việc sử dụng kỹ thuật matting là rất cần thiết. Matting là một kỹ thuật làm mịn ranh giới giữa đối tượng phân đoạn và background bằng cách tính toán các giá trị alpha.

# Method ( Algorithms)

Bất kỳ đầu vào người dùng nào chỉ định nền trước và nền được coi là gắn nhãn cứng có nghĩa là chúng sẽ không thay đổi trong tiến trình. Máy tính thực hiện ghi nhãn ban đầu tùy thuộc vào dữ liệu chúng ta đã cung cấp. Nó gắn nhãn các pixel nền trước và nền sau (hoặc nhãn cứng).

Một Mô hình Hỗn hợp Gaussian (GMM) [2] được sử dụng để mô hình hóa nền trước và nền sau. Tùy thuộc vào dữ liệu chúng ta đã cung cấp, GMM học và tạo phân phối pixel mới. Tức là, các pixel không xác định được gắn nhãn hoặc nền trước có thể xảy ra hoặc có thể xảy ra tùy thuộc vào mối quan hệ của nó với các pixel được gắn nhãn cứng khác về số liệu thống kê màu (Nó giống như phân cụm).

Biểu đồ được tạo từ phân phối pixel này. Các nút trong biểu đồ tương ứng với pixel. từ đó tạo thêm hai nút đặc biệt, nút Nổi và nút Chìm. Mỗi pixel trong biểu đồ phải kết nối với nút Nổi và nút Chìm. Mỗi pixel nền trước được kết nối với nút Nổi và mỗi pixel nền được kết nối với nút Chìm. Để phân đoạn ảnh thì nút Nổi và nút Chìm phải được tách ra.

Hàm năng lượng được kết hợp vào biểu đồ dưới dạng trọng số giữa những nút pixel và trọng số giữa pixel và nút Nổi hoặc nút Chìm. Trọng số giữa những pixel được xác định bởi thông tin về cạnh (edge information) trong ảnh. Nếu có sự khác biệt lớn về màu pixel, các nút giữa chúng những pixel sẽ có trọng số thấp. Thông tin về vùng (region information) sẽ xác định trọng số giữa những pixel với nút Nổi và nút Chìm. Những trọng số này được tính bằng cách xác định xác suất giữa các pixel thuộc về nền trước (foreground region) hoặc nền sau (background region).

Sau đó, thuật toán Min-Cut/Max-Flow được sử dụng để phân đoạn biểu đồ. Nó cắt biểu đồ thành hai nút Nổi và nút chìm với hàm chi phí tối thiểu.

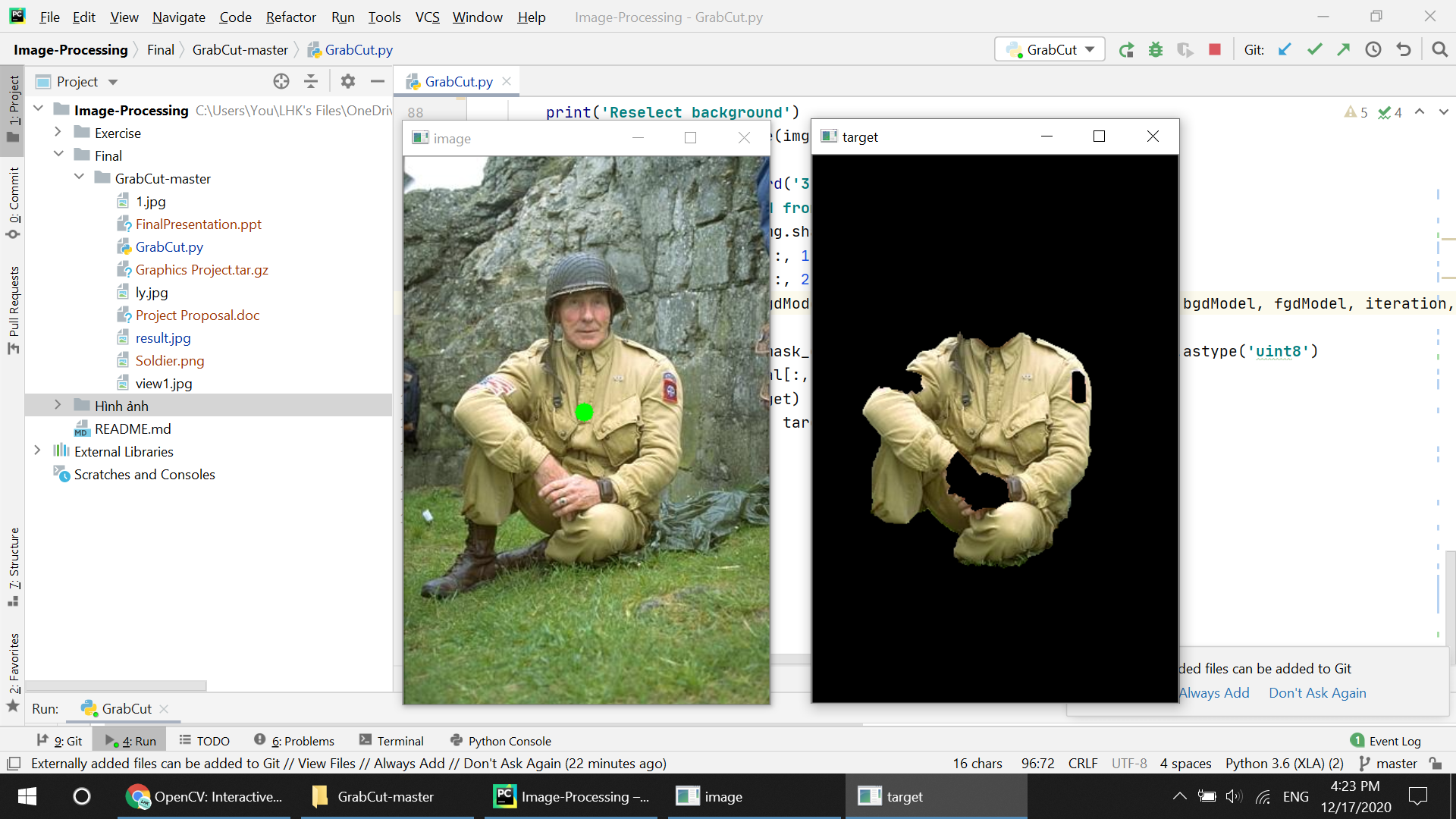
Hàm chi phí là tổng của tất cả các trọng số của các cạnh được cắt. Sau khi các nút Nổi và Chìm được tách ra, tất cả các pixel kết nối với nút Nổi trở thành nền trước (foreground) và những điểm kết nối với nút Chìm trở thành nền sau (background).

Quá trình này được tiếp tục cho đến khi phân loại hội tụ.

Các pixels trong ảnh số luôn có những giá trị alpha riêng lẻ. Mỗi giá trị alpha là biểu hiện của sự mờ đi (opacity) mà mỗi pixel sở hữu và có giá trị từ 0 đến 255 [1]. Ta có thể hiểu với giá trị alpha bằng 0 thì ảnh sẽ hoàn toàn trong suốt (transparent), còn giá trị alpha bằng 255 thì sẽ hoàn toàn mờ. Nếu không có kỹ thuật matting, đối tượng được phân đoạn sẽ có giá trị alpha mặc định trong mỗi pixel là 255. Do đó, kỹ thuật matting là rất cần thiết để làm mịn các cạnh của đối tượng được phân đoạn bằng cách điều chỉnh giá trị alpha từ 0 đến 255.

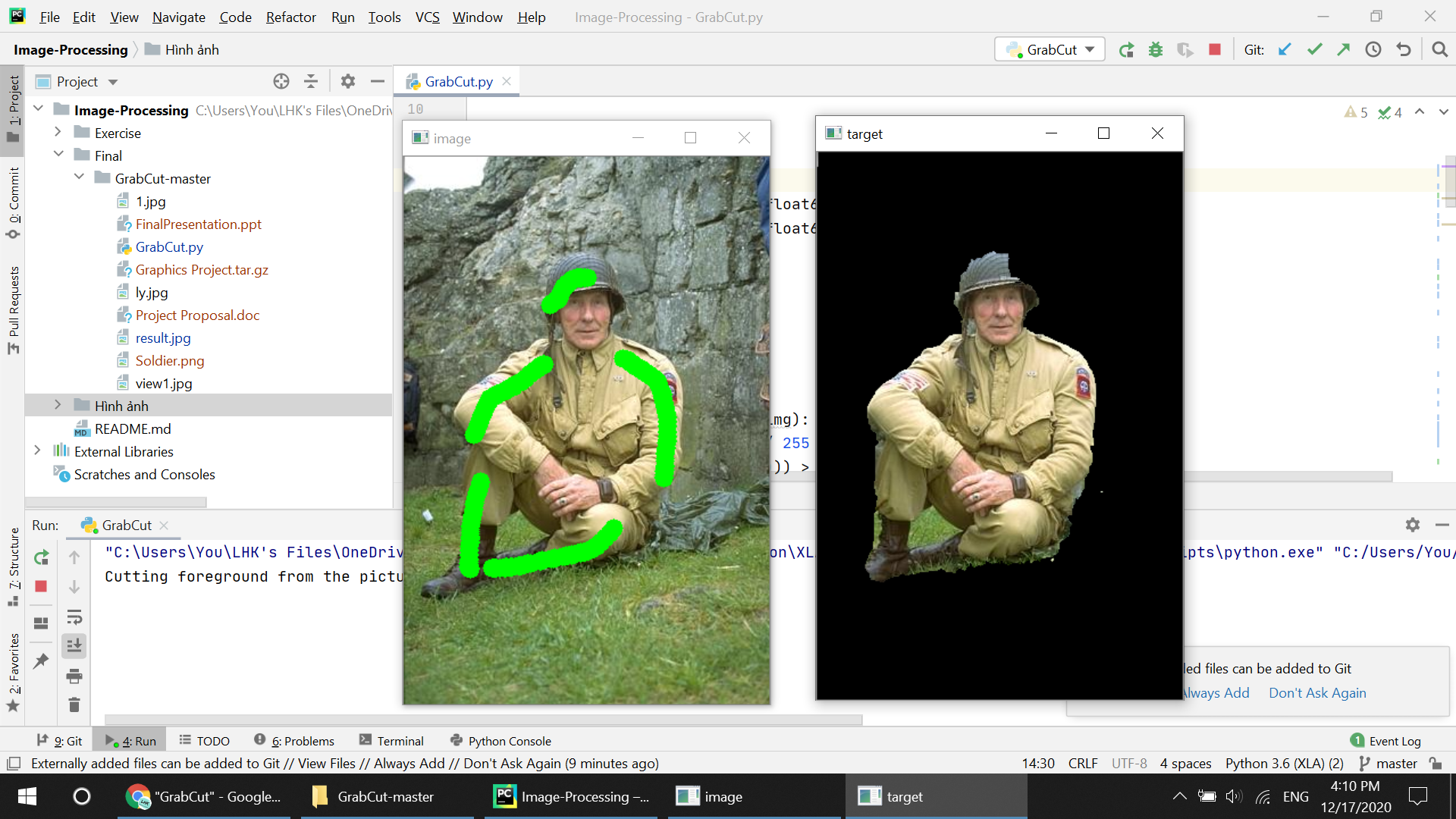
# Experiments and Analysis

## Vẽ 1 chấm định vị force



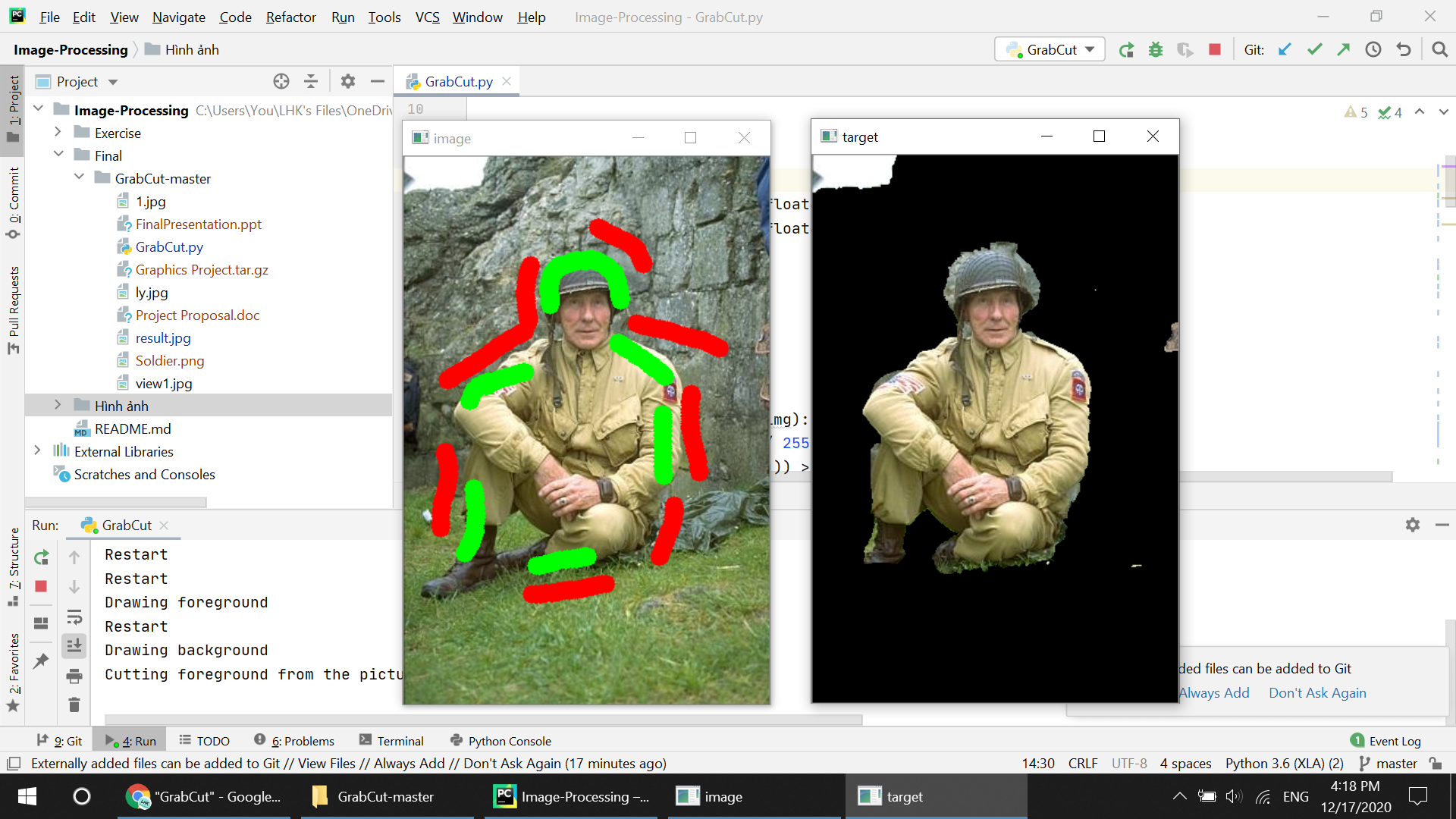
Nhìn chung, thuật toán định vị được các pixel có giá trị tương đồng.

## Vẽ vài đường Nổi (định vị Force)



Ta có thể thấy thuật toán chạy khá tốt ở phần pixel phân biệt rõ (phần vai phải). Tuy nhiên phần pixel Nổi có giá trị gần với giá trị pixel Nền (phần nón) thuật toán còn chưa phân biệt được chính xác.

## Vẽ vài đường Nổi (định vị Force) kết hợp với đường Nền (định vị Background)



## Kết luận ban đầu

Kết quả tốt khi chỉ ra được các pixel của force, tránh chỉ ra nhầm lên các pixel nền khiến thuật toán hiểu nhầm.

Các kỹ thuật phía trên là đủ cho việc phân đoạn ảnh, nhưng đôi lúc phần viền cạnh bên ngoài của Object chưa thực sự tốt đôi khi còn bị mất hoặc mờ đi. Vì thế để giải quyết vấn đề trên, việc sử dụng kỹ thuật matting là rất cần thiết. Matting là một kỹ thuật làm mịn ranh giới giữa đối tượng được phân đoạn và background bằng cách tính toán các giá trị alpha.

Trong ví dụ dưới đây, alpha mờ màu trắng là một alpha có giá trị 255, màu đen có giá trị bằng 0, và những giá trị giữa 0 đến 255 được hiển thị dưới dạng ảnh xám.



# Acknowledgement

Nhóm xin gửi lời cảm ơn sâu sắc đến Thầy Ngô Quốc Việt đã tận tình chỉ dạy trong suốt khóa học đầu tiên về Xử lý Ảnh. Sự nhiệt huyết của Thầy đã truyền động lực rất lớn cho nhóm hoàn thành môn học và hứng thú với Xử lý Anh,

Xin chân thành cảm ơn Thầy.

# References

1. Implementing the "GrabCut" Segmentation Technique as a Plugin for the GIMP [[*https://www.cs.ru.ac.za/research/g02m1682/*](https://www.cs.ru.ac.za/research/g02m1682/) ]
2. Interactive Foreground Extraction using GrabCut Algorithm

[https://docs.opencv.org/master/d8/d83/tutorial\_py\_grabcut.html]

1. Q. Hou, M. M. Cheng, X. Hu, A. Borji, Z. Tu and P. H. S. Torr, “Deeply Supervised Salient Object Detection with Short Connections,”  IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018