

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ  
KHOA CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG  
BỘ MÔN CÔNG NGHỆ THÔNG TIN**



**LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC  
NGÀNH CÔNG NGHỆ THÔNG TIN**

**Đề tài**

**Tách phong đối tượng người trong ảnh sử dụng  
kỹ thuật phân vùng ngữ nghĩa FCN-CRFs**

**Giảng viên hướng dẫn:  
TS. Thái Minh Tuấn**

**Sinh viên thực hiện: Nguyễn Nhựt Tín  
Mã số: B1507321  
Khoá: 41**

**Cần Thơ, 4/2019**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ  
KHOA CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG  
BỘ MÔN CÔNG NGHỆ THÔNG TIN**



**LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC  
NGÀNH CÔNG NGHỆ THÔNG TIN**

**Đề tài**

**Tách phong đối tượng người trong ảnh sử dụng  
kỹ thuật phân vùng ngữ nghĩa FCN-CRFs**

**Giảng viên hướng dẫn:  
TS. Thái Minh Tuấn**

**Sinh viên thực hiện: Nguyễn Nhựt Tín  
Mã số: B1507321  
Khoá: 41**

**Cần Thơ, 4/2019**

[illegible]

## **LỜI CAM ĐOAN**

Tôi xin cam đoan tất cả nội dung và số liệu được trình bày trong luận văn đều minh bạch, trung thực được trích dẫn nguồn rõ ràng và được thực hiện bởi tôi dưới sự hỗ trợ, giúp đỡ của giảng viên hướng dẫn Tiến sĩ Thái Minh Tuấn.

*Cần Thơ, ngày ... tháng ... năm ...*

**Người viết**

Nguyễn Nhựt Tín

## **LỜI CẢM ƠN**

Luận văn được thực hiện dưới hỗ trợ tận tình của Tiến sĩ Thái Minh Tuấn, Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ.

Xin gửi lời cảm ơn chân thành và ý nghĩa!

## MỤC LỤC

<b>1</b>	<b>GIỚI THIỆU</b>	<b>1</b>
1.1	Tổng quan về tách phoneme . . . . .	1
<b>2</b>	<b>CƠ SỞ LÝ THUYẾT</b>	<b>2</b>
2.1	Mạng tích chập đầy đủ (Fully Convolutional Network): . . . . .	2
2.1.1	Bộ giảm mẫu - mã hóa (Downsampling - Encoder): . . . . .	2
<b>3</b>	<b>KẾT QUẢ THỰC HIỆN</b>	<b>3</b>
3.1	Phân vùng ngữ nghĩa và ứng dụng tách phoneme: . . . . .	3
3.1.1	Sơ đồ chức năng ứng dụng: . . . . .	3
<b>4</b>	<b>KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN</b>	<b>4</b>
4.1	Nhận xét kết quả đạt được: . . . . .	4
4.2	Hạn chế: . . . . .	4
4.3	Hướng phát triển: . . . . .	4

## **DANH SÁCH HÌNH VẼ**

1.1	Phát hiện đối tượng thông thường (Nguồn: <a href="http://www.cgtrader.com">www.cgtrader.com</a> ) . . . . .	1
2.1	Mạng nơ-ron tích chập truyền thống [Nguồn: <a href="http://towardsdatascience.com">towardsdatascience.com</a> )] . .	2

## **DANH SÁCH BẢNG**

3.1	Các tập dữ liệu phổ biến cho phân vùng ngữ nghĩa ảnh . . . . .	3
-----	--	---



## TÓM TẮT

Kỹ thuật phân vùng ngữ nghĩa có nhiều ứng dụng trong lĩnh vực đồ họa kỹ thuật số,

...

Trong bài báo cáo này, tác giả tận dụng phương pháp tính toán của mạng tích chập đầy đủ ...

**Từ khóa:** *phân vùng ngữ nghĩa, mạng tích chập đầy đủ, trường điều kiện ngẫu nhiên.*

## ABSTRACT

The semantic image segmentation technique has many applications in digital graphics, ...

In this report, the author uses the calculation method of Fully Convolutional Network ...

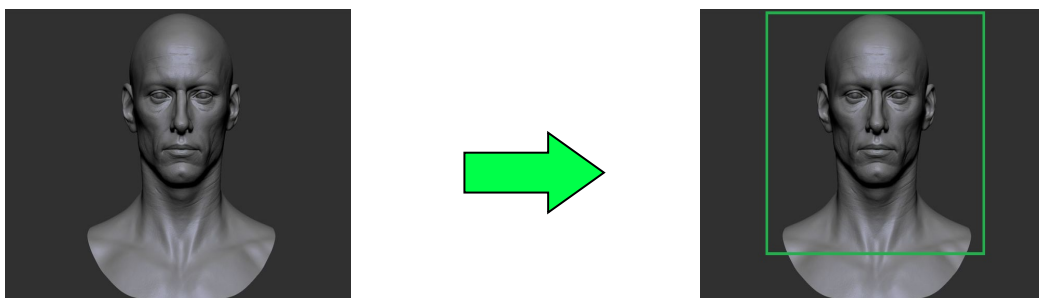
**Keywords:** *semantic image segmentation, fully convolutional network, conditional random fields.*

# CHƯƠNG 1: GIỚI THIỆU

## 1.1 Tổng quan về tách phong ảnh

Hiện nay, các tập đoàn công nghệ không ngừng đầu tư và ứng dụng các loại công nghệ trong đó có tách phong ảnh vào các sản phẩm công ty của họ đem lại doanh thu cũng như thu hút lượng khách hàng rất lớn, mang lại trải nghiệm mới mẻ cho người dùng. Cụ thể, tập đoàn Bkav thay vì tốn chi phí cho phần cứng phải đầu tư máy ảnh kép, họ tiết kiệm chi phí bằng cách tích hợp khả năng tách phong vào máy chụp ảnh trên điện thoại cho phép người dùng làm mờ cảnh xung quanh người, làm nổi bật đối tượng được chụp. Ứng dụng SNOW<sup>[1]</sup> ra đời cho phép ghép ảnh, thêm các chi tiết như râu, nón, mắt kiếng,... lên từng bộ phận đã được tách phong trên mặt người trực tiếp khi quay rất độc đáo. Ngoài ra, phải kể đến thành công của tách phong ảnh khi làm nổi bật các bộ phận trong cơ thể con người như tim, gan, phổi,... từ ảnh X-quang hoặc tách phong thiết bị máy móc y học khi hoạt động phẫu thuật trong cơ thể người.

Tách phong ảnh là quá trình cho phép tạo ra những hình ảnh mới từ ảnh gốc với những vùng chi tiết quan trọng được giữ lại, xóa bỏ những vùng nền không có tác dụng phục vụ rất nhiều cho y tế, phần mềm công nghệ máy ảnh, đồ họa kỹ thuật số và tất nhiên chúng đòi hỏi thách thức cao khi thực hiện. Với những lợi ích đem lại cũng cho thấy tính cần thiết của vai trò tách phong ảnh cần được nghiên cứu, đầu tư nhiều hơn. Tách phong ảnh thực sự là ứng dụng của bài toán phân vùng ngữ nghĩa đối tượng người trong ảnh, chúng giải quyết vấn đề xác định vị trí chính xác của đối tượng chiếm trong ảnh, khắc phục nhược điểm của bài toán phát hiện đối tượng khi chỉ có khả năng xác định khung chứa đại khái đối tượng.



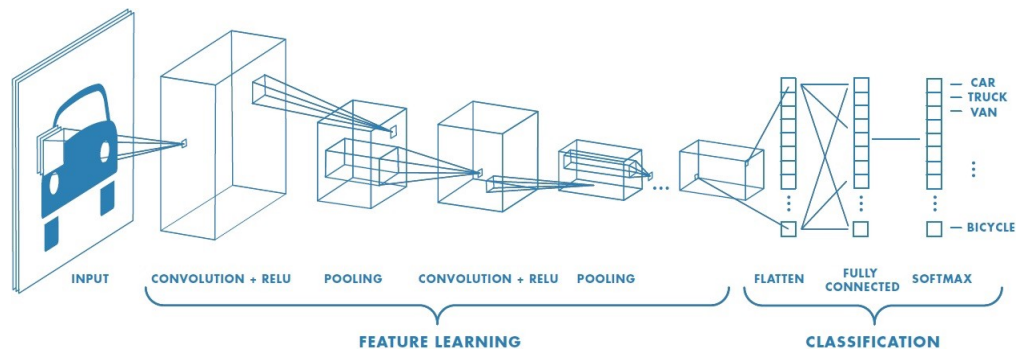
Hình 1.1: Phát hiện đối tượng thông thường (Nguồn: [www.cgtrader.com](http://www.cgtrader.com))

## CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

### 2.1 Mạng tích chập đầy đủ (Fully Convolutional Network):

Ứng dụng phân vùng ngữ nghĩa được xây dựng thao tác trực tiếp với ảnh và mạng nơ-ron tích chập nguyên thủy CNN (Convolutional Neural Network) là giải pháp rất thành công ở thời điểm hiện tại khi thao tác với các loại ảnh. Chính vì vậy, CNN thường được sử dụng làm điểm tựa để nghiên cứu về vấn đề trên.

Theo lý thuyết, mạng nơ-ron tích chập nguyên thủy<sup>[2]</sup> là sự tổ hợp hợp lý các lớp tích chập (Convolution layer), lớp phi tuyến (Nonlinear layer) và lớp thăm dò (Pooling layer) với mỗi lớp sẽ đảm nhiệm những vai trò tính toán chuyên biệt nhằm rút lấy đặc trưng (Feature Learning), sau đó sẽ đưa qua bộ phân lớp (Classification) để xuất ra một véc-tơ chứa xác suất hay khả năng đối tượng đầu vào thuộc từng lớp đó.



Hình 2.1: Mạng nơ-ron tích chập truyền thống [Nguồn: towardsdatascience.com)]

#### 2.1.1 Bộ giảm mẫu - mã hóa (Downsampling - Encoder):

## CHƯƠNG 3: KẾT QUẢ THỰC HIỆN

### 3.1 Phân vùng ngữ nghĩa và ứng dụng tách phong:

#### 3.1.1 Sơ đồ chức năng ứng dụng:

Ứng dụng xây dựng trên bộ công cụ lập trình giao diện Tkinter của Python được sử dụng trong bài báo cáo để làm bộ phận hiển thị cho người dùng và kết hợp với các tính năng phân vùng ảnh ngữ nghĩa của hệ thống, đồng thời ghép ảnh nền vào ảnh đã tách phong, tạo ra những hình ảnh thú vị.

Ta có công thức tính hàm mất mát trên toàn bộ dữ liệu như sau:

$$J(W; X, Y) = - \sum_{i=1}^N \sum_{j=1}^C y_{ji} \log(a_{ji}) \quad (3.1)$$

với hàm Softmax:

$$a_{ji} = \frac{e^{W_j^T x_i}}{\sum_{k=1}^C e^{W_k^T x_i}} \quad (3.2)$$

Tập dữ liệu	Chủ đề	Số lượng	Năm ra đời	Mô tả
<i>Pascal Voc 2012</i>	Chung	17125	2012	Tập dữ liệu chứa nhiều đối tượng như máy bay, người, xe, tàu lửa, động vật,...
<i>Unite the People (UP-S31)</i>	Dáng người	8515	2017	Hình ảnh toàn thân con người khi di chuyển.
<i>Part Labels (Labeled Faces in the Wild - LFW)</i>	Chân dung người	2927	2013	Gán nhãn mặt, đầu và phong nền.
<i>Face/Headseg (FH)</i>	Chân dung người	75	2018	Tập dữ liệu được trích từ 19002 ảnh gán nhãn mặt, mũi, tóc, tai, mắt, lông mày và phong nền.
<i>SVCNTT-2019</i>	Chân dung người	130	2019	Tập dữ liệu được tác giả thu thập trực tiếp từ các sinh viên khoa Công nghệ thông tin và truyền thông (10 tấm/1 sinh viên).

Bảng 3.1: Các tập dữ liệu phổ biến cho phân vùng ngữ nghĩa ảnh

## **CHƯƠNG 4: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

### **4.1 Nhận xét kết quả đạt được:**

Sau thời gian nghiên cứu, tìm hiểu về đề tài dưới sự hỗ trợ của thầy cô bạn bè, báo cáo đã hoàn thành các mục tiêu được đặt ra và đạt kết quả tương đối tốt.

Quá trình giảm mẫu của FCN-8s gồm tích chập thuần và thăm dò được tận dụng khá hiệu quả để rút trích đặc trưng của ảnh và giảm mẫu ảnh về dạng xếp tầng. Quá trình tăng mẫu của FCN-8s gồm tích chập chuyển vị và chuyển tiếp nối kết, với kỹ thuật đầu tiên - tích chập chuyển vị được sử dụng để khôi phục ảnh thay cho kỹ thuật tích chập giãn nở (Dilated Convolution) kết hợp với kỹ thuật thứ hai - chuyển tiếp nối kết vẫn đem lại kết quả như mong muốn.

### **4.2 Hạn chế:**

Với sự nỗ lực hết mức nhưng luận văn vẫn không thể tránh khỏi những thiếu sót do cả lý do khách quan lẫn chủ quan làm giảm tính chất lượng và trải nghiệm của người dùng.

### **4.3 Hướng phát triển:**

Hướng phát triển của bài luận văn với rất nhiều ý tưởng và phương pháp cần xét tới nhưng vẫn chú trọng vào 2 điểm chính.

Một là, nghiên cứu cải thiện những khuyết điểm, hạn chế đã nêu bằng cách gia tăng qui mô của tập dữ liệu với nhiều đối tượng, đa dạng số lượng ảnh, bên cạnh đó áp dụng kỹ thuật từ nghiên cứu của Yu Liu<sup>[3]</sup> chuyên sử dụng cho kỹ thuật DeepLab để cải thiện kỹ thuật đang xét của luận văn, đồng thời cũng sử dụng các loại công nghệ lập trình giao diện ẩn tượng hơn để hoàn thiện về mặt thẩm mỹ cho ứng dụng tách phong.

Hai là, ứng dụng ý tưởng kỹ thuật của chúng đi kèm tập dữ liệu chuyên biệt vào những hệ thống đặc biệt phục vụ người dùng như hệ thống thương mại điện tử có tích hợp thử đồ trực tuyến, ứng dụng ghép ảnh trên điện thoại thông minh, hoặc là hệ thống tách phong các thiết bị máy móc khi hoạt động phẫu thuật cơ thể.

## TÀI LIỆU THAM KHẢO

- [1] *SNOW mobile application*, Camp Mobile, 9 2015, online; accessed 02-April-2019. [Online]. Available: [https://en.wikipedia.org/wiki/Snow\\_\(app\)](https://en.wikipedia.org/wiki/Snow_(app))
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems* 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- [3] Y. Liu and M. S. Lew, “Improving the discrimination between foreground and background for semantic segmentation,” pp. 1272–1276, 09 2017.

## **PHỤ LỤC: CÀI ĐẶT VÀ SỬ DỤNG CHƯƠNG TRÌNH**

### **1. Cài đặt thư viện:**

Các thư viện sử dụng đều được tạo trên hệ điều hành Windows 7 và python 3.6 hoặc lớn hơn, tải về qua liên kết <https://www.python.org/> .

- Sử dụng câu lệnh python để biên dịch và chạy một tập tin theo ngôn ngữ python.
- Cài đặt câu lệnh pip và sử dụng chúng để tải các nội dung khác.
- Các câu lệnh trực tiếp được chạy trên cửa sổ dòng lệnh:
  - python -m pip install -U pip setuptools
  - python -m pip install --upgrade matplotlib numpy pandas scipy scikit-learn opencv-python tensorflow keras imutils

Một trong những lỗi tốn thời gian nhất khi cài đặt thư viện opencv-python trên Windows là "import cv2 ImportError DLL load failed the specified module could not be found". Quá trình giải quyết thông thường là bổ sung các phần mềm hỗ trợ Microsoft Visual Studio để biên dịch opencv nhưng nếu vẫn không thể giải quyết thì cần cập nhật bổ sung Windows bằng chính chức năng Windows Update từ Control Panel của hệ điều hành hoặc giảm phiên bản opencv-python sẽ giải quyết vấn đề trên.

### **2. Sử dụng chương trình:**