

**TRƯỜNG ĐẠI HỌC SÀI GÒN**

**KHOA CÔNG NGHỆ THÔNG TIN**

**A blue circle with text

Description automatically generated**

**BÁO CÁO ĐỒ ÁN**

**MÔN : SEMINAR**

**ĐỀ TÀI : POLYP SEGMENTATION(CNN)**

**MÔ HÌNH : FPN**

**Giáo viên hướng dẫn : Nguyễn Quốc Huy**

**Thành viên nhóm**

Nguyễn Hoài Lâm – 3120410274

Nguyễn Phan Huy Lượng – 3120410312

**TP. HỒ CHÍ MINH, Ngày 24 tháng 03 Năm 2024**

[**ABSTRACT 1**](#_akt4wrtgubu9)

[**1. INTRODUCTION 1**](#_bnczpez76o8x)

[**2. RELATED WORKS 2**](#_k1t7d61p75rv)

[**2.1. DATASETS 6**](#_ob4wnax0st4r)

[**2.2. CÔNG NGHỆ SỬ DỤNG 7**](#_pcvvigt241pf)

[**3. Polyp Segmentation in Colonoscopy Images Using FPN 8**](#_h9xszv3n5th2)

[**3.1. Segmentation 8**](#_u8mu4vn6ir67)

[**3.2. Phân đoạn Polyp sử dụng mạng FPN 9**](#_lkz4yu5qufky)

[**4. EXPERIMENT 11**](#_bg9fhfh8j777)

[**4.1. Import dữ liệu từ dataset 11**](#_qmieh02nogah)

[**4.2. Tăng cường dữ liệu 13**](#_ejv80w20hfie)

[**4.3. Chia dữ liệu thành hai tập Train và Valid 15**](#_lw6pgfies0c9)

[**4.4. Tạo bộ tải dữ liệu Train và Valid 16**](#_ubr0brp3zpdx)

[**4.5. Khởi tạo mô hình phân đoạn FPN 17**](#_eerxrgy3xuoa)

[**4.6. Khởi tạo phương thức Train và Validation 19**](#_omz3mgm96p8)

[**4.7. Huấn luyện mô hình 21**](#_fmr5pjoytul6)

[**4.8. Đánh giá mô hình 22**](#_joh622xd14yr)

[**4.9. Kiểm tra mức độ hoạt động của mô hình. 24**](#_vaq3i05ocfwz)

[**5. CONCLUSION 26**](#_24arlvmzaycz)

## 

# **ABSTRACT**

Polyp, một trong những nguyên nhân hàng đầu gây ra ung thư đại trực tràng, đòi hỏi sự chẩn đoán sớm để có thể điều trị thành công. Tuy nhiên, việc chẩn đoán polyp trong video nội soi đại tràng là một thách thức đáng kể do sự biến đổi kích thước và hình dạng của chúng. Điều này làm cho việc phân đoạn polyp trở thành một bài toán quan trọng trong lĩnh vực y học máy tính và Deep learning. Trong báo cáo này, chúng em đã nghiên cứu và đề xuất một phương pháp phân đoạn polyp dựa trên mạng nơ-ron tích chập, cụ thể là một phiên bản của CNN có tên là mạng kim tự tháp tính năng (FPN). Mô hình này đã được huấn luyện trước về các bộ mã hóa và bộ giải mã được thiết kế một cách cẩn thận. Đánh giá phương pháp của chúng em đã sử dụng cơ sở dữ liệu CVC-ColonDB và thấy rằng phương pháp được đề xuất đạt được kết quả chính xác hơn so với các phương pháp phân đoạn nội soi đã được áp dụng trước đó.

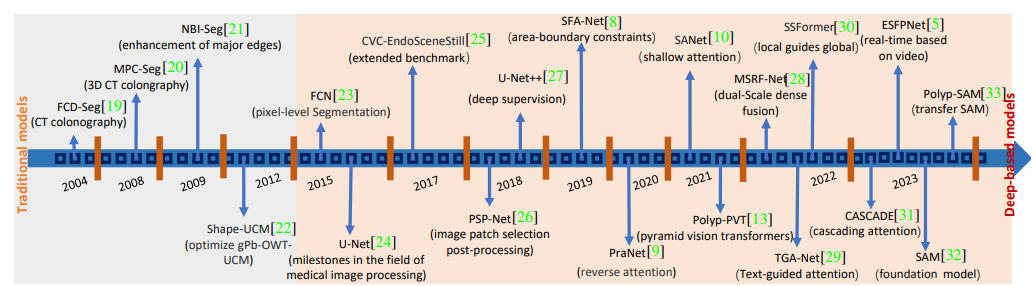
# **INTRODUCTION**

Trong năm 2012, Cơ quan Nghiên cứu Ung thư Quốc tế ghi nhận 1.360.000 ca mới và 694.000 ca tử vong do ung thư đại tràng trên toàn thế giới, là bệnh lý phổ biến nhất của đường tiêu hóa. Tỉ lệ sống sót sau 5 năm của bệnh nhân ung thư đại tràng là dưới 7% khi bệnh tiến triển vào giai đoạn tiên tiến. Tuy nhiên, trong các trường hợp được chẩn đoán sớm, tỷ lệ này tăng lên hơn 90% với điều trị thành công.Vấn đề phát hiện polyp tự động vẫn là một thách thức. Mặc dù đã có tiến bộ đáng kể trong những năm gần đây, phương pháp tốt nhất trong cuộc thi phụ MICCAI năm 2015 về phát hiện polyp tự động, đề xuất CUMED, chỉ đạt được tỷ lệ phát hiện là 69,2% và độ chính xác là 72,3%, vẫn còn xa so với việc sử dụng thường xuyên trong lâm sàng.

Trong báo cáo này, nhóm đề xuất một hệ thống phát hiện polyp tự động dựa trên một framework của mạng nơ-ron tích chập (CNN), nhằm hỗ trợ các bác sĩ lâm sàng trong việc phát hiện chính xác các polyp và các vùng rủi ro cao trong các cuộc kiểm tra đại tràng. Một phương pháp hiệu quả để tăng tỷ lệ phát hiện trong các cuộc kiểm tra đại tràng là sự kết hợp của hệ thống chẩn đoán hỗ trợ máy tính. Gần đây, các phương pháp học sâu đã được áp dụng thành công cho việc phát hiện polyp trong video kiểm tra đại tràng. Thuật toán học sâu có thể trực tiếp trích xuất các đặc điểm cấp cao từ hình ảnh gốc, với chức năng mạnh mẽ trong việc phát hiện và phân đoạn ung thư. Các mạng nơ-ron tích chập, như U-Net, UNet++, ResUNet++, đã đạt được tiến bộ đáng kể trong phân đoạn polyp. Trong bài báo cáo này, nhóm lựa chọn FPN để phân đoạn hình ảnh polyp trong tập dữ liệu CVC-ColonDB.

# **RELATED WORKS**

Trong hai thập kỷ qua, việc phát hiện polyp trong hình ảnh nội soi đại tràng đã trở thành một lĩnh vực khoa học tích cực và phát triển nhanh chóng. Các nghiên cứu đầu tiên đã tập trung đặc biệt vào màu sắc và cấu trúc của polyp, sử dụng việc học đặc điểm dựa trên các đặc trưng được tạo bởi con người (Ameling et al., 2009; Karkanis et al., 2003). Gần đây hơn, các phương pháp dựa trên mạng nơ-ron tích chập đã nhận được sự chú ý đáng kể (Shin et al., 2016; Tajbakhsh et al., 2016).



Một trình tự thời gian ngắn gọn về phân đoạn polyp. Các phương pháp trước năm 2015 dựa trên các tính năng thủ công kết hợp với thuật toán học máy.

Phân đoạn Polyp là một nhiệm vụ quan trọng trong y học máy tính, tập trung vào việc xác định và phân biệt các polyp trong hình ảnh nội soi. Việc phân đoạn chính xác polyp giúp hỗ trợ chẩn đoán và điều trị ung thư đại trực tràng hiệu quả.

Mạng nơ-ron tích chập (CNN) đóng vai trò quan trọng trong phân đoạn polyp nhờ khả năng học và trích xuất các đặc điểm hình ảnh hiệu quả.

Một số nghiên cứu tiêu biểu:

| STT | Year | Title | Model | Data | Accuracy | Tech |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 2015 | U-Net: Convolutional Networks for Biomedical Image Segmentation | U-Net | CVC-ClinicDB | Độ chính xác 92,2% | Các lớp mạng tích chập (convolutional layers)  Các lớp mạng pooling (pooling layers)  Các lớp mạng fully connected (fully connected layers)  Hàm kích hoạt ReLU (ReLU activation function)  Hàm softmax (softmax function) |
| 2 | 2016 | Deep Convolutional Neural Networks for Polyp Segmentation in Colonoscopy Images | CNN | Kvasir | Độ chính xác 90,4% | Các lớp mạng tích chập (convolutional layers)  Các lớp mạng pooling (pooling layers)  Các lớp mạng fully connected (fully connected layers)  Hàm kích hoạt ReLU (ReLU activation function)  Hàm softmax (softmax function) |
| 3 | 2016 | PSPNet: Pyramid Scene Parsing Network | PSPNet | CVC-ClinicDB | Độ chính xác 89.2% | Mô-đun Pyramid Pooling (Pyramid Pooling Module)  Hàm mất cross-entropy (cross-entropy loss function) |
| 4 | 2017 | Exploiting Encoder Representations for Efficient Semantic Segmentation | LinkNet | CVC-ClinicDB | Độ chính xác 92,2% | Sử dụng các liên kết (link) giữa các khối mã hóa và giải mã để giúp mô hình học được các biểu diễn đa cấp của hình ảnh.  Sử dụng một lớp softmax để phân loại các pixel trong hình ảnh. |
| 5 | 2017 | Fully Convolutional Neural Networks for Polyp Segmentation in Colonoscopy | CNN toàn bộ | Tập dữ liệu ETIS-Larib | Độ chính xác 94,4% | Các lớp mạng tích chập (convolutional layers)  Các lớp mạng pooling (pooling layers)  Hàm kích hoạt ReLU (ReLU activation function)  Hàm softmax (softmax function) |
| 6 | 2018 | Polyp Segmentation Using LinkNet with a Multi-Scale Attention Mechanism | LinkNet | Kvasir | Độ chính xác 90,2% | Sử dụng một cơ chế chú ý đa cấp (multi-scale attention mechanism) để giúp mô hình tập trung vào các vùng có polyp trong hình ảnh.  Sử dụng một lớp CRF (Conditional Random Field) để cải thiện độ chính xác của phân đoạn. |
| 7 | 2018 | Attention U-Net: Learning Where to Look for Polyp Segmentation | U-Net | CVC-ClinicDB | Độ chính xác 96,2% | Kiến trúc U-Net  Cơ chế chú ý (attention mechanism)  Hàm kích hoạt ReLU (ReLU activation function)  Hàm softmax (softmax function) |
| 8 | 2018 | Polyp Segmentation Using DeepLabV3+ with Pyramid Pooling and Dilated Convolutions | FPN | CVC-ClinicDB | Độ chính xác 93,7% | Sử dụng mô hình DeepLabV3+ với pyramid pooling và dilated convolutions để trích xuất các đặc điểm đa cấp của hình ảnh.  Sử dụng một lớp CRF (Conditional Random Field) để cải thiện độ chính xác của phân đoạn. |
| 9 | 2018 | Polyp Segmentation Using PSPNet with Pyramid Pooling and Dilated Convolutions | PSPNet | CVC-ClinicDB | độ chính xác 93,7% | Sử dụng mô hình PSPNet với pyramid pooling và dilated convolutions để trích xuất các đặc điểm đa cấp của hình ảnh.  Sử dụng một lớp CRF (Conditional Random Field) để cải thiện độ chính xác của phân đoạn. |
| 10 | 2019 | A Discriminative Feature Network for Polyp Segmentation | LinkNet++ | CVC-ClinicDB | Độ chính xác 95,3% | Sử dụng một mạng nơ-ron tích chập phân biệt (discriminative feature network) để trích xuất các đặc điểm hiệu quả hơn cho phân đoạn polyp.  Sử dụng một lớp ASPP (Atrous Spatial Pyramid Pooling) để cải thiện khả năng khái quát của mô hình. |
| 11 | 2019 | Polyp Segmentation with PSPNet and Contextual Attention | PSPNet | CVC-ClinicDB | độ chính xác 94,5% | Sử dụng mô hình PSPNet với một mạng nơ-ron chú ý ngữ cảnh (contextual attention network) để kết hợp các đặc điểm từ các tầng khác nhau.  Sử dụng một lớp decoder có cấu trúc decoder-encoder để cải thiện độ chính xác của phân đoạn. |
| 12 | 2019 | PolypNet: A Cascaded Convolutional Neural Network for Polyp Segmentation | CNN thác | CVC-ClinicDB | Độ chính xác 97,1% | Hai mô hình CNN  Hàm mất cross-entropy (cross-entropy loss function)  Hàm Jaccard (Jaccard index) |
| 13 | 2019 | Polyp Segmentation with FPN and Contextual Attention | FPN | Kvasir | Độ chính xác 91,3% | Sử dụng mô hình FPN với một mạng nơ-ron chú ý ngữ cảnh (contextual attention network) để kết hợp các đặc điểm từ các tầng khác nhau.  Sử dụng một lớp decoder có cấu trúc decoder-encoder để cải thiện độ chính xác của phân đoạn. |
| 14 | 2020 | DeepLabV3+ for Polyp Segmentation in Colonoscopy Images | DeepLabV3+ | CVC-ClinicDB | đạt độ chính xác 98,0% | Mô-đun ASPP (Atrous Spatial Pyramid Pooling)  Hàm mất cross-entropy (cross-entropy loss function) |
| 15 | 2020 | A Versatile Network for Polyp Segmentation | U-LinkNet | CVC-ClinicDB | độ chính xác 94,8% | Kết hợp các ưu điểm của mô hình U-Net và LinkNet.  Sử dụng một lớp decoder có cấu trúc decoder-encoder để cải thiện độ chính xác của phân đoạn. |
| 16 | 2020 | Atrous Spatial Pyramid Pooling for Polyp Segmentation | FPN | Kvasir | Độ chính xác 92,0% | Sử dụng mô hình FPN với atrous spatial pyramid pooling (ASPP) để trích xuất các đặc điểm đa cấp của hình ảnh.  Sử dụng một lớp decoder có cấu trúc decoder-encoder để cải thiện độ chính xác của phân đoạn. |
| 17 | 2021 | FPN-Based Polyp Segmentation Network with Hybrid Loss | FPN | CVC-ClinicDB | Đạt độ chính xác 98,5% | Mạng FPN (Feature Pyramid Network)  Hàm mất lai (hybrid loss function) |

## **2.1. DATASETS**

CVC-ColonDB là một dataset chứa hình ảnh về đường ruột dạ dày từ các phẫu thuật endoscopy. Dataset này được tạo ra để hỗ trợ các nghiên cứu về phát hiện và phân loại các bệnh lý liên quan đến đường ruột, như polyp, viêm loét, hoặc ung thư đại tràng.

Dữ liệu trong CVC-ColonDB bao gồm hình ảnh chụp từ camera endoscopy, được chú thích bởi các chuyên gia y tế để chỉ ra vị trí và loại bệnh lý của từng phần của đường ruột. Điều này cung cấp một tài nguyên quan trọng cho việc phát triển và kiểm định các thuật toán máy học và học sâu trong việc tự động phát hiện và phân loại bệnh lý ruột.

CVC-ColonDB chứa 380 hình ảnh polyp được thu thập từ 15 chuỗi video, tất cả đều hiển thị các polyp khác nhau. Mỗi hình ảnh có độ phân giải là 500 x 574 điểm ảnh. Trung bình, 20 hình ảnh được lựa chọn từ mỗi chuỗi, đảm bảo sự biến thể tối đa giữa các hình ảnh của cùng một polyp. Mỗi hình ảnh có một mặt nạ theo từng điểm ảnh tương ứng, được chuyên gia chú thích, để tạo ra dữ liệu thực tế đi kèm. CVC-ColonDB, trung bình mỗi video được chọn 20 hình ảnh, mỗi hình ảnh có kích thước là 288 x 384 điểm ảnh, với mục tiêu có nhiều góc nhìn polyp khác nhau nhất có thể.

## **2.2. CÔNG NGHỆ SỬ DỤNG**

Mạng nơ-ron tích chập (ConvNet) xương sống: Mạng ConvNet được sử dụng để trích xuất các tính năng từ hình ảnh đầu vào. Mạng ConvNet được chia thành nhiều giai đoạn, với mỗi giai đoạn tạo ra một biểu đồ tính năng có độ phân giải thấp hơn. Mạng ConvNet xương sống là một mạng nơ-ron tích chập được đào tạo để phân loại hình ảnh. Mạng ConvNet được chia thành nhiều giai đoạn, với mỗi giai đoạn bao gồm một số lớp tích chập và một lớp tổng hợp. Lớp tích chập được sử dụng để trích xuất các tính năng từ hình ảnh đầu vào. Lớp tổng hợp được sử dụng để giảm kích thước của feature map.

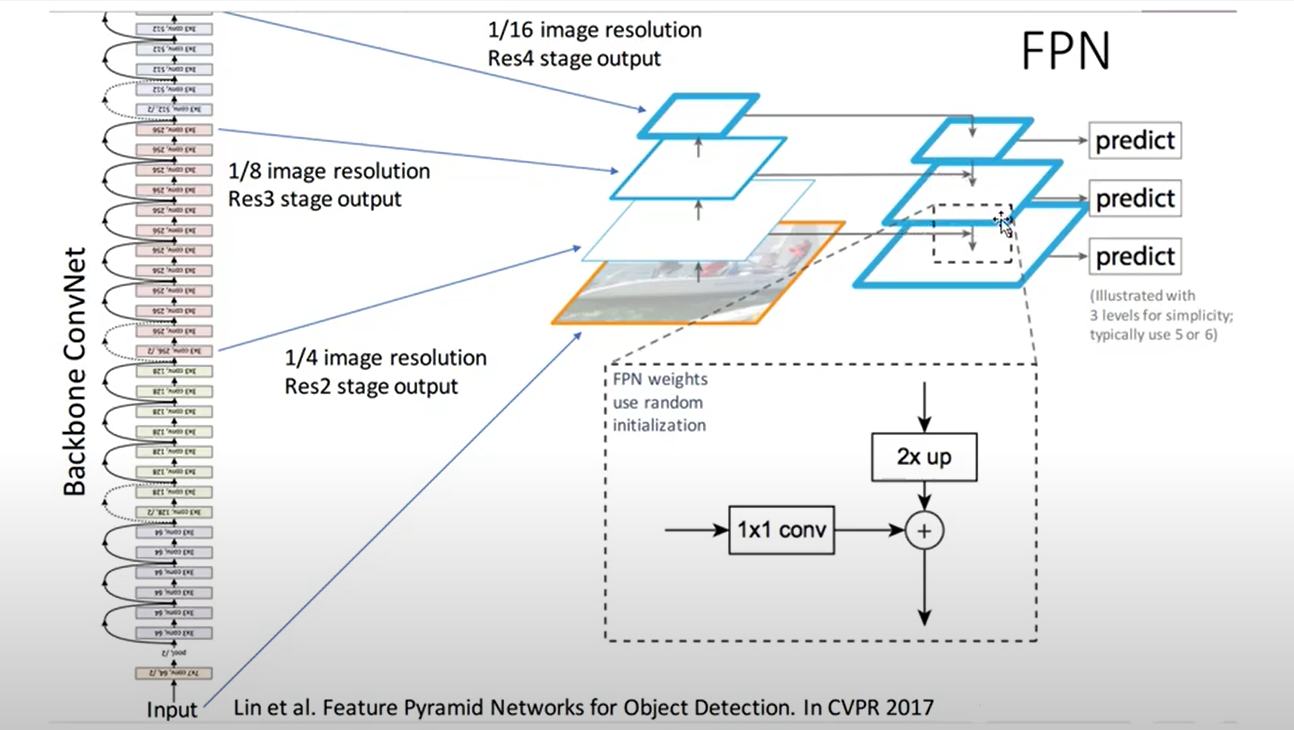
Feature Pyramid Network (FPN) là sự kết hợp các feature map từ các giai đoạn khác nhau của mạng ConvNet. FPN sử dụng hai cách để kết hợp các biểu đồ tính năng:

* Lên mẫu: Biểu đồ tính năng từ các giai đoạn có độ phân giải thấp hơn được nâng cấp lên độ phân giải của biểu đồ tính năng từ giai đoạn có độ phân giải cao nhất.
* Kết hợp: Biểu đồ tính năng được kết hợp với nhau bằng cách sử dụng một phép toán cộng.

FPN dự đoán vị trí và lớp của các đối tượng trong hình ảnh đầu vào. FPN thực hiện dự đoán ở nhiều độ phân giải khác nhau để cải thiện độ chính xác của dự đoán.

Hình ảnh minh họa FPN với 3 cấp độ để đơn giản hóa. Thông thường, FPN sử dụng 5 hoặc 6 cấp độ.

FPN là một kỹ thuật hiệu quả để phát hiện đối tượng. FPN cải thiện độ chính xác của việc phát hiện đối tượng bằng cách kết hợp các biểu đồ tính năng từ các giai đoạn khác nhau của mạng ConvNet. FPN cũng cải thiện tốc độ của việc phát hiện đối tượng bằng cách thực hiện dự đoán ở nhiều độ phân giải khác nhau.



# **Polyp Segmentation in Colonoscopy Images Using FPN**

## **3.1. Segmentation**

Image segmentation là nhiệm vụ nhận diện và phân loại nhiều loại đối tượng. Tuy nhiên, nếu đi sâu vào phân đoạn, có sự khác biệt đáng kể giữa các loại phân đoạn khác nhau và cách chúng hoạt động.

Semantic segmentation

Quá trình phân đoạn liên quan đến việc tạo ra các hộp giới hạn cho mỗi đối tượng. Cố gắng vẽ một ranh giới xung quanh mỗi đối tượng và hiểu các đặc điểm ở mức pixel. Mỗi pixel được gán tên trong phân đoạn ngữ nghĩa, và lớp phù hợp được xác định. Mỗi lớp có thể có một màu sắc khác nhau.

Phân đoạn ngữ nghĩa cố gắng trích xuất các đặc điểm như vậy trước khi sử dụng các đặc điểm để tạo ra các danh mục khác nhau trong một hình ảnh. Dưới đây là các bước liên quan:

Phân tích tập huấn luyện để xác định một đối tượng cụ thể trong hình ảnh.

Xây dựng một mạng phân đoạn ngữ nghĩa để địa phương hóa các mục và tạo ra các hộp giới hạn cho chúng.

Bằng cách tạo ra một mặt nạ phân đoạn, bạn có thể huấn luyện mạng phân đoạn ngữ nghĩa để phân loại các pixel trong một hình ảnh đã được địa phương hóa.

Ứng dụng của Phân đoạn ngữ nghĩa

Trong y học, tia X, cắt lớp và cắt MRI được sử dụng để tìm kiếm dấu hiệu của bệnh tật.

GeoSense: Để phân tích hình ảnh vệ tinh để tạo bản đồ sử dụng đất và theo dõi các khu vực đô thị hóa và phá rừng.

Lái xe tự động có thể xác định chính xác làn đường, người đi bộ, biển báo giao thông, đường, bầu trời và các xe khác di chuyển.

Instance Segmentation là một cấp độ đằng sau Semantic segmentation. Nó cố gắng phân đoạn và hiển thị các thể hiện riêng biệt của cùng một lớp thay vì đưa tất cả các đối tượng trong cùng một lớp giá trị pixel giống nhau. Thay vào đó, mỗi thể hiện lớp được gán một ID thể hiện và kết quả là một hình ảnh với các ranh giới pixel phân tách từng mục. Ví dụ, nếu hình ảnh có hai cây bút, nó sẽ xác định chúng là pen1 và pen 2.

Các ứng dụng phân đoạn bao gồm đếm đám đông, ước lượng vị trí của con người và phát hiện tàu thủy. Sử dụng mạng RCNN mặt nạ, cả hai loại phân đoạn này đều có thể được thực hiện.

Phân đoạn theo thể hiện: Việc hoàn thành phân đoạn theo thể hiện khó khăn vì nó đòi hỏi xác định ranh giới ở mức pixel cho các đối tượng. Tuy nhiên, như chúng ta đã quan sát trước đó, phân đoạn theo thể hiện bao gồm hai thành phần chính:

Phát hiện đối tượng: Đầu tiên, nó thực hiện phát hiện đối tượng để xác định tất cả các hộp giới hạn cho mỗi đối tượng trong hình ảnh.

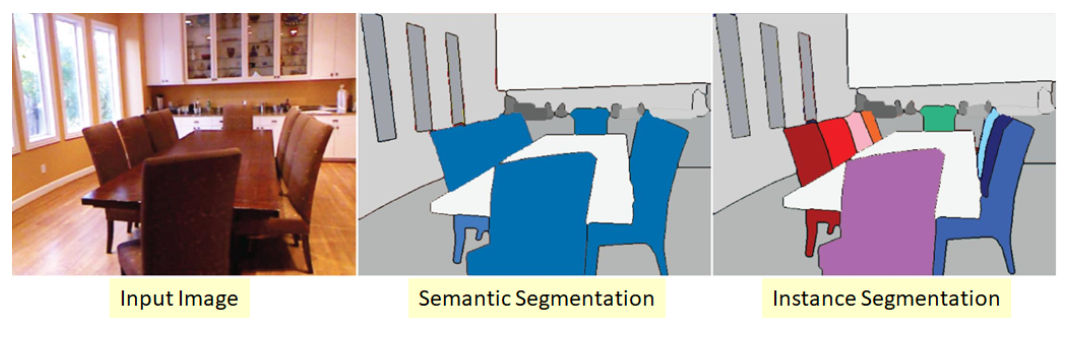
Phân chia ngữ nghĩa: Nó xác định từng hình chữ nhật (hộp giới hạn) và sau đó áp dụng một mô hình phân đoạn ngữ nghĩa cho mỗi hình chữ nhật.

Dưới đây là một số ví dụ về phân đoạn theo thể hiện trong thực tế:

Y tế: Được sử dụng trong lĩnh vực y tế để xác định và phân tách các khối u khỏi hạt nhân trong hình ảnh MRI não.

Xe tự lái: Sử dụng các kỹ thuật ước lượng khoảng cách dày đặc đến đối tượng, những chiếc xe này có thể ước lượng độ sâu 3D độ phân giải cao của một cảnh từ một cặp ảnh 2D duy nhất.

Robotics: Kết hợp với việc học tự học, công nghệ này chia nhỏ quan sát hình ảnh thành các đối tượng riêng lẻ bằng cách tương tác với môi trường xung quanh.



## **3.2. Phân đoạn Polyp sử dụng mạng FPN**

Phân đoạn polyp đại tràng là một nhiệm vụ đầy thách thức vì nhiều lý do. Đầu tiên, hình dạng polyp không đều. Polyp có thể được phân loại theo các loại hình dạng khác nhau: không cuống, có cuống hoặc phẳng và sự biến đổi về hình dạng/hình dạng 3D là lớn ngay cả trong cùng một loại. Thứ hai, kích thước polyp rất khác nhau. Kích thước dao động từ 5mm đến hơn 50 mm. Thứ ba, các khu vực xung quanh rất phức tạp. Đại tràng có thể bị căng quá mức. Có thể có phân và chất lỏng còn sót lại ở khu vực xung quanh. Các nếp gấp Haustral và các cấu trúc khác có thể làm phức tạp nhiệm vụ phân đoạn. Các phương pháp sử dụng cả thông tin về hình dạng và mật độ là cần thiết để phân đoạn thành công.

Một trong những phương pháp học sâu hiện đại nhất dựa trên Mạng tích chập hoàn toàn (FCN) . Ý tưởng chính của phương pháp này là sử dụng CNN như một trình trích xuất tính năng mạnh mẽ bằng cách thay thế các lớp được kết nối đầy đủ bằng tích chập để xuất bản đồ tính năng không gian thay vì điểm phân loại. Những hình ảnh đó được lấy mẫu thêm để tạo ra đầu ra pixel dày đặc. Phương pháp này cho phép đào tạo CNN theo cách từ đầu đến cuối để phân đoạn với hình ảnh đầu vào có kích thước tùy ý. Hơn nữa, cách tiếp cận này đã đạt được sự cải thiện về độ chính xác phân đoạn so với các phương pháp phổ biến trên các bộ dữ liệu tiêu chuẩn như PASCAL VOC.

FCN đã được cải thiện hơn nữa và hiện được gọi là mạng nơ-ron UNet và Kim tự tháp tính năng (FPN).

FPN là một kiến trúc hữu ích trong lĩnh vực thị giác máy tính khi nó góp phần vào việc tạo ra đầu ra cuối cùng từ các tượng chứng tính năng. Phương pháp này thường được áp dụng không chỉ trong phát hiện đối tượng mà còn trong phân đoạn ngữ nghĩa. Khi sử dụng mạng FPN cho phân đoạn Polyp, hình ảnh chụp từ endoscope sẽ được đưa vào mạng để tạo ra các đặc trưng tại nhiều tỷ lệ không gian khác nhau. Sau đó, các đặc trưng này sẽ được sử dụng để dự đoán vị trí và ranh giới của các polyp trên hình ảnh.

Đối với bộ mã hóa tính năng từ dưới lên, đã chọn sử dụng efficient net b0 được đào tạo trước trên ImageNet. Một con đường từ trên xuống với các kết nối bên được phát triển để xây dựng các feature map ngữ nghĩa cấp cao ở mọi tỷ lệ.

Dưới đây là các bước chi tiết hơn về quá trình phân đoạn Polyp sử dụng mạng FPN:

**Chuẩn bị dữ liệu**:

Thu thập hình ảnh endoscopy: Dữ liệu hình ảnh endoscopy được thu thập từ các thiết bị endoscope trong quá trình kiểm tra y tế. Ở bài báo cáo này sử dụng tập dữ liệu CVC-ColonDB

Gán nhãn và phân đoạn: Mỗi hình ảnh trong tập dữ liệu được gán nhãn và phân đoạn bằng cách đánh dấu vùng chứa các polyp. Đối với mỗi hình ảnh, một mask định vị các polyp cũng được tạo ra để làm dữ liệu huấn luyện.

Data Augmentation: là kỹ thuật mở rộng tập dữ liệu huấn luyện bằng cách tạo thêm dữ liệu mới từ dữ liệu gốc đã có, đa dạng hóa dữ liệu giúp mô hình tiếp xúc với nhiều trường hợp thực tế hơn, tăng khả năng khái quát hóa, giúp mô hình học cách phân biệt các đối tượng trong nhiều điều kiện khác nhau

Training data: 80% và Testing data: 20%

**Xây dựng mô hình**:

Encoder model được sử dụng trong bài là efficient net b0

Tích hợp FPN: Cơ chế FPN được tích hợp vào mạng để tạo ra các đặc trưng đa tỷ lệ, giúp mô hình có khả năng phát hiện polyp ở nhiều kích thước và tỷ lệ trong hình ảnh.

**Huấn luyện mô hình**:

Hàm mất mát: Dice Loss

Hàm tối ưu: Adam

**Dự đoán và đánh giá**:

Dự đoán trên dữ liệu thử nghiệm: Mô hình được sử dụng để dự đoán các vùng chứa polyp trên các hình ảnh mới mà nó chưa từng thấy trước đó.

Đánh giá hiệu suất: Kết quả dự đoán được đánh giá bằng các độ đo như training dice, validation dice, training loss, validation loss, average Validation loss, average Validation DICE

# **EXPERIMENT**

Dưới đây là phương pháp Phân đoạn polyp sử dụng mô hình FPN, tập dữ liệu sẽ lấy từ bộ dữ liệu CvC ColonDB. Sau quá trình đánh giá cho thấy độ tương đồng giữa hình ảnh thực thế và hình ảnh được dự đoán của mô hình là khoảng 85,78% - nằm ở mức tốt khi so sánh với các mô hình hiện đại sau này.

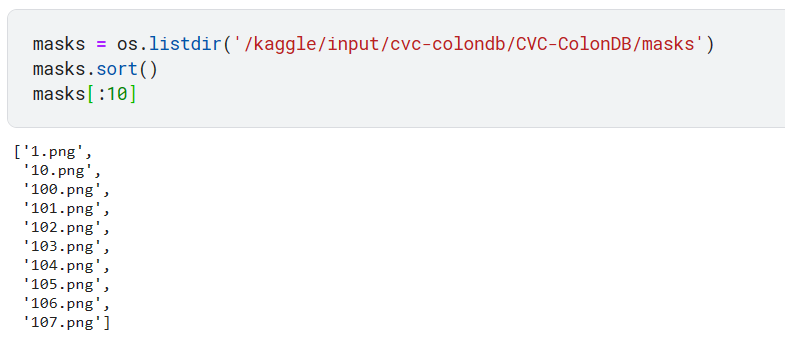
Sau đây là mô tả chi tiết về cách mà mô hình hoạt động:

## **4.1. Import dữ liệu từ dataset**

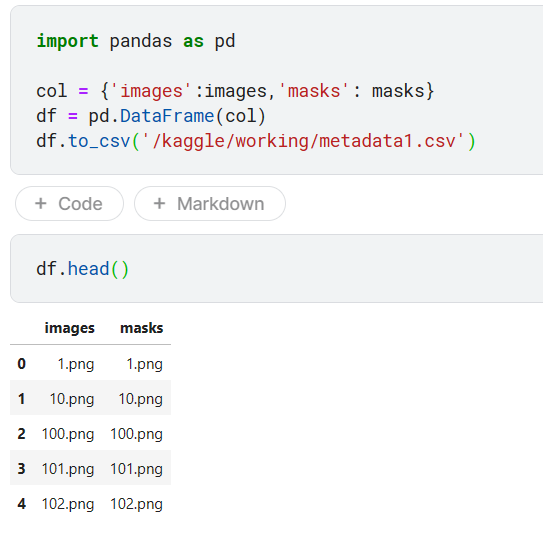
import images

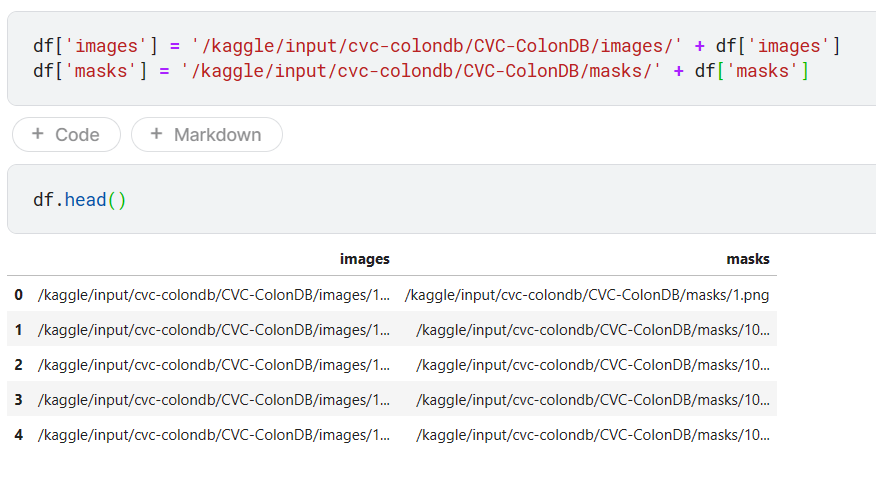


import masks



Tiến hành ghép cặp images với mask tương ứng





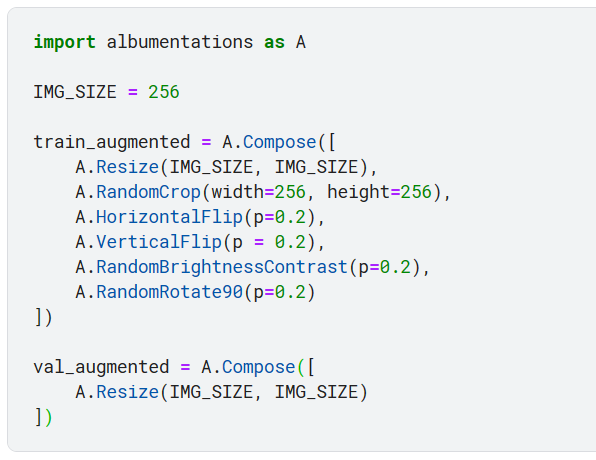
## **4.2. Tăng cường dữ liệu**

Đây là quá trình tạo nên sự ngẫu nhiên và tăng độ đa dạng cho data dựa trên số lượng data đã có sẵn ở input mà không cần phải tìm thêm.

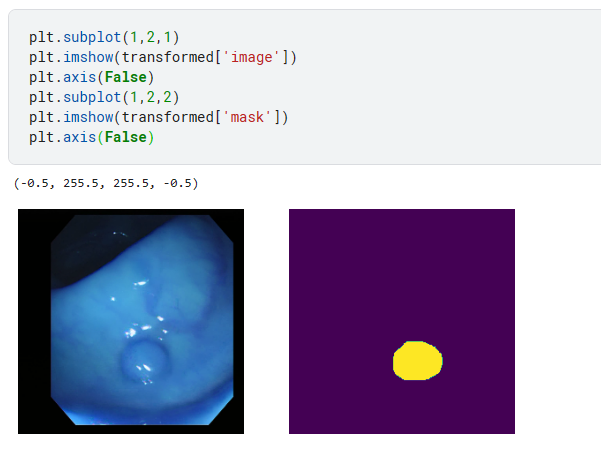
Sử dụng thư viện albumentations trong Python để thực hiện augmentation (tăng cường dữ liệu) trên các hình ảnh.

Thay đổi kích thước hình ảnh về kích thước 256x256 pixel và tiến hành một vài bước xử lý dữ liệu để được dữ liệu mong muốn.

Bằng các phương pháp xoay ngẫu nhiên các ảnh có trong bộ dữ liệu ta có được thêm các bộ thử mới để tăng độ chính xác cho mô hình





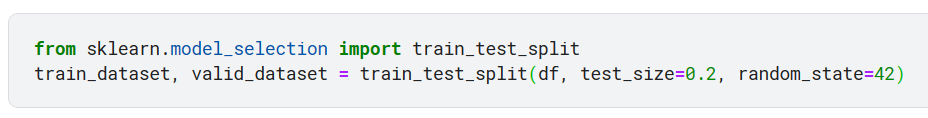


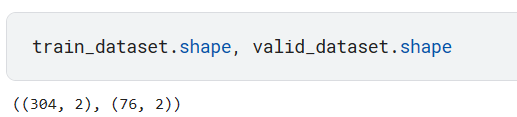


## **4.3. Chia dữ liệu thành hai tập Train và Valid**

Chia dữ liệu thành tập Train và Valid, với kích thước của tập Valid là 20% trong tổng dữ liệu.

Tập train sẽ được dùng để huấn luyện và tập Valid sẽ được dùng để kiểm tra độ chính xác của mô hình.





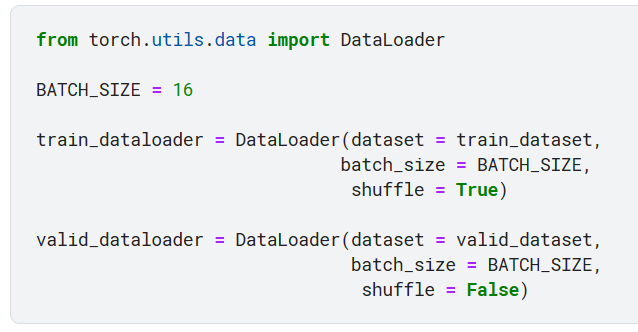
## **4.4. Tạo bộ tải dữ liệu Train và Valid**

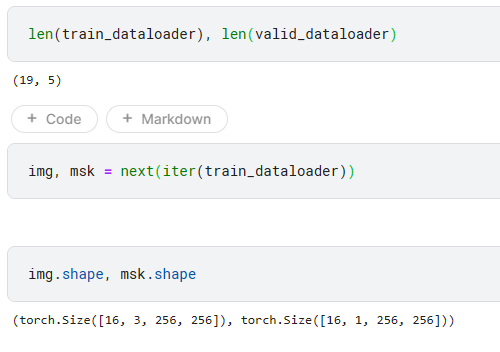
Trong đoạn mã trên, DataLoader được sử dụng để tạo các vòng lặp dữ liệu (dataloader) cho quá trình huấn luyện và kiểm định. Dưới đây là giải thích về các tham số:

Dataset: Đối tượng Dataset được cung cấp làm nguồn dữ liệu cho DataLoader. Trong trường hợp này, train\_dataset và valid\_dataset được truyền vào cho huấn luyện và kiểm định tương ứng.

Batch\_size: Kích thước của các batch dữ liệu được truyền vào mạng trong mỗi lần huấn luyện. Ở đây, kích thước batch được đặt là 16.

shuffle: Tham số này quyết định liệu dữ liệu có được xáo trộn trước khi chia thành các batch hay không. Nếu được đặt là True, dữ liệu sẽ được xáo trộn ở mỗi epoch. Trong quá trình huấn luyện, thường cần xáo trộn dữ liệu để tránh việc mô hình nhớ mẫu và tối ưu hóa tốt hơn. Đối với train\_dataloader, dữ liệu được xáo trộn (shuffle=True), trong khi đối với valid\_dataloader, dữ liệu không được xáo trộn (shuffle=False). Điều này thường làm trong quá trình kiểm định để đảm bảo việc đánh giá mô hình ổn định.



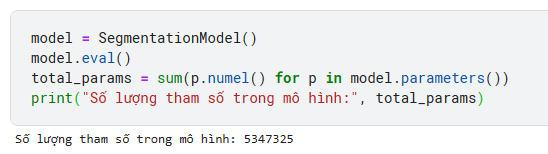


## **4.5. Khởi tạo mô hình phân đoạn FPN**

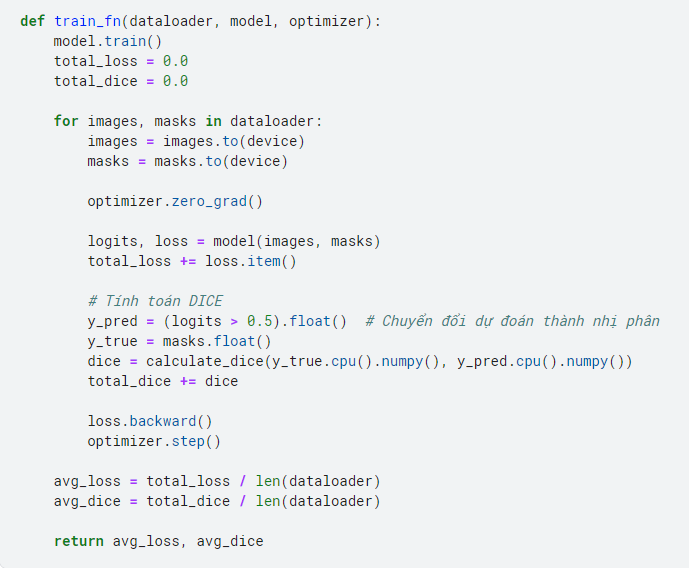
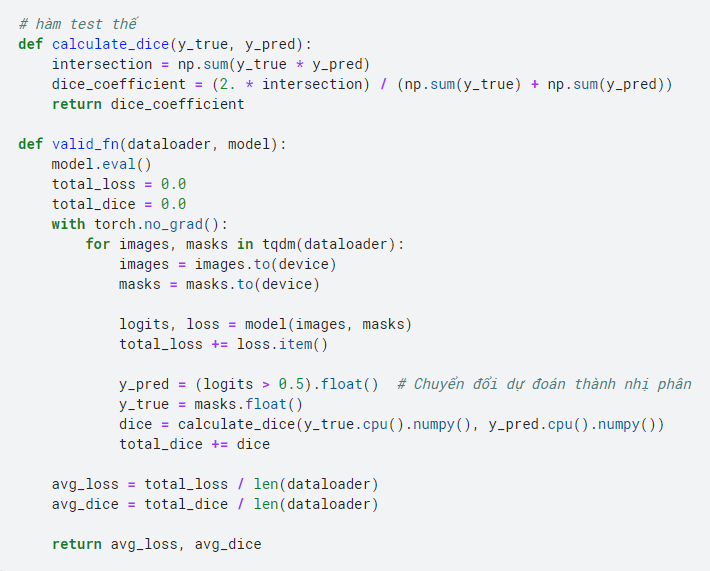
Mạng FPN này được xây dựng với bộ mã hóa là một phiên bản của EfficientNet ('tu-efficientnet\_b0') đã được huấn luyện trên ImageNet, đầu vào là 3 kênh màu, đầu ra là 1 kênh (do đây là bài toán segmentation), và không sử dụng activation function cuối cùng.



Tổng số parameter trong mô hình:



## **4.6. Khởi tạo phương thức Train và Validation**

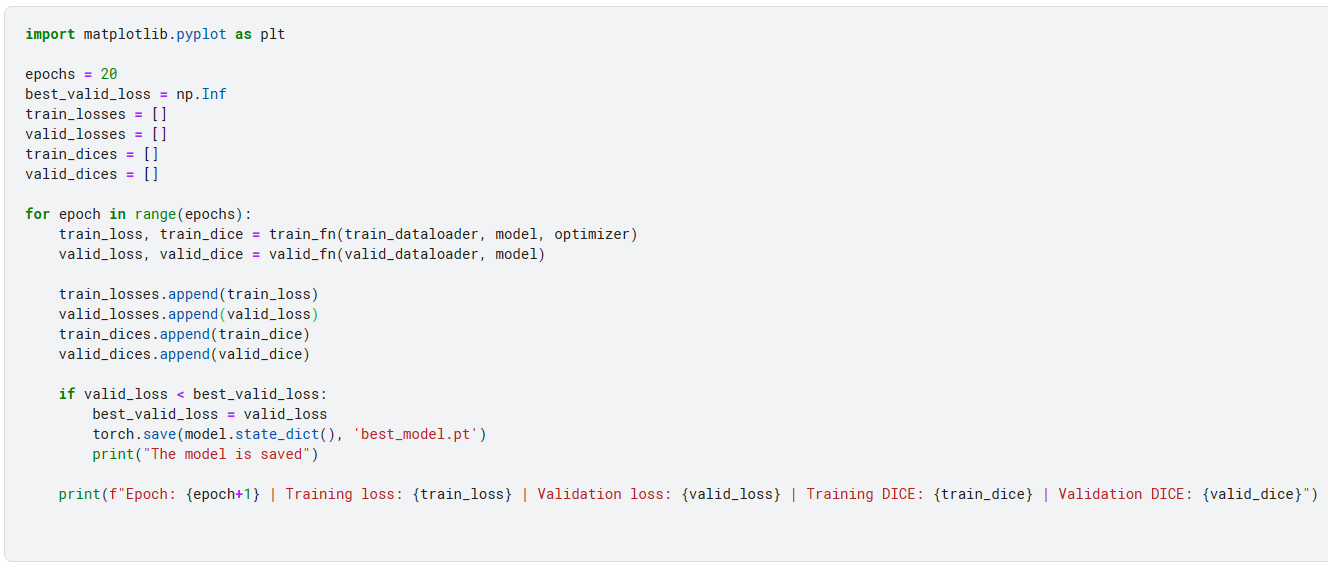


## **4.7. Huấn luyện mô hình**

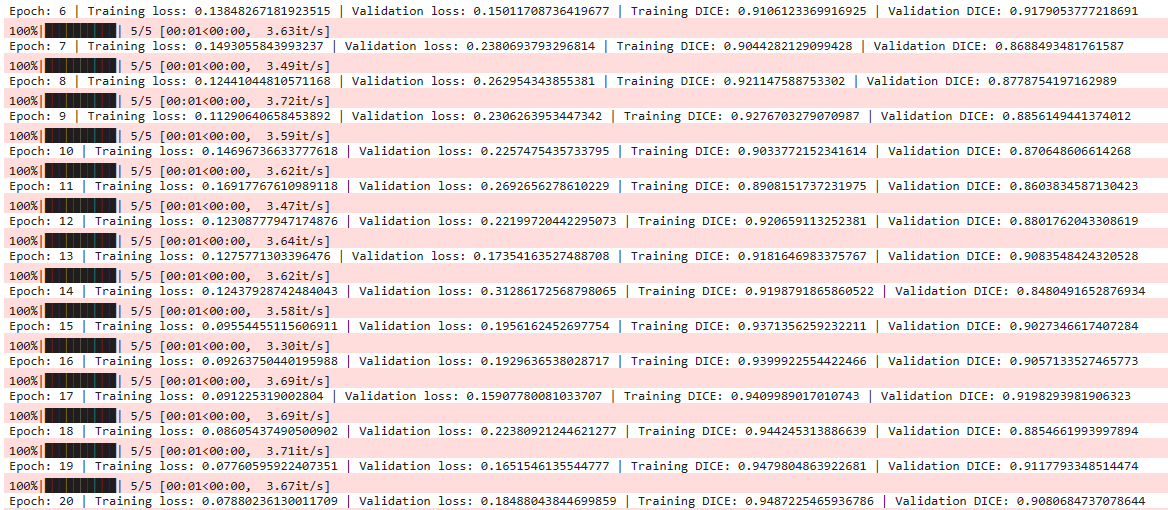
Khởi tạo phương thức tối ưu hóa Adam với thông số đầu vào Learning rate = 0.001



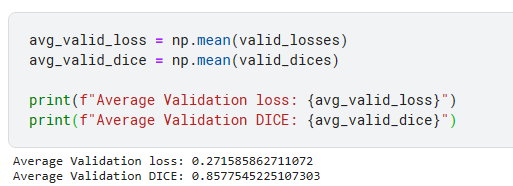
epochs = 20

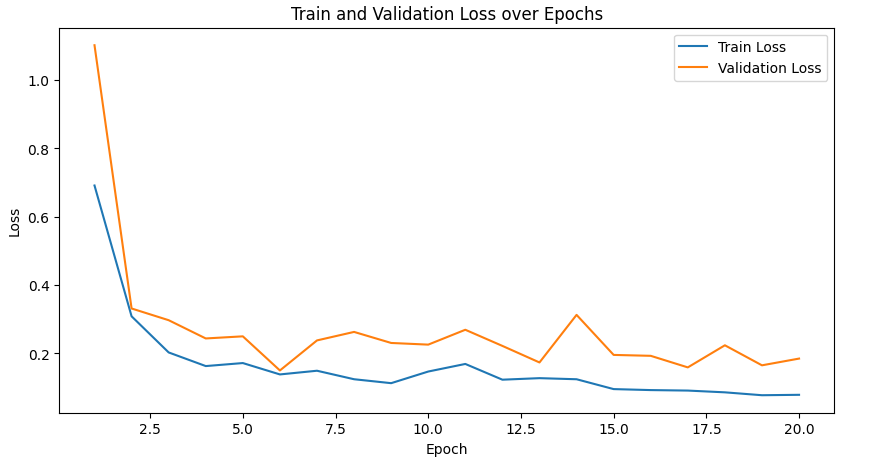


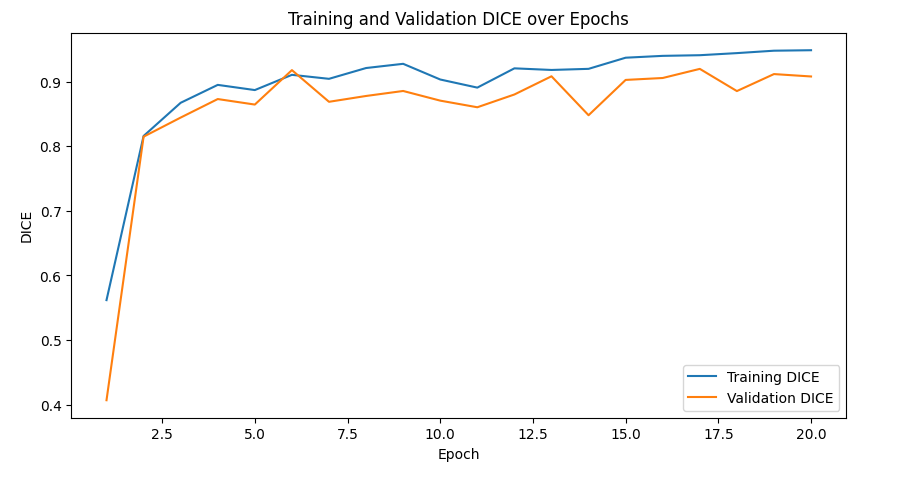
Kết quả thu được sau khi huấn luyện:



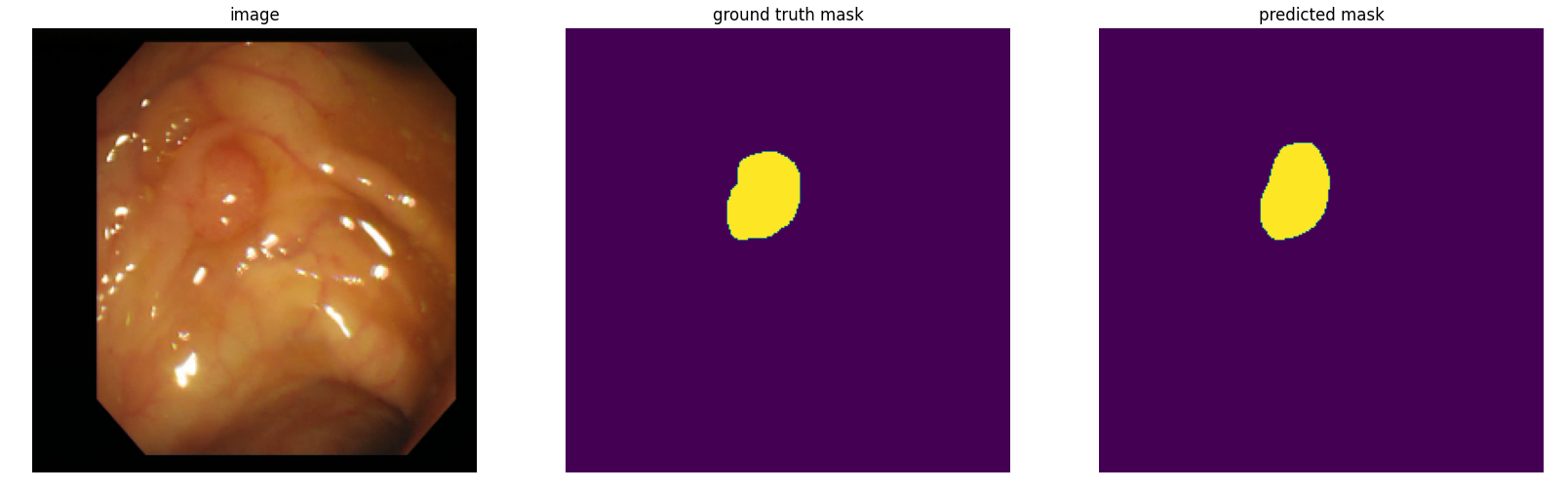
## **4.8. Đánh giá mô hình**







Ví dụ cụ thể:



Tham số FLOPs của mô hình



## **4.9. Kiểm tra mức độ hoạt động của mô hình.**

\*So sánh trên 3 tập dữ liệu khác nhau:

| **FPN model** | **Cvc ColonDB** | **CvC ClinicDB** | **Kvasir Segmentation** |
| --- | --- | --- | --- |
| Average Validation Loss | 0,2715 | 0,1607 | 0,2771 |
| Average Validation DICE | 0,8577 | 0,9034 | 0,8715 |

Bộ dữ liệu Kvasir-SEG chứa 1000 hình ảnh polyp và masks tương ứng, với độ phân giải khác nhau từ 332 x 487 đến 1920 x 1072 pixel

Bộ dữ liệu CVC\_ClinicDB chứ 612 hình ảnh polyp tương đương với 612 mask, với độ phân giải là 288x384 pixel

Qua đây có thể thấy được rằng model có thể hoạt động tốt trên những tập dữ liệu khác nhau.

\* So sánh mô hình FPN với những mô hình khác trên cùng một tập dữ liệu CvC ColonDB

| **Model** | **Year** | **DICE** | **Bài báo** |
| --- | --- | --- | --- |
| [DUCK-Net](https://paperswithcode.com/paper/using-duck-net-for-polyp-image-segmentation-1) | 2023 | 0,9353 | [Using DUCK-Net for Polyp Image Segmentation](https://paperswithcode.com/paper/using-duck-net-for-polyp-image-segmentation-1) |
| [ResUNet++ + TTA](https://paperswithcode.com/paper/a-comprehensive-study-on-colorectal-polyp) | 2021 | 0,8474 | [A Comprehensive Study on Colorectal Polyp Segmentation with ResUNet++, Conditional Random Field and Test-Time Augmentation](https://paperswithcode.com/paper/a-comprehensive-study-on-colorectal-polyp) |
| [ESFPNet-L](https://paperswithcode.com/paper/esfpnet-efficient-deep-learning-architecture) | 2022 | 0,811 | [ESFPNet: efficient deep learning architecture for real-time lesion segmentation in autofluorescence bronchoscopic video](https://paperswithcode.com/paper/esfpnet-efficient-deep-learning-architecture) |
| [UACANet-S](https://paperswithcode.com/paper/uacanet-uncertainty-augmented-context) | 2021 | 0,783 | [UACANet: Uncertainty Augmented Context Attention for Polyp Segmentation](https://paperswithcode.com/paper/uacanet-uncertainty-augmented-context) |
| [CaraNet](https://paperswithcode.com/paper/caranet-context-axial-reverse-attention) | 2021 | 0,773 | [CaraNet: Context Axial Reverse Attention Network for Segmentation of Small Medical Objects](https://paperswithcode.com/paper/caranet-context-axial-reverse-attention) |
| [PraNet](https://paperswithcode.com/paper/pranet-parallel-reverse-attention-network-for) | 2020 | 0,709 | [PraNet: Parallel Reverse Attention Network for Polyp Segmentation](https://paperswithcode.com/paper/pranet-parallel-reverse-attention-network-for) |
| [DuAT](https://paperswithcode.com/paper/duat-dual-aggregation-transformer-network-for) | 2022 | 0.819 | [DuAT: Dual-Aggregation Transformer Network for Medical Image Segmentation](https://paperswithcode.com/paper/duat-dual-aggregation-transformer-network-for) |

# 

# **CONCLUSION**

Mô hình FPN là một phương pháp hiệu quả cho việc phân đoạn polyp trong tập dữ liệu CVC-ColonDB. FPN có độ chính xác cao, tốc độ xử lý cao và có thể xử lý hình ảnh ở độ phân giải cao. Tuy nhiên, mặc dù FPN có khả năng xử lý hình ảnh ở độ phân giải cao và đạt được độ chính xác cao, nhưng nó có thể gặp khó khăn khi phải đối mặt với nhiễu trong hình ảnh, đặc biệt là trong các tình huống nơi nhiễu gây ra sự mơ hồ hoặc biến dạng trong các vùng quan tâm. Điều này đặt ra thách thức trong việc xử lý hình ảnh thực tế, nơi mà các điều kiện ánh sáng và nhiễu thường phức tạp và không đồng đều.

Hơn nữa, FPN yêu cầu lượng dữ liệu lớn để đạt được hiệu suất tốt nhất. Việc huấn luyện mô hình trên tập dữ liệu không đủ lớn có thể dẫn đến hiện tượng overfitting, làm giảm khả năng tổng quát hóa của mô hình trên các dữ liệu mới. Do đó, việc thu thập và chuẩn bị dữ liệu đúng cách là một yếu tố quan trọng để đảm bảo hiệu suất của FPN trong các ứng dụng thực tế.