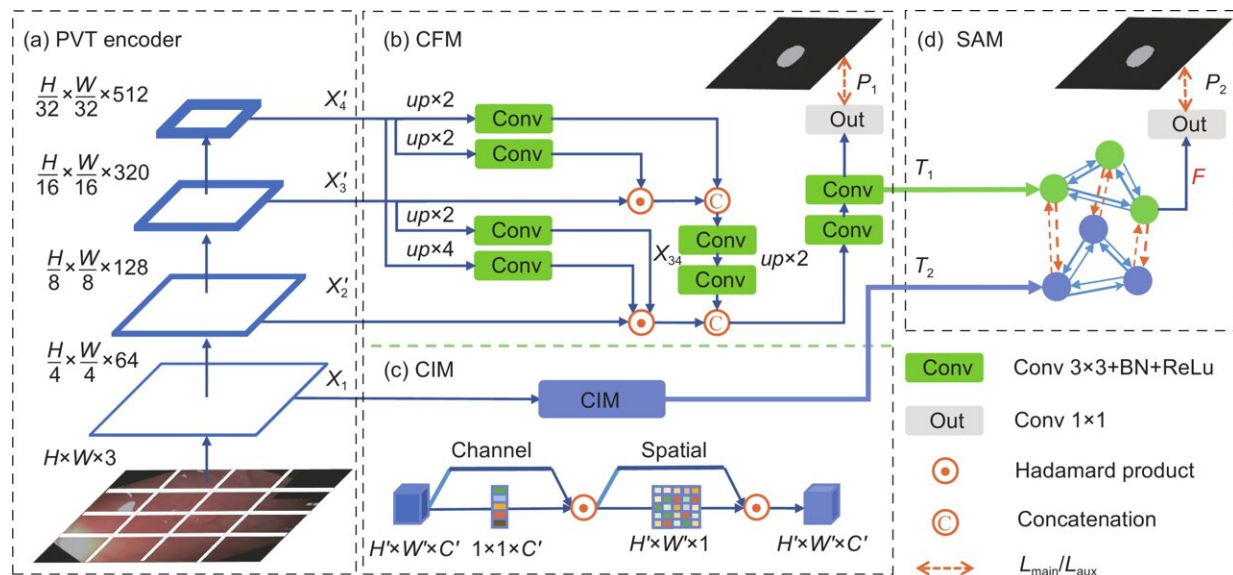# New deep-learning approach gets to the bottom of colonoscopy

August 4 2023

Framework of our Polyp-PVT, which consists of (a) a pyramid vision transformer (PVT) as the encoder network, (b) cascaded fusion module (CFM) for fusing the high-level feature, (c) camouflage identification module (CIM) to filter out the low-level information, and (d) similarity aggregation module (SAM) for integrating the high- and low-level features for the final output. Credit: *CAAI Artificial Intelligence Research* (2023). DOI: 10.26599/AIR.2023.9150015

Researchers have developed a pair of modules that gives a boost to the use of artificial neural networks to identify potentially cancerous growths in colonoscopy imagery, traditionally plagued by image noise

resulting from the colonoscopy insertion and rotation process itself.

A paper describing the approach was published in the journal *CAAI Artificial Intelligence Research* on June 30.

Colonoscopy is the gold standard for detecting colorectal growths or 'polyps' in the inner lining of your colon, also known as the large intestine. Via analysis of the images captured by a colonoscopy camera, medical professionals can identify polyps early on before they spread and cause rectal cancer.

The identification process involves what is called 'polyp segmentation,' or differentiating the segments within an image that belong to a polyp from those segments of the image that are normal layers of mucous membrane, tissue and muscle in the colon.

Humans traditionally performed the whole of the image analysis, but in recent years, the task of polyp segmentation has become the purview of computer algorithms that perform pixel-by-pixel labeling of what appears in the image. To do this, computational models mainly rely on characteristics of the colon and polyps such as texture and geometry.

"These algorithms have been a great aid to medical professionals, but it is still challenging for them to locate the boundaries of polyps," said Bo Dong, a computer scientist with the College of Computer Science at Nankai University and lead author of the paper.

With the application of deep learning in recent years, polyp segmentation has achieved great progress over cruder traditional methods. But even here, there remain two main challenges.

First, there is a great deal of image 'noise' that polyp segmentation deep learning efforts struggle with. When capturing images, the colonoscope

lens rotates within the intestinal tract to capture polyp images from various angles. This rotational movement often leads to motion blur and reflection issues. This complicates the segmentation task by obscuring the boundaries of the polyps.

The second challenge comes from the inherent camouflage of polyps. The color and texture of polyps often closely resemble that of the surrounding tissues, resulting in low contrast and strong camouflage. This similarity makes it difficult to distinguish polyps from the background tissue accurately. The lack of distinctive features hampers the identification process and adds complexity to the segmentation task.

To address these challenges, the researchers developed two deep learning modules. The first, a 'Similarity Aggregation Module', or SAM, tackles the rotational noise issues, and the second, a 'Camouflage Identification Module', or CIM, addresses camouflage.

The SAM extracts information from both individual pixels in an image, and via "semantic cues" given by the image as a whole. In computer vision, it is important not merely to identify what objects are in an image, but also the relationships between objects.

For example, if in a picture of a street, there is a red, three-foot high, cylindrical object on a sidewalk next to the road, the relationships between that red cylinder and both the sidewalk and road give the viewer additional information beyond the object itself that aid in identification of the object as a fire hydrant. Those relationships are semantic cues. They can be represented as a series of labels that are used to assign a category to each pixel or region of pixels in an image.

The novelty of the SAM however is that it extracts both local pixel information and these more global semantic cues via use of non-local and graph convolutional layers. Graph convolutional layers in this case

consider the mathematical structure of relationships between all parts of an image, and non-local layers are a type of node in a [neural network](#) that assesses more long-range relationships between different parts of an image.

The SAM enabled the researchers to achieve a 2.6 percent increase in performance compared to other state-of-the-art polyp segmentation models when tested on five different colonoscopy image datasets widely used for deep learning training.

To overcome the camouflage difficulties, the CIM captures subtle polyp clues that are often concealed within low-level image features—the fine-grained visual information that is present in an image, such as the edges, corners, and textures of an object.

However, in the context of polyp segmentation, low-level features can also include noise, artifacts, and other irrelevant information that can interfere with accurate segmentation. The CIM is able to identify the low-level information that is not relevant to the segmentation task, and filters it out. With the integration of the CIM, the researchers were able to achieve an additional 1.8% improvement compared to other state-of-the-art polyp segmentation models.

The researchers now want to refine and optimize their approach to reduce its significant computational demand. By implementing a range of techniques including model compression, they hope to reduce the computational complexity sufficient for application in real-world medical contexts.

Proposed mode: github.com/DengPingFan/Polyp-PVT

Provided by Tsinghua University Press