**1.Kiến trúc Mạng :**

Đối với các mạng U-Net truyền thống, đầu vào và đầu ra không nhất quán (kích thước dữ liệu đầu vào là 572 × 572, và kích thước dữ liệu đầu ra là 388 × 388). Điều này dẫn đến mất thông tin biên [25]. Sau khi chồng chất các lớp tích chập, kích thước của đầu vào và đầu ra rất khác nhau, và các pixel gần biên của hình ảnh huấn luyện bị thiếu, dẫn đến huấn luyện không tốt các đặc điểm thông tin [22]. Để giải quyết vấn đề này, mạng hoàn toàn tích chập (FCN) sử dụng phương pháp nội suy song tuyến tính để thực hiện lấy mẫu. Tuy nhiên, phương pháp này quá đơn giản để xử lý các đặc điểm chi tiết một cách hiệu quả. Kể từ khi Zeiler và Fergus [24] đề xuất khái niệm giải mã tích chập, và sau khi phát triển thêm tích chập chuyển vị, vấn đề trên đã được giải quyết, và thuật toán phân đoạn ngữ nghĩa dựa trên mạng nơ-ron trở nên ngày càng phổ biến.

Trái ngược với phương pháp được đề cập ở trên, mạng trong bài báo này, như được hiển thị trong Hình 2, sử dụng padding (đệm), do đó kích thước đầu vào và đầu ra khớp nhau và thông tin biên của hình ảnh đầu vào được giữ lại. Hoạt động upsampling (lấy mẫu tăng) được thực hiện như sau. Đầu tiên, đệm giá trị 0 xung quanh bản đồ đặc trưng của W × H để có được bản đồ đặc trưng của .2W + 1/ × .2H + 1/. Sau đó, tích chập với kích thước bộ lọc 2 × 2 được sử dụng để tăng kích thước của bản đồ đặc trưng đầu ra. Cuối cùng, sau một vài lớp tích chập và upsampling, kích thước của hình ảnh phân đoạn có thể được mở rộng sao cho bằng kích thước của hình ảnh đầu vào.

Cấu trúc của CNN đề xuất bao gồm một đường co lại (bên trái) và một đường mở rộng (bên phải), như được hiển thị trong Bảng 1, trong đó Conv đại diện cho lớp tích chập, Max pooling đại diện cho lớp lấy mẫu tối đa, Upsampling đại diện cho lớp giải tích chập, và Merge đại diện cho lớp nối bỏ qua. Xem xét rằng chỉ có hai loại, gồm nền và các dải, bài báo này giảm một số lớp ẩn dựa trên mô hình U-net để chỉ còn tổng cộng 16 lớp tích chập và ba lớp pooling, cũng như ba lớp upsampling và ba cấu trúc kết nối bỏ qua. Trên đường co lại, có sáu lớp tích chập với bộ lọc kích thước 3 × 3 pixel và ba lớp pooling tối đa.