

* **Pre-trained ResNet0:**
* **Block 1, Block 2, Block 3, Block 4**: Đây là các khối của ResNet50 đã được huấn luyện trước trên một tập dữ liệu lớn (như ImageNet). Chúng trích xuất các đặc trưng từ hình ảnh đầu vào qua nhiều tầng tích chập (convolutional layers) và các tầng chuẩn hóa (normalization layers).
* **Output của từng khối**: Đầu ra từ mỗi khối sẽ được đưa vào các bước xử lý tiếp theo. Các đặc trưng được trích xuất sẽ ngày càng phức tạp và trừu tượng hơn khi tiến sâu vào mạng.
* **Residual Transformer Block (RT Block):**
* **Conv (1x1), BatchNorm, LeakyReLU**: Đầu tiên, đặc trưng từ ResNet50 được đưa qua một lớp tích chập 1x1 để giảm số chiều của đặc trưng, sau đó là lớp chuẩn hóa (Batch Normalization) và lớp kích hoạt LeakyReLU.
* **Transformer Encoder Block**: Đây là một khối mã hóa của Transformer, bao gồm các lớp Attention và Feed Forward. Nó giúp mô hình học được các mối quan hệ không gian và ngữ cảnh phức tạp trong dữ liệu hình ảnh.
* **Conv (1x1), BatchNorm**: Sau khối Transformer, đặc trưng tiếp tục được điều chỉnh qua lớp tích chập 1x1 và chuẩn hóa.
* **Add, LeakyReLU, Residual Block**: Tầng cộng (Add) kết hợp đầu ra của các lớp trước với đầu vào ban đầu của khối để duy trì thông tin. Sau đó, đặc trưng được đưa qua lớp LeakyReLU và khối residual để giữ lại thông tin cần thiết và giúp quá trình huấn luyện hiệu quả hơn.
* **Residual Block**: Các khối residual giúp duy trì và kết hợp các đặc trưng đã học trong mô hình, đảm bảo rằng các đặc trưng quan trọng không bị mất mát qua các tầng.
* **Concatenate và Upsample:**
* **Concatenate:** Các đặc trưng từ các tầng khác nhau của ResNet được kết hợp lại. Điều này giúp tận dụng thông tin từ các tầng sâu hơn và tầng nông hơn của mạng.
* **Upsample:** Các lớp này tăng độ phân giải của đặc trưng lên để phù hợp với kích thước của đầu ra mong muốn (mặt nạ đầu ra). Quá trình này có thể bao gồm các phép nội suy (interpolation) hoặc lớp deconvolution.
* **Residual Block:**
* **Conv (1x1), BatchNorm:** Thêm một lớp tích chập 1x1 và chuẩn hóa để điều chỉnh và tinh chỉnh các đặc trưng trước khi tạo đầu ra cuối cùng.
* **LeakyReLU, Residual Block:** Các tầng residual cuối cùng giúp duy trì các đặc trưng đã học, đảm bảo rằng thông tin quan trọng không bị mất trong quá trình truyền qua các tầng.
* **Output Mask:** Sau khi qua tất cả các bước xử lý và kết hợp đặc trưng, mô hình xuất ra một mặt nạ (mask) dự đoán cho hình ảnh đầu vào. Mặt nạ này có thể sử dụng cho các tác vụ như phân đoạn ảnh (image segmentation), nhận dạng đối tượng (object detection), hoặc các ứng dụng khác trong thị giác máy tính (computer vision).