

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KỸ THUẬT MÁY TÍNH



BÁO CÁO ĐỒ ÁN MÔN HỌC
TRÍ TUỆ NHÂN TẠO CHO HỆ THỐNG NHÚNG

Đề tài số 10: RescueNet Dataset for Semantic Segmentation

Nhóm thực hiện: Soloveling

Thành viên: Nguyễn Phan Anh Minh - 22520878

Lớp: CE340.P21

Giảng viên phụ trách: Nguyễn Thanh Thiện

MỤC LỤC

Chương 1: Giới thiệu	5
1.1. Bối cảnh và động lực nghiên cứu	5
1.2. Mục tiêu đề tài	5
Chương 2: Thiết kế mô hình	6
2.1. Tổng quan	6
2.2. Nội dung chi tiết	6
a. Tiền xử lý dữ liệu	6
b. Xây dựng mô hình	6
c. Cài đặt huấn luyện	8
d. Đánh giá mô hình.....	8
e. Trực quan hóa kết quả.....	8
Chương 3: Thực nghiệm và đánh giá.....	9
3.1. Thực nghiệm	9
3.2. Kết quả	9
a. Kết quả với tập train.....	9
b. Kết quả trên tập test	9
3.3. Đánh giá và nhận xét	9
Tài liệu tham khảo.....	10

Danh sách hình vẽ

Hình 1. Minh họa mô hình Unet tích hợp Attention Gate	7
Hình 2. Quy trình trong một bước upsample của decoder	7
Hình 3. Trực quan hóa kết quả.....	8
Hình 4. Kết quả trên tập test	9

Danh mục từ viết tắt

CNN: convolutional neural network (mạng nơ – ron tích chập)

JPG: Joint Photographic Experts Group

PNG: Portable Network Graphics

Chương 1: Giới thiệu

1.1. Bối cảnh và động lực nghiên cứu

Trong những năm gần đây, các hệ thống thị giác máy tính đã trở thành một phần quan trọng trong nhiều lĩnh vực ứng dụng thực tiễn như xe tự hành, giám sát đô thị, phân tích nông nghiệp và đặc biệt là hỗ trợ tìm kiếm cứu nạn trong các khu vực bị ảnh hưởng bởi thiên tai. Một trong những kỹ thuật cốt lõi để hiểu ngữ cảnh của hình ảnh trong các ứng dụng này là semantic segmentation – bài toán phân loại từng điểm ảnh của hình ảnh đầu vào thành một nhãn tương ứng.

Tuy nhiên, trong các môi trường khắc nghiệt như khu vực bị động đất hoặc thiên tai, hình ảnh thu được thường có chất lượng kém, nhiễu nhiều, đối tượng bị che khuất hoặc phân bố không đều. Điều này đòi hỏi các mô hình phân đoạn phải có khả năng nhận diện tốt các đặc trưng không gian và tập trung vào các vùng quan trọng của ảnh. Mô hình Attention U-Net – một phiên bản cải tiến của U-Net – ra đời nhằm mục tiêu đó bằng cách tích hợp cơ chế attention để tăng cường khả năng trích xuất đặc trưng ở các vùng có ý nghĩa cao.

Tập dữ liệu RescueNet là một bộ dữ liệu thực tế, mô phỏng các khu vực đô thị sau thiên tai, cung cấp nhiều lớp nhãn như đường, xe, đồng cỏ nát, cây cối và tòa nhà. Việc nghiên cứu và áp dụng Attention U-Net trên tập dữ liệu này không chỉ giúp đánh giá hiệu quả mô hình trong bối cảnh khó khăn mà còn có tiềm năng đóng góp vào các hệ thống hỗ trợ cứu nạn, phản ứng khẩn cấp và quy hoạch sau thiên tai trong tương lai.

1.2. Mục tiêu đề tài

- Nghiên cứu và ứng dụng mô hình semantic segmentation.
- Đánh giá hiệu suất mô hình trên tập dữ liệu RescueNet.

Chương 2: Thiết kế mô hình

2.1. Tổng quan

Quá trình thực hiện mô hình gồm 5 bước chính: tiền xử lý dữ liệu, xây dựng kiến trúc mô hình Attention U-Net, huấn luyện mô hình, đánh giá hiệu suất, và trực quan hóa kết quả.

- Tiền xử lý dữ liệu: Bao gồm resize ảnh, chuẩn hóa và augmentation nhằm tăng độ đa dạng cho tập huấn luyện.
- Xây dựng mô hình Attention – Unet : Thiết kế kiến trúc mạng tích hợp cơ chế attention vào U-Net để mô hình có khả năng tập trung vào các vùng quan trọng.
- Huấn luyện mô hình: Sử dụng optimizer Adam và hàm mất mát CrossEntropy để tối ưu hóa quá trình học.
- Đánh giá mô hình: Đo lường hiệu suất bằng các chỉ số như mIoU, F1 trên tập kiểm tra.
- Trực quan hóa kết quả: So sánh kết quả phân đoạn đầu ra với nhãn thực để đánh giá trực quan.

Việc lựa chọn xây dựng theo các bước như vậy, vì đây là quy trình chuẩn để xây dựng và huấn luyện một mô hình. Quy trình được xây dựng theo từng bước có tính tuần tự và logic nhằm đảm bảo hiệu quả cả về kỹ thuật lẫn kết quả đầu ra, đồng thời đảm bảo khả năng mở rộng và cải tiến trong tương lai

Việc lựa chọn mô hình Attention U-Net là do khả năng xử lý hiệu quả các đối tượng nhỏ và phức tạp trong ảnh – điều rất quan trọng trong các tình huống khẩn cấp như sau thiên tai. Thêm vào đó, RescueNet là tập dữ liệu đa lớp, nhiều chi tiết phức tạp, nên một mô hình có khả năng tập trung vào vùng quan trọng như Attention U-Net là phù hợp hơn các mô hình phân đoạn cổ điển khác.

2.2. Nội dung chi tiết

a. Tiền xử lý dữ liệu

Trong tập dataset Rescuenet có 3 tập thư mục dữ liệu là train, val, test. Để huấn luyện cho mô hình cần chọn đường dẫn đến đúng thư mục có tên là train. Trong thư mục train có 2 thư mục cho ảnh và nhãn, ảnh có định dạng jpg, nhãn có định dạng png. Sau khi có được đường dẫn của image và mask tương ứng. Image và mask sẽ được transform bằng cách sử dụng hàm của thư viện albumentations [\[3\]](#).

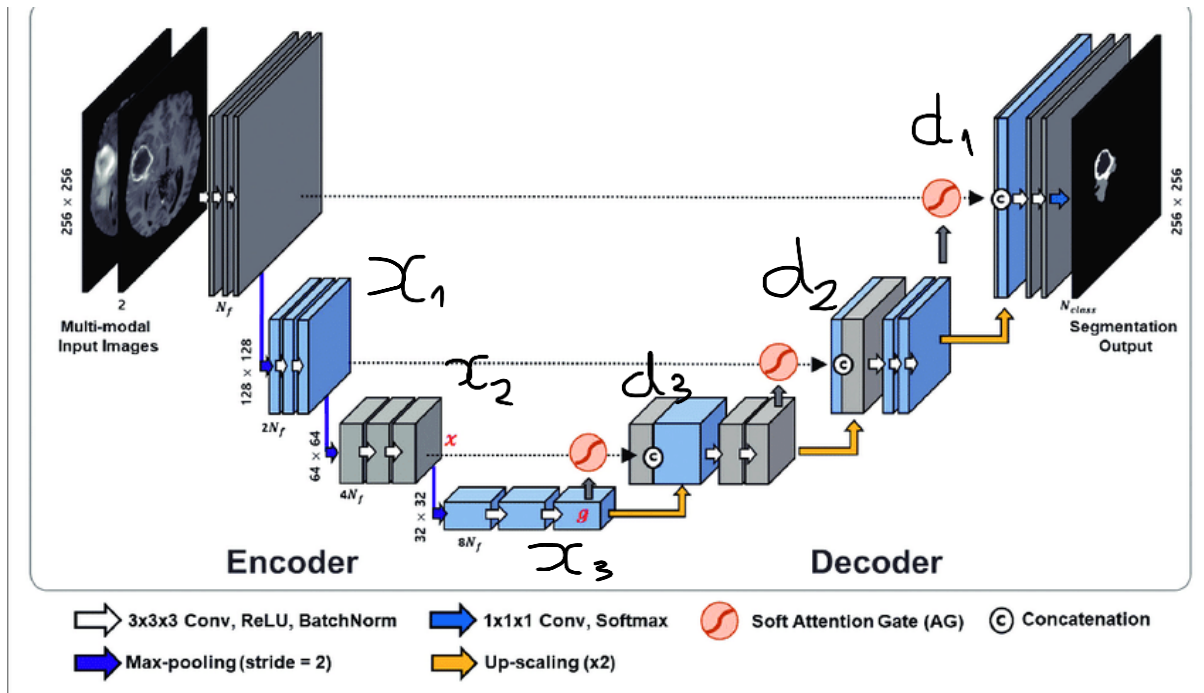
Transform:

- Ảnh resize về 320 x 320, chuẩn hóa giá trị pixel với mean bằng (0.485, 0.456, 0.406), std bằng (0.229, 0.224, 0.225), áp dụng các kỹ thuật biến đổi ảnh như lật ngang ảnh, thay đổi ngẫu nhiên độ sáng và độ tương phản, làm mờ ảnh bằng kernel, làm sắc nét ảnh bằng cách tăng cường các chi tiết cạnh, sau đó chuyển thành tensor.
- Nhãn (mask) được resize 320x320, chuẩn hóa giá trị pixel với mean bằng (0.485, 0.456, 0.406), std bằng (0.229, 0.224, 0.225). sau đó chuyển thành tensor.

Bước này đảm bảo đầu vào đồng nhất cho mô hình và tăng đa dạng dữ liệu, giúp mô hình học tốt hơn và giảm overfitting.

b. Xây dựng mô hình

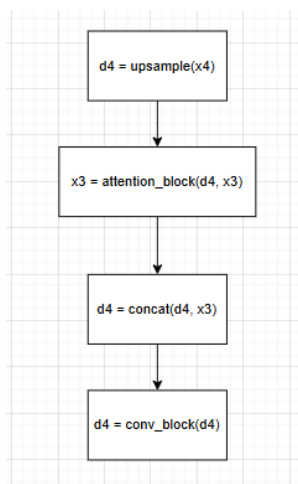
Mô hình được xây dựng theo kiến trúc Unet và được tích hợp với Attention Gate [\[1\]](#)



Hình 1. Minh họa mô hình Unet tích hợp Attention Gate

Mô hình bao gồm:

- Encoder: Trích xuất đặc trưng từ ảnh đầu vào, giảm kích thước không gian (spatial) nhưng tăng chiều sâu đặc trưng. Encoder sử dụng 4 convolution làm giảm kích thước và cho ra 512 kênh đặc trưng ở convolution cuối cùng
- Decoder: Tái tạo ảnh đầu ra cùng kích thước với đầu vào thông qua các bước upsampling và skip connection. Ở mỗi bước upsample, ví dụ đầu tiên d3 là upsamle của x3, sau đó ta sẽ cho d3 và x2 qua khối Attention và gán vào biến x3, sau đó concat đặc trưng d3 đã qua upsample và đặc trưng x3 từ khối Attention. Tiếp theo cho kết quả thu được qua khối conv_block để giảm số chiều.



Hình 2. Quy trình trong một bước upsample của decoder

- Attention Block: Chèn vào trước khi nối đặc trưng từ encoder và decoder giúp làm nổi bật vùng quan trọng hơn, giảm nhiễu từ các đặc trưng không liên quan.

Tổng số tham số của mô hình: 8652276

c. Cài đặt huấn luyện

- Sử dụng hàm mất mát CrossEntropy Loss
- Optimizer Adam
- Batch size: 16
- Số epoch huấn luyện: 37

d. Đánh giá mô hình

Mô hình được đánh giá qua mỗi epoch trong lúc huấn luyện để xem khả năng học của mô hình

Sử dụng các độ đo mIoU, F1

e. Trực quan hóa kết quả

Sau khi huấn luyện đủ số lượng epoch mong muốn. Kiểm tra mô hình có thực sự tốt bằng cách trực quan hóa kết quả, in ra ảnh gốc, nhãn thực (ground truth) và ảnh dự đoán.



Hình 3. Trực quan hóa kết quả

Chương 3: Thực nghiệm và đánh giá

3.1. Thực nghiệm

Sau khi có tập dataset đã được transform, ta chia tập dữ liệu thành từng batch nhỏ theo chỉ số batch mà ta cho. Sử dụng dataloader của pytorch để chia dữ liệu theo batch ngoài ra dataloader còn giúp xáo trộn dữ liệu và tải dữ liệu song song bằng nhiều luồng.

Kiểm tra định dạng của ảnh và mask được đọc từ dataloader để đưa vào mô hình.

Các tham số sử dụng cho huấn luyện:

- Hàm mất mát: CrossEntropy Loss
- Optimizer: Adam
- Learning rate: $1e-4$
- Epochs: 37
- Batch size: 16
- Num_workers: 2

Môi trường thực nghiệm và huấn luyện:

- Google colab (runtime T4)

Các thư viện sử dụng: numpy, opencv2, matplotlib, torch, torchmetrics, albumentations, tqdm, glob

3.2. Kết quả

a. Kết quả với tập train

- Epoch: 37
- Loss: 0.4371
- IoU: 0.4842

b. Kết quả trên tập test

- Average FPS: 0.336
- F1: 0.266
- Score: 0.007

```
./RescueNet/segmentation-validationset/val-org-img/11714.jpg  
Average FPS: 0.330  
F1: 0.266  
Score: 0.007
```

Hình 4. Kết quả trên tập test

3.3. Đánh giá và nhận xét

Kết quả giữa lúc train và khi test cho thấy một sự chênh lệch. Kết quả trên tập test nói chung vẫn có thể chấp nhận được, vì mô hình vẫn chưa được huấn luyện nhiều và học được tốt. Trong tương lai nếu có thể sẽ cho mô hình học được nhiều hơn thì kết quả có thể sẽ tốt hơn.

Tài liệu tham khảo

- [1] J. S. L. L. F. M. L. M. H. K. M. K. M. S. M. N. Y. H. B. K. B. G. D. R. Ozan Oktay, “Attention U-Net: Learning Where to Look for the Pancreas,” [Online]. Available: <https://arxiv.org/abs/1804.03999>.
- [2] T. C. & R. M. Maryam Rahnemoonfar, “RescueNet: A High Resolution UAV Semantic Segmentation Dataset for Natural Disaster Damage Assessment,” [Online]. Available: <https://www.nature.com/articles/s41597-023-02799-4>.
- [3] “Albumentations,” [Online]. Available: <https://albumentations.ai/>.