International Relations Team
291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
Tel: +82-42-350-2441~9, Fax: +82-42-350-4930, E-mail: irt@kaist.ac.kr
http://io.kaist.ac.kr

KISS Track 2: Research Report

Student Name	Vijit Dua	**Home University**	**UC Davis**
Student ID No.	[redacted]	**Degree**	BS , Computer Science
Department	School of Computing	**Academic Advisor**	Prof. Sungho Jo
Participation Period	2024.07.01-2024.07.26	**Laboratory Name**	Neuro-machine augmented intelligence lab (NMAIL)
Project Title	Road Infrastructure Identification from satellite images using learning-based methods		

Overview of Research

1. Research Abstract

This research aims to develop a deep-learning model to identify roads from satellite images and minimize the manual labor required for road mapping. For this process, we used UNet & Attention UNet models modified to suit our needs. We tested them on various environments with different domains & testing environments, including mixed-domain and cross-domain testing. The models were initially developed and tested on Massachusetts dataset and later evaluated using space-net Shanghai & Vegas satellite data.

2. Research Participation Records

Week	Research Participation
1st week	Understood deep-learning from scratch with no prior experience, solved simple deep-learning problems (MNist numbers dataset, CIFAR-10), etc. Gained an understanding of convolutional neural networks, and various model architectures such as the UNet and ResNet models.
2nd week	Modified open-source code to fit our needs. Debugged / re-route utilities like data-loading. Then performed initial testing of UNet model.
3rd week	Tested model with different hyper-parameters and settings on the Massachussets data-base. Noted poor performance, improved optimizer settings, and fixed issues in dataloading (satellite image lacked data that masks showed, meaning that masks weren't cropped in the provided data-set appropriately). Began experimenting with Space-Net data-sets, which contained far more satellite images in varying settings (Las Vegas, Shanghai regions).
4th week	Fixed evaluation criteria (added Precision & Recall, previously only dice scores were used). Built Attention-UNet model to test on different domains. Started running evaluations. Noticed poor performance with Attention-UNet in same domain / cross domain, but noticed higher performance using Attention-UNet in mixed domain testing. Concluded research and began to run final evaluations for research report with different test/train sets to confirm mixed domain performance hypothesis.

Confirmation by Student, Academic Supervisor, and Mentor-Student

By signing below, I hereby confirm that the information contained in this form is true, correct, and complete.

Date : 2024.08.01. **Name of the student :** Vijit Dua **Signature:** [redacted]

Date : 2024.08.01 **Name of the advisor :** Prof. Sungho Jo **Signature:**

By signing below, I hereby confirm that I provided support for the student's research partici ation.

Date : 2024.08.01 **Name of the mentor-student :** Hochang Lee **Signature:** [redacted]

Road Identification From Satellite Images using UNet & Attention UNet

Vijit Dua

Contents

1 Abstract

This research aims to develop a deep-learning model to identify roads from satellite images and minimize the manual labor required for road mapping. For this process, we used UNet & Attention UNet models modified to suit our needs. We tested them on various environments with different domains & testing environments, including mixed-domain and cross-domain testing. The models were initially developed and tested on the Massachusetts dataset and later evaluated using space-net Shanghai & Vegas satellite data.

2 Introduction

Road maps are an essential tool for navigation, planning, and traffic control. In today's fast-paced digital world, there is an essential need for accurate and constantly improved and updated maps. Traditional methods of marking up roads from satellite data require 8+ hours of labor to map $1km^2$ [1]. Current methods for automated mapping lack sophistication and, among other inaccuracies, struggle to detect roads that are obstructed by nearby objects. Therefore, a more accurate automated solution is required to reduce the manual labor involved in mapping these roads.

3 Methodology and Objective

The main objective of this research is to create a deep-learning model, taking inspiration from the image segmentation methods such as UNet architecture [5], to accurately identify road infrastructure from satellite images. Our goal is to develop a road identification model that can accurately identify roads under various conditions, such as obstruction and varying lengths, from satellite images and minimize the manual labor required for road-mapping.

4 Datasets

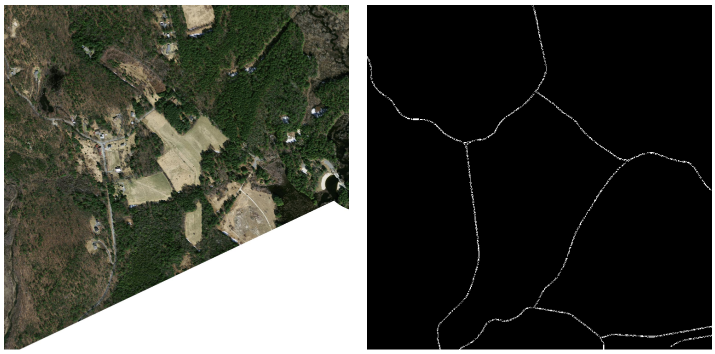
4.1 Dataset Selection

For the purposes of this research, we utilized the Massachusetts Roads Dataset [2] and the SpaceNet dataset [6]. Given the popularity of these data-sets within the coding and research communities, this gave us access to extensive resources online through our research along with well-documented repositories. The SpaceNet dataset [6] was particularly valuable as it offered data from multiple regions, giving us the opportunity to work on cross domain analysis which is crucial to make our model more generalized.

5 Data Handling

5.1 Massachusetts Road Dataset

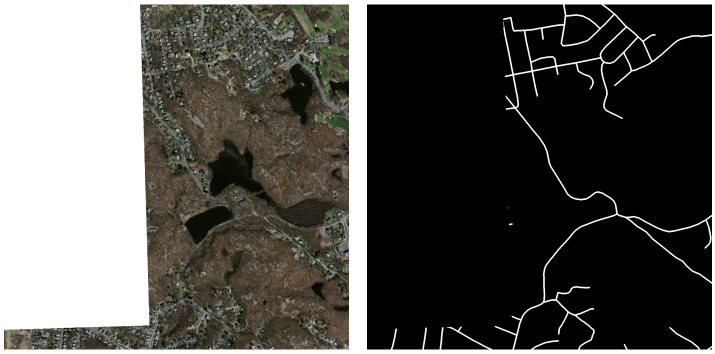
Incorporating the Massachusetts Road dataset [2] presented a set of unique challenges. There were several images in the database where images were cropped, yet the corresponding masks were not adjusted, resulting in inaccurate training data. As a solution, we implemented a prepossessing step that removes regions from the masks corresponding to all-white or all-black areas in the satellite images (indicating cropping). This ensured that the data we utilized was correct for training.



(a) Satellite image before pre-processing

(b) Mask image before pre-processing

Figure 1: Images before pre-processing



(a) Satellite image after pre-processing (b) Mask image after pre-processing

Figure 2: Images after pre-processing

5.2 SpaceNet Dataset

For the SpaceNet Dataset [6], in addition to the pre-processing methods we utilized for the Massachusetts Road Dataset [2], we used a version of the data-set where the TIFF mask images had been converted to 8-bit images (suitable for processing with our python PIL Library) on Kaggle. We manually split the data into an 80-20 ratio for training and testing. On further inspection of the Massachusetts Dataset [2] we realized there were instances where the visible roads were not marked on the masks, leading to inaccuracies in the training of the model. Noticing this, we decided to use the Space-Net dataset for final evaluations due to its accuracy.

6 Models & Code

6.1 UNet

UNet [5] is a type of convolution neural network that is utilized for image segmentation. It features encoding to progressively down-sample an image and a decoder that up-samples the image with contextual information with feature maps to construct the segmentation map. The key to UNet is the skip connections between the corresponding encoding and decoding layers that help preserve contextual information.

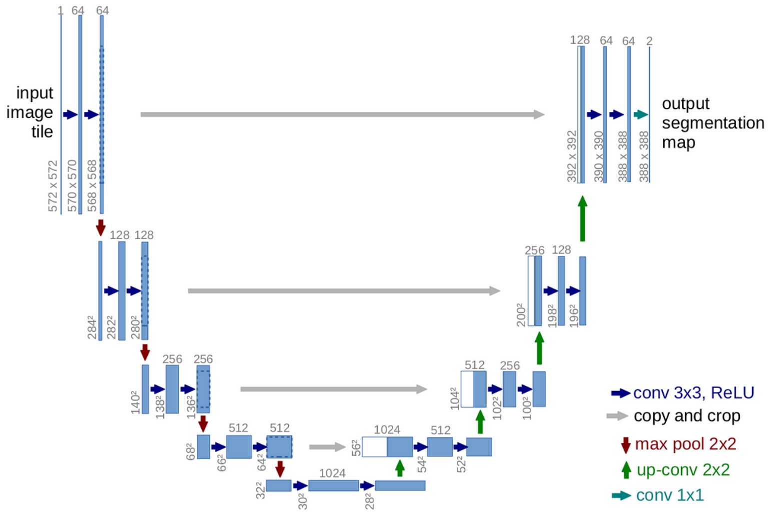


Figure 3: UNet Architecture

6.2 Attention-UNet

Attention-UNet adds more to the UNet architecture by incorporating attention mechanics. These mechanics help the model focus precisely on relevant features of the input image. This is useful to the model to distinguish between closely related classes. In Attention-UNet, attention gates are placed at the levels where previously only skip connections were placed, improving the accuracy of segmentation.

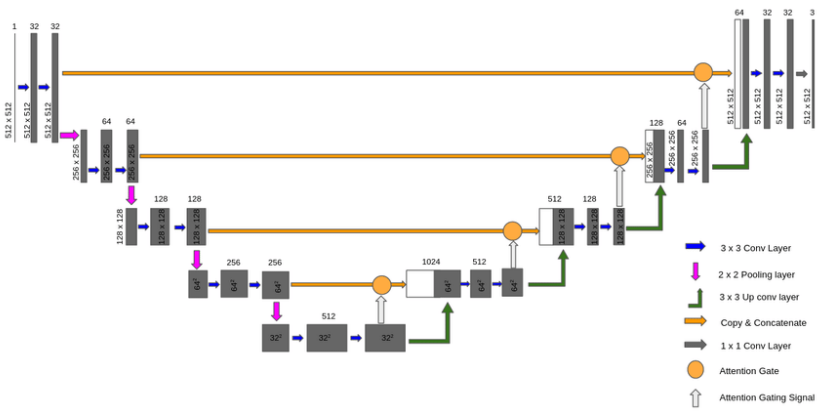


Figure 4: Attention-UNet Architecture

6.3 Implementation

Our models were implemented by modifying the code from MileSail's UNet model for Carvana GitHub repository [4], which was originally aimed at using the Carvana [3] dataset. The repository provided a robust starting point for further development and customization tailored to our needs for satellite image segmentation.

7 Calculation Metrics

For the purposes of evaluation, we used some key metrics: F1 Score, Precision, Recall, and Dice Score.

7.1 Precision

The ratio of true positives to the total predicted positives/

$$\text{Precision} = \frac{TP}{TP + FP}$$

where TP stands for True Positives, and FP stands for False Positives.

7.2 Recall

The ratio of true positives, to the ratio of true positives and false negatives.

$$\text{Recall} = \frac{TP}{TP + FN}$$

where FN stands for False Negatives.

7.3 F1 Score

The harmonic mean of precision and recall.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

7.4 Dice Score

The measure of overlap between prediction and ground truth.

$$\text{Dice Score} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

8 Results

To evaluate model performance, we tried three types of training for each model.

1. **Vegas on Vegas**: Trained on Vegas and Tested on the Vegas data.

2. **Shanghai on Shanghai**: Trained on Shanghai and Tested on Shanghai data.

3. **Combined**: Trained on both datasets combined, and Tested on both datasets combined.

Each training was ran for 15 epochs and the epoch with the best F1 score was selected for further cross-domain evaluation by utilized the saved weights from this epoch.

8.1 UNet Results

Trained On	**Tested On**	**F1 Score**	**Precision**	**Recall**	**Validation Dice Score**
Vegas	Vegas	0.739	0.792	0.698	0.716
Shanghai	Shanghai	0.516	0.763	0.390	0.507
Combined	Combined	0.341	0.850	0.214	0.401
Vegas	Shanghai	0.130	0.402	0.077	0.191
Vegas	Combined	0.243	0.519	0.165	0.425
Shanghai	Vegas	0.371	0.724	0.251	0.401
Shanghai	Combined	0.504	0.762	0.377	0.459
Combined	Shanghai	0.314	0.852	0.193	0.322
Combined	Vegas	0.466	0.848	0.324	0.498

Table 1: Performance metrics for UNet model across different training and testing scenarios.

8.2 Attention UNet Results

Trained On	**Tested On**	**F1 Score**	**Precision**	**Recall**	**Validation Dice Score**
Vegas	Vegas	0.729	0.784	0.683	0.707
Shanghai	Shanghai	0.478	0.719	0.359	0.466
Combined	Combined	0.415	0.809	0.280	0.473
Vegas	Shanghai	0.127	0.291	0.081	0.135
Vegas	Combined	0.236	0.420	0.168	0.390
Shanghai	Vegas	0.354	0.656	0.244	0.403
Shanghai	Combined	0.470	0.718	0.349	0.438
Combined	Shanghai	0.378	0.801	0.247	0.365
Combined	Vegas	0.592	0.828	0.464	0.607

Table 2: Performance metrics for Attention UNet model across different training and testing scenarios.

9 Conclusion

These results suggest that for mixed dataset training (Vegas and Shanghai), using the Attention UNet model enhances the performance). However, for singular domain experience, Attention UNet didn't quite achieve the same results. This indicates the possibility that Attention UNet is effective for generalization however traditional UNet is better for specific domain tasks. Noticing the performance of the model in areas such as Shanghai where architecture and road styles are not consistent vs. the performance in Vegas indicates that the model could be utilized for areas such as Vegas, however, for areas similar to Shanghai, more work needs to be done on improving the models before the models can be utilized.

10 Future Work

Given the limitation of datasets used in our study, further research is required to confirm whether these trends are co-incidental or consistent. Additionally, exploring models that, while not achieving the best F1 score on their own domain, might offer superior cross-domain performance should be investigated. Attempting to use more complex models, using data augmentation techniques like supervised and un-supervised learning to enhance generalization should be explored.

References

[1] Alexander Buslaev, Serhii Seferbekov, Vladimir Iglovikov, and Alexey Shvets. Fully convolutional network for automatic road extraction from satellite imagery. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 207–217, 2018.

[2] Kaggle. Massachusetts road dataset.

[3] Kaggle. Carvana image masking challenge, 2017.

[4] milesail. Pytorch-unet, 2019.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597*, 2015.

[6] SpaceNet. Spacenet road dataset.