IEEE GRSS Methodology Summary

George Wang[1]
dept. Computer Science
Duke University
Durham, USA
zhiyi.wang@duke.edu

Caleb Kornfein[1]
dept. Computer Science
Duke University
Durham, USA
caleb.kornfein@duke.edu

Matthew Feder[1]
dept. Computer Science
Duke University
Durham, USA
matthew.feder@duke.edu

Lorne Zhang[1]
dept. Computer Science
Duke University
Durham, USA
lorne.zhang@duke.edu

Jordan Malof[1]
dept. Electrical and Computer Engineering
Duke University
Durham, USA
jordan.malof@duke.edu

Lesia Semenova[1]
dept. Computer Science
Duke University
Durham, USA
lesia@cs.duke.edu

Bohao Huang[1]
dept. Computer Science
Duke University
Durham, USA
bohao.huang@duke.edu

Cynthia Rudin*
dept. Computer Science
Duke University
Durham, USA
cynthia@cs.duke.edu

***Abstract*—This manuscript details the methods employed for our entry in the 2021 IEEE GRSS Data Fusion Contest Track DSE. Our team utilizes a variety of techniques, including normalization, feature selection, Fully Convolutional Networks, and U-net based encoder-decoder schemes to best predict when a given area contains a human settlement without electricity.**

***Index Terms*—Computer Vision, Machine learning, Segmentation, Settlements Detection**

I. Introduction

The 2021 IEEE GRSS Data Fusion competition builds off of previous geospatial competitions, including the 2018 DeepGlobe challenge [1]–[3]. This iteration is unique in its push to promote research into resource-efficient automatic detection of human settlements without electricity. Whereas DeepGlobe and other challenges have largely been concerned with semantic segmentation problems involving limited number of input channels, the 2021 IEEE GRSS Data Fusion competition provides label maps at a lower spatial resolution than the input imagery, and contains data spanning 98 channels.

II. Data

A. Data Specifications

The 2021 IEEE GRSS Data Fusion competition provides 98 channels of satellite imagery over 60 training land tiles. Each land tile represents an 8 kilometer by 8 kilometer region. Each of the 98 images per tile is at an 800x800 resolution, making the satellite imagery at a 10m GSD. Satellite bands containing higher or lower resolutions are resampled to the 10m GSD level. The 98 channels comprise a combination of satellite imagery from Sentinel-1, Sentinel-2, Landsat 8, and VIIRS datasets. Label maps for each tile are 16x16, meaning each label corresponds to a 50x50 pixel area on the image. Submission involves creating a 16x16 binary classification map indicating wither the given area contains a human settlement without electricity.

B. Data Preprocessing

To prevent the model from over-emphasizing layers that have large intensity values, the pixel values in the images are first normalized. Since some of the satellite channels contain multitemporal data, normalization occurs at the level of the channel. Histogram plots of the pixel intensities by label show that important information exists along the tails of distributions, likely signifying light emitted from human sources. Thus, we normalized the data by subtracting the median and dividing by the mean-absolute-deviation in order to best preserve this information. Prior to median-absolute-normalization, we add one to the pixel intensities and take a logarithmic transform. We also tried several alternative transformations, such as the fourth root.

Since the dimensions of the input channels are 800×800, which requires downsampling in order for them to be directly supplied into most architectures and may result in the loss of information, the input images were first split into 50×50 and 200×200 arrays before being fed into FCN and Unet architectures.

Finally, we considered seven channels instead of the original 98 in the provided dataset. More specifically, we took the RGB channels along with the NIR and two SWIR channels from the Sentinel-2 dataset and the DNB imagery from the VIIRS dataset. For all of these channels, we took only one day worth of data. We chose these channels because we had prior knowledge that led us to believe that these channels would

possess the most valuable information about the outcome. Also, most available pre-trained encoders consider input data with three RGB channels. In order to utilize their weights, we would have had to stack all 98 channels and then reduce the number of channels down to three, which would have led to a loss of a lot of information. Having only seven channels helps us preserve more valuable information while utilizing existing encoders; it makes training models from scratch more manageable.

III. Methodology

Three separate architectures were applied to the processed data: Resnet152V2, Resnet50, and U-net.

A. FCN – Resnet152V2, Resnet50

Our first two architectures utilize Resnet152V2 and Resnet50 encoders. These models use a combination of batch-normalization, padding, convolutional, and bilinear upsampling layers to transform the data batches into the necessary dimensions to be fed through a pretrained encoder. For the data we chose 50×50 patches with seven channels as discussed above. This data was upsampled and passed through a 2D convolutional layer to feed into the pretrained encoder. Fully connected layers at the end are trained to classify the samples into two classes (settlements with and without electricity). Loss is computed using binary cross entropy. We also use a warm-up procedure. First, the initial model was trained by freezing the weights of the pretrained encoder. After training the initial model, the whole model is retrained by unfreezing the weights of the pretrained encoder.

B. U-net

The UNet architecture is believed to be a competent candidate for image segmentation [4]. We took advantage of the encoder-decoder structure of U-net, as shown in Figure 1, and thus created a mask with equal dimension for each 200×200 instance. Based on the sum of labels of the output mask, the model then decides whether a 50×50 area is a human settlement without electricity or not. Due to the mismatch of dimensionality between the input from a pretrained model that support RGB images, which has 3 instead of 98 channels, we trained this model from scratch.

IV. CodaLab Accounts (username, e-mail)

A. Team CodaLab Information

The team's CodaLab Accounts are as follows:

- georgezywang, georgezywang@outlook.com
- Calebk1299, caleb.kornfein@duke.edu
- mfeder, matthew.feder@duke.edu
- lorneez, lornz17@gmail.com

Teammates Rudin, Semenova, Huang, and Malof are not coding, but are contributing to strategy.

B. Account to be used for Test Phase:

Username: Calebk1299, E-mail: caleb.kornfein@duke.edu

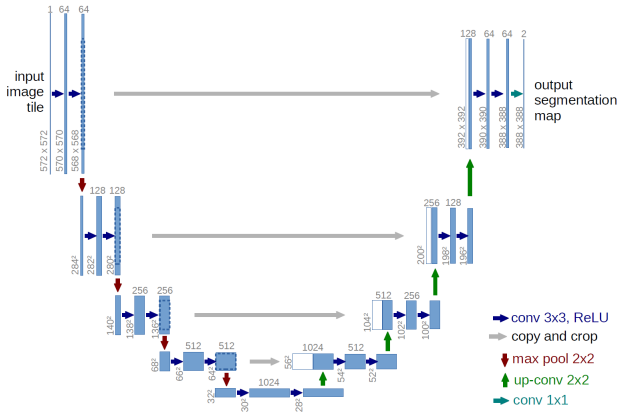

Fig. 1. The Unet Architecture [4]

V. Acknowledgement

We would like to express our sincerest gratitude to Dr. Kyle Bradbury for his valuable advice.

References

[1] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar, "Deepglobe 2018: A challenge to parse the earth through satellite images," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[2] V. I. Iglovikov, S. S. Seferbekov, A. V. Buslaev, and A. Shvets, "Ternausnetv2: Fully convolutional network for instance segmentation," *CoRR*, vol. abs/1806.00844, 2018.

[3] L. Zhou, C. Zhang, and M. Wu, "D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 192–1924, 2018.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015.