Using Generative and Supervised Neural Networks for Thermal Image Analysis in an Urban Heat Island Study

Shaunak Sharma[1], Taseen Islam[2], Kip Nielsen[3], Ashley Grey[4], Audrey Lofthouse[5], Hamidreza Norouzi[6], Reginald Blake[6]

[1] South Brunswick High School, [2] CUNY Macaulay Honors College, [3] University of Kansas, [4] Baruch College Campus High School, [5] Brigham Young University, [6] CUNY City College of Technology

INTRODUCTION

While thermal cameras are essential for quantifying temperature variations across different applications, their high cost and coarse resolution present limitations. In an urban heat island study that attempts to correlate land surface temperature (LST) and air temperature, thermal imagery is constricted to bounded street-level imagery due to regulations prohibiting drone-based platforms. Additionally, in order to differentiate surface types for further analysis, thermal images' visual counterparts must be manually segmented, which is labor intensive. This study proposes two neural networks: a Pix2Pix conditional Generative Adversarial Network (cGAN) to generate highly granular and realistic synthetic thermal images and a Region-based Convolutional Neural Network (Mask R-CNN) to segment corresponding visual images based on feature types. Unlike previous urban heat studies, this work pioneers deep learning to generate realistic synthetic thermal imagery with the segmented visual images providing novel training data. In label-studio, a Python integration, we hand-segmented 150 visual images for training and exported a COCO-format JSON file for training the Mask R-CNN model. Our Mask R-CNN is composed of an 80-10-10 split for trained, validated, and tested images, respectively, and it is the most fine-tuned with a predictor value between 0.75 and 0.90. Synthetic thermal images, produced by the Pix2Pix cGAN, were generated in 30-40 seconds, on average, for 1000 steps; 554 images were trained on. Overlaid synthetic thermal images on the Mask R-CNN's predicted segmentations were used to determine average LST values for cement, pavement, vegetation, and bare dirt. When hand-segmented thermal overlays, hand-segmented synthetic thermal overlays, Mask R-CNN-segmented thermal overlays, and Mask R-CNN-segmented synthetic thermal overlays were compared for the same bounded scene, temperature values varied, on average, by a small RMSE. Future work will utilize and test visual satellite, airplane-sourced, and drone imagery within dense urban settings. Leveraging the performance of the trained neural networks, transfer learning will pioneer new capabilities in localized weather prediction by adapting these advanced models to enhance existing k-nearest neighbor architectures, unlocking hyper-accurate forecasting and enabling high-resolution predictive capabilities.

MOTIVATIONS

After conducting two field campaigns—across Manhattan and in Bedford-Stuyvesant, Brooklyn — we wanted to utilize our gathered thermal images effectively. While we did account for different land surface types using infrared thermometers, our thermal imagery offers greater insights into color, shade, and surface variations. Machine learning is a powerful tool for resolving the analysis component. Previous studies for the urban heat island effect have not utilized generative machine learning for analysis and synthetic image production.

1a


1b


1c


Figures 1a - 1c
1a) FLIR Infrared camera
1b) Map of transects in Bedford,-Stuyvesant campaign (Brooklyn)
1c) Map of New York City with neighborhood of Bedford-Stuyvesant, Brooklyn highlighted in red outline

BACKGROUND

There are **two** neural networks being utilized in this study:

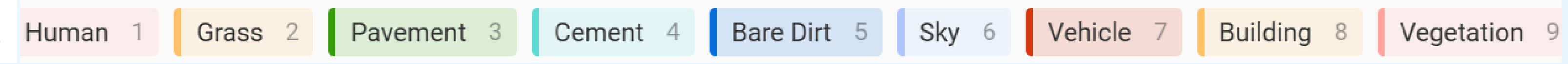
- **Mask R-CNN (Region-Based Convolutional Neural Network)**: Unlike a conventional CNN, which only classifies an entire image, a Mask R-CNN is able to segment specific boundaries within an image to determine a multi-faceted, holistic understanding of what is depicted. This constitutes supervised learning.
- **Pix2Pix cGAN (conditional Generative Adversarial Network)**: This constitutes unsupervised learning with continuous data. Unlike a regular GAN, which primarily serves to enhance and improve existing images, a Pix2Pix cGAN offers two advantages:
  - Conducting image translation: it is able to generate a highly realistic and accurate version of an outputted image that does not exactly match its input but corresponds to its input.
  - Accepting conditionals: it accepts conditional inputs to aid in the process of creating a synthetic image. If needed, it can also output a corresponding conditional that relates to the synthetic image that it produces as well.

We chose to use **neural networks** over **classical ML** for increased data scalability and eventual higher accuracy.

DATA + RESULTS

2a


2b

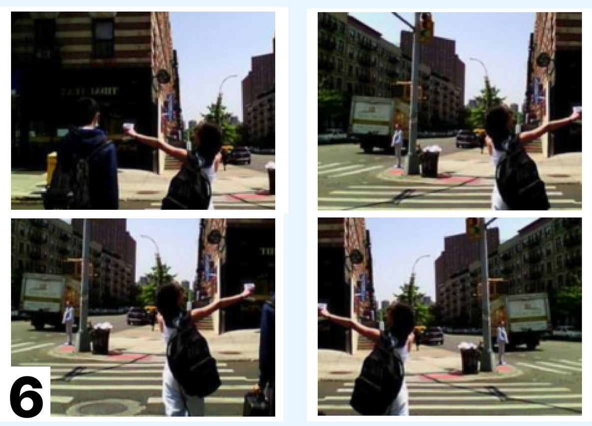

3


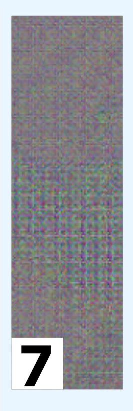
4a

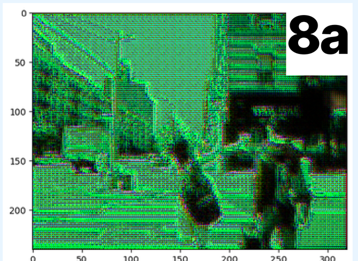

4b

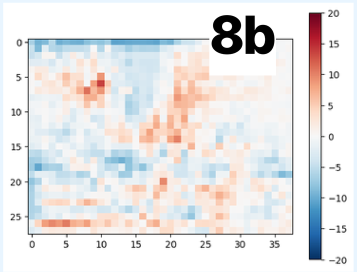

5a

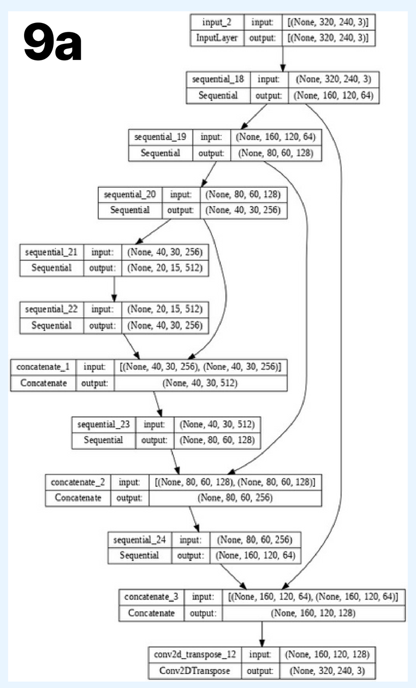

5b


6


7


8a


8b


9a


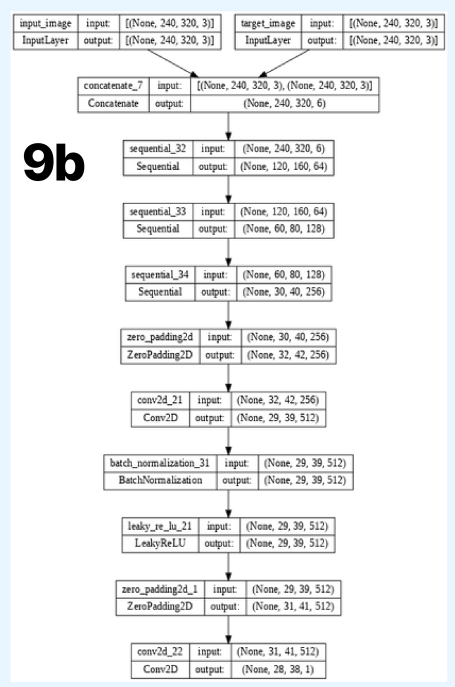
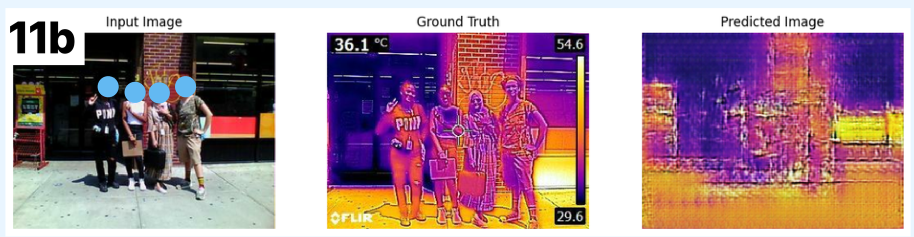
9b


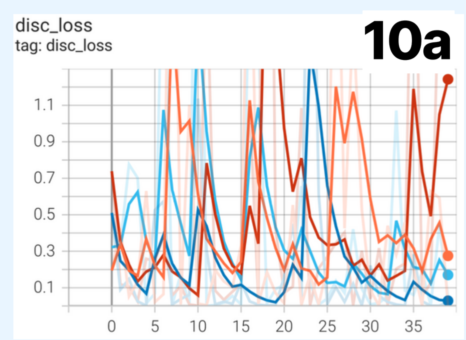
Figure 2a/b: Two visual image examples from Spring field campaign

Figure 3: Feature label types

Figure 4a/b: Label-studio segmentation annotations

11a


11b


10a


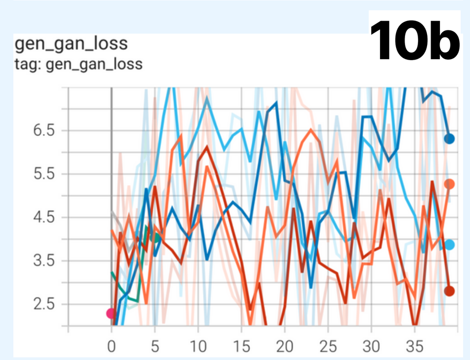
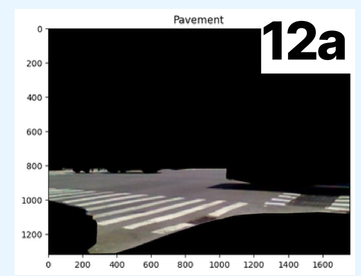
10b


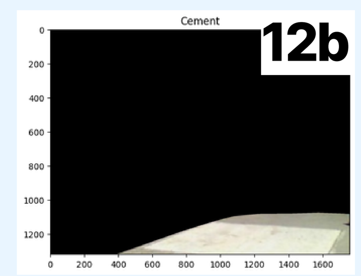
Figure 5a/b: Two thermal image examples from Spring 2023 field campaign
Figure 6: Resizing and randomizing images for data augmentation
Figure 7: Initial synthetic image after 1/40,000 steps completed
Figure 8a/b: Generator/discriminator initial testing
Figure 9a/b: Initial generator/discriminator framework
Figure 10a/b/c: Discriminator loss and generator loss plots
Figure 11a/b: Three-image layout (real visual, real thermal, synthetic thermal) after 16,000/40,000 steps and 20,000/40,000 steps completed in cGAN
Figure 12: Segmentation annotations in *pycocotools.*

METHODOLOGY

For the **Mask R-CNN approach**, we:

- used a Python integration called *label-studio*. It allows the user to manually segment images and create a COCO-format dataset, which contains:
  - a file, which contains annotation boundaries and image info. Feature label types are displayed as well.
  - all visual images.
- used Pytorch with the preexisting *maskrcnn_resnet50_fpn* architecture. Images/annotations were resized, and a training loop was run for *30* epochs with a *lr=0.001*.
  - The pycocotools library was utilized to print the annotations within matplotlib.
  - Testing was done on 15 images, and validation was conducted on another 15 images. Predictor value of 0.82 yielded boundaries that were neither over-fitted nor inaccurate.

12a


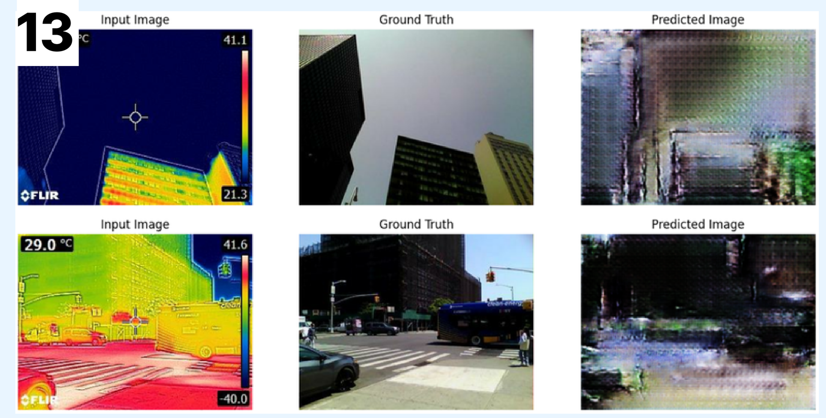
12b


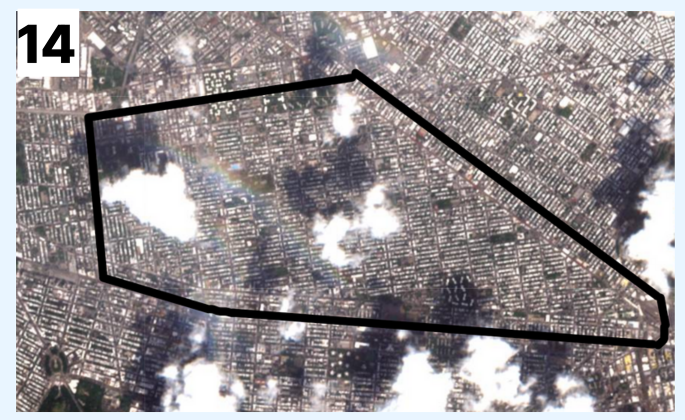
12c


For the **cGAN approach**, we:

- Initially used a regular **cGAN**, which only yielded inaccurate "noisy" images. After revisiting and conducting a minor literature review, we determined that the **Pix2Pix methodology** would work the best for us.
  - Pix2Pix is utilized via TensorFlow, and in order to train the model, we used 414 image sets. Testing utilized 70 of both, and validation also used 70, totaling 554.
  - Images were resized using the *flir_image_extractor.* It works to extract temperature data and visual imagery from the thermal image metadata. Images were resized and concatanated using *Pillow;* file names were renamed using *os.*
  - Testing involved the creation of two methods; training was 40k steps. *Data augmentation techniques* (such as resizing and random cropping) were used for enhancing the datasets.

CONCLUSIONS + NEXT STEPS

13


14


We will expand our dataset—which currently contains 554 sets of images for the Pix2Pix cGAN and 150 annotated images for the Mask R-CNN—to include more information from preexisting thermal image datasets. We hope to implement multi-threading and other forms of encoding/decoding to accurately generate corresponding ".csv" files which contain thermal-temperature values (**Figure 13** shows the reverse cGAN). Furthermore, **Figure 14** shows a recent satellite image from Sentinel-2 L2A; with this, we hope to create "thermalized" heat map images that can help us understand the temperature variations from an aerial perspective. This will also include drone imagery.

REFERENCES

Batchuluun, G., Y. W. Lee, D. T. Nguyen, T. D. Pham, and K. R. Park, 2020: Thermal image reconstruction using deep learning. IEEE Access, 8, 126839–126858, doi:10.1109/access.2020.3007896.

He, K., G. Gkioxari, P. Dollár, and R. Girshick, 2018: Mask R-CNN. arXiv.org,. https://arxiv.org/abs/1703.06870.

Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros, 2018: Image-to-image translation with conditional adversarial networks. arXiv.org,. https://arxiv.org/abs/1611.07004.

ACKNOWLEDGEMENTS

We would like to thank Dr. Norouzi and Dr. Blake for their continual support throughout this summer, as they provided valuable insights and directions for all of our unique segments within the overall urban heat island study. Also, a big thank you to Mary Ann Woody, Mary Zaradich, and other program coordinators of the INSPIRE program. Finally, thank you to Matthew Pearce for his excellent mentorship within the CCRI program.