SSC22-WKVIII-09

Bronco Ember an Edge Computing Acceleration Platform with Computer Vision

Zachary Gaines, Maximum Wilder-Smith, Cristian Rodriguez
Bronco Space
3670 W Temple Ave Pomona CA 91768 Suite 278; 909-904-7595
zegaines@cpp.edu

ABSTRACT

Bronco Ember is a nascent wildfire detection system that leverages edge computing capabilities, multi-spectral imaging, and artificial intelligence to greatly increase the performance of small satellite remote sensing payloads. The core hardware onboard is a SWIR InGaAs camera imaging in the 900nm to 1700nm wavelength and a GPU enabled single board computer. Artificial intelligence is used for fire detection and analysis using computer vision and neural networks being able to detect fires only filling a few pixels in each image. The system is based on traditional CNN networks and includes time series analysis that gives the system an 85% success rate in being able to detect wildfires with about a 50m diameter from a high-altitude balloon technology demonstration flight. The neural net is trained to monitor the movement and spread of the fire compared to prediction maps. This greatly reduces the number of false positive detected. The development of this payload has been supported through the NASA TechLeap Autonomous Observation Challenge No. 1 that has pushed the technology from concept to test flight in less than one calendar year. The system acts a rapid response remote sensing technology.

INTRODUCTION

Bronco Ember is a project that was supported through the NASA Techleap Prize for the Autonomous Observation Challenge No.1. This challenge focuses on the development of new earth observing technologies embracing the SWAP-C criteria and small satellite form factor design. The challenge was met through the creation of an observation platform that is designed to look at nascent wildfires using a SWIR (short wave infrared camera) and a 2 DOF (degrees of freedom) gimbal that enables large area coverage for the optical system. The current system is designed for a high-altitude balloon flight that will act as a technology demonstration for the system capabilities before transitioning to on-orbit performance.

Small satellite systems are limited greatly by power and communication bandwidths. Their small size naturally puts limits on the amount of power generation for sustainability and thus communication bandwidth becomes limited. This is a large challenge to overcome for a small satellite industry that is revolving around remote sensing technologies and missions. Most missions have the goal of just acquiring raw data from orbit and then being reliant on being able to downlink mass amounts of images and other metadata down to ground stations to then be analyzed and studied. This process is not ideal since there are long delays between data acquisition and results. The new age of compact, low-power, and efficient SBCs (single board computers) allows for small satellites to benefit from the powers of edge computing. Edge computing technologies allow for data processing and autonomous rapid response to changing situations. Images are some of the most sought-after pieces of data though these are large files that can be hard to downlink in full quality and in large amounts meaning that engineers are either limiting the amount of data collected or the amount processed to overcome the bandwidth shortcoming. Though in the small satellite industry with each system more often than not being designed for a single-use case implementation of reliable and accurate data processing systems is the ideal for overcoming the complexity of trying to manage so many different mission profiles but rather streamlining the commands and feedback to simple but effective.

The GOES (Geostationary Operational Environmental Satellite) has been observing Earth for decades from a geosynchronous orbit and aids greatly in weather prediction and analysis. The satellite does this through the use of multiple sensor systems, one being infrared cameras that have helped to produce images of all over the United States in this wavelength. The Aerospace Corporation has already proven the reliability and capability of multispectral imaging technologies in a CubeSat form factor. The CUMULOS mission has proven orbit durability while providing images in VIS, SWIR, and LWIR bandwidths and these images have been used to identify and monitor earth events such as the wildfires. The SWIR camera provides profound nighttime imagery making the high infrared signatures stand out with even more contrast than expected, illuminating the immense detection capabilities of remote sensing CubeSats.

SYSTEM OVERVIEW

Bronco Ember is a technology demonstration mission for that currently houses a few components which include: a 2 DOF gimbal, an InGaAs SWIR camera, and a NVIDIA Jetson TX2 NX. These three components drive and enable all operations.

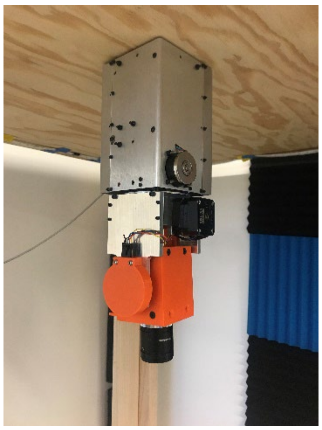


Figure 1: Bronco Ember System

Gimbal

The gimbal has been designed and built to be very compact and space ready. The gimbal houses the SWIR camera giving it greatly increased FOV, quick and consistent area coverage, and precision movements. It uses two stepper motor drive systems roll the system 360 degrees and pitching ±30 degrees. The roll is constructed of worm gear that provides a 61:1 gear ratio and the pitch uses a linear drive train with a gear ration of 35:1 This allows the gimbal as a whole to have a maximum pointing accuracy of 11.4 arc seconds for pitch and 6.6 arc seconds for roll. It is made of low cost and weight metal and give it a tight form factor being able to fit into a 3U form factor.

SWIR Camera

The main optical instrument is the Allied Vision Goldeye G-034. This is an inGaAs SWIR (short wave infrared) camera taking images in the 900nm to 17000nm wavelength range. It has a focal plane array of 636x508 pixels taking images at up to 303 fps (frames per second) at 0.3 MP resolution. For the application of fire detection, the SWIR wavelength is optimal since it allows for high contrast from the rest of the ambient environment that makes hotspots appear very distinctly. Also, this wavelength is minimally affected by cloud cover and other weather conditions that makes it possible to always have continuous coverage of key areas no matter the conditions.

The lens that used is the LM50HX-SW a 1" 50mm SWIR C mount lens. This is a lens that gives a 50mm focal length while being optimized for imaging in the SWIR wavelength ranging between 800nm and 1900nm. This lens also has an F stop aperture ranging from 1.4-16 with the largest setting being used for all testing. With lens at the altitude of a high altitude balloon the field of view covers 5.7km by 4.5km surface area with a ground sample distance of 9 meters.

Figure 2: Goldeye G-034

TX2 NX

The NVIDIA Jetson TX2 NX serves as the primary flight computer on board as it runs all functional and computational software. The module provides a very small form factor, low power draw, but exceptional performance. This module was used in conjunction with the SEED A203 Carrier board that offers a smaller size than the traditional developer boards for the module.

Table 1: Technical Specifications of TX2 NX Module

AI Performance	1.33 TOPS
GPU	NVIDIA Pascal™ Architecture GPU with 256 CUDA cores
CPU	Dual-core NVIDIA Denver 2 64-bit CPU and quad Core ARM A57
Memory	4GB 128-bit LPDDR4, 1600Mz-51.2 GBs
Storage	16 GB eMMC 5.1 Flash storage
Power	7.5W | 15W
Mechanical	69.6 mm x 45 mm

The combination of the TX2 NX module and the SEED carrier board make a space an power optimized solution while still delivering all of the key connections and performance. The gimbal system is operated through 40 pin header that is able to connect to the motor drivers and receive positional feedback from the optical encoders and the IMUs onboard. The gigabit ethernet port allows for connection to the Gig-E SWIR camera and simplifies the wiring harness for the system with the use of an ethernet switch to provide both power and data over the ethernet link.

ARTIFICAL INTELLIGENCE SOFTWARE

At the core of the fire detection system is two interconnected neural networks. The first performs basic CNN-style processing on each of the raw frames from the sensors. This data is then processed in a time series where the network considers the most recent frame as well as previous ones in order to filter out prediction noise and combine both data streams.

Per-frame Prediction

The first step in processing the per-frame data is to ensure that the different cameras are properly aligned. To do this, the fundamental matrices were calculated for the specific camera models, and the extrinsic transformation matrices were generated to align both camera images. These matrices were calculated with Direct-Linear-Transformation (DLT) where a calibration checkerboard was placed in front of the cameras in the final configuration [4]. The relation we model with DLT is shown below

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R \mid T] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

Where the left side represents a point in homogeneous image coordinates (u,v) while the right side represents a point in world coordinates (x,y,z). The matrices we want to derive are K, R and T, the intrinsic, rotation and translation matrix respectively. Multiplying out the components we get 11 equations which means we need a minimum of 6 3D-2D correspondence points as each point constrains 2 equations. Using a calibration grid, we can extract these points and derive the intrinsic and extrinsic matrices to map both data streams to common image coordinates. As each camera has a different lens and resolution, we use the intrinsic matrices to map each view to a planar representation. To handle the different resolutions, the output of SWIR network is scaled up to 1000x1300 while the output of the VIS network is scaled down to this resolution.

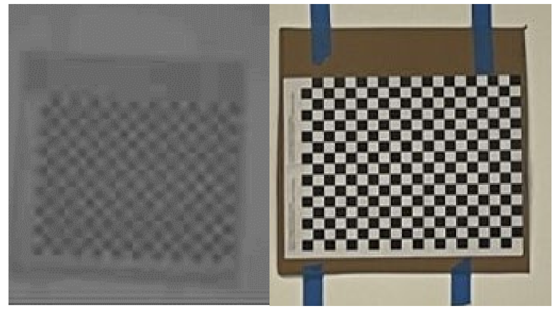


Figure 3: Image of SWIR/VIS Camera Calibration

Once the images have been properly aligned and processed, the precomputed transformation matrices can be used on the fly for live alignment of images during the mission. Now that they share a common 2D coordinate space, each image can be processed by their respective neural networks. Each data stream is processed by a convolution based neural network with a U-Net architecture [5]. U-Net is a segmentation network that down samples the image while increasing the filter size in order to encode fine detail features in the encoder half of the network. After encoding the features the results are upscaled back to the original size in order to generate a per-pixel prediction of where the target features are.

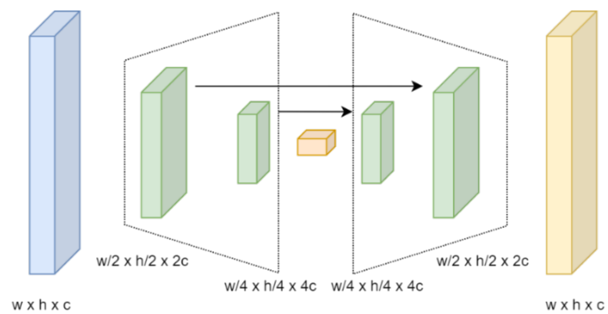


Figure 4: U-Net architecture

For the SWIR input data, the network tries to identify fire based on its appearance in the SWIR spectrum. As the largest potential cause of false positives in this case is reflections and bright artificial light sources, we ensured to capture these in the training data as negative examples. Through qualitative analysis we found the biggest difference between natural fire and the artificial light or sun reflections was border contrast. Fire has more of a spread at the edge while the other sources have a sharp change in brightness.

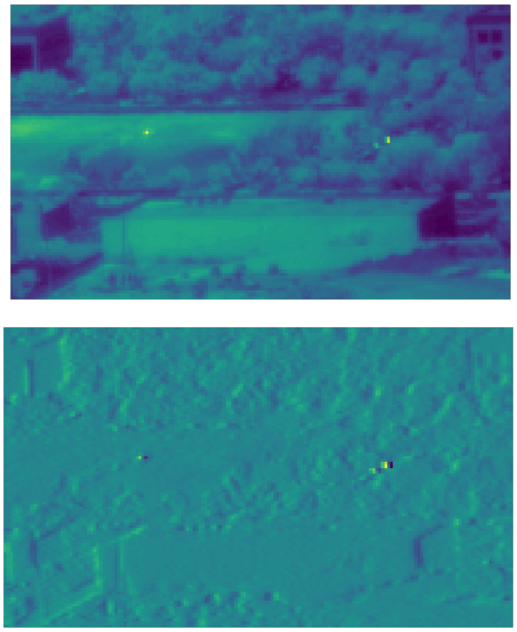
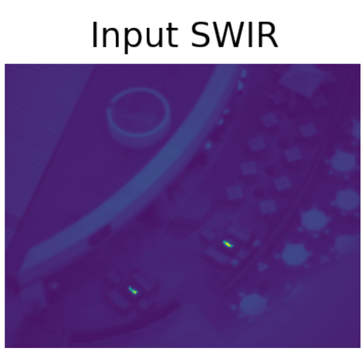


Figure 5: Image of fire and reflection and their gradients

In order to better enable classification of these cases the 1-channel grey scale is augmented by appending the x and y gradients to the second and third channels of the input image. In Figure 5 above the dot on the left is the fire which is more spread out, as shown in its gradient on the right. The dot on the left is a reflection of the sun which is a sharp point with a much more dramatic gradient. The gradient of the image captures the change in value across each axis, allowing the network to more easily identify the sharp vs smooth transition in brightness in fire vs light sources.

After training, the network was able to label and highlight the expected area of fire for each image. This labeled output was then fed into the second stage of the network.



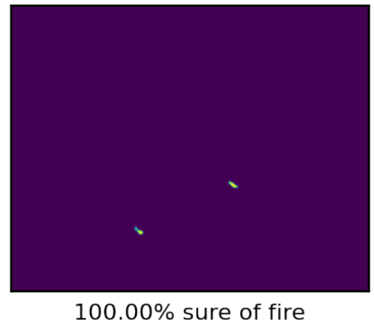


Figure 6: Image of SWIR fire and predicted fire location

For RGB input, the U-Net is tasked with identifying and labeling the location of smoke plumes and their fire. The architecture is similar to that of the SWIR data except the raw input has three channels (red, green and blue) and as the gradient has not proven to increase performance, therefore it is not computed. Once this model identifies the smoke plume this data is passed on to the second stage of the network.

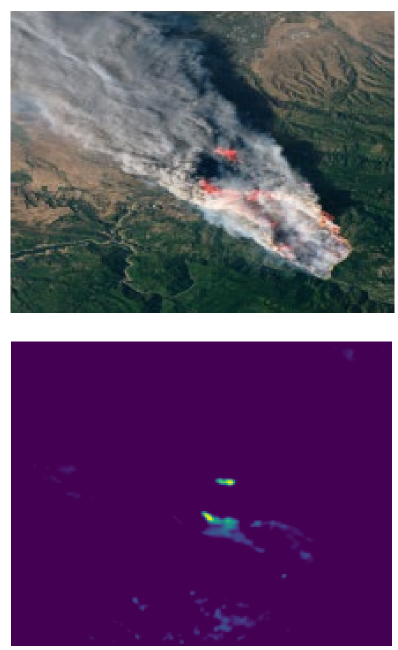


Figure 7: Smoke plume analysis

Time Series

Once per-frame segmentations have been performed they need to be combined and compared to the previous frames in order to create a more accurate prediction of if and where the fire is. This step also adds user-defined parameters to adjust the sensitivity of the program to help prefer certain outcomes. Additionally, in cases where one data steam is unavailable, such as thick cloud cover preventing RGB data or smaller smokeless fires, this temporal processing unit will appropriately weigh the predictions of both cameras.

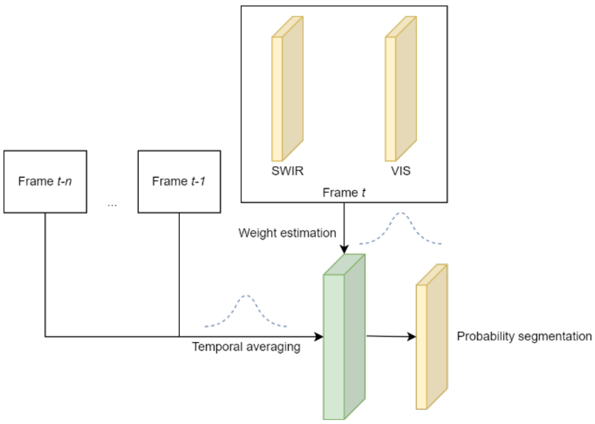


Figure 8: Time Series Framework

In order to perform this processing, we use a Bayesian neural network in order to directly model outcome probability distributions based on the input data. In generating these new posterior predictions, we perform a temporal averaging that reduces false positives from noise. This ensures that if a single frame has data in both cameras that would set off the detector, it isn't properly flagged unless it persists in a way that is consistent with actual fire. There are a few parameters here for the user to set in order to fine tune the system. Such parameters are the number of previous frames to consider and positive detection threshold.

Fire Spread Prediction

In addition to the segmentation and localization networks described above, this deployment includes a fire spread predictor. This predict uses agglomerates a series of data sources and types to predict the spread of wildfires in a particular area. This multimodal neural network uses as input the canopy density, elevation map, canopy type, and canopy height in order to predict how the fire will behave in a particular area. For training and testing CalFire's publicly available data was used. This prediction is used to cross reference the initially detected point source fires and ensure continued accuracy over longer periods of time. The map is able to propagate the required data based on having onboard terrain and vegetation maps that can be referenced after geolocation of the fire based on the gimbal position at detection.

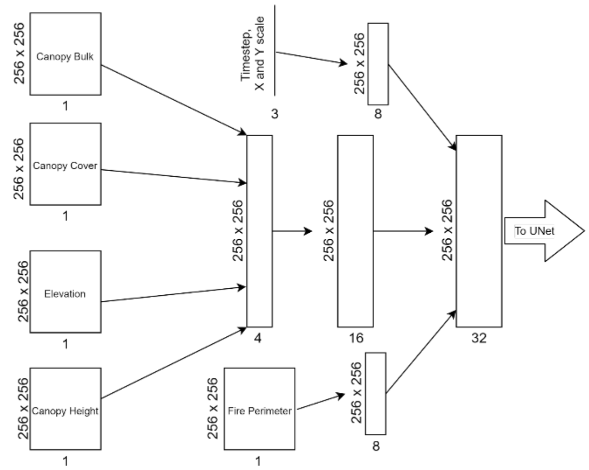


Figure 9: Input model for fire spread

For testing, this network predicted the spread of the 2020 California Bobcat fire over its 12 day burn. Predictions and actual fire spread is shown below.

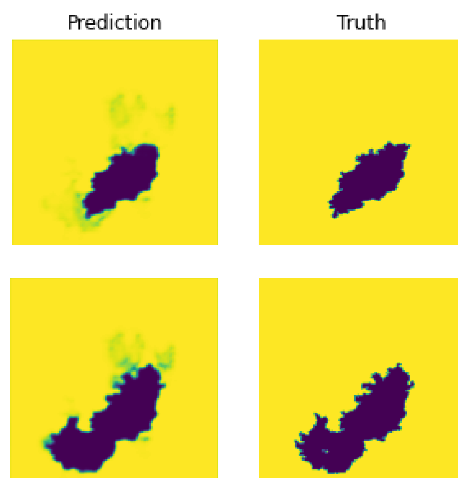


Figure 10: Bobcat fire spread prediction

Results

This system will be flown on a high altitude balloon technology demonstration flight in South Dakota where it will be completely tested. This test is scheduled to occur the last week of June 2022. Results from that flight will be synthesized and added to this section.

Future Work

In order to take advantage of some of the latest advances in time-based machine learning, we are investing the use of transformers to replace the time series and fire spread systems. Transformers are a newer architecture designed to transform time-based data of one form and output time-based data of another form. In this case we are looking at using a transformer for segmentation to take previous scan data as well as past predictions [6]. This will combine the features of the time series Bayesian network with the fire spread prediction system. The new network would use transfer learning to express the previously learned data from the prediction model, but would instead be designed to predict purely based on the SWIR, VIS and past predictions. This would remove the need for accurate geolocation and lookup of vegetation and elevation maps for a given area.

Conclusion

Bronco Ember brings together a plethora of technologies with a specific focus on infrared remote sensing, edge computing, and artificial intelligence. The software and optical technologies have allowed for new methodologies to come together to create high accuracy identification of nascent wildfires appearing as small pixel points on sensors at high altitude. The system is looking to be continually tested further wanting to jump to proving the technology on orbit for long duration and real situations. Bronco Ember could help to greatly improve response times to wildfire but also on a broader scope introduce a new methodology for remote sensing platforms of the future to build off of.

References

1. Finney, M. A.; Ryan, K. C. 1995. Use of the FARSITE fire growth model for fire prediction in U.S. National Parks. In: *Proceedings of the International Emergency Management and Engineering Conferenc*e; 1995 May.
2. Keane, R. E.; Mincemoyer, S. A.; Schmidt, K. M.; Long, D. G.; Garner, J. L. 2000. Mapping vegetation and fuels for fire management on the Gila National Forest Complex, New Mexico, General Technical Report RMRS-GTR-46-CD. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
3. Stratton, R. D; 2006; Guidance on spatial wildland fire analysis: models, tools, and techniques. General Technical Report RMRS-GTR-183. Ft. Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.
4. Zhang, Z; 1998; A Flexible New Technique for Camera Calibration. Technical Report MSR-TR-98-71. Redmon, WA: Microsoft Research.
5. Ronneberge, O; 2015; U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR.
6. Strudel R; 2021; Segmentor: Transformer for Semantic Segmentation. CoRR.