
OPTICAL FLOW FOR INTERMEDIATE FRAME INTERPOLATION OF MULTISPECTRAL GEOSTATIONARY SATELLITE DATA

A PREPRINT

Thomas Vandal

NASA Ames Research Center

Bay Area Environmental Research Institute

Moffett Field, CA, United States

thomas.vandal@nasa.gov

Ramakrishna Nemani

NASA Ames Research Center

Moffett Field, CA, United States

rama.nemani@nasa.gov

September 27, 2019

ABSTRACT

Applications of satellite data in areas such as weather tracking and modeling, ecosystem monitoring, wildfire detection, and landcover change are heavily dependent on the trade-offs related to the spatial, spectral and temporal resolutions of the observations. For instance, geostationary weather tracking satellites are designed to take hemispherical snapshots many times throughout the day but sensor hardware limits data collection. In this work we tackle this limitation by developing a method for temporal upsampling of multi-spectral satellite imagery using optical flow video interpolation deep convolutional neural networks. The presented model, extends Super SloMo (SSM) from single optical flow estimates to multichannel where flows are computed per wavelength band. We apply this technique on up to 8 multispectral bands of GOES-R/Advanced Baseline Imager mesoscale dataset to temporally enhance full disk hemispheric snapshots from 15 minutes to 1 minute. Through extensive experimentation with a multi-terabyte dataset, we show SSM greatly outperforms the linear interpolation baseline and that multichannel optical flows improves performance on GOES/ABI. Furthermore, we discuss challenges and open questions related to temporal interpolation of multispectral geostationary satellite imagery.

1 Introduction

Every second satellites around the earth are generating valuable data to monitor weather, land-cover, oceans, infrastructure, and human activity. These satellites capture reflectance intensities at designated spectral wavelength, spatial, and temporal resolutions. Properties of the sensors, including wavelengths and resolutions, are optimized for particular applications. Most commonly, satellites are built to capture the visible wavelengths, which are essentially RGB images. Other sensors capture a larger range of wavelengths, such as micro, infrared, and thermal waves, providing information for many applications such as storm tracking and wildfire detection. However, sensing more wavelengths is technologically more complex and hence applies further constraints of temporal and spatial resolution. Similarly, a higher temporal frequency requires specific orbital dynamics which then affects the spatial resolution due to its distance from earth.

NASA and other agencies have developed satellites with varying spectral, spatial, and temporal resolutions, each for specific applications. For instance, the most current satellite as part of NASA's Landsat Continuity Mission, Landsat-8 [Roy et al.2014], captures the entire earth every 16-days with 11 spectral bands. This Landsat-8 data is used to monitor crop yields [Skakun et al.2016] and changing land cover [Yuan et al.2005]. The Moderate Resolution Imaging Spectroradiometer (MODIS) [Pagano and Durham1993] is another NASA satellite that monitors Earth everyday at an optimal spatial resolution of 250m and in 36 different spectral bands. The higher temporal resolution of MODIS allows for different applications such as wildfire [Justice et al.2002] monitoring. NASA and NOAA's most recent set of geostationary satellites as part of the GOES-R program, GOES-16/17, improve the frequency further to 15 minute hemispheric coverage (full disk), 5 minutes Continental United States (CONUS) coverage, and up to 30 second flexible mesoscale coverage of 16 spectral bands at 0.5-2km spatial resolution [Schmit et al.2017]. This improved frequency

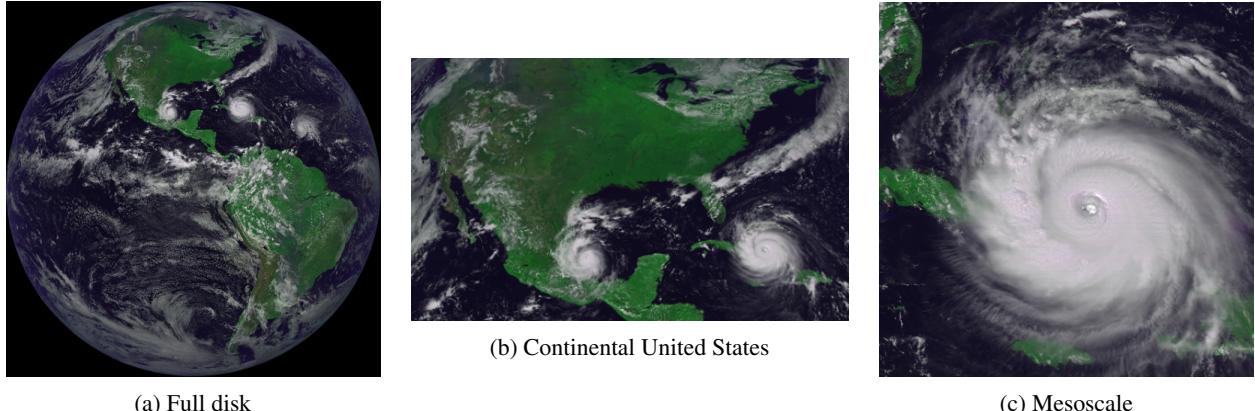


Figure 1: GOES-16 snapshot on September 9, 2017 at 2:00pm EST at three scales, full disk, continental United States, and mesoscale. The mesoscale is taking 1 minute snapshots of a snow-storm moving north up the eastern United States.

allows for near real-time detection and observation of wildfires, hurricane tracking, transportation safety, flood risks, and others.

Improved resolution of data from each of these satellites can benefit nearly all earth science applications and has been widely studied on a variety of satellite based and atmospheric modeling datasets. Satellite imagery processing techniques such as unmixing, pan-sharpening, and data fusion leverage the multitude of available datasets for resolution enhancement [Ghassemian2016]. Recent advances in computer vision and video processing provides opportunities to drastically improve the current state of the art techniques in various resolution enhancement approaches [Vandal et al.2017]. While satellite datasets contain more spectral bands than typically exist in images, the vast amount of data and pattern similarity can be leveraged in deep learning frameworks. In particular, as we will explore in this work, video intermediate frame interpolation models can be further developed to improve temporal resolution of geostationary multi-spectral satellite imagery by extending Super SloMo (SSM) [Jiang et al.2018] to the multivariate case.

The remainder of this paper is outlined as follows. Section 2 discusses related work including data fusion and interpolation in the earth sciences and the current state of video intermediate frame interpolation. Section 3 introduces the GOES-R dataset and section 4 details the methodology. Experiments on a large scale dataset and case study are presented in section 5. Lastly, section 6 concludes with challenges and further work.

2 Related Work

In this section we begin by reviewing previous work in the areas of data fusion and resolution enhancement as applied generally to remote sensing satellite imagery as well as some recent successes of deep learning in the area. Secondly, we provide a brief review of video intermediate frame interpolation techniques.

2.1 Data Fusion and Interpolation in Earth Sciences

Earth science datasets are complex and often require extensive preprocessing and domain knowledge to effectively render itself useful for large-scale applications or monitoring. Such datasets may contain frequent missing values due to sensor limitations, low quality pixel intensities, incomplete global coverage, and contaminated with atmospheric processes related to cloud and aerosols. Techniques to handle these challenges have been developed and are widely applied across the remote sensing community.

Liebmann et al. presented the first linearly interpolated datasets filling in missing and erroneous longwave radiation to improve global coverage [Liebmann and Smith 1996]. In more recent datasets, interpolation techniques are widely applied to unprocessed level 1 data to construct easier to use levels 2, 3, and 4 products, often named assimilated datasets [Rodell et al. 2004]. These higher level products may not contain raw observations but simplify the analysis process to non-domain experts.

In recent years, signal processing and data driven approaches to learn from satellite imagery has become an active area of research. Data fusion is one area of particular interest to the remote sensing community where two or more datasets are fused to generate an enhanced product [Hall and Llinas1997]. The Spatial and Temporal Adaptive Reflectance Fusion (STARFM) algorithm is one example which uses Landsat and MODIS to produce a daily 30-meter reflectance product

by using a spectral wise weighting model [Zhu et al.2010]. Scharlemann et al. presented a temporal fourier processing approach to MODIS data for more accurate analysis of ecological and epidemiology studies [Scharlemann et al.2008]. Nearest neighbor analog multiscale patch-decomposition data driven models are used as state-of-the-art interpolation techniques for developing global sea surface temperature (SST) datasets [Fablet, Viet, and Lguensat2017]. Neural networks are beginning to aid in filling in reconstructing spatio-temporal SST using Kalman filters to represent 2D geophysical dynamics [Ouala et al.2018].

Advances in deep learning and convolutional neural networks have not penetrated interpolation processes for data assimilation of satellite based datasets. Of the few examples of deep learning applied to data fusion, recurrent and convolutional neural networks have been shown to produce effect results on assimilating multiple satellite images [Benedetti et al.2018]. In other areas of remote sensing, deep learning has presented a remarkable ability to detect and classify phenomena. DeepSat showed that normalized deep belief networks tuned where able to outperform traditional techniques for image classifications [Basu et al.2015]. Convolutional neural networks have been shown to effectively classify land use in remotely sensed images, from urban areas [Castelluccio et al.2015] to crop types [Kussul et al.2017]. In this work, we will extend the applicability of deep learning to interpolate and generate higher temporal resolution satellite datasets with inspiration from video processing.

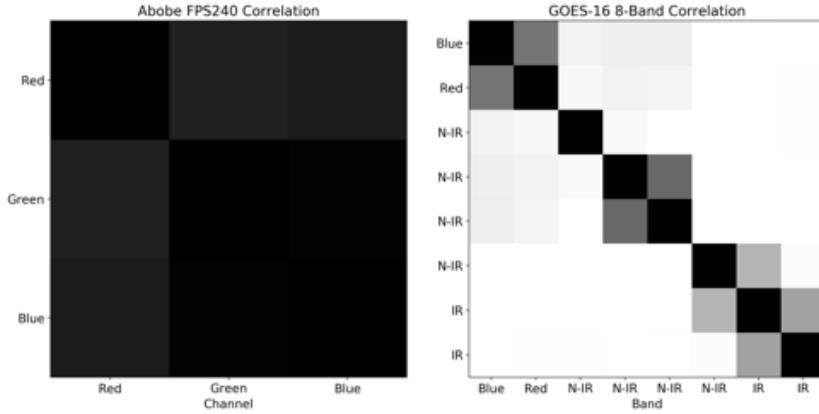


Figure 2: Cross-correlation of band-wise temporal differences.

2.2 Video Intermediate Frame Interpolation

Video interpolation techniques have shown a high skill at generating slow motion footage by generating intermediate frames in spatially and temporally coherent sequences [Liu et al.2017] [Jiang et al.2018]. These approaches are designed to minimize occlusion while still generating sharp and accurate intermediate frames. Typically, video interpolation techniques focus on *single frame* interpolation, meaning that a single frame is estimated between two consecutive frames [Liu et al.2017]. However, when interpolating satellite imagery, time-dependent and multi-frame estimation is preferred.

For multi-frame interpolation, classical approaches estimate optical flows between consecutive frames and warp the frames dependent on time [Barron, Fleet, and Beauchemin1994]. Computing optical flows is an expensive task and can be approximated using neural network architectures using both supervised and unsupervised learning frameworks [Dosovitskiy et al.2015]. Minimizing occlusion between frames is another important task of video interpolation which optical flows cannot handle alone [Baker et al.2011]. Furthermore, optical flows have proven to be a valuable tool in understanding movement in satellite imagery for tasks such as estimating global displacement [Cohen and Herlin1996]. In satellite imagery, these principles allow us to estimate time-dependent intermediate frames via optical flows while capturing occlusions such as clouds covering land-cover.

Jiang et al. presented Super SloMo which combines both optical flow and occlusion models for time-dependent estimation between consecutive frames [Jiang et al.2018]. The time-dependent nature of this approach produces spatially and temporally coherent predictions of any time between 0 and 1. In their experiments, Jiang et al. shows that 240-fps video clips can be estimated from 30-fps inputs. Further details of this work will be presented in Section 4 where we apply their architecture with an extension to multivariate optical flows.

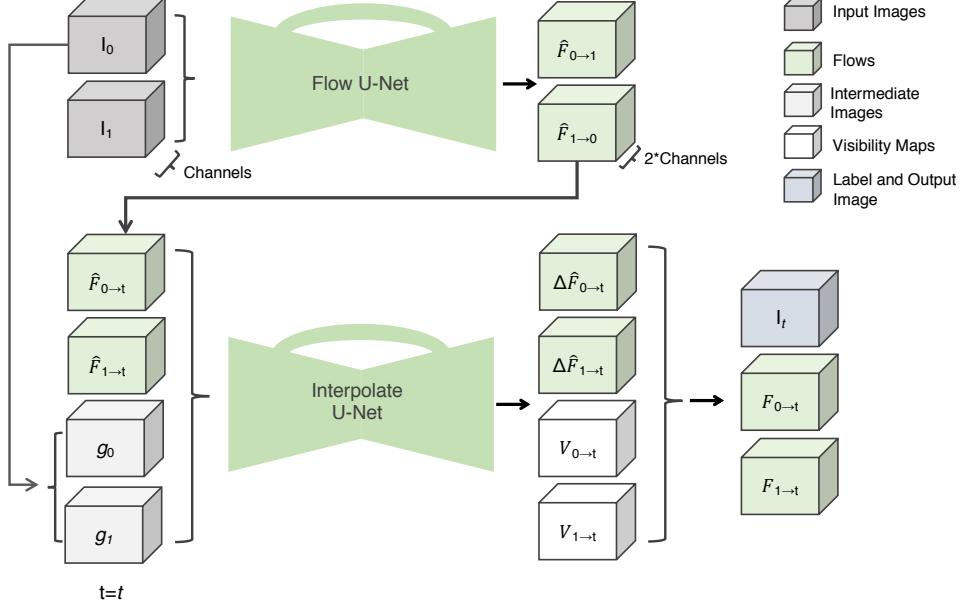


Figure 3: Multivariate SloMo Architecture with Flow and Interpolation Networks.

3 GOES-R Satellite Dataset

Geostationary satellites are synchronized in orbit with earth's spin to hover over a single location. Given this location, the sensor, measuring radiation as often as possible, can frequently capture data over a continuous and large region. This feature makes geostationary satellites ideal for capturing environmental dynamics. The GOES-R series satellites, namely GOES-16/17, operated by NASA and NOAA provides scientists with unprecedented temporal frequency enabling real-time environmental monitoring using the Advanced Baseline Imager (ABI) [Schmit et al.2017]. GOES-16/17 collect 16-bands of data with band 2 at 500m, bands 1, 2, and 4 at 1km, and the remainder at 2km. Three data products are derived from the GOES-R series, 1. Full-disk covering the western hemisphere every 15-minutes, 2. Continental US every 5-minutes, and 3. Mesoscale user directed 1000km by 1000km subregion every 30 seconds (often 2 subregions every 60 seconds). ABI's 16 spectral bands includes two visible channels, four near-infrared channels, and ten infrared channels enabling a suite of applications. Compared to the previous generation, the ABI gives three times more spectral information, four times better spatial resolution, and five times faster coverage. Similarly, the Japanese Space Agency operates Himawari-8 with 15 identical bands, spatial, and temporal resolutions [Bessho2016].

These geostationary satellites are particularly useful in tracking weather, monitoring high-intensity events, estimating rainfall rates, fire detection, and many others at near real-time. Mesoscale mode gives forecasters the ability to "point" the satellite at a user specific subregion for near constant monitoring of events such as wildfires and hurricanes. For example, GOES-16 provided emergency response units tools for decision making during the 2018 California wildfires. However, this high frequency data also provides valuable information of environmental dynamics and retrospective analysis. Furthermore, mesoscale data can be used to inform techniques to produce higher temporal resolution CONUS and full-disk coverage. In this work, we develop a model to improve the temporal resolutions of CONUS and full-disk by learning an optical flow model to interpolate between consecutive frames. With this, we are able to generate 1-minute full-disk artificially enhanced data.

4 Methodology

As discussed above, the fundamentals of optical flow are closely aligned with typical dynamical models used to model earth science processes. In this section, we will describe an adaptation of an optical flow approach for time-dependent interpolation of intermediate video frames, Super-SloMo (SSM) [Jiang et al.2018], to multi-spectral satellite imagery. Our adaptation removes the assumption of high cross-correlation between spectral bands, which as can be seen in Figure 2, is not satisfied in geostationary satellite imagery. Furthermore, we discuss a stochastic approach to train our networks more efficiently on large-scale datasets and how Bayesian optimization can be applied to optimize hyper-parameters.

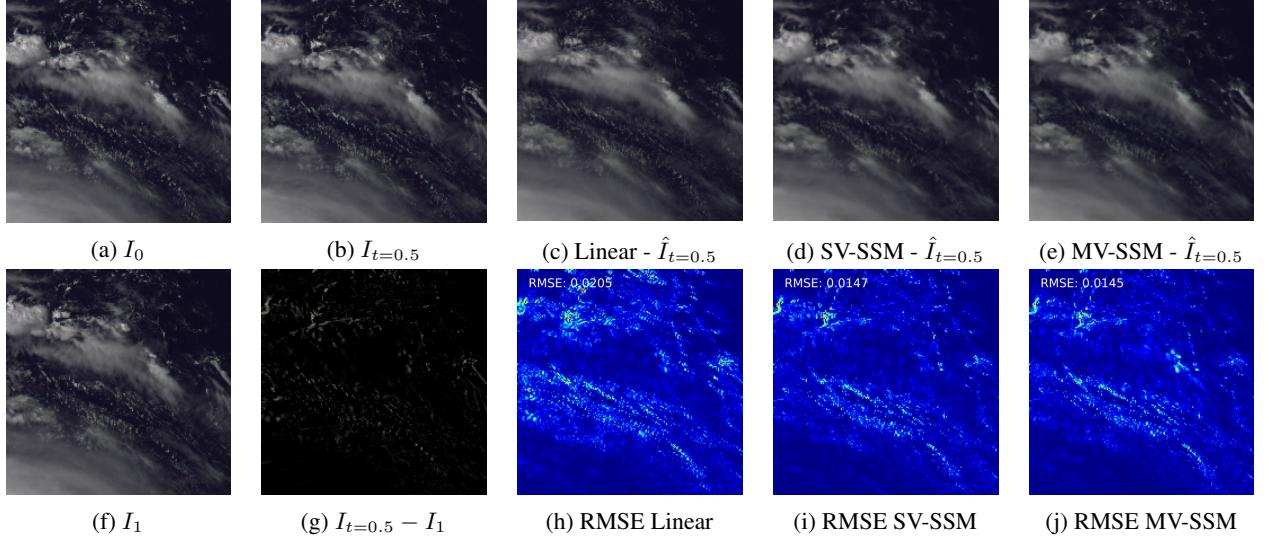


Figure 4: Zoomed in interpolation comparison during Hurricane Irma on first 3 bands with linear interpolation, SV-SSM, and MV-SSM. Input images, I_0 and I_1 on the left, $I_{t=0.5}$ is the intermediate frame to predict. (h), (i), and (j) present RMSEs per pixel for each respective model.

4.1 Intermediate Frame Interpolation

SSM estimation of intermediate frames considers the case of interpolating between two images with RGB channels (3 spectral bands). In the case of satellite based datasets, the number of spectral bands can reach the hundreds. With a minor change to the optical flow and interpolation networks, we extend their framework to model the flows per band. The overall architecture can be seen in Figure 3.

Following the notation from [Jiang et al.2018], let $I_0, I_1, I_t \in \mathcal{R}^{H \times W \times C}$ where $t \in (0, 1)$, H as image height, W as image width, and C a number of spectral bands. The goal is then to construct an intermediate frame $I_{t,c}$, where c is a particular channel, with a linear combination of warped $I_{0,c}$ and $I_{1,c}$ as defined by:

$$\hat{I}_{t,c} = \alpha \cdot g(I_{0,c}, F_{0 \rightarrow t,c}) + (1 - \alpha) \cdot g(I_{1,c}, F_{1 \rightarrow t,c}) \quad (1)$$

where $F_{0 \rightarrow t,c}$ and $F_{t \rightarrow 1,c}$ are the optical flows from $I_{0,c}$ to $I_{t,c}$ and $I_{t,c}$ to $I_{1,c}$, respectively, for channel c . g is defined as the *backward warping* function, implemented with bilinear interpolation, and α represents a scalar weight coefficient to enforce temporal consistency and allow for occlusion reasoning. In the case of high temporal resolution satellite imagery, the interpolation is virtually estimating the state of atmospheric variables (clouds, water vapor, etc.) over a static land surface. If a given pixel in I_0 captures land surface but the same pixel in I_1 sees a cloud, the occlusion principle is used to estimate at what time t the cloud covers the pixel. As applied in [Jiang et al.2018], visibility maps, $V_{t \leftarrow 0,c}, V_{t \leftarrow 1,c} \in (0, 1)^{H \times W}$, can be modeled to capture occluded pixels. Equation 1 can then be redefined as:

$$\hat{I}_{t,c} = \frac{1}{Z} \cdot ((1 - t) \cdot V_{t \leftarrow 0,c} \cdot g(I_{0,c}, F_{0 \rightarrow t,c}) + t \cdot V_{t \leftarrow 1,c} \cdot g(I_{1,c}, F_{1 \rightarrow t,c})) \quad (2)$$

where $Z = (1 - t) \cdot V_{t \leftarrow 0} + t \cdot V_{t \leftarrow 1}$ is a normalization factor. Intermediate optical flows, $F_{0 \rightarrow t,c}$ and $F_{1 \rightarrow t,c}$, are approximated using forward and backward flows between I_0 and I_1 , $F_{0 \rightarrow 1,c}$ and $F_{1 \rightarrow 0,c}$ formulated as:

$$\begin{aligned} \hat{F}_{t \rightarrow 0,c} &= -(1 - t)tF_{0 \rightarrow 1,c} + t^2F_{1 \rightarrow 0,c} \\ \hat{F}_{t \rightarrow 1,c} &= (1 - t)^2F_{0 \rightarrow 1,c} - t(1 - t)F_{1 \rightarrow 0,c} \end{aligned} \quad (3)$$

Two U-Net convolutional neural networks (CNN) H_{flow} and H_{Interp} are learned in an end-to-end manner estimating the optical flows and visible maps. Each of H_{flow} and H_{Interp} are defined with identical architectures with varying input and

output dimensions. The architecture includes 4 down- and up-sampling layers starting with 64 kernels and doubled up to 512 at the low-dimensional embedding. Our flow network, H_{flow} , approximates the forward and backward optical flows between I_0 and I_1 and is defined as follows:

$$\hat{F}_{0 \rightarrow 1}, \hat{F}_{1 \rightarrow 0} = H_{\text{flow}}(I_0, I_1). \quad (4)$$

Following, for each channel c and a time t we compute intermediate image estimations, $g_{0,c} = g(I_{0,c}, \hat{F}_{0 \rightarrow t,c})$ and $g_{1,c} = g(I_{1,c}, \hat{F}_{1 \rightarrow t,c})$. As noted in [Jiang et al.2018], these intermediate approximations perform poorly in occluded and non-smooth regions, both common in satellite imagery, which can be mitigated by incorporating a interpolation network. This model is defined as follows:

$$\begin{aligned} V_{t \leftarrow 0}, V_{t \leftarrow 1}, \Delta \hat{F}_{0 \rightarrow t}, \Delta \hat{F}_{1 \rightarrow t} = \\ H_{\text{Interp}}(I_0, I_1, \hat{F}_{0 \rightarrow t}, \hat{F}_{1 \rightarrow t}, g_0, g_1). \end{aligned} \quad (5)$$

where $\Delta \hat{F}_{0 \rightarrow t}$ and $\Delta \hat{F}_{1 \rightarrow t}$ are optical flow residuals used to better approximate non-smooth regions. Hence, we then have:

$$\begin{aligned} F_{0 \rightarrow t,c} &= \hat{F}_{0 \rightarrow t,c} + \Delta \hat{F}_{0 \rightarrow t,c} \\ F_{1 \rightarrow t,c} &= \hat{F}_{1 \rightarrow t,c} + \Delta \hat{F}_{1 \rightarrow t,c} \end{aligned} \quad (6)$$

Lastly, plugging in the optical flows $F_{0 \rightarrow t,c}$ and $F_{1 \rightarrow t,c}$ and visibility maps $V_{t \rightarrow 0}$ and $V_{t \rightarrow 1}$ into Equation 2 constructs and estimation of $I_{t,c}$. Applying H_{Interp} to each channel c synthesizes an intermediate multispectral prediction I_t .

4.2 Training

Channels	Multivariate	λ_w	λ_s
3	False	0.01	1.30
3	True	0.01	0.03
8	False	0.01	0.30
8	True	0.01	1.01

Table 1: Hyper-parameters found by applying Constrained Bayesian Optimization.

As all variables in the architecture are differentiable, the model can be learned in an end-to-end manner. Given two inputs frames I_0 and I_1 with N intermediate frames $\{I_{t_i}\}_{i=1}^N$ and corresponding predictions $\{\hat{I}_{t_i}\}_{i=1}^N$ a loss function can be defined as a weighted combination of reconstruction, warping, and smoothness losses such that:

$$l = \lambda_r l_r + \lambda_w l_w + \lambda_s l_s. \quad (7)$$

We note that [Jiang et al.2018] includes a fourth term for perception of image classes which are not available for this satellite dataset. Similarly, we employ L_1 loss functions for each loss terms unless noted otherwise.

The *reconstruction loss* is defined as the euclidean distance between observed and predicted intermediate frames:

$$l_r = \frac{1}{N} \sum_{i=1}^N \|\hat{I}_{t_i} - I_{t_i}\|_2. \quad (8)$$

A *warping loss* is used to optimize estimated optical flows between input and intermediate frames for a channel c :

$$\begin{aligned} l_w^{(c)} &= \|I_{0,c} - g(I_{1,c}, F_{0 \rightarrow 1,c})\| + \|I_{1,c} - g(I_{0,c}, F_{1 \rightarrow 0,c})\| + \\ &\quad \frac{1}{N} \sum_{i=1}^N \|I_{t_i,c} - g(I_{0,c}, F_{0 \rightarrow t_i,c})\| + \\ &\quad \frac{1}{N} \sum_{i=1}^N \|I_{t_i,c} - g(I_{1,c}, F_{1 \rightarrow t_i,c})\| \end{aligned} \quad (9)$$

such that $l_w = \sum_{c=1}^C l_w^{(c)}$ for C multi-spectral channels.

A *smoothness loss* is applied to forward and backward flows from I_0 to I_1 to satisfy the smoothness assumption of optical flows in the first network such that:

$$l_s = \|\nabla F_{0 \rightarrow 1}\|_1 + \|\nabla F_{1 \rightarrow 0}\|_1 \quad (10)$$

In practice, this training setup requires optimization over multiple hyper-parameters including λ_r , λ_s , λ_w , and a learning rate. Constrained Bayesian optimization using Monte Carlo simulations [Letham et al. 2019] was applied over the hyper-parameter space (keeping $\lambda_r = 1$ constant) by using the open-source Ax library [Ax]. The optimization was ran for 20 total trials each with 1 epoch through the training dataset. The weights, λ_s and λ_w , shown in Table 1 were found through Bayesian optimization.

5 Experiments

As discussed in Section 3, the GOES-16 geostationary satellite produces unique mesoscale data on a minute by minute basis. In this section, we show how high temporal resolution mesoscale data can be used to train effective interpolation models of multispectral imagery. We begin by describing data processing and training data generation followed by experimental results on held out test data. The NEX platform on NASA's Pleiades super-computing system was used for all data processing and model training. Each model was trained using 4 Nvidia V100 GPUs.

5.1 Data

Mesoscale GOES-16 data takes one snapshot every 30 seconds of selected areas, typically of interesting weather events. Most often, these snapshots are selected as adjacent tiles for greater spatial coverage, producing one minute data. This one minute data is used for training and testing our approach. Training data was selected as every 5 days from March 16 2017 (first day of data availability) to December 31 2018 between the hours of 12pm and 12am (GMT). Each day consists of 14GB of data with a total of 129 days and 1.8TB of training data. Data processing each snapshot involves the following steps:

1. Each band is mapped to 2km using bilinear interpolation to a size of 500 by 500 (1000km by 1000km).
2. Pixels are normalized between 0 and 1 using scaling factors found in Table 4 of the "GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document For Cloud and Moisture Imagery Product" [goe].
3. Bands are stacked into a $H \times W \times K$ dimensional array with corresponding geographic coordinates.

Training examples are then generated by selecting 134 by 134 sub-images and randomly cropped to 128 by 128 during training. Temporally, 12 consecutive time-steps are extracted per example, of which 9 are randomly selected during training. Furthermore, each image is randomly rotated and flipped during training to reduce model artifacts generated by U-Net. This produces, depending on missing values, on average 1,000 images per day for training. In total, this process produces approximately 200,000 training examples for a heterogeneous dataset containing all seasons and a variety of weather events.

Similarly, test data was selected from every 5 days of 2019 between the hours of 12pm (GMT) and 12am (GMT), and hence, held out from the training dataset. This totalled 30 days and 420GB of test data. Furthermore, we select a major tropical cyclone, Hurricane Irma, approaching Florida on September 8, 2017. This is not included in the training set. For testing, data processing enumerated above is applied.

5.2 Results

Experiments are executed on 3 and 8 multivariate channels. For each experiment, two models are trained, 1. with a single optical flow and 2. with channel-wise optical flows, to better understand the applicability of learning multiple optical flows. Linear interpolation is selected as a baseline method. Root mean squared error (RMSE) is used for comparison where interpolated pixels are compared to the ground truth 1-minute mesoscale data. Unless otherwise noted, RMSE values are computed by averaging RMSEs of each test example.

To begin, we look at the overall RMSEs and per experiment, method, and spectral band in Table 2. We note that for each experiment, the optical flow interpolation with SloMo greatly outperform a simple linear regression. It can be seen that in the 3 band experiment a univariate optical flow outperforms the multivariate model. This corresponds with Figure 2 where bands 1 and 2 are highly correlated and hence the enforcement of a relatively high smoothness loss coefficient in Table 1 found with Bayesian optimization. In the 8 band experiment, multivariate flows improve performance for

Band	3 Band Models			8 Band Models		
	Linear	SV-SSM	MV-SSM	Linear	SV-SSM	MV-SSM
1	0.0232	0.0173	0.0180	0.0231	0.0175	0.0173
2	0.0329	0.0262	0.0268	0.0329	0.0265	0.0263
3	0.0288	0.0219	0.0225	0.0287	0.0221	0.0218
4	–	–	–	0.0095	0.0061	0.0057
5	–	–	–	0.0214	0.0170	0.0168
6	–	–	–	0.0137	0.0101	0.0099
7	–	–	–	0.0017	0.0012	0.0011
8	–	–	–	0.0026	0.0020	0.0017
3 Band	0.0283	0.0218	0.0224	0.0282	0.0221	0.0218
8 Band	–	–	–	0.0180	0.0138	0.0136

Table 2: Test set inference results - Root mean squared error (RMSE) computed for each band for every 5 evenly spaced days in 2019

each band. Furthermore, the first 3 bands of the 8 band experiment have the same performance as in the 3 band model. Overall, and as expected, this suggests multivariate flows are suitable with a high number of spectral bands.

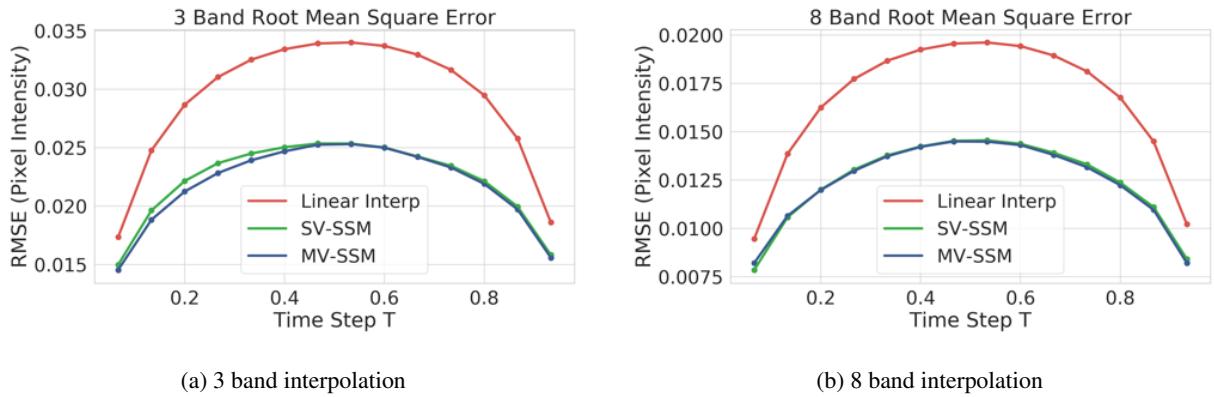


Figure 5: Root mean squared error (RMSE) computed temporally for every minute between 1 and 15, scaled to $t \in (0,1)$ for linear interpolation (red), SV-SloMo (green), and MV-SloMo (blue).

Temporally, the error relative to the distance between images is of interest when using the interpolated data. In Figure 5 we test this by looking at RMSEs over 14 evenly spaced timesteps, t , between 0 and 1. As expected, the error is largest directly between two input images at $t = 0.5$ for each method and experiment. Linear interpolation error increases quicker than the proposed optical flow approaches. In the 8 band experiment, the multivariate model's error is below univariate throughout the temporal domain, while the opposite is true in the 3 band experiment.

The optical flows learned from these experiments resemble high-level dynamics found in physical models (dynamical). In Figure 6a we show a quiver plot where the brightness is flow intensity and arrow direction average pooled over every 10 pixels. Hurricane Irma is in the south east where the flows are pointing in a circular motion, a characteristic of tropical cyclones. The northern side of the tropical cyclone shows intense water vapor movement. Similarly, Hurricane Jose is seen on the coast of Mexico. Furthermore, the jet stream moving east is clear, particularly the clouds moving to the northeast east. On the west coast the jet stream is moving north into Canada. In the center of the country, where water vapor is limited, we see little movement.

5.2.1 Case Study: Hurricane Irma

Here we focus on a particular event to test and visualize flows at a more fine grain resolution. Depicted in Figure 4j, Hurricane Irma makes landfall in Puerto Rico on September 8, 2017. Figure ?? shows the difference after 15-minutes. Data from 12pm (GMT) to 12am (GMT) are selected for analysis as presented in Table 3. We find that, similar to above, the optical flow methods outperform linear interpolation by 30%. However, multivariate flows do not improve the 8 band univariate experiment. Visually, in Figure 4 we show an example of temporal interpolation with fast moving clouds on the hurricane's edge. Plots (h), (i), and (j), depict the RMSE values per pixel over the first 3 bands. We see that the brightness, high error, in linear interpolation is more widespread in cloudy pixels whereas optical flow has less error in the same regions.

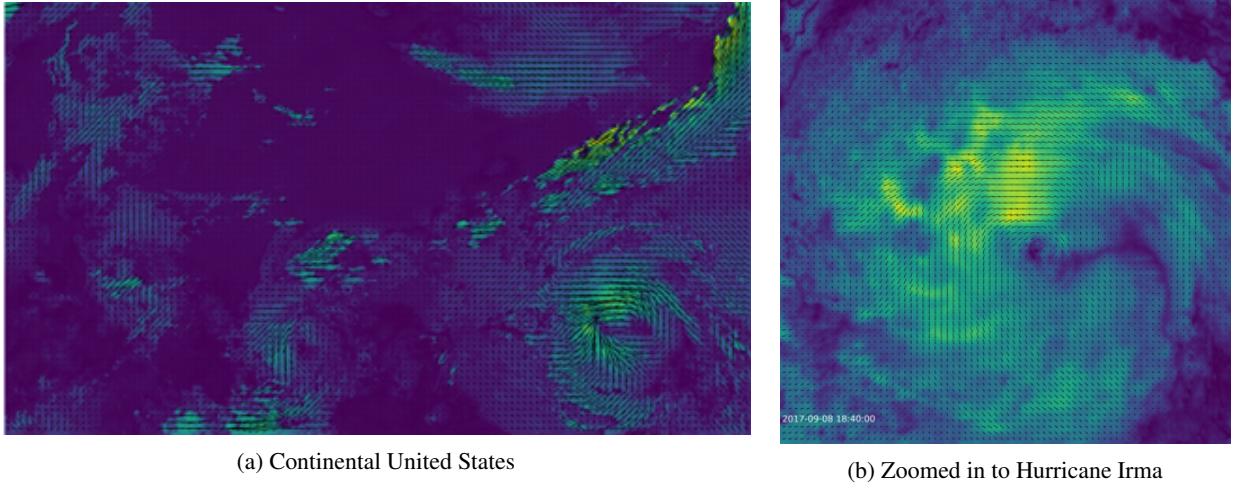


Figure 6: Hurricane Irma approaches Florida on September 9, 2017 at 2:09pm EST. Water vapor (band 8) flow intensity (color brightness/vectors) and direction, \hat{F}_{01} .

	3 Band Models			8 Band Models		
	Linear	SV-SSM	MV-SSM	Linear	SV-SSM	MV-SSM
1	0.0258	0.0193	0.0190	0.0258	0.0191	0.0193
2	0.0356	0.0284	0.0282	0.0356	0.0283	0.0286
3	0.0300	0.0229	0.0224	0.0300	0.0225	0.0226
4	—	—	—	0.0086	0.0055	0.0056
5	—	—	—	0.0196	0.0156	0.0155
6	—	—	—	0.0118	0.0089	0.0088
7	—	—	—	0.0017	0.0012	0.0012
8	—	—	—	0.0019	0.0014	0.0013
3 Band	0.0305	0.0236	0.0232	0.0305	0.0233	0.0235
8 Band	—	—	—	0.0184	0.0140	0.0140

Table 3: Hurricane Irma on September 8, 2017 root mean squared error (RMSE) per band.

In Figure 6b we visualize the water vapor optical flow vectors, $\hat{F}_{01,8}$, where the brightness is flow intensity with arrows corresponding to direction. These flows come from the 8 band multivariate model. The eye of the cyclone is evident where intensity is low and arrows are pointing in circles. Water vapor is moving at high intensity around the eye, particularly to the north.

6 Conclusion

The application of video interpolation techniques and optical flow to generating higher temporal coverage of geostationary data is promising. In this work we show that the Super SloMo architecture with optical flow is well suited for this task and can be generalized to handle multispectral satellite imagery. We present a large scale analysis and test on evenly spaced days in 2019 to show that optical flow is able to well capture optical flows between two images and apply these to predict intermediate frames. Furthermore, predicting optical flows per spectral band generally improves performance as the number of spectral bands increases.

This capability will allow data providers to provide one minute temporal coverage of the continental United States at a 2km spatial resolution with 16 spectral bands using the GOES-R series of satellites. Furthermore, this approach can also be scaled to improve the temporal resolution of the 15-minute fulldisk observations. These data products will allow for improved analysis of atmospheric and physical processes such as hurricane dynamics and wildfire modeling.

The visualizations presented in Figures 6a and 6b open questions regarding the use of optical flow in satellite imagery. While not validated, they show that general atmospheric patterns seem to be captured sufficiently. At the higher resolution, we can see where the most intense movement is occurring within Hurricane Irma. Further research will

focus on understanding these flow vectors in relation to physical processes as well as testing underlying assumptions of optical flow in satellite imagery.

References

- [AxF] Ax. <https://github.com/facebook/Ax>. Accessed: 2019-06-20.
- [Baker et al.2011] Baker, S.; Scharstein, D.; Lewis, J.; Roth, S.; Black, M. J.; and Szeliski, R. 2011. A database and evaluation methodology for optical flow. *International Journal of Computer Vision* 92(1):1–31.
- [Barron, Fleet, and Beauchemin1994] Barron, J. L.; Fleet, D. J.; and Beauchemin, S. S. 1994. Performance of optical flow techniques. *International journal of computer vision* 12(1):43–77.
- [Basu et al.2015] Basu, S.; Ganguly, S.; Mukhopadhyay, S.; DiBiano, R.; Karki, M.; and Nemani, R. 2015. Deepsat: a learning framework for satellite imagery. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 37. ACM.
- [Benedetti et al.2018] Benedetti, P.; Ienco, D.; Gaetano, R.; Ose, K.; Pensa, R. G.; and Dupuy, S. 2018. m^3 fusion: A deep learning architecture for multiscale multimodal multitemporal satellite data fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11(12):4939–4949.
- [Bessho2016] Bessho, K. e. a. 2016. An introduction to himawari-8/9—japan’s new-generation geostationary meteorological satellites. *Journal of the Meteorological Society of Japan. Ser. II* 94(2).
- [Castelluccio et al.2015] Castelluccio, M.; Poggi, G.; Sansone, C.; and Verdoliva, L. 2015. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092*.
- [Cohen and Herlin1996] Cohen, I., and Herlin, I. 1996. Optical flow and phase portrait methods for environmental satellite image sequences. In *European Conference on Computer Vision*, 141–150. Springer.
- [Dosovitskiy et al.2015] Dosovitskiy, A.; Fischer, P.; Ilg, E.; Hausser, P.; Hazirbas, C.; Golkov, V.; Van Der Smagt, P.; Cremers, D.; and Brox, T. 2015. Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, 2758–2766.
- [Fablet, Viet, and Lguensat2017] Fablet, R.; Viet, P. H.; and Lguensat, R. 2017. Data-driven models for the spatio-temporal interpolation of satellite-derived sst fields. *IEEE Transactions on Computational Imaging* 3(4):647–657.
- [Ghassemian2016] Ghassemian, H. 2016. A review of remote sensing image fusion methods. *Information Fusion* 32:75–89.
- [goe] Goes-r advanced baseline imager (abi) algorithm theoretical basis document for cloud and moisture imagery product (cmip). <https://www.star.nesdis.noaa.gov/goesr/docs/ATBD/Imagery.pdf>. Accessed: 2019-06-24.
- [Hall and Llinas1997] Hall, D. L., and Llinas, J. 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE* 85(1):6–23.
- [Jiang et al.2018] Jiang, H.; Sun, D.; Jampani, V.; Yang, M.-H.; Learned-Miller, E.; and Kautz, J. 2018. Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 9000–9008.
- [Justice et al.2002] Justice, C.; Giglio, L.; Korontzi, S.; Owens, J.; Morisette, J.; Roy, D.; Descloitres, J.; Alleaume, S.; Petitcolin, F.; and Kaufman, Y. 2002. The modis fire products. *Remote Sensing of Environment* 83(1-2):244–262.
- [Kussul et al.2017] Kussul, N.; Lavreniuk, M.; Skakun, S.; and Shelestov, A. 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters* 14(5):778–782.
- [Letham et al.2019] Letham, B.; Karrer, B.; Ottoni, G.; Bakshy, E.; et al. 2019. Constrained bayesian optimization with noisy experiments. *Bayesian Analysis* 14(2):495–519.
- [Liebmann and Smith1996] Liebmann, B., and Smith, C. A. 1996. Description of a complete (interpolated) outgoing longwave radiation dataset. *Bulletin of the American Meteorological Society* 77(6):1275–1277.
- [Liu et al.2017] Liu, Z.; Yeh, R. A.; Tang, X.; Liu, Y.; and Agarwala, A. 2017. Video frame synthesis using deep voxel flow. In *Proceedings of the IEEE International Conference on Computer Vision*, 4463–4471.
- [Ouala et al.2018] Ouala, S.; Fablet, R.; Herzet, C.; Chapron, B.; Pascual, A.; Collard, F.; and Gaultier, L. 2018. Neural network based kalman filters for the spatio-temporal interpolation of satellite-derived sea surface temperature. *Remote Sensing* 10(12):1864.

- [Pagano and Durham1993] Pagano, T. S., and Durham, R. M. 1993. Moderate resolution imaging spectroradiometer (modis). In *Sensor Systems for the Early Earth Observing System Platforms*, volume 1939, 2–18. International Society for Optics and Photonics.
- [Rodell et al.2004] Rodell, M.; Houser, P.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.-J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. 2004. The global land data assimilation system. *Bulletin of the American Meteorological Society* 85(3):381–394.
- [Roy et al.2014] Roy, D. P.; Wulder, M.; Loveland, T. R.; Woodcock, C.; Allen, R.; Anderson, M.; Helder, D.; Irons, J.; Johnson, D.; Kennedy, R.; et al. 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote sensing of Environment* 145:154–172.
- [Scharlemann et al.2008] Scharlemann, J. P.; Benz, D.; Hay, S. I.; Purse, B. V.; Tatem, A. J.; Wint, G. W.; and Rogers, D. J. 2008. Global data for ecology and epidemiology: a novel algorithm for temporal fourier processing modis data. *PloS one* 3(1):e1408.
- [Schmit et al.2017] Schmit, T. J.; Griffith, P.; Gunshor, M. M.; Daniels, J. M.; Goodman, S. J.; and Lebair, W. J. 2017. A closer look at the abi on the goes-r series. *Bulletin of the American Meteorological Society* 98(4):681–698.
- [Skakun et al.2016] Skakun, S.; Kussul, N.; Shelestov, A. Y.; Lavreniuk, M.; and Kussul, O. 2016. Efficiency assessment of multitemporal c-band radarsat-2 intensity and landsat-8 surface reflectance satellite imagery for crop classification in ukraine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9(8):3712–3719.
- [Vandal et al.2017] Vandal, T.; Kodra, E.; Ganguly, S.; Michaelis, A.; Nemani, R.; and Ganguly, A. R. 2017. Deepsd: Generating high resolution climate change projections through single image super-resolution. In *23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- [Yuan et al.2005] Yuan, F.; Sawaya, K. E.; Loeffelholz, B. C.; and Bauer, M. E. 2005. Land cover classification and change analysis of the twin cities (minnesota) metropolitan area by multitemporal landsat remote sensing. *Remote sensing of Environment* 98(2-3):317–328.
- [Zhu et al.2010] Zhu, X.; Chen, J.; Gao, F.; Chen, X.; and Masek, J. G. 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment* 114(11):2610–2623.