Temporal Interpolation of Geostationary Satellite Imagery with Task Specific Optical Flow

Thomas Vandal
NASA Ames Research Center
Bay Area Environmental Research Institute
Moffett Field, USA

Ramakrishna Nemani
NASA Ames Research Center
Moffett Field, USA

ABSTRACT

Applications of satellite data in areas such as weather tracking and modeling, ecosystem monitoring, wildfire detection, and land-cover change are heavily dependent on the trade-offs to spatial, spectral and temporal resolutions of observations. In weather tracking, high-frequency temporal observations are critical and used to improve forecasts, study severe events, and extract atmospheric motion, among others. However, while the current generation of geostationary satellites have hemispheric coverage at 10-15 minute intervals, higher temporal frequency observations are ideal for studying mesoscale severe weather events. In this work, we apply a task specific optical flow approach to temporal up-sampling using deep convolutional neural networks. We apply this technique to 16-bands of GOES-R/Advanced Baseline Imager mesoscale dataset to temporally enhance full disk hemispheric snapshots of different spatial resolutions from 15 minutes to 1 minute. Experiments show the effectiveness of task specific optical flow and multi-scale blocks for interpolating high-frequency severe weather events relative to bilinear and global optical flow baselines. Lastly, we demonstrate strong performance in capturing variability during a convective precipitation events.

KEYWORDS

Optical flow, temporal interpolation, remote sensing

ACM Reference Format:
Thomas Vandal and Ramakrishna Nemani. 2020. Temporal Interpolation of Geostationary Satellite Imagery with Task Specific Optical Flow. In *Proceedings of 1st ACM SIGKDD Workshop on Deep Learning for Spatiotemporal Data, Applications, and Systems (DeepSpatial '20).* ACM, New York, NY, USA, 9 pages.

1 INTRODUCTION

Every second satellites around the earth are generating valuable data to monitor weather, land-cover, infrastructure, and human activity. Satellite sensors capture reflectance/radiance intensities at designated spectral wavelengths, 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. Scientific specific sensors capture a larger range of wavelengths, such as micro, infrared, and thermal waves, providing information to many applications such as storm tracking and wildfire detection. However, sensing a greater number of wavelengths is technologically more complex and applies further constraints of temporal and spatial resolution. Similarly, a higher temporal frequency requires high altitude orbital dynamics which then affects the spatial resolution due to it's distance from earth.

NASA and other agencies have developed satellites designed for a variety of applications in both polar and geostationary orbits. Polar orbiting satellites cross south and north poles each revolution around the earth. These satellites have relatively low altitude orbits which allow for high spatial resolution but with an optimal revisit interval of 1-day. NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) [28] and Landsat-8 [29] satellites follow a polar orbit with 1- and 8-day revisit times, respectively. Data provided by MODIS and Landsat are widely used for quantifying effects of climate change, land-cover usage, and air pollution, among others, but are not well suited to monitoring high-frequency events. On the other hand, geostationary satellites are well suited for sub-daily events such as tracking weather events and understanding diurnal cycles. The geostationary orbit keeps satellites in a consistent point 35, 786km above Earth's equator. While the high altitude reduces spatial resolution, the current generation of geostationary satellites is able to provide minute-by-minute data, enabling immense opportunity for understanding atmospheric, land-cover, and oceanic dynamics.

Within a few years, a constellation of geostationary satellites by multiple international institutions will provide global coverage of earth's state. Latest generation of geostationary satellites includes NOAA/NASA's GOES-16/17 [2], Japan's Himwari-8/9 [7], China's Fengyun-4 [34], and Korea's GEO-KOMPSAT-2A with future plans in development. Full-disk coverage from such satellites have revisit times of 10-15 minutes allowing applications to real-time detection and observation of wildfires [32], hurricane tracking, air flood, precipitation estimation, flood risk, and others [31]. Further, given improved spectral and spatial resolution in current generation sensors, geostationary satellites open opportunities to incorporate and learn from less frequent observations from polar orbiters.

While 10-15 minute revisit times is temporally sufficient for many applications, higher frequency snapshots can aid a variety of tasks. For instance, understanding rapidly evolving convective events is a high priority for improving atmospheric models, which are notoriously poor at simulating heavy precipitation and as highlighted in NASA's Earth Science Decadal Survey [8]. However, data for analyzing such events is often not available at the desired frequency.

DeepSpatial '20, August 2020, Virtual

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

Similarly, comparing multiple satellite observations is dependent on their corresponding timestamps. This leads to an interpolation task between observations in a multi-spectral spatio-temporal sequence, similar to that of video interpolation.

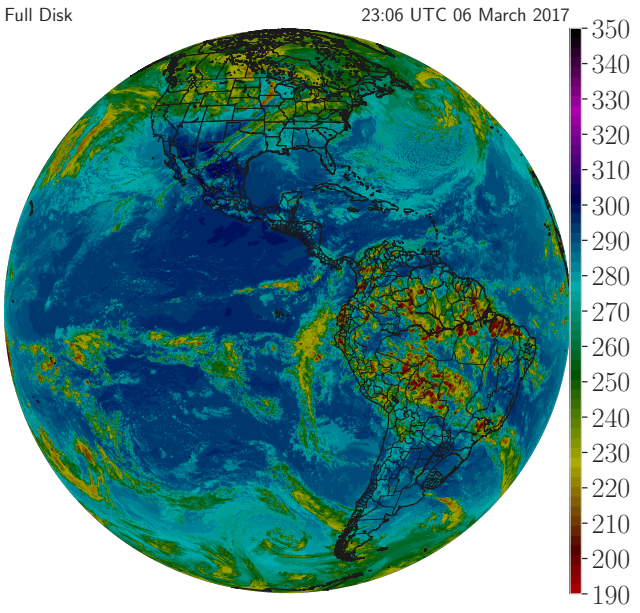
Optical flow is a problem of tracking apparent motion by estimating partial derivatives between images. Optical flow is the basis for top performing video interpolation methods [4, 17, 24]. In this work, we adapt Superslomo (SSM) [17], a video interpolation method, to the problem of temporal interpolation between geostationary images. We compare properties of global, task specific, and multi-scale block SSM models with the traditional linear interpolation. Our experiments show that task specific SSM is capable of interpolating high-frequency severe atmospheric events. Further, visual analysis suggests the learned optical flows resemble atmospheric with dynamic visibility maps.

The remainder of this paper is outlined as follows. Section 2 discusses related work including resolution enhancement 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 a severe storm case study are presented in section 5. Lastly, section 6 concludes with challenges and further work.

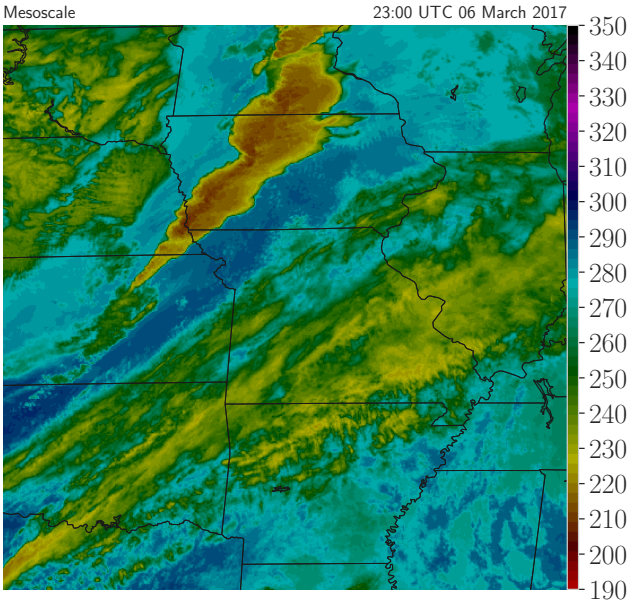
2 GOES-R SATELLITE DATASET

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 (East and West side of the Americas), operated by NASA and NOAA provides scientists with unprecedented temporal frequency enabling real-time environmental monitoring using the Advanced Baseline Imager (ABI) [30]. GOES-16/17 senses 16-bands of data which are listed in Table 1 with central central wavelength, spatial resolution, and band name. Three data products are derived from each GOES-16/17; 1. Full-disk covering the western hemisphere every 15-minutes (figure 1a), 2. Continental US every 5-minutes, and 3. Mesoscale user directed 1000km by 1000km sub-region every at an optimal 30 seconds (figure 1b). ABI's 16 spectral bands includes two visible (1-2), four near-infrared (3-6), and ten infrared (7-16) bands enabling a suite of applications.

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 sub-region for near constant monitoring of severe events. 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, such as studying convective events [14]. 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.



(a)



(b)

Figure 1: Cloud-top temperature from "clean" IR long-wave window Band 13 Full-disk (1a) and Mesoscale (1b) coverage.

3 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

Band	Central Wavelength (μm)	Spatial Resolution (km)	Name
1	0.47	1	Blue
2	0.64	0.5	Red
3	0.86	1	Veggie
4	1.37	1	Cirrus
5	1.6	1	Snow/Ice
6	2.24	2	Cloud Particle Size
7	3.9	2	Shortwave Window
8	6.2	2	Upper-level Water Vapor
9	6.9	2	Mid-level Water Vapor
10	7.3	2	Low-Level Water Vapor
11	8.4	2	Cloud-Top Phase
12	9.6	2	Ozone
13	10.3	2	"Clean" IR Longwave
14	11.2	2	IR Longwave
15	12.3	2	"Dirty" IR Longwave
16	13.3	2	CO_2 Longwave IR

Table 1: GOES-R Series Bands

of deep learning in the area. Secondly, we provide a brief review of video intermediate frame interpolation techniques.

3.1 Resolution Enhancement of Satellite Data

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. Further, spatial and temporal resolution enhancement is often applied to improve analysis precision. Techniques to handle these challenges have been developed and are widely applied across the remote sensing community.

Many statistical and machine learning methodologies for improving spatial resolution have been explored and is an active area of research. Data fusion is one area where two or more datasets are *fused* to generate an enhanced product, often with both higher spatial and temporal resolutions [15]. The Spatial and Temporal Adaptive Reflectance Fusion (STARFM) algorithm, for example, uses Landsat and MODIS to produce a daily 30-meter reflectance product by using a spectral wise weighting model [35]. Similarly, 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 [13]. In recent years, super-resolution techniques have presented state-of-the-art results for spatial enhancement of satellite images [20, 22, 33].

Approaches for temporal resolution enhancement of individual satellite observations have not been as well studied. Liebmann et al. presented the first linearly interpolated datasets filling in missing and erroneous longwave radiation many days apart to improve global coverage [23]. Similarly, [18] presented a comparison of multiple methods for interpolating between MODIS observations to generate a synthetic leaf area index dataset. A number of statistical techniques including long-term climatology measures and time-series decomposition were applied to smooth observation and fill gaps. [11] presented an approach using linear interpolation on sub-daily geostationary imagery to match timestamps between multiple satellites. However, given more frequent observations by the recent generation of geostationary observations, more complex methods beyond linear interpolation may be more applicable and accurate in the temporal domain.

Our work proposes to apply deep learning methodologies to optimize the interpolation problem. In recent years, a number of applications in processing and learning from satellite data have shown state-of-the-art results using deep learning. For example, [6] showed that recurrent and convolutional neural networks effectively assimilate multiple satellite images. [20] presented a global deep learning super-resolution approach for Sentinal-2 with a 50% improvement beyond traditional techniques. In terms of classification, DeepSat showed that normalized deep belief networks tuned where able to outperform traditional techniques for image classifications [5]. Convolutional neural networks have been shown to effectively classify land use in remotely sensed images, from urban areas [9] to crop types [19].

While many studies have explored resolution enhancement spatially, and temporally, the authors are not aware of any prior work exploring temporal interpolation at the minute-to-minute scale. Prior approaches on longer time scales have applied linear interpolation and nearest neighbor techniques. We will explore the applicability of a more complex optical flow approach to temporal interpolation at very high resolutions and use linear interpolation as our baseline, as applied in prior work.

3.2 Video Intermediate Frame Interpolation

Video interpolation techniques have shown high skill at generating slow motion footage by generating intermediate frames in spatially and temporally coherent sequences [4, 17, 24, 25]. These approaches are designed to learn the dynamics by inferring displacement of spatial structure between consecutive images. Optical flow is widely used for this task which estimates pixel displacement by comparing two images and interpolating appropriately. For RGB imagery, this task is equivalent to estimating movement of objects. In recent years, deep learning architectures have shown promising results for both optical flow and video interpolation. Flownet presented an encoder-decoder architecture for optical flow with correlation operations and skip-connections for supervised learning [12]. Following studies have extended this work with more complex architectures such as stacking networks for large and small displacements [16] and unsupervised learning [26]. Deep Voxel Flow [25], Superslomo [17], cyclic frame generation [24], and others have shown deep learning optical flow techniques to be well suited for video interpolation.

However, many video interpolation techniques focus on *single frame* interpolation, meaning that a single frame is estimated between two consecutive frames [24, 25, 27]. However, when interpolating satellite imagery, time-dependent and multi-frame estimation is preferred for more flexibility. Jiang et al. presented Superslomo (SSM) which combines both optical flow and occlusion models for

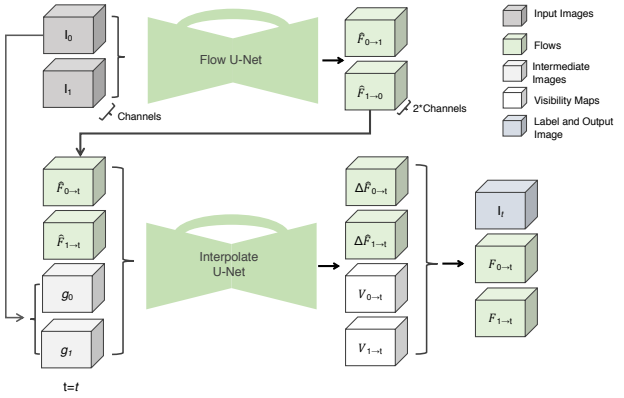


Figure 2: SuperSlomo Architecture with Flow and Interpolation Networks.

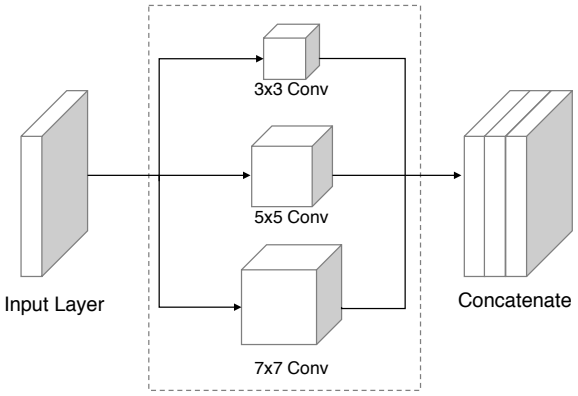


Figure 3: Multi-scale block

time-dependent estimation between consecutive frames [17]. The time-dependent nature of this approach produces spatially and temporally coherent predictions of any time between 0 and 1. In their experiments, [17] 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 multi-scale optical flows.

High-frequency satellite imagery can take advantage of these techniques to extract dynamics of different physical processes. We study how SSM can be effectively applied to this problem by experimenting with global and task specific models.

4 METHODOLOGY

Temporal up-sampling of geostationary satellite data is a near identical problem as intermediate video frame interpolation with domain specific characteristics. In video interpolation, the goal is to estimate an intermediate frame given two or more consecutive RGB images. A single set of optical flows are sufficient for interpolating between RGB images as objects captured in the visible spectrum are reasonably consistent across frames. However, as discussed above, satellite imagery often consists of 10's or even 100's of spectral channels with varying spatial resolutions. Further, each channel captures different physical properties with heterogeneous motion including severe events such as convection leading to heavy precipitation and tornadoes. The goals of the proposed methodologies include interpolating to a user defined point in time, capturing varying spatial dynamics, and computational efficiency at scale. In this section, we review the SSM framework for temporal up-sampling with optical flow, as presented by [17]. Next, we discuss task specific SSM models and the chosen network architecture with multi-scale blocks.

4.1 Intermediate Frame Interpolation

SSM intermediate frame interpolation considers the case of frame estimation at a user defined point in continuous time [17]. To ensure smooth transitions and structural similarity between frames, SSM is designed to predict optical flows between two input images as a function of time. The approach, which can be seen in Figure 2, consists of two deep neural networks. The first estimates forward and backward flows between two input images. The second network, depending on time, updates the forward and backward flows and generates visibility maps to handle occlusion. These features of SSM are well suited to geostationary data by enabling arbitrary temporal up-sampling and synchronization of multiple datasets.

Following the notation from [17], let $I_0, I_1, I_t \in \mathcal{R}^{H \times W}$ where $t \in (0, 1)$, H as image height and W as image width, and C a number of spectral bands. In our case, $C = 1$. The goal is then to construct an intermediate frame I_t with a linear combination of warped I_0 and I_1 as defined by:

$$\hat{I}_t = \alpha \cdot g(I_0, F_{0\to t}) + (1 - \alpha) \cdot g(I_1, F_{1\to t}) \tag{1}$$

where $F_{0\to t}$ and $F_{t\to 1}$ are the optical flows from I_0 to I_t and I_t to I_1, respectively. 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. Further, atmospheric dynamics cause physical characteristics to change over time. One example is convection such that warm/cold air vertically and rapidly mixes in the atmosphere causing severe weather events. In the context of interpolating, dynamics between I_0 and I_1 cause cloud temperature to rapidly decrease, leading to a drastic change brightness intensity and breaking assumptions of optical flow. However, visibility maps, $V_{0t}, V_{1t} \in (0, 1)^{H \times W}$, weight brightness importance to account for both occlusion and intensity changes. Equation 1 is then be redefined as:

$$\hat{I}_t = \frac{1}{Z} \cdot \big((1 - t) \cdot V_{0t} \cdot g(I_0, F_{0\to t}) + t \cdot V_{1t} \cdot g(I_1, F_{1\to t})\big) \tag{2}$$

where $Z = (1-t) \cdot V_{0t} + t \cdot V_{1t}$ is a normalization factor. Forward and backward optical flows, $(F_{0\to t}, F_{1\to t})$, at time t are estimated by a sequence of two flow networks, G_{flow} and G_{Interp}, as presented in

Figure 2. The first network, $G_{\text{flow}}(I_0, I_1)$, infers backward and forward optical flows, $(F_{0\to1}, F_{1\to0})$, between two input images. After generating approximate intermediate flows, $(\hat{F}_{t\to0}, \hat{F}_{t\to1})$, intermediate images are generated. The interpolation network, G_{Interp}, predicts visibility maps (V_{0t}, V_{1t}) and final flows $(F_{t\to0}, F_{t\to1})$ as a function of a concatenation of input images, intermediate flows, and intermediate warped images.

4.2 Task Specific Interpolation

As discussed in Section 2, GOES-16 consists of 16 channels with resolutions between 500m and 2km. Flows between images with different spatial resolutions will have flows of varying intensity. Clouds in 500m images will cover 4x more pixels than a corresponding 2km image. We explore the use of task specific networks by learning separate SuperSlomo models for each spectral channel and compare with a single global model. While requirements for GPU computation multiplies, we will show that improved performance of task specific models improves results substantially.

4.3 Network Architecture

Deep neural networks with encoding and decoding are well suited to model both local and global spatial structure. Architectures of this type include Flownet [12] and U-Net [17] which have been shown to perform well in the task of optical flow. We follow this approach using a U-Net architecture for each of the flow and interpolation networks. The U-Net architecture applied has 4 down-sampling layers followed by 4 up-sampling layers with skip connections between each corresponding layer. A convolution layer maps the input to 64 channels with a kernel size of 7. The following downsampling layers are of size 128, 256, 512, and 512 with kernel sizes 5, 5, 3 and 3. Each downsampling layer performs. average pooling and two convolutions with rectified linear unit (ReLu) activations. Upsampling layers of size 256, 128, 64, and 32 all with kernel sizes of 3 is then applied. Each layer performs bilinear interpolation followed by two convolutions with rectified linear unit (ReLu) activations. Lastly, 32 channels in the last hidden layer are mapped to the number of output channels using a convolution operation of kernel size 3. Flow and interpolation networks use the same architecture with different input and output dimensions as discussed above.

Tracking both small and large displacements continues to be a challenge, even with encoder-decoder network architectures. Other approaches have shown that using a stack of networks performing small and large displacement perform well [16]. In this work, we explore the applicability of multi-scale hidden layers to track local and global features. We follow a similar approach applied in [10] where hidden layers are defined to have multiple convolution operations with different sized kernels followed by a concatenation layer, as shown in Figure 3. In our networks, kernels of size 3, 5, and 7 conserve high-frequency spatial details while abstracting global motion for improved optical flows and visibility maps.

4.4 Training Loss

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

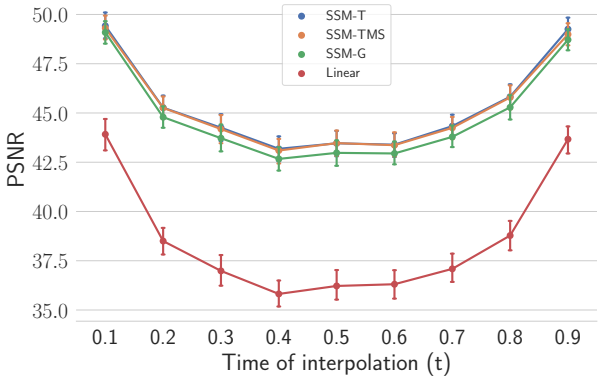


Figure 4: Interpolation Error as a function of time.

combination of reconstruction, warping, and smoothness losses such that:

$$l = \lambda_r l_r + \lambda_w l_w + \lambda_s l_s. \tag{3}$$

We note that [17] 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 distance between observed and predicted intermediate frames:

$$l_r = \frac{1}{N}\sum_{i=1}^{N} ||\hat{I}_{t_i} - I_{t_i}||. \tag{4}$$

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

$$l_w = ||I_0 - g(I_1, F_{0\to1})|| + ||I_1 - g(I_0, F_{1\to0})|| + \frac{1}{N}\sum_{i=1}^{N} ||I_{t_i} - g(I_0, F_{0\to t_i})|| + \frac{1}{N}\sum_{i=1}^{N} ||I_{t_i} - g(I_1, F_{1\to t_i})||. \tag{5}$$

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:

$$l_s = ||\nabla F_{0\to1}||_1 + ||\nabla F_{1\to0}||_1 \tag{6}$$

In practice, this training setup requires optimization over multiple hyper-parameters including λ_r, λ_s, λ_w, and a learning rate.

5 EXPERIMENTS

We demonstrate the effectiveness of a set of SSM models on a large-scale dataset using a high-performance computing system with a cluster of GPUs. The goal of our experiments is to show that optical flow is highly applicable for temporal interpolation of satellite imagery and compare to the baseline of linear interpolation, as traditionally applied. The following sub-sections outline the training process, compare methodologies, and study the effectiveness on a severe convective precipitation event.

Model Band	PSNR ↑				RMSE ↓				SSIM ↑			
	Linear	SSM-G	SSM-T	SSM-TMS	Linear	SSM-G	SSM-T	SSM-TMS	Linear	SSM-G	SSM-T	SSM-TMS
1	37.795	36.828	**38.282**	37.818	0.178	0.198	**0.160**	0.185	0.719	0.682	**0.734**	0.722
2	37.408	37.006	**37.748**	37.583	0.185	0.186	**0.169**	0.177	0.637	0.616	**0.649**	0.644
3	41.808	40.544	**41.350**	41.135	0.100	0.112	**0.099**	0.108	0.760	0.712	0.731	**0.737**
4	60.519	60.838	**62.598**	61.925	0.012	0.011	**0.008**	0.009	0.969	0.974	**0.983**	0.982
5	56.097	55.129	**56.044**	55.703	0.018	0.019	**0.018**	**0.018**	0.932	0.928	**0.937**	0.935
6	55.076	58.316	58.693	**58.758**	0.373	0.255	**0.242**	**0.242**	0.747	0.884	0.893	**0.895**
7	40.721	46.084	**46.591**	46.496	1.825	0.972	**0.917**	0.932	0.766	0.899	**0.907**	0.905
8	50.669	57.656	**58.432**	58.135	0.613	0.374	**0.226**	0.358	0.747	0.907	**0.913**	0.912
9	47.476	55.287	56.014	**56.015**	0.904	0.336	0.306	**0.305**	0.756	0.924	**0.929**	**0.929**
10	44.601	52.535	**53.226**	53.120	1.222	0.582	**0.418**	0.550	0.748	0.919	**0.924**	**0.924**
11	38.530	44.753	45.184	**45.243**	2.335	1.071	**1.020**	1.013	0.770	0.922	**0.929**	**0.929**
12	43.560	49.568	50.023	**50.030**	1.314	0.626	**0.594**	**0.594**	0.762	0.913	**0.921**	0.920
13	38.667	44.925	**45.439**	45.343	2.286	1.177	**0.991**	1.130	0.782	0.925	**0.933**	0.932
14	38.167	44.594	44.996	**45.090**	2.392	1.080	1.036	**1.024**	0.770	0.924	0.931	**0.932**
15	38.163	44.699	**45.284**	45.252	2.387	1.185	**0.991**	1.118	0.762	0.921	0.929	**0.930**
16	40.578	47.313	47.731	**47.778**	1.819	0.786	0.751	**0.745**	0.721	0.892	0.898	**0.899**

Table 2: Model comparison results from 200 randomly selected samples in 2019. Bold highlights the top performing model and * highlights the second best.

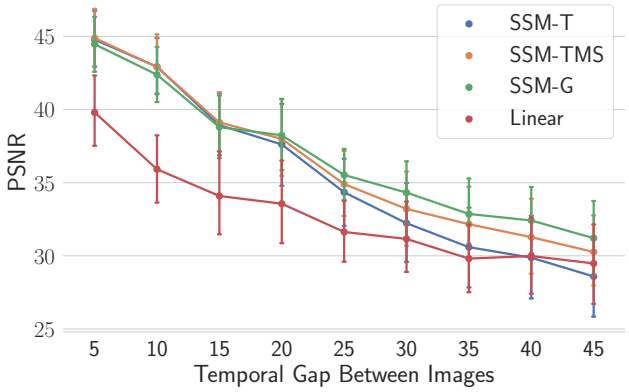


Figure 5

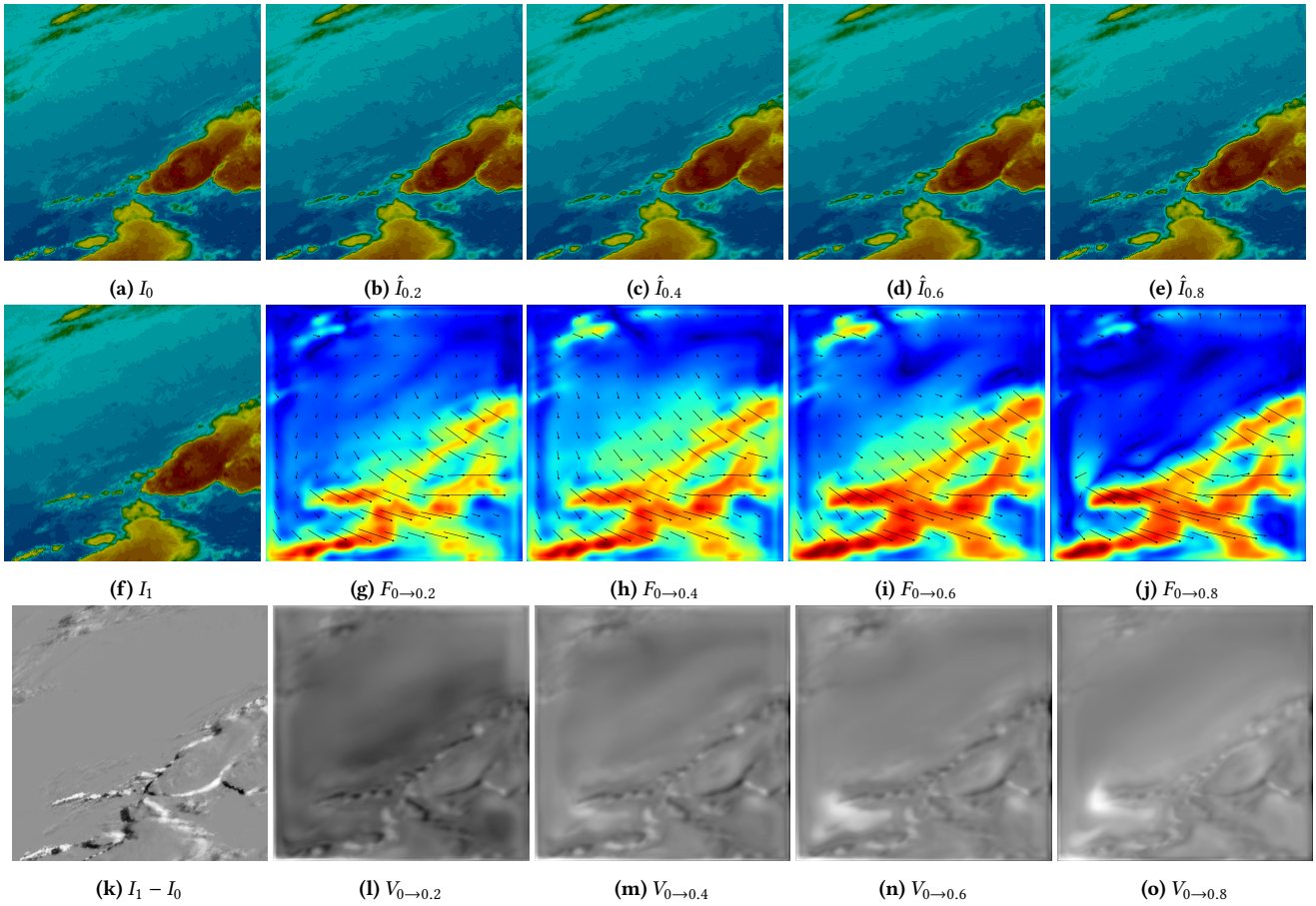
5.1 Training

Data for training and testing was taken from the GOES-16 Mesoscale 1-minute imagery. These images are of identical spatial and spectral resolution as North America (available every 5 minutes) and full-disk imagery (available every 10 minutes) so the learned models are directly applicable to these datasets. Training data was selected using all samples for every 5-days of the year 2018 and testing data on a randomly selected set of examples from 2019. Samples were generated as 264x264 sub-images and randomly cropped to 256x256 during training. Standardized normalization was applied independently to each channel to ensure similar pixel intensity distributions across bands. Temporally, samples are selected from a sequence of 15 time-steps such that inputs (I_0, I_1) are 10-minutes apart with a random label I_t in-between. Further, during training, images are randomly flipped and rotated to improve generality in the U-Net architecture. A random training/validation split of 20% was used to monitor learning. We select cloud top temperature tracked by band 13 ($10.3\mu m$) in ablation and demonstration experiments as used in studies of convection and atmospheric motion vectors. Experiments for this study leveraged NASA's Pleiades Supercomputer and the NASA Earth Exchange to process large-scale GOES-16 data and train individual networks for each of the 16 channels.

Adam optimization is used to minimize Equation 3 with default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and eps=1e-8 in PyTorch. We found that learning is sensitive to hyper-parameters λ_s and λ_w and are optimized using probabilistic grid search, constrained Bayesian optimization [21]. Constrained Bayesian optimization applies efficient randomized Monte Carlo simulations over λ_s and λ_w holding $\lambda_r = 1$. We perform this process using the open-source Ax library [1] for 20 trials on band 1 with SuperSlomo and find $\lambda_s = 0.23$ and $\lambda_w = 0.65$ minimized reconstruction loss on the validation set. These hyper-parameters are applied to all following bands and experiments. Training the suite of models was executed on multi-node GPU cluster of V100's with 1 model per band. Multi-gpu training was used for serial hyper-parameter optimization and global models.

5.2 Model Comparison

This section compares variations of SuperSlomo with a linear interpolation baseline for interpolation of geostationary images. Linear interpolation between frames is performed by taking a linear combination of two input images weighted by time, $\hat{I}_t = (1-t) * I_0 + t * I_1$. A set of three SuperSlomo models are explored including global (SSM-G), task specific (SSM-T), and task specific with multi-scale layers (SSM-TMS). We note, that due to the multi-scale blocks, SSM-TMS has fewer parameters than SSM-T. SSM-T models are trained for each band separately. SSM-G is trained using training data from all bands and hence a substantially larger training set. Root mean



(a) I_0 **(b)** $\hat{I}_{0.2}$ **(c)** $\hat{I}_{0.4}$ **(d)** $\hat{I}_{0.6}$ **(e)** $\hat{I}_{0.8}$

(f) I_1 **(g)** $F_{0\to0.2}$ **(h)** $F_{0\to0.4}$ **(i)** $F_{0\to0.6}$ **(j)** $F_{0\to0.8}$

(k) $I_1 - I_0$ **(l)** $V_{0\to0.2}$ **(m)** $V_{0\to0.4}$ **(n)** $V_{0\to0.6}$ **(o)** $V_{0\to0.8}$

Figure 6: A severe convective event on May 23, 2019 from 2:00-2:10 UTC taken from GOES-16 Mesoscale. 6a and 6f are in the input images and their difference 6k. 6b-6e show SSM-T interpolated predictions, flow intensity and direction in 6g-6j, and visibility maps in 6l-6o

square error (RMSE), peak to signal noise ratio (PSNR), and self similarity measure (SSIM) are used to evaluate performance.

We first study inherent properties of SSM on Band 13 including time dependence and sensitivity to larger displacements. Interpolation between two frames are expected to have smooth transitions from one frame to another. Generally interpolation will have the largest error where the distance to frames is maximum (ie. directly between the input frames). In figure 4 we compare PSNR as a function of $t \in [0, 1]$ between models and see this effect. The gap between linear and SSM models is pronounced. Between SSM models, SSM-T and SSM-TMS have similar performance. SSM-G which is a more generalized model does not perform quite as well as SSM-T and SSM-TMS, suggesting task specific models across bands perform better. Figure 5 shows PSNR at $t = 0.5$ while increasing the gap between I_0 to I_1 from 5 to 45-minutes. A 45-minute gap contains 9x more displacement than a 5-minute gap making the optical flow problem more difficult. Over the first 15-minutes SSM models perform similarly and better than linear. As the gap widens, SSM-TMS and SSM-G begin performing better tham SSM-T. This suggests that SSM-TMS multi-scale layers may be capturing more motion. SSM-G's more diverse dataset includes 500m data which has larger displacements than the 2-km band 13.

In table 2 we present the results for each of the 16 bands of GOES-16 in 200 randomly sampled 10-minute intervals from 2019. Metrics are computed for each sample at $t = 0.5$, where the error is largest, and averages over all examples. As a whole, our results find that task specific SSM models, SSM-T and SSM-TS, outperform linear interpolation and a single global interpolation network, SSM-G. Interpolation of the visible and near-infrared bands (1-6) with optical flow provided modest improvements in all metrics. Interpolation with SSM of infrared, or thermal bands, drastically improves performance on all metrics. We find that task specific models outperform the global model throughout even with the reduced training data size. SSM-T and SSM-TMS perform similarly with SSM-TMS having many fewer parameters.

5.3 Severe Weather Event

This section studies an example of a convective precipitation event visualized in figure 6a. In the context of severe weather, convection

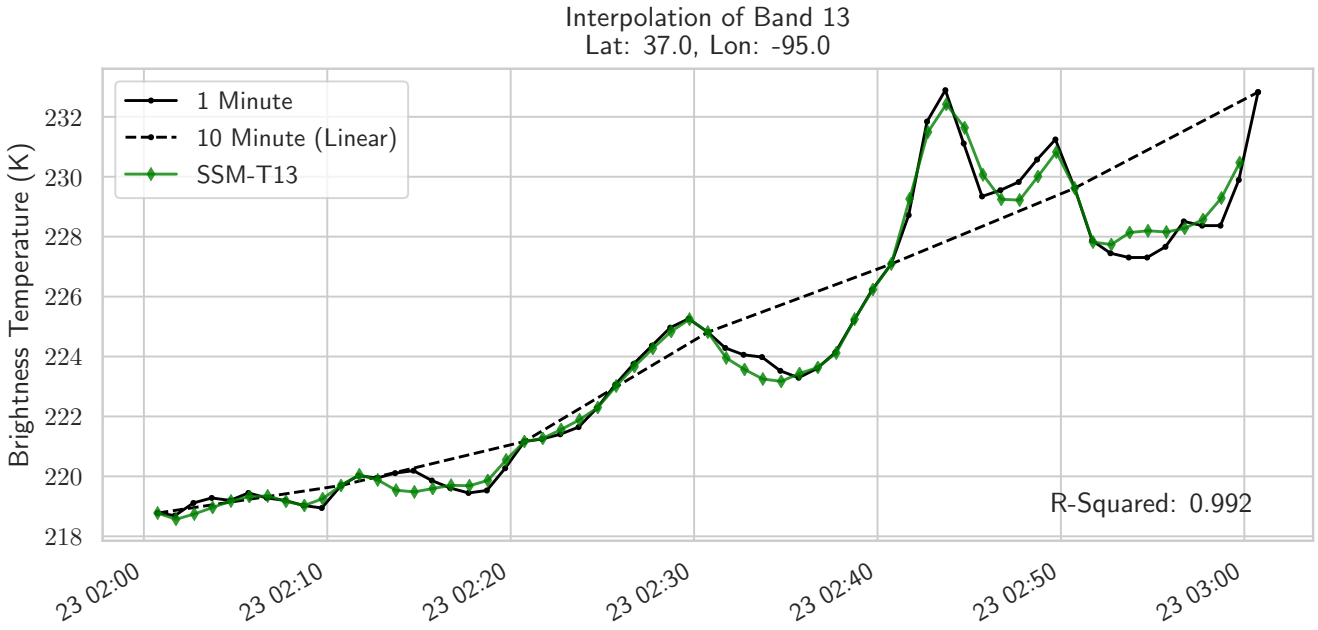


Figure 7: Cloud-top temperature during a convective event

is vertical motion in the atmosphere that occurs when warm air on the surfaces forces cold air in the atmosphere down often causing super-cells and heavy precipitation. For the first time, [3] studied this process using GOES-14 1-minute imagery for a set of super-cells. The authors found that atmospheric motion can help define signatures of super-cell events to better inform weather forecasting models. Here, we show that cloud top brightness can be interpolated from 10 to 1-minute during a convective event.

One minute mesoscale (M1) data from May 23, 2019 from 2:00 to 3:00 UTC at -95° longitude and 37° latitude is used for analysis. In this region, a convective storm is occuring and moving east. The data is down-sampled to 10-minutes interpolated back to a 1-minute time-series. Figure 6a shows the region of interest with predictions (I_t), optical flows ($F_{0\to t}$), and visibility maps ($V_{0\to t}$) between time 0 and t. The optical flows show the storm moving east and slightly rotating with maximum displacement around the storm edges. According to the flows, horizontal cloud movement in the center of the storm is less than nearby areas. Visibility maps show if the corresponding pixel in I_0 occurs in I_t. Visibility pixels correspond to edges of clouds which allows 1 to be a non-linear combination relative to time.

Figure 7 presents these time-series over the defined 1-hour time-frame at (-95°, 37°). The time-series shows cloud top brightness increasing as warmer air rises in the atmosphere. A dashed line shows the 10-minute time-series and is equivalent to linear interpolation. SSM-T is overlayed the observation and well captures the variability of a drastic 12°K temperature increase. These results suggest that optical flow may be a promising approach for interpolating geostationary imagery for applications to severe events. For reference, we include two more examples of extreme events in the supplement.

6 CONCLUSION

This work proposes that temporal interpolation with optical flow is capable of modeling high-frequency events between geostationary images with high-accuracy by learning from mesoscale rapid-scan observations. Experiments showed that learning independent weights of SSM for each band improve performance beyond one global SSM as well as the linear interpolation baseline. Multi-scale blocks in SSM-TMS have fewer parameters, performs well for larger displacements, and comparable to SSM-T overall. Interpolation well captured temporal variability of cloud top brightness during a severe convective event. This interpolation has direct applications to improved precipitation estimation and weather variability.

While further analysis is necessary, our results suggest that dynamics of atmospheric motion is learned by the network using displacement flows and visibility maps which would have direct implications to weather forecasting. The learned optical flows derive dense atmospheric motion vectors that can be used to initialize weather models and analyze large scale winds. Secondly, internal dynamics captured may provide knowledge on how to predict future states as applied for video-frame prediction. In future work we will explore the accuracy of optical flow to estimating atmospheric motion relative to large-scale observations as well as model interpretability to better understand which physical dynamics are captured.

REFERENCES

[1] [n.d.]. Ax. https://github.com/facebook/Ax. Accessed: 2019-06-20.
[2] [n.d.]. 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.
[3] Jason M Apke, John R Mecikalski, and Christopher P Jewett. 2016. Analysis of mesoscale atmospheric flows above mature deep convection using super rapid scan geostationary satellite data. *Journal of Applied Meteorology and Climatology*

55, 9 (2016), 1859–1887.
[4] Wenbo Bao, Wei-Sheng Lai, Chao Ma, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang. 2019. Depth-aware video frame interpolation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3703–3712.
[5] Saikat Basu, Sangram Ganguly, Supratik Mukhopadhyay, Robert DiBiano, Manohar Karki, and Ramakrishna Nemani. 2015. DeepSat: a learning framework for satellite imagery. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 37.
[6] Paola Benedetti, Dino Ienco, Raffaele Gaetano, Kenji Ose, Ruggero G Pensa, and Stéphane Dupuy. 2018. M^3Fusion: 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 (2018), 4939–4949.
[7] Kotaro Bessho, Kenji Date, Masahiro Hayashi, Akio Ikeda, Takahito Imai, Hidekazu Inoue, Yukihiro Kumagai, Takuya Miyakawa, Hidehiko Murata, Tomoo Ohno, et al. 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 (2016), 151–183.
[8] Space Studies Board, Engineering National Academies of Sciences, Medicine, et al. 2019. *Thriving on our changing planet: A decadal strategy for Earth observation from space*. National Academies Press.
[9] Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva. 2015. Land use classification in remote sensing images by convolutional neural networks. *arXiv preprint arXiv:1508.00092* (2015).
[10] Lian Ding, Kai Zhao, Xiaodong Zhang, Xiaoying Wang, and Jue Zhang. 2019. A lightweight U-net architecture multi-scale convolutional network for pediatric hand bone segmentation in X-ray image. *IEEE Access* 7 (2019), 68436–68445.
[11] David R Doelling, Norman G Loeb, Dennis F Keyes, Michele L Nordeen, Daniel Morstad, Cathy Nguyen, Bruce A Wielicki, David F Young, and Moguo Sun. 2013. Geostationary enhanced temporal interpolation for CERES flux products. *Journal of Atmospheric and Oceanic Technology* 30, 6 (2013), 1072–1090.
[12] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick Van Der Smagt, Daniel Cremers, and Thomas Brox. 2015. Flownet: Learning optical flow with convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*. 2758–2766.
[13] Ronan Fablet, Phi Huynh Viet, and Redouane Lguensat. 2017. Data-driven Models for the Spatio-Temporal Interpolation of satellite-derived SST Fields. *IEEE Transactions on Computational Imaging* 3, 4 (2017), 647–657.
[14] Thomas Fiolleau and Rémy Roca. 2013. An algorithm for the detection and tracking of tropical mesoscale convective systems using infrared images from geostationary satellite. *IEEE transactions on Geoscience and Remote Sensing* 51, 7 (2013), 4302–4315.
[15] David L Hall and James Llinas. 1997. An introduction to multisensor data fusion. *Proc. IEEE* 85, 1 (1997), 6–23.
[16] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. 2017. Flownet 2.0: Evolution of optical flow estimation with deep networks. In *IEEE conference on computer vision and pattern recognition (CVPR)*, Vol. 2. 6.
[17] Huaizu Jiang, Deqing Sun, Varun Jampani, Ming-Hsuan Yang, Erik Learned-Miller, and Jan Kautz. 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.
[18] Sivasathivel Kandasamy, Frederic Baret, Aleixandre Verger, Philippe Neveux, and Marie Weiss. 2013. A comparison of methods for smoothing and gap filling time series of remote sensing observations–application to MODIS LAI products. *Biogeosciences* 10, 6 (2013), 4055–4071.
[19] Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrii Shelestov. 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters* 14, 5 (2017), 778–782.
[20] Charis Lanaras, José Bioucas-Dias, Silvano Galliani, Emmanuel Baltsavias, and Konrad Schindler. 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing* 146 (2018), 305–319.
[21] Benjamin Letham, Brian Karrer, Guilherme Ottoni, Eytan Bakshy, et al. 2019. Constrained Bayesian optimization with noisy experiments. *Bayesian Analysis* 14, 2 (2019), 495–519.
[22] Feng Li, Lei Xin, Yi Guo, Dongsheng Gao, Xianghao Kong, and Xiuping Jia. 2017. Super-resolution for GaoFen-4 remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 15, 1 (2017), 28–32.
[23] Brant Liebmann and Catherine A Smith. 1996. Description of a complete (interpolated) outgoing longwave radiation dataset. *Bulletin of the American Meteorological Society* 77, 6 (1996), 1275–1277.
[24] Yu-Lun Liu, Yi-Tung Liao, Yen-Yu Lin, and Yung-Yu Chuang. 2019. Deep video frame interpolation using cyclic frame generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 8794–8802.
[25] Ziwei Liu, Raymond A Yeh, Xiaoou Tang, Yiming Liu, and Aseem Agarwala. 2017. Video frame synthesis using deep voxel flow. In *Proceedings of the IEEE International Conference on Computer Vision*. 4463–4471.
[26] Simon Meister, Junhwa Hur, and Stefan Roth. 2018. UnFlow: Unsupervised learning of optical flow with a bidirectional census loss. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
[27] Simon Niklaus, Long Mai, and Feng Liu. 2017. Video frame interpolation via adaptive separable convolution. In *Proceedings of the IEEE International Conference on Computer Vision*. 261–270.
[28] Thomas S Pagano and Rodney M Durham. 1993. Moderate resolution imaging spectroradiometer (MODIS). In *Sensor Systems for the Early Earth Observing System Platforms*, Vol. 1939. International Society for Optics and Photonics, 2–18.
[29] David P Roy, MA Wulder, Thomas R Loveland, CE Woodcock, RG Allen, MC Anderson, D Helder, JR Irons, DM Johnson, R Kennedy, et al. 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote sensing of Environment* 145 (2014), 154–172.
[30] Timothy J Schmit, Paul Griffith, Mathew M Gunshor, Jaime M Daniels, Steven J Goodman, and William J Lebair. 2017. A closer look at the ABI on the GOES-R series. *Bulletin of the American Meteorological Society* 98, 4 (2017), 681–698.
[31] Timothy J Schmit, SCOTT S Lindstrom, JORDAN J Gerth, and MATHEW M Gunshor. 2018. Applications of the 16 spectral bands on the Advanced Baseline Imager (ABI). (2018).
[32] Guang Xu and Xu Zhong. 2017. Real-time wildfire detection and tracking in Australia using geostationary satellite: Himawari-8. *Remote Sensing Letters* 8, 11 (2017), 1052–1061.
[33] Daiqin Yang, Zimeng Li, Yatong Xia, and Zhenzhong Chen. 2015. Remote sensing image super-resolution: Challenges and approaches. In *2015 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE, 196–200.
[34] Jun Yang, Zhiqing Zhang, Caiying Wei, Feng Lu, and Qiang Guo. 2017. Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4. *Bulletin of the American Meteorological Society* 98, 8 (2017), 1637–1658.
[35] Xiaolin Zhu, Jin Chen, Feng Gao, Xuehong Chen, and Jeffrey G Masek. 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment* 114, 11 (2010), 2610–2623.