Estimating Optical Flows in Satellite Imagery

Working Group Meeting for Space Lidar Winds

Presenter: **Thomas Vandal**
Contributions from: Ramakrishna Nemani

NASA Ames / BAERI

July 11 2019

NASA Earth Exchange

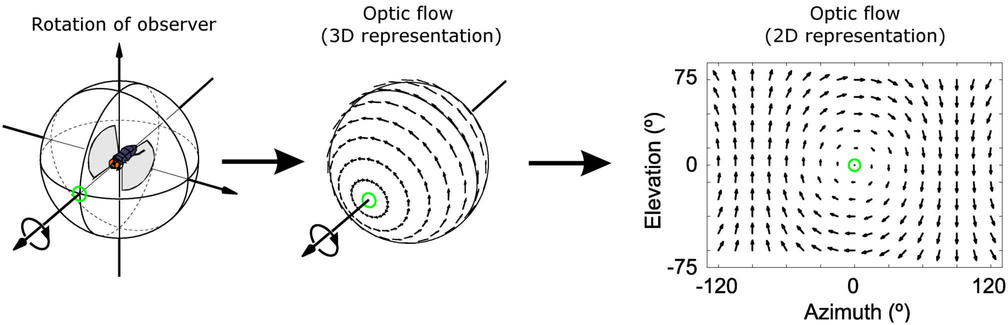
Roadmap

Introduction to Optical Flow

Introduction to Optical Flow

Optical Flow is the distribution of apparent velocities of movement of brightness patterns in an image. [Horn and Schunck, 1981]

Arises from the relative movement of an object from a viewer.



Optical Flow Assumptions

Let $E(x, y, t)$ denote the brightness of pixel (x, y) at time t in a sequence of 2D images.

Assume the brightness of a particular point is constant, $\frac{dE}{dt} = 0$

Using the chain rule,

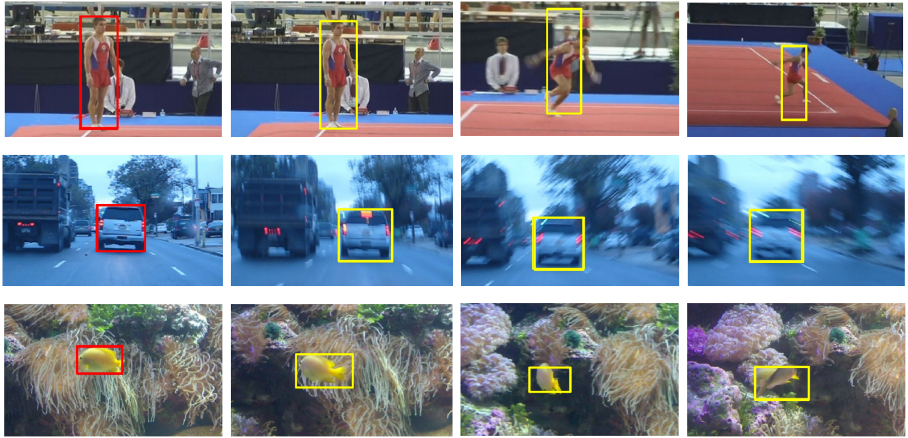
$$\frac{\partial E}{\partial x}\frac{dx}{dt} + \frac{\partial E}{\partial y}\frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \quad (1)$$

The goal is to estimate the velocities in the x and y directions: $\frac{dx}{dt}$ and $\frac{dy}{dt}$

Derivatives can be estimated using sequential images with numerical approximations.

Application I: Object Tracking [Bertinetto et al., 2016]

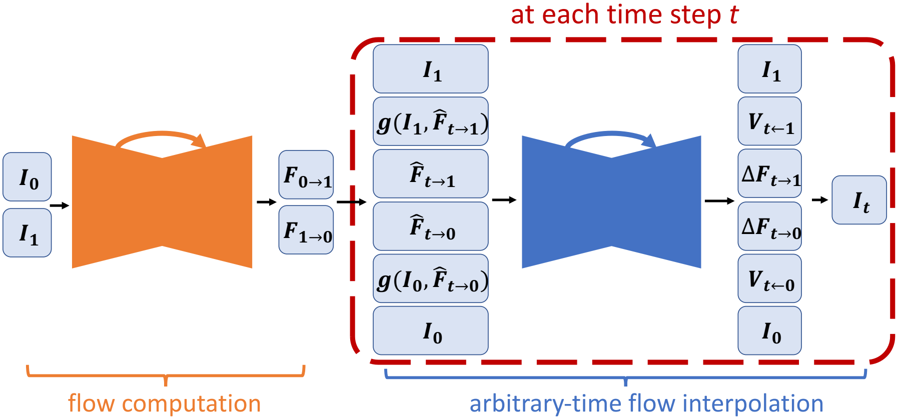
Goal: Tracking a moving object over a sequence of images. Optical flow is used to ensure the same object is tracked through multiple images.



Application II: Video Frame Interpolation [Jiang et al., 2018]

Link: `https://www.youtube.com/watch?v=MjViy6kyiqs`

Super SloMo



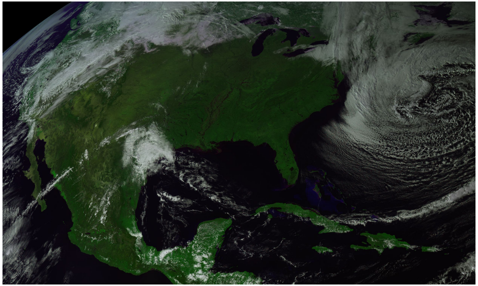
[Jiang et al., 2018]

Temporal Interpolation of Satellite Imagery

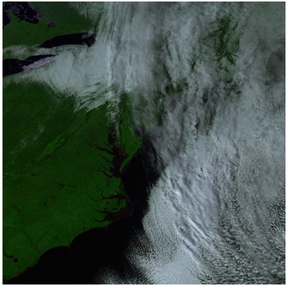
Spatial and Temporal Resolutions of GOES-16

Full Disk

Continental United States (CONUS)



Mesoscale



Spatial - 500 meters (visible), 1 km (near infrared), 2km (infrared)

Temporal - 10/15 min full disk, 5 minute CONUS, 30-60 second mesoscale

Temporal Interpolation of GOES-16

Mesoscale

- Flex mode allows the satellite to capture a user defined region to monitor major weather and environmental events.
- Coverage of 1000km by 1000km every 30 seconds (or two boxes every 1 minute).

Problem: Can we generate 1 minute full-disk coverage using machine learning and optical flow?

Approach: Apply the Super SloMo optical methodology by learning optical flows from mesoscale to interpolate between sequences of images.

Intermediate Frame Interpolation

Let $I_0, I_1, I_t \in \mathcal{R}^{H\times W\times C}$ such that $t \in (0, 1)$.
Construct an intermediate frame $\hat{I}_t$ from I_0 and I_1:

$$\hat{I}_t = \alpha \cdot g(I_0, F_{0\leftarrow t}) + (1-\alpha) \cdot g(I_1, F_{1\leftarrow t}) \tag{2}$$

$F_{0\leftarrow t}$ = Flow from I_t to I_0
$F_{1\leftarrow t}$ = Flow from I_t to I_1
g = backward warping function

Occlusion

Occlusion reasoning can be used to estimate the states of atmospheric variables over a static land surface by applying visbility maps, $V_{t\to 0}$ and $V_{t\to 1}$.

ie. For a given intermediate frame and pixel, is there cloud cover?

$$\hat{I}_t = \frac{1}{Z} \cdot \left((1-t) \cdot V_{t\to 0} \cdot g(I_0, F_{0\leftarrow t}) + t \cdot V_{t\to 1} \cdot g(I_1, F_{1\leftarrow t})\right) \quad (3)$$

where $Z = (1-t) \cdot V_{t\to 0} + t \cdot V_{t\to 1}$ is a normalization factor.

Deep neural networks are currently the state of the art for estimating $V_{t\to 0}, V_{t\to 1}, F_{0\leftarrow t}$, and $F_{0\leftarrow t}$.

Model Setup

Flow Network:

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

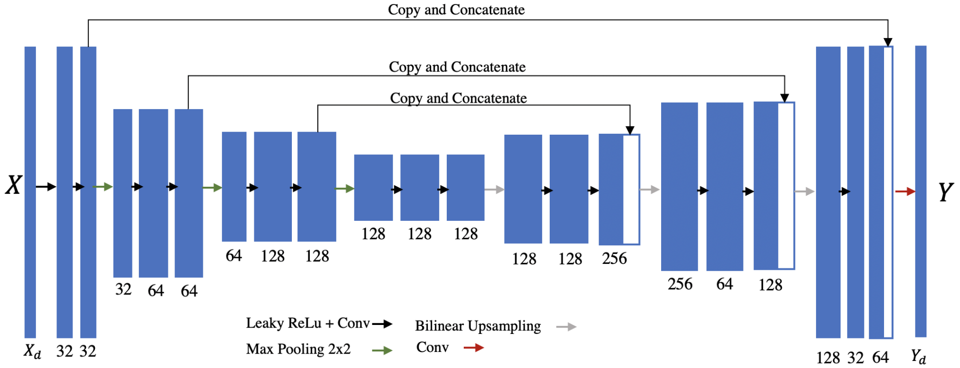
$$\begin{aligned} \hat{F}_{0\leftarrow t} &= -(1-t)tF_{0\leftarrow 1} + t^2 F_{1\leftarrow 0} \\ \hat{F}_{1\leftarrow t} &= (1-t)^2 F_{0\leftarrow 1} - t(1-t)F_{1\leftarrow 0} \end{aligned} \tag{5}$$

Interpolation Network:

$$V_{t\to 0}, V_{t\to 1}, \Delta\hat{F}_{0\leftarrow t}, \Delta\hat{F}_{1\leftarrow t} = H_{\text{Interp}}(I_0, I_1, \hat{F}_{0\leftarrow t}, \hat{F}_{1\leftarrow t}, g_0, g_1). \tag{6}$$

$$\begin{aligned} F_{0\leftarrow t} &= \hat{F}_{0\leftarrow t} + \Delta\hat{F}_{0\leftarrow t} \\ F_{1\leftarrow t} &= \hat{F}_{1\leftarrow t} + \Delta\hat{F}_{1\leftarrow t} \end{aligned} \tag{7}$$

Neural Network Architecture - UNet



Learning

Overall loss consists of a combination of reconstruction error, warping error, and smoothness regularization:

$$l = \lambda_r l_r + \lambda_w l_w + \lambda_s l_s. \tag{8}$$

Reconstruction loss is the euclidean distance between observed and predicted intermediate frames:

$$l_r = \frac{1}{N} \sum_{i=1}^{N} ||\hat{I}_{t_i} - I_{t_i}||_2. \tag{9}$$

Warping loss is used to optimize estimated optical flows between input and intermediate frames:

$$\begin{aligned} l_w = {} & ||I_0 - g(I_1, F_{0\to1})|| + ||I_1 - g(I_0, F_{1\to0})|| \\ & + \frac{1}{N} \sum_{i=1}^{N} ||I_{t_i} - g(I_0, F_{0\to t_i})||_2 + \frac{1}{N} \sum_{i=1}^{N} ||I_{t_i} - g(I_1, F_{1\to t_i})||_2 \end{aligned} \tag{10}$$

Smoothness loss ensures locally smooth flows:

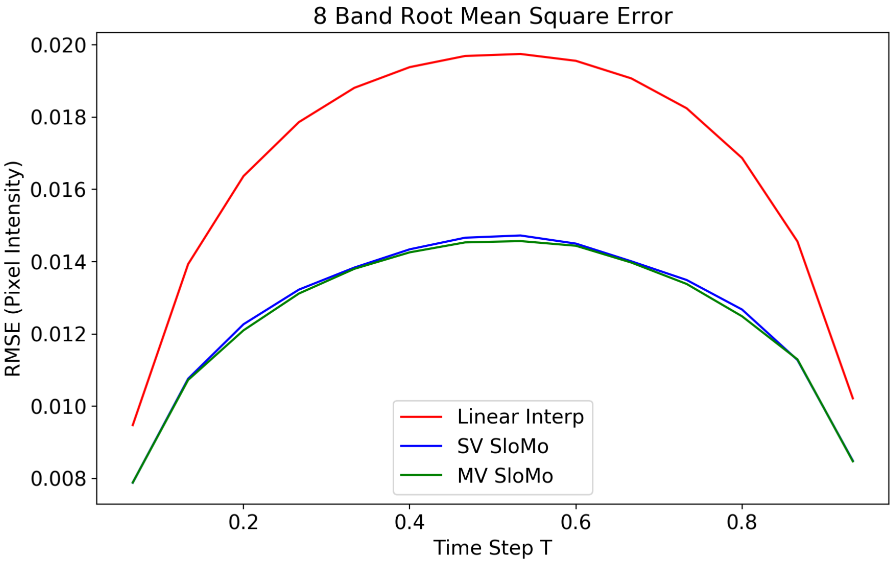
$$l_s = ||\Delta F_{0\to1}||_1 + ||\Delta F_{1\to0}||_1 \tag{11}$$

Results I

Table: Root mean square error (RMSE) over a held out test set of every 5 days in 2019 mesoscale data for a 15 minute temporal enhancement.

index	Linear	SV-SloMo	MV-SloMo	Linear	SV-SloMo	MV-SloMo
1	0.0232	**0.0172**	**0.0172**	0.0231	0.0171	**0.0170**
2	0.0329	**0.0260**	0.0261	0.0329	**0.0261**	**0.0261**
3	0.0288	**0.0218**	**0.0218**	0.0287	0.0217	**0.0216**
4	–	–	–	0.0095	0.0059	**0.0058**
5	–	–	–	0.0214	0.0168	**0.0167**
6	–	–	–	0.0137	0.0100	**0.0098**
7	–	–	–	0.0017	0.0012	**0.0011**
8	–	–	–	0.0026	0.0019	**0.0018**
3 Band Mean	0.0283	**0.0217**	**0.0217**	0.0282	**0.0216**	**0.0216**
8 Band Mean	–	–	–	0.0180	0.0136	**0.0135**

Results II


8 Band Root Mean Square Error
RMSE (Pixel Intensity)
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.2
0.4
0.6
0.8
Time Step T
Linear Interp
SV SloMo
MV SloMo

Video

Link: `https://www.youtube.com/watch?v=NeMXPQw3CJU`

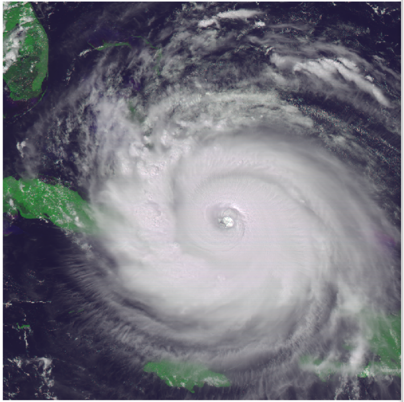
Extracting Pixel-wise Flow Vectors

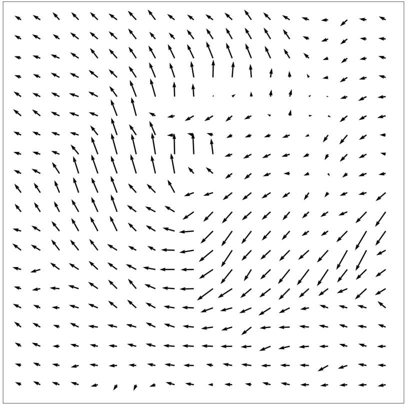
Flow Vectors

Flow vectors, $F_{0 \leftarrow t}$ and $F_{1 \leftarrow t}$, learned from the interpolation model has the following properties:

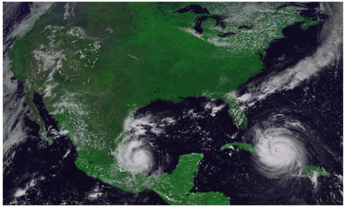
- Each F consists of u and v components representing horizontal and vertical velocities
- Direction is extracted from u and v
- Locally smooth vector magnitude and direction
- Each pixel is 2km, temporal period is 15 minutes $(u * 2/15 * 4 = km/hour)$

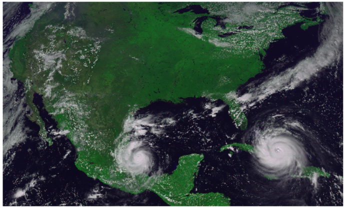
Mesoscale - Hurricane Irma - September 8 2017



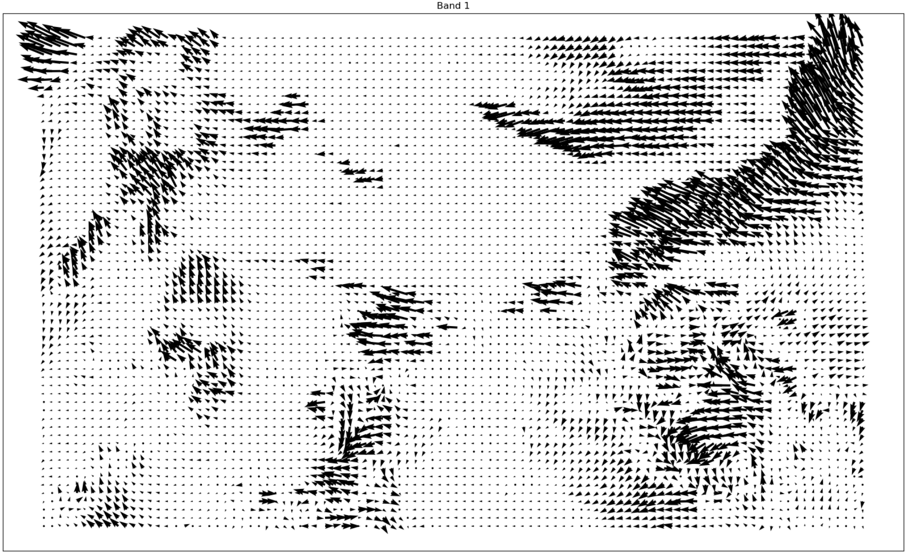


CONUS - September 8 2017

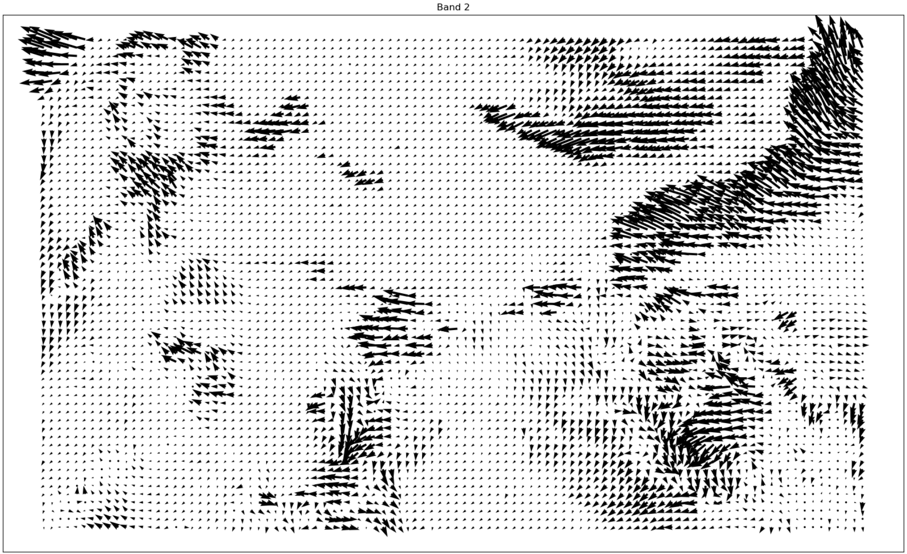




Band 1 - Visible - Blue

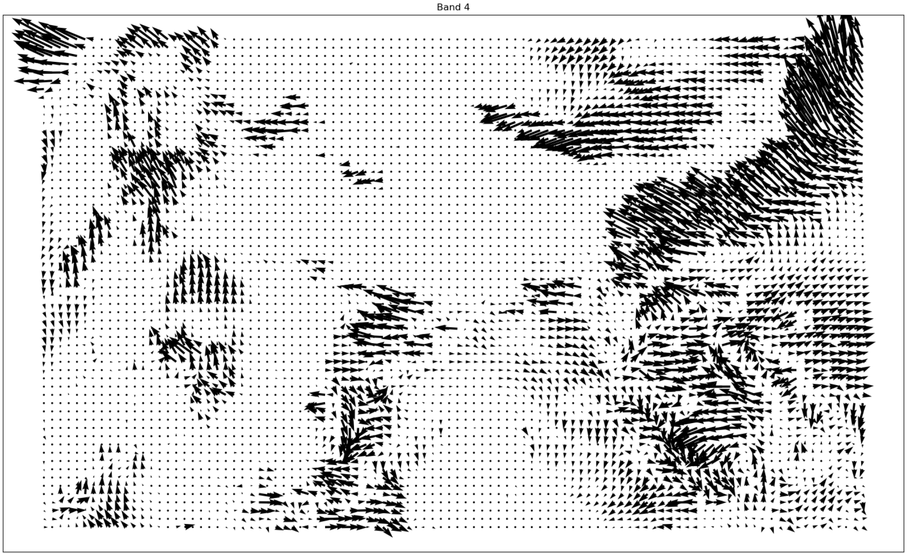


Band 2 - Visible - Red

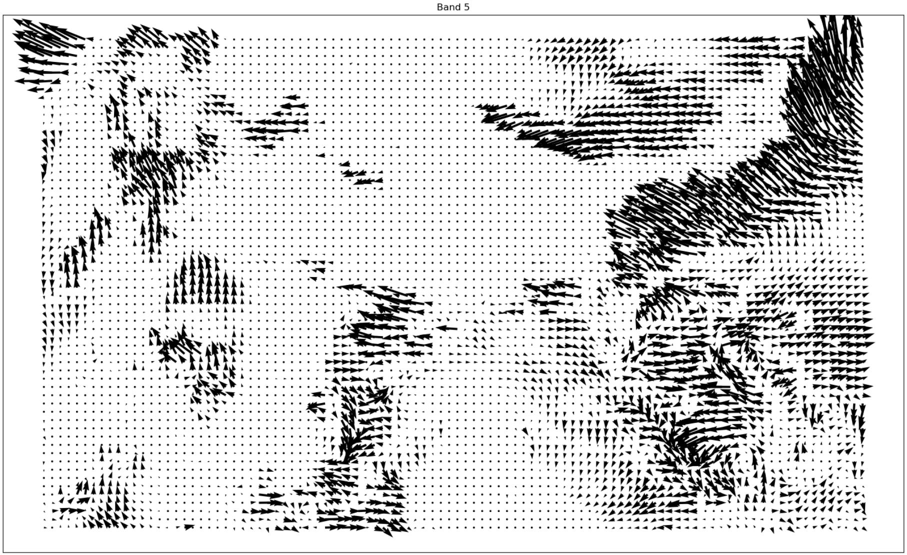


Band 3 - Near-IR

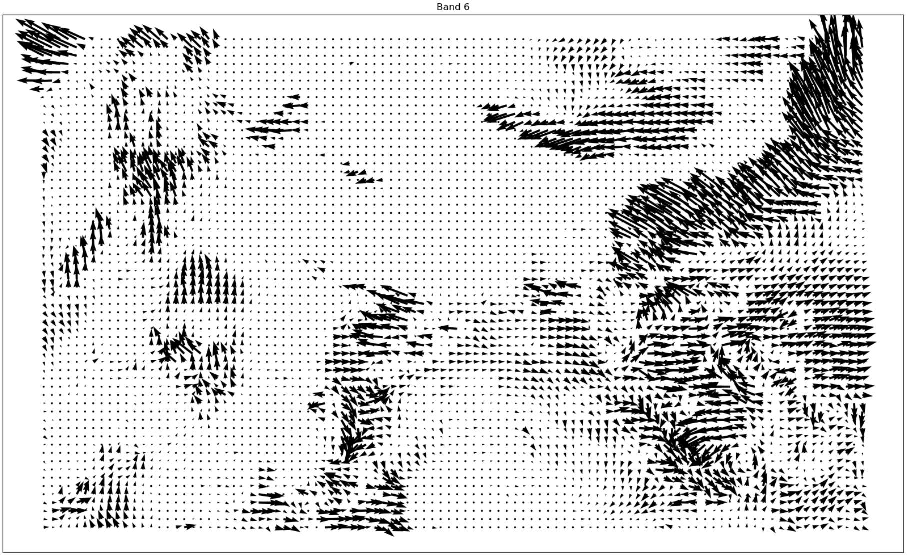
Band 4 - Near-IR



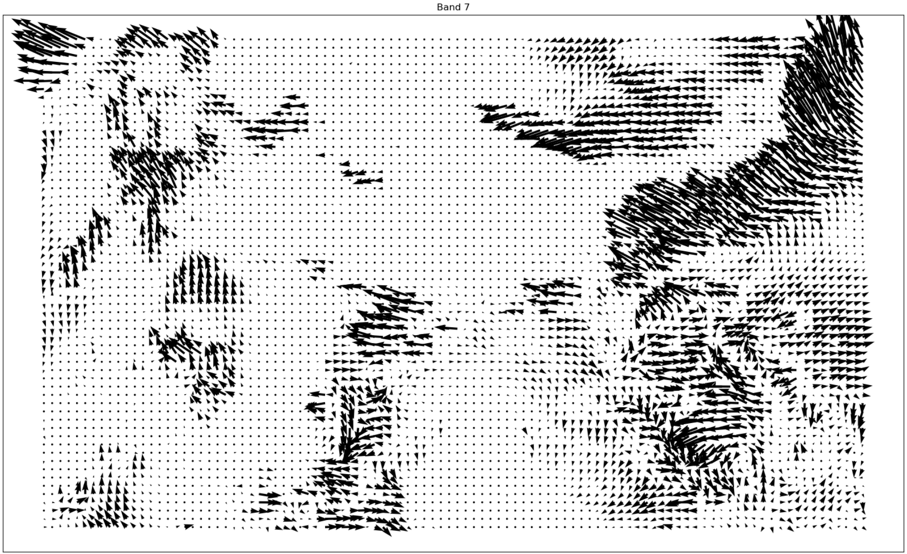
Band 5 - Near-IR



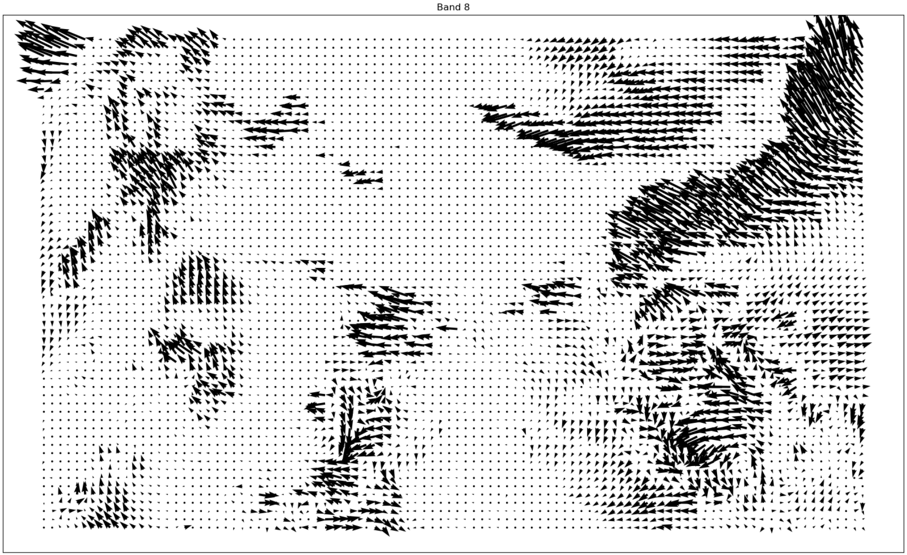
Band 6 - Near-IR



Band 7 - IR



Band 8 - Near-IR - Upper-Level Tropospheric Water Vapor



Conclusions

Conclusions

1. Key Points
   - Optical flows are numerically estimated with deep neural networks
   - Flow vectors are used to track the movement of objects in images, such as clouds
   - The intermediate frame interpolation approach can estimate flow vectors for any time in the domain
2. Next Steps
   - How are the flow vectors related to wind? What exactly is being captured?
   - Use ground truth data (Aeolus?) to understand the vectors
   - Simplify optical flow model
3. Optical flow applied to nowcasting of geostationary data is another promising direction.

References I

Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., and Torr, P. H. (2016).
Fully-convolutional siamese networks for object tracking.
In *European conference on computer vision*, pages 850–865. Springer.

Horn, B. K. and Schunck, B. G. (1981).
Determining optical flow.
Artificial intelligence, 17(1-3):185–203.

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*, pages 9000–9008.