

# Intermediate and Future Frame Prediction of Geostationary Satellite Imagery With Warp and Refine Network

Minseok Seo<sup>1</sup>, Yeji Choi<sup>1\*</sup>, Hyungon Ryu<sup>2</sup>, Heesun Park<sup>3</sup>, Hyungkun Bae<sup>1</sup>, Hyesook Lee<sup>3</sup>, Wanseok Seo<sup>2</sup>,

<sup>1</sup> SI Analytics, <sup>2</sup> NVIDIA, <sup>3</sup> National Institute of Meteorological Sciences (NIMS)  
\*yejichoi@si-analytics.ai

## Abstract

Geostationary satellite imagery has applications in climate and weather forecasting, planning natural energy resources, and predicting extreme weather events. For precise and accurate prediction, higher spatial and temporal resolution of geostationary satellite imagery is important. Although recent geostationary satellite resolution has improved, the long-term analysis of climate applications is limited to using multiple satellites from the past to the present due to the different resolutions. To solve this problem, we proposed warp and refine network (WR-Net). WR-Net is divided into an optical flow warp component and a warp image refinement component. We used the TV-L1 algorithm instead of deep learning-based approaches to extract the optical flow warp component. The deep-learning-based model is trained on the human-centric view of the RGB channel and does not work on geostationary satellites, which is gray-scale one-channel imagery. The refinement network refines the warped image through a multi-temporal fusion layer. We evaluated WR-Net by interpolation of temporal resolution at 4 min intervals to 2 min intervals in large-scale GK2A geostationary meteorological satellite imagery. Furthermore, we applied WR-Net to the future frame prediction task and showed that the explicit use of optical flow can help future frame prediction.

## Introduction

Weather prediction using geostationary satellite imagery indirectly contributes to the earth's environment in various fields such as planning natural energy resources (Pathak et al. 2022) and extreme weather event prediction. Physically-based numerical weather prediction models have traditionally been used to analyze and forecast weather and climate, but recent developments in deep learning-based models have made data-driven methods possible to use for weather prediction. In the era of the climate crisis, accurate weather prediction is essential to prepare the sudden disasters and to maximize the use of renewable energy such as wind and solar power. Therefore, the spatial and temporal resolution of forecasting results is important to appropriate countermeasures. Recently, a geostationary-satellite resolution has been improved with 1-2 minute intervals with a

pixel size of 1-2 km. However, to integrate multiple satellites from the past to the present for long-term analysis for climate applications, there is a limit to high-resolution analysis because most of the past geostationary satellites have a lower resolution of 10 to 15 minutes.

To solve this problem, Vandal and Nemani (Vandal and Nemani 2021) proposed a temporal interpolation of geostationary satellite imagery with optical flow approach based on Super SloMo (Jiang et al. 2018). In the Super SloMo based method, when images  $I_t$  and  $I_{t+1}$  are given, the interpolation frame  $I_i$  is predicted through optical flow estimation and warp between them. This method achieved state-of-the-art, but does not extract the optical flow properly when a sample that has not been learned in the training phase is input and is vulnerable to intensity changes such as brightness changes.

To solve this problem, video frame interpolation methods using pre-trained weights for PWC-Net (Sun et al. 2018) and FlowNet (Dosovitskiy et al. 2015) have been proposed. However, PWC-Net and FlowNet trained on RGB channels of human-centric view do not work in geostationary satellite imagery.

In this paper, we propose Warp and Refine Network (WR-Net). WR-Net is divided into optical flow estimation and warp component and refinement component. In the optical flow estimation and warp component, since deep learning-based supervised optical flow estimation methods cannot be used, we estimate and warp the optical flow through the total variation-L1 algorithm, a non-parametric approach. After that, the refinement network refines the intensity change through a multi-temporal fusion of the warp frame with  $I_t$  and  $I_{t+1}$  to refine the intensity change, such as the brightness change. We verified the WR-Net in large-scale GK2A geostationary meteorological satellite imagery, and as a result achieved state-of-the-art.

Furthermore, we extend WR-Net to future frame prediction. Our motivation is that optical flow includes information such as movement direction and speed, and this will be effective not only for video frame interpolation but also for future frame prediction. We predicted the frame after 90 min through WR-Net and evaluated it qualitatively. As a result of qualitative analysis, it was shown that the explicit use of optical flow is helpful for future frame prediction. Our contributions can be summarized as follows:

\*corresponding author

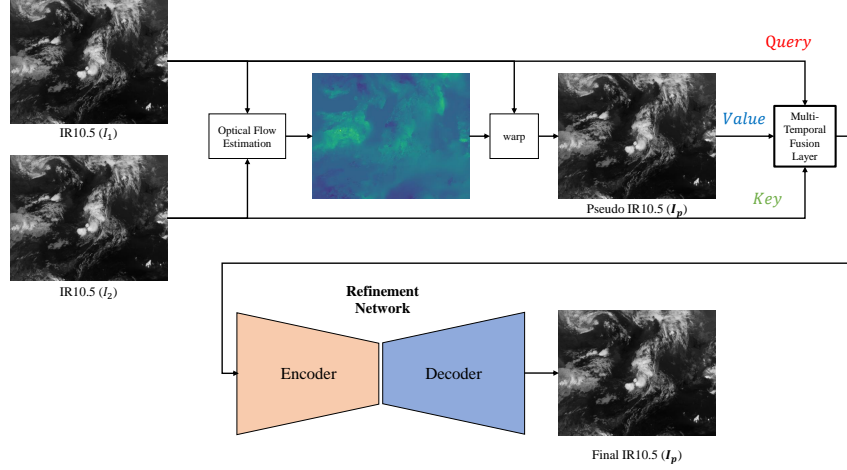


Figure 1: An overview of our proposed WR-Net. WR-Net consists of an optical flow warp component and a refinement component. The optical flow is extracted through the TV-L1 algorithm. The image warp with the extracted optical flow is refined through a multi-temporal fusion network and a refinement network.

- We propose WR-Net to extract optical flow using TV-L1 algorithm and warp and refine it.
- Our WR-Net achieves state-of-the-art in geostationary satellite imagery video frame interpolation.
- To the best of our knowledge, for the first time, optical flow was explicitly applied to deep learning-based weather forecasting.

## Method

In this section, we divide our proposed Warp-and-Refine Networks (WR-Net) into three components and describe them in detail. The first component, optical flow warp, estimates the optical flow between the  $I_t$  image and the  $I_{t+1}$  image, and warps the  $I_t$  or  $I_{t+1}$  image with the estimated optical flow to predict intermediate or future frames. Therefore, we propose and use a multi-temporal fusion layer to correct  $I_{inter}$ ,  $I_{future}$  by referring to the intensity values of highly correlated parts in  $I_t$  and  $I_{t+1}$  images. The multi-temporal fusion feature is input to the U-Net based Refine Network to refine the intensity value.

### Optical Flow Warp

The optical flow warp method is a recent learning approach in the field of video frame interpolation. However, as shown in Fig. 2-(a), the existing optical flow estimation network is designed and trained on the human-centric view of RGB images, so the performance of geostationary satellite imagery is greatly reduced. Therefore, we estimate the optical flow by using the GPU accelerated total variation-L1 (TVL1) (Wedel et al. 2009) algorithm as shown in Fig. 2-(b) instead of the deep learning-based optical flow estimate network.

When there is  $I_t$  frame and  $I_{t+1}$ , the intermediate frame  $I_i$  and future frame  $I_f$  can be synthesized through Eq. (1) and Eq. (2), respectively.

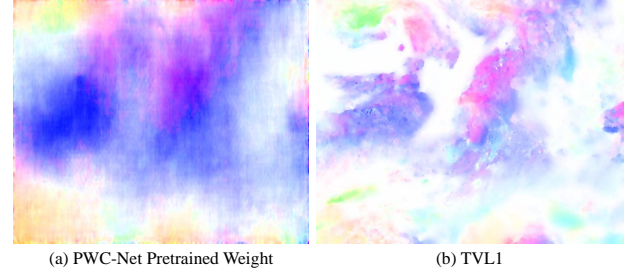


Figure 2: Qualitative comparison of optical flow extracted with PWC-Net and TV-L1 algorithms.

$$I_i = g(I_t, Op(I_{t+1}, I_t) \times \alpha), \quad (1)$$

$$I_f = g(I_{t+1}, Op(I_{t+1}, I_t) \times \alpha), \quad (2)$$

where  $Op(\cdot, \cdot)$  is a TV-L1 optical flow estimation algorithm. The  $g(\cdot)$  is a backward warping function. Parameter  $\alpha$  is the interval between time  $t$  and  $t+1$ . For example, if  $\alpha$  is set 0.5, an intermediate frame  $I_{t+0.5}$  can be generated, or a future frame  $I_{t+1.5}$  can be generated.

### Multi Temporal Fusion Network

The warp based on optical flow is not an optimal value because  $I_i$  or  $I_f$  are generated by linear interpolation of  $I_t$  or  $I_{t+1}$ . Therefore, we need to refine the pixel intensity values that cannot be expressed by linear interpolation. Therefore, we refine the intensity value of  $I_i$  or  $I_f$  using the information of  $I_t$  and  $I_{t+1}$ .

We propose a multi-temporal fusion layer based on Non-local Neural Networks to converge multi-temporal information. Figure 3 shows the structure of a multi-temporal fusion

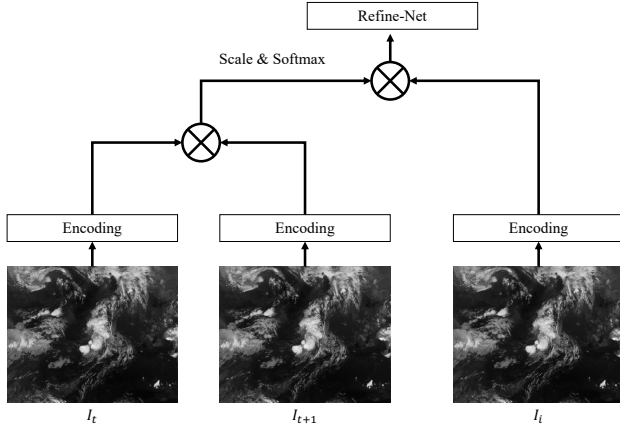


Figure 3: Overview of multi-temporal fusion layers. In video frame interpolation,  $I_t$  and  $I_{t+1}$  are input as query and key. In future frame prediction, replace  $I_i$  with the predicted future frame.

layer. As shown in the Fig. 3, the features with high correlation between  $I_t$  and  $I_{t+1}$  are weighted and multiplied by  $I_i$ . Through this process,  $I_i$  can be modeled for correlation between  $I_t$  and  $I_{t+1}$ .

### Refinement Network

Multi-temporal fusion features are input to U-Net (Ronneberger, Fischer, and Brox 2015) based Refinement Network. Refinement Network restores the input feature to the ground truth  $I_{igt}$  or  $I_{fgt}$ . When the input images  $I_t$  and  $I_{t+1}$  are given, the intermediate frame prediction’s objective function of Warp and Refine Network is as follows:

$$l_r = ||I_{igt} - Refine(MTF(I_{t+1}, I_t, I_i))||_1, \quad (3)$$

where  $MTF(., ., .)$  is a Multi temporal fusion network. The  $Refine(.)$  is a refinement network. For future frame prediction, simply change  $I_i$  to  $I_f$  in Eq. (3).

## Experiments

**Dataset** We trained and evaluated WR-Net on a large-scale GK2A geostationary satellite weather observation imagery to evaluate the intermediate and future frame prediction performance. The GK2A (GEO-KOMPSAT-2A) satellite dataset (Chung et al. 2020) from August 2020 to July 2021 is used for training and validation. The dataset is in 2-minute intervals over the East-Asia regions with 2 km spatial resolution. Since the GK2A dataset is very large-scale, only 1, 6, 11, 16, 21, and 26 days were selected and used.

We strictly followed the GK2A official user manual for pre-processing the physical value data of GK2A. The image size of the GK2A dataset is  $1950 \times 1550$ , and in our experiments, only channels of infrared ray (IR)  $10.5 \mu m$ , short wave infrared ray (SW)  $0.38 \mu m$ , and water vapor (WV)  $0.69 \mu m$  were used.

**Training** We set the time step parameter  $\tau$  to 0.25, the attachment weight parameter  $\lambda$  to 0.15, and the tightness pa-

Method	Channels	PSNR $\uparrow$	SSIM $\uparrow$
Linear	IR	38.667	0.745
SSM-T	IR	43.285	0.831
WR-Net (Warp Only)	IR	44.213	0.904
WR-Net (Full)	IR	<b>46.527</b>	<b>0.934</b>
Linear	SW	38.381	0.719
SSM-T	SW	45.935	0.827
WR-Net (Warp Only)	SW	48.237	0.922
WR-Net (Full)	SW	<b>50.538</b>	<b>0.936</b>
Linear	WV	43.526	0.766
SSM-T	WV	51.842	0.895
WR-Net (Warp Only)	WV	56.191	0.929
WR-Net (full)	WV	<b>58.350</b>	<b>0.973</b>

Table 1: Quantitative experimental results of video frame interpolation of WR-Net on the GK2A geostationary satellite weather observation dataset.

rameter  $\theta$  to 0.3 when using the optical flow with the TV-L1 algorithm. As for augmentation, we used random crop augmentation with a size of  $975 \times 775$  and random rotate  $90^\circ$  augmentation. Each input channel of the multi-modal fusion network is embedded with 512 channels, and the U-Net based on VGG16 is used as the refine network.

All experiments were performed on NVIDIA A100 GPU $\times 8$ , batch size was 64, adaptive moment estimation (Adam) was used as the optimizer, and the learning rate was set to  $1e-4$ . Note that the hyperparameters of the TV-L1 algorithm has a significantly effect on the performance, so it must be adjusted according to the dataset.

**Video frame interpolation results** The scale of weather tracking and modeling, ecosystem monitoring and climate change tracking depend on the spatial and temporal resolutions of observations.

However, when we use multiple satellites, the temporal resolution needs to be set to the past geostationary (GEO) satellite resolution at intervals of 10-15 min, although we have a higher resolution dataset with 1-2 min intervals. To solve this problem, we use WR-Net to interpolate the temporal resolution of geostationary (GEO) satellites and evaluate their performance through peak signal-to-noise ratio (PSNR) and structural similarity index map (SSIM).

We interpolated satellite images at 4min intervals in 2min increments and compared their performance with SSM-T (state-of-the-art method (Vandal and Nemani 2021)) and linear interpolation.

Table 1 shows the results of video frame interpolation experiments. As shown in Tab 1, WR-Net showed the highest performance in IR  $10.5 \mu m$ , SW  $0.38 \mu m$ , and WV  $0.69 \mu m$  channels. In addition, the performance of video frame interpolation by extracting the optical flow through the TV-L1 algorithm and simply warping the optical flow was higher than that of SSM-T. These experimental results indicate that the video frame interpolation method of superslomo series, which extracts and wraps the optical flow as an unsupervised learning method, is inefficient in the weather observation geostationary satellite.

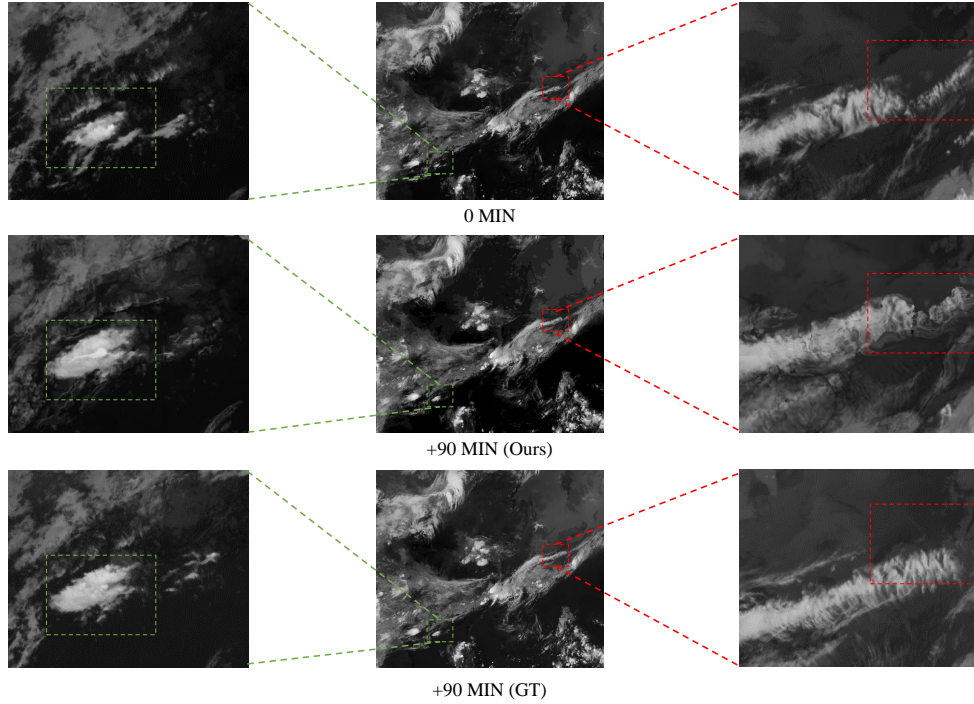


Figure 4: Qualitative result of future frame prediction of WR-Net. WR-Net receives 0min and +4min as input and predicts +8min, and inputs between +4min and predicted +8min again into WR-Net to predict +12min. Repeat this process to predict up to +90min.

Finally, the performance of using the refinement network was higher than the optical flow warping of WR-Net alone. These experimental results indicate that the refinement network operated significantly.

**Future frame prediction results** Future frame prediction is a very important field that can be used in various fields such as nowcasting, tracking temperature change, disaster forecasting, and climate change forecasting. According to this importance, deepmind’s deep generative model of rain (DGMR) (Ravuri et al. 2021) or FourCastNet has been proposed. All of these methods modeled multi-temporal information, but they were limited to either concatenating and inputting multi-temporal frames to the deep learning model or using LSTM.

Optical flow is a modality that includes movement direction and velocity between two input frames. Especially in the short term, the direction of movement is rarely reversed. Because of these characteristics, movement direction and speed are very important analysis factors in weather forecasting.

Therefore, we applied WR-Net to verify the usability of optical flow in the field of future frame prediction. Figure 4 is the qualitative result of predicting the frame after 90 minutes through WR-net.

As shown in the green box in Fig. 4, the cloud expansion direction and size were predicted similarly to the actual ground truth. These quantitative experimental results indicate that optical flow is very helpful in predicting cloud expansion and direction, and we confirm the possibility of pre-

dicting clouds generations with the refinement network (the green box in Fig. 4). However, in some cases, it has limitations to generate newly developed clouds (the red box in Fig. 4).

## Limitation and Future Work

Our study performed video frame interpolation or future frame prediction using only single channels with short timestamps. However, tasks such as weather forecasting and climate forecasting can be performed with high performance by fusion of information from various channels as well as a single channel. In our future work, we plan to study a method using multi-channel information. Also, we will verify WR-Net with tropical cyclone cases to check the generation and dissipation of cloud in the future study.

## Conclusion

We solved the problem that PWC-Net or FlowNet did not work in weather observation satellites through TV-L1 algorithm and proposed WR-Net to achieve state-of-the-art in the field of video frame interpolation. Also, as far as we know, it showed the possibility of using optical flow for the first time in the field of future frame prediction. Although our work has only been evaluated by weather observation satellites, we believe that our method has applications in a variety of fields, including climate change and disaster prediction. We hope that our research will have widespread applications in earth science as well as weather observations.

## References

- Chung, S.-R.; Ahn, M.-H.; Han, K.-S.; Lee, K.-T.; and Shin, D.-B. 2020. Meteorological products of geo-KOMPSAT 2A (GK2A) satellite.
- 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.
- 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.
- Pathak, J.; Subramanian, S.; Harrington, P.; Raja, S.; Chattopadhyay, A.; Mardani, M.; Kurth, T.; Hall, D.; Li, Z.; Azizzadenesheli, K.; et al. 2022. FourCastNet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*.
- Ravuri, S.; Lenc, K.; Willson, M.; Kangin, D.; Lam, R.; Mirowski, P.; Fitzsimons, M.; Athanassiadou, M.; Kashem, S.; Madge, S.; et al. 2021. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878): 672–677.
- Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.
- Sun, D.; Yang, X.; Liu, M.-Y.; and Kautz, J. 2018. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8934–8943.
- Vandal, T. J.; and Nemani, R. R. 2021. Temporal interpolation of geostationary satellite imagery with optical flow. *IEEE Transactions on Neural Networks and Learning Systems*.
- Wedel, A.; Pock, T.; Zach, C.; Bischof, H.; and Cremers, D. 2009. An improved algorithm for tv-l1 optical flow. In *Statistical and geometrical approaches to visual motion analysis*, 23–45. Springer.