Super-Resolution for Sparse Climate Data

Meshal Alharbi
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
meshal@mit.edu

Abstract

When considering the optimal placement of renewable energy resources, having access to high spatial-resolution data is critical for accurate assessment. In this project, we consider the task of video super-resolution on sparse wind speed data in the United States. We investigate the application of image super-resolution methods on this task and propose new modifications that quantitatively and qualitatively improve performance. To the best of our knowledge, this is the first work in the literature that considers video super-resolution on wind speed data at a high temporal resolution of 5 minutes.

1. Introduction

Climate data plays a crucial role in the planning and development of current and future renewable energy projects. Yet, there exists a disparity between the spatial resolution demanded by these initiatives and the data provided by existing climate models and satellite sensing [1]. For instance, tasks such as market analysis of wind power necessitate wind data at a spatial resolution of approximately 2 km, which is roughly 5 to 50 times higher than what is currently available through prevailing climate models [4].

This project aims to study the task of performing 4× Video Super-Resolution (VSR) on sparsely sampled wind speed data in the United States (Figure 1). Our main data source will be the Wind Integration National Dataset (WIND) provided by the National Renewable Energy Lab (NREL) [3]. To the best of our knowledge, this is the first work in the literature that considers the VSR task on this dataset.

Our goal is to investigate the performance of existing image super-resolution models on the VSR wind speed task. Furthermore, we aim to propose modifications to these algorithms that improve their qualitative and quantitative behavior. Ultimately, we hope to develop an approach that achieves sufficient accuracy while requiring reasonable computational resources.

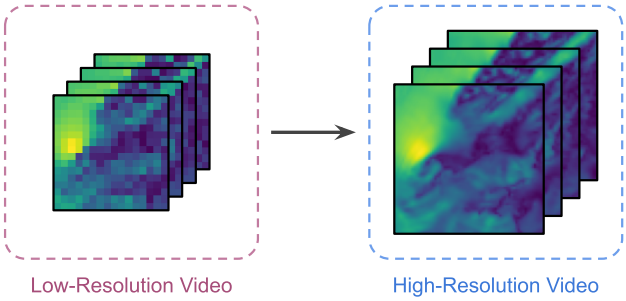

Figure 1. VSR aims to find a map from a low-resolution video space to a high-resolution video space. In this project, we consider the VSR task on sparse wind speed data.

This report is structured as follows. The remainder of this section is dedicated to surveying the most relevant literature. In Section 2, we formally describe the SR and VSR tasks. Section 3 introduce the data used in this project and the preprocessing steps used in data generation. Then, in Section 4, we present the baseline models, describe our proposed approach, and introduce the evaluation metrics used in this study. Section 5 is dedicated to experiments results and discussions. Finally, Section 6 conclude this report.

1.1. Related Literature

Super Resolution (SR) methods generate High-Resolution (HR) images or video from Low-Resolution (LR) inputs. Traditional signal-processing approaches for this task rely on interpolation and produce outputs that are too smooth (*e.g.*, bicubic interpolation) or too pixelated (*e.g.*, nearest neighbor interpolation) [12]. SR is challenging because the problem is ill-posed; several HR outputs can be valid for any given LR input.

In the context of machine learning, SR and VSR tasks have been formulated as a dictionary learning problem [6, 20] and as a regression problem [2, 5]. Further, training SR and VSR regression models with Generative Adversarial Networks (GANs) has been demonstrated to work well in practice [16, 17, 21]. More recently, diffusion-based

models have been used directly for SR and VSR tasks or to complement other more traditional methods [10, 13, 15]. See the work of Wang et al. [18] for a general survey on deep learning methods for SR and the work of Liu et al. [11] for a survey on deep learning methods for VSR.

Recently, multiple works in the literature have proposed the use of SR methods to enhance the spatial resolution of sparse climate data [8, 9]. Specifically, Kurinchi-Vendhan et al. [9] utilizes a GAN-based method to perform SR on wind speed and solar radiation data while Kumar et al. [8] utilizes a more traditional convolutional neural network to upscale sparse wind data. Both of these examples operate only on images and consider climate data that is spare in time (*e.g.*, hourly data). The data that we plan to use in this project will be at a 5-minute temporal resolution making it suitable for renewable energy market analysis [1].

2. Problem Statement

The goal of SR is to scale up a LR image $\mathbf{x} \in \mathbb{R}^{w \times h}$ to a HR image $\mathbf{y} \in \mathbb{R}^{w_\uparrow \times h_\uparrow}$, where $w_\uparrow = r \cdot w$ and $h_\uparrow = r \cdot h$ for a scale factor $r \in \mathbb{N}_+$. Note that we focus on single-channel images as they are the norm in climate data. A video of length $t \in \mathbb{N}_+$ consists of a sequence of t images (frames). For VSR, the goal is to map form a LR video $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_t) \in \mathbb{R}^{t \times w \times h}$ to a HR video $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_t) \in \mathbb{R}^{t \times w_\uparrow \times h_\uparrow}$.

We use $\mathcal{M} : \mathbb{R}^{w \times h} \to \mathbb{R}^{w_\uparrow \times h_\uparrow}$ to denote a SR model and $\mathcal{V} : \mathbb{R}^{t \times w \times h} \to \mathbb{R}^{t \times w_\uparrow \times h_\uparrow}$ to denote a VSR model. We consider parameterized models (*e.g.*, $\mathcal{M}(\cdot\,; \theta)$) and the goal is to find a set of parameters $\hat{\theta}$ that minimizes a loss function $\mathcal{L}$:

$$\hat{\theta}_{\mathcal{M}} = \arg\min_{\theta} \mathcal{L}_{\text{M}}(\mathcal{M}(\mathbf{x}; \theta), \mathbf{y}) \tag{1}$$

$$\hat{\theta}_{\mathcal{V}} = \arg\min_{\theta} \mathcal{L}_{\text{V}}(\mathcal{V}(\mathbf{X}; \theta), \mathbf{Y}) \tag{2}$$

where $\mathcal{L}_{\text{M}} : \mathbb{R}^{w_\uparrow \times h_\uparrow} \times \mathbb{R}^{w_\uparrow \times h_\uparrow} \to \mathbb{R}$ is a SR loss function and $\mathcal{L}_{\text{V}} : \mathbb{R}^{t \times w_\uparrow \times h_\uparrow} \times \mathbb{R}^{t \times w_\uparrow \times h_\uparrow} \to \mathbb{R}$ is VSR loss function. The choice of $\mathcal{L}_{\text{M}}$ and $\mathcal{L}_{\text{V}}$ heavily influences the behavior and quality of the SR and VSR models [7, 14]. For convenience, we use $N_{\mathbf{z}}$ to denote the cardinality of $\mathbf{z}$ and $\Omega_{\mathbf{z}}$ to denote the set of all valid indices in $\mathbf{z}$. For instance, if $\mathbf{z} \in \mathbb{R}^{w \times h}$, we have:

$$N_{\mathbf{z}} = w \cdot h \tag{3}$$

$$\Omega_{\mathbf{z}} = \{(i, j) \in \mathbb{N}_+^2 \,|\, i \leq w,\, j \leq h\} \tag{4}$$

3. Dataset

The WIND dataset covers a spatial grid of 2 km across the continental United States, with a temporal resolution of 5 minutes. It was compiled by integrating data from ground-based measurement stations, satellite sensing, and numerical weather prediction models. For further information on the dataset, refer to Draxl et al. [3]. Our project specifically focuses on the wind speed magnitude subset at a height of 100 meters above the Earth's surface.

3.1. Preprocessing

While a part of the dataset at hourly temporal resolution is preprocessed by NREL[1], this work requires a dataset at the 5 minutes temporal resolution for the VSR task. Hence, we conduct our own preprocessing to produce the data utilized herein. Specifically, we select 40 locations around major U.S. cities and extract data from four weeks in 2018 (two weeks from January and two from July). For each location, we create a uniform spatial grid and perform bilinear interpolation from the original spatial grid. The details of these steps and visualization of the data are provided in Appendix C and Appendix A respectively.

All in all, we generated 13,440 videos each consisting of 48 frames at 64×64 pixel resolution. The total size of this dataset is 20 GB. We consider this dataset our HR outputs and the LR inputs are obtained by downsampling by a factor of 4 (*i.e.*, an LR video will consist of 48 frames at 16×16 pixel resolution).

4. Methodology

4.1. Network Architecture

We use the Residual-in-Residual Dense Block (RRDB) architecture proposed in [16, 17]. RRDB consists of a series of convolutional and non-linear layers (without batch normalization) with dense connections (*i.e.*, concatenation) in between. Therefore, the number of channels in a given RRDB grows linearly with depth, allowing the block to capture complex features and patterns. However, the final layer in an RRDB maps back to the original number of channels in the input layer of the block. A network of this architecture consists of multiple RRDBs with skip connections in between operating in the LR space followed by shallow upsampling layers.

4.2. Models

As baselines in our study, first, we consider bicubic interpolation $\mathcal{I}_{\text{Bicubic}}$, which fits a bicubic function to a 4×4 grid surrounding each target pixel to interpolate and uses the function value as an estimate. Second, we adapt the image SR architecture used in [17] and apply it to videos. Using our notations, let $\mathcal{F}_{\text{RRDB}}$ be the RRDB network architecture described above, this baseline will be given by:

$$\mathcal{V}_{\text{Baseline}}(\mathbf{X}; \theta) = (\mathcal{F}_{\text{RRDB}}(\mathbf{x}_1; \theta), \ldots, \mathcal{F}_{\text{RRDB}}(\mathbf{x}_t; \theta)) \tag{5}$$

[1] https://github.com/NREL/hsds-examples/blob/master/datasets/wtk-us.md

	$\mathcal{V}_{\text{Baseline}}$	$\mathcal{V}_{\text{Proposed}}$
# of RRDB	8	
Maximum # Channels	192	640
Batch Size[a]	16	64
Optimizer	AdamW	
Learning Rate	1×10^{-4}	

[a] Measured in numbers of videos.

Table 1. Hyperparameters for each model.

We propose two modifications to this baseline; Firstly, utilizing the fact that bulks of RRDB computations are done on the LR space, we propose treating the videos $\mathbf{X}$ as multi-channel images and using an RRDB network to map directly from the LR video space to the HR video space. Our hypothesis is that this approach would alleviate the ill-posed nature of the problem. This is because the data in this project originate from physical models that adhere to conservation laws, and information from future frames exhibits a high correlation with information from historical frames. Secondly, instead of mapping directly to the HR video space, we propose learning a residual map on top of bilinear interpolation $\mathcal{I}_{\text{Bilinear}}$. Therefore, the model is optimized to learn high-frequency details that are absent in the interpolation. In notation, both of these two modification translate to:

$$\mathcal{V}_{\text{Proposed}}(\mathbf{X};\theta) = \mathcal{F}_{\text{RRDB}}(\mathbf{X};\theta) + \mathcal{I}_{\text{Bilinear}}(\mathbf{X}) \tag{6}$$

Finally, we mention that SR and VSR models commonly undergo training utilizing adversarial losses, along with data augmentation of multiple degradations in the process of downsampling HR outputs to LR inputs. However, owing to computational limitations, we opt for standard L1 and L2 loss objectives in our training and we also skip the degradation augmentations.

4.3. Evaluation Metrics

To evaluate the performance of the VSR models, we consider a set of metrics commonly used in the SR and VSR literature. Let $\hat{\mathbf{Y}}$ be an approximation of the ground truth $\mathbf{Y}$. First, we have the typical Mean Squared Error (MSE) and Mean Absolute Error (MAE) metrics:

$$\text{MSE}(\hat{\mathbf{Y}}, \mathbf{Y}) \coloneqq \frac{1}{N_{\mathbf{Y}}} \sum_{p \in \Omega_{\mathbf{Y}}} (\hat{\mathbf{Y}}_p - \mathbf{Y}_p)^2 \tag{7}$$

$$\text{MAE}(\hat{\mathbf{Y}}, \mathbf{Y}) \coloneqq \frac{1}{N_{\mathbf{Y}}} \sum_{p \in \Omega_{\mathbf{Y}}} |\hat{\mathbf{Y}}_p - \mathbf{Y}_p| \tag{8}$$

Furthermore, we consider the Peak Signal-to-Noise Ratio (PSNR), which is the ratio between the dynamic range L of the ground truth $\mathbf{Y}$ and MSE between $\hat{\mathbf{Y}}$ and $\mathbf{Y}$ measured on a log-scale. Formally:

$$\text{PSNR}(\hat{\mathbf{Y}}, \mathbf{Y}) \coloneqq 10 \log_{10} \frac{L^2}{\text{MAE}(\hat{\mathbf{Y}}, \mathbf{Y})} \tag{9}$$

MSE, MAE, and PSNR are considered pixel-wise evaluation metrics, and they might not always correlate well with the perceived visual quality of images and videos [19]. A metric with a better correlation to human quality assessments is the Structural Similarity Index (SSIM). SSIM measures degradation in structural information (*e.g.*, luminance and contrast) over a window of an image or video. If u is a window from $\hat{\mathbf{Y}}$ and v is a window from $\mathbf{Y}$ with common size and location, then the SSIM between u and v is given by:

$$\text{SSIM}(u, v) \coloneqq \frac{(2\mu_u\mu_v + c_1)(2\sigma_{uv} + c_2)}{(\mu_u^2 + \mu_v^2 + c_1)(\sigma_u^2 + \sigma_v^2 + c_2)} \tag{10}$$

where μ and σ are the typical mean and standard deviation operators and c_1 and c_2 are small constants to avoid instability. The final SSIM between $\hat{\mathbf{Y}}$ and $\mathbf{Y}$ is obtained by averaging all the windows that segment the $\hat{\mathbf{Y}}$ and $\mathbf{Y}$.

5. Experimental Results

5.1. Training Details

We utilize 80% (10,752 videos) of our dataset for training and validation and 20% (2,688 videos) for testing. In Table 1, we list the hyperparameters used in the training of each model. We note that although our proposed method has a larger number of features, its input is larger (by a factor of 48) than the input of the baseline. With these hyperparameters, both models utilized similar VRAM (around 14 GB) during training. The training was conducted using an NVIDIA Tesla V100 GPU. On average, the baseline model required 112 minutes to train, whereas our proposed method trains in 28 minutes.

5.2. Results and Discussions

We summarize the performance on the test dataset for each model in Table 2. Each entry in the table is an average of four different seeds. For the loss function $\mathcal{L}_{\text{V}}$, each model is trained with L1 and L2 objectives. As we can see from the table, our proposed model with the L1 objective outperforms the baseline and the bicubic interpolation in all metrics. Moreover, for our proposed approach, we notice that the model trained with the L1 objective obtains lower L2 error than the model trained with the L2 objective. We hypothesize that the gradient vanishes too quickly for small pixel-wise errors when we train with the L2 objective. Further, this effect is exacerbated because we are trying to learn a residual model on top of bilinear interpolation.

	$\mathcal{I}_{\text{Bilinear}}$	$\mathcal{V}_{\text{Baseline}}$		$\mathcal{V}_{\text{Proposed}}$	
		L1 Loss	L2 Loss	L1 Loss	L2 Loss
MSE ↓	0.1332 ± 0	$0.0935 \pm 1.2 \times 10^{-4}$	$0.0929 \pm 3.6 \times 10^{-4}$	$\mathbf{0.0870} \pm 1.1 \times 10^{-4}$	$0.0894 \pm 2.9 \times 10^{-4}$
MAE ↓	0.2298 ± 0	$0.1866 \pm 1.5 \times 10^{-4}$	$0.1892 \pm 4.1 \times 10^{-4}$	$\mathbf{0.1827} \pm 1.8 \times 10^{-5}$	$0.1883 \pm 3.8 \times 10^{-4}$
PSNR ↑	26.590 ± 0	$28.172 \pm 4.6 \times 10^{-3}$	$28.150 \pm 1.5 \times 10^{-2}$	$\mathbf{28.486} \pm 2.8 \times 10^{-3}$	$28.320 \pm 1.0 \times 10^{-2}$
SSIM ↑	0.6722 ± 0	$0.7404 \pm 2.8 \times 10^{-4}$	$0.7358 \pm 1.4 \times 10^{-3}$	$\mathbf{0.7534} \pm 2.7 \times 10^{-4}$	$0.7428 \pm 5.1 \times 10^{-4}$

Table 2. Models performance on the test data set. Each figure is an average of four independent runs and $\pm$ shows the standard deviation between the runs.

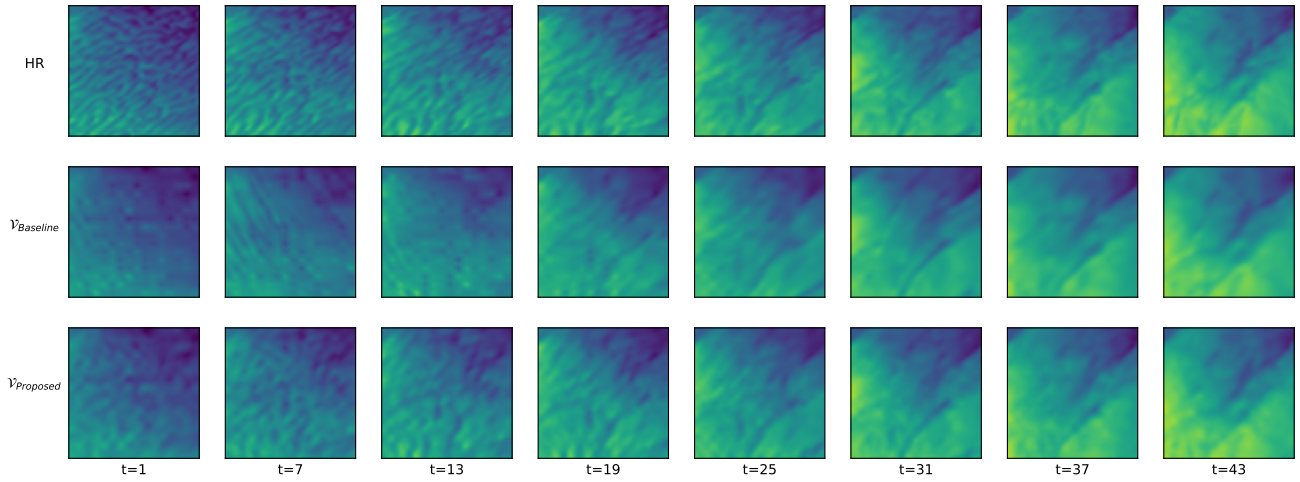

Figure 2. Example of the models outputs on the test dataset. Our proposed model achieves better temporal coherence on the small wind turbulence between frames 1 and 19.

In Figure 2, we showcase an example of the models outputs on the test dataset. In this example, we can notice that our proposed model achieves better temporal coherence on the small wind turbulence. Specifically, the direction of the high-frequency turbulence flips between consecutive frames in the baseline while it remains consistent in our proposed model. We believe as our proposed model has access to multiple LR frames, it is able to correctly identify the direction of the small turbulence in this example. The full figure for this example is provided in Appendix B along with other examples with similar behavior (Figure 5 to Figure 7).

Moreover, we showcase other examples in Appendix B that demonstrate the overall sharpness of each model (Figure 8 to Figure 10). From these examples, we notice that our proposed model recovers sharper details than the baseline. We believe this is likely due to the larger frame buffer that the proposed model has, as details that might be obscured at the current LR frame might be revealed in historical or future frames.

Finally, we note that the models developed here can be applied to larger spatial areas than 16×16 pixels. For instance, to perform 64×64 to 256×256 super-resolution, we can segment the 64×64 input into four segments each of size 16×16, and pass them individually to the models. However, to ensure coherence on the boundaries, we might need to overlap the segments. We can use a similar idea to extend the temporal axis.

6. Conclusions

In this project, we investigated video super-resolution on sparse wind speed data in the United States. We have shown that a larger frame buffer in the regression mapping helps improve both temporal coherence and sharpness in this task. Further, our results indicate that models trained with L1 loss appear to behave quantitatively and qualitatively better than models trained with L2 loss. Future work includes considering other loss functions and/or other models architecture such as diffusion models.

References

[1] Sarah L Cox, Anthony J Lopez, Andrea C Watson, Nicholas W Grue, and Jennifer E Leisch. Renewable Energy Data, Analysis, and Decisions: A Guide for Practitioners. Technical report, National Renewable Energy Lab.(NREL),

Golden, CO (United States), 2018. URL https://www.osti.gov/biblio/1427970.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image Super-Resolution Using Deep Convolutional Networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. URL https://ieeexplore.ieee.org/abstract/document/7115171.

[3] Caroline Draxl, Andrew Clifton, Bri-Mathias Hodge, and Jim McCaa. The Wind Integration National Dataset (WIND) Toolkit. *Applied Energy*, 151:355–366, 2015. URL https://doi.org/10.1016/j.apenergy.2015.03.121.

[4] Filippo Giorgi. Thirty Years of Regional Climate Modeling: Where Are We and Where Are We Going Next? *Journal of Geophysical Research: Atmospheres*, 124(11):5696–5723, 2019. URL https://doi.org/10.1029/2018JD030094.

[5] Xiaodan Hu, Mohamed A Naiel, Alexander Wong, Mark Lamm, and Paul Fieguth. RUNet: A Robust UNet Architecture for Image Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. URL https://ieeexplore.ieee.org/document/9025499.

[6] Xiao-Yuan Jing, Xiaoke Zhu, Fei Wu, Xinge You, Qinglong Liu, Dong Yue, Ruimin Hu, and Baowen Xu. Super-Resolution Person Re-Identification With Semi-Coupled Low-Rank Discriminant Dictionary Learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 695–704, 2015. URL https://ieeexplore.ieee.org/document/7298669.

[7] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016. URL https://arxiv.org/abs/1603.08155.

[8] Ashutosh Kumar, Tanvir Islam, Jue Ma, Takehiro Kashiyama, Yoshihide Sekimoto, and Chris Mattmann. WindSR: Improving Spatial Resolution of Satellite Wind Speed through Super-Resolution. *IEEE Access*, 2023. URL https://ieeexplore.ieee.org/abstract/document/10174644.

[9] Rupa Kurinchi-Vendhan, Björn Lütjens, Ritwik Gupta, Lucien Werner, and Dava Newman. WiSoSuper: Benchmarking Super-Resolution Methods on Wind and Solar Data. *arXiv preprint arXiv:2109.08770*, 2021. URL https://arxiv.org/abs/2109.08770.

[10] Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. Srdiff: Single Image Super-Resolution with Diffusion Probabilistic Models. *Neurocomputing*, 479:47–59, 2022. URL https://doi.org/10.1016/j.neucom.2022.01.029.

[11] Hongying Liu, Zhubo Ruan, Peng Zhao, Chao Dong, Fanhua Shang, Yuanyuan Liu, Linlin Yang, and Radu Timofte. Video Super-Resolution Based on Deep Learning: A Comprehensive Survey. *Artificial Intelligence Review*, 55(8):5981–6035, 2022. URL https://doi.org/10.1007/s10462-022-10147-y.

[12] Brian B Moser, Federico Raue, Stanislav Frolov, Sebastian Palacio, Jörn Hees, and Andreas Dengel. Hitchhiker's Guide to Super-Resolution: Introduction and Recent Advances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. URL https://ieeexplore.ieee.org/abstract/document/10041995.

[13] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image Super-Resolution via Iterative Refinement. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):4713–4726, 2022. URL https://ieeexplore.ieee.org/abstract/document/9887996.

[14] Sadaf Salehkalaibar, Truong Buu Phan, Jun Chen, Wei Yu, and Ashish Khisti. On the Choice of Perception Loss Function for Learned Video Compression. *Advances in Neural Information Processing Systems*, 36, 2024. URL https://arxiv.org/abs/2305.19301.

[15] Shuyao Shang, Zhengyang Shan, Guangxing Liu, LunQian Wang, XingHua Wang, Zekai Zhang, and Jinglin Zhang. ResDiff: Combining CNN and Diffusion Model for Image Super-Resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8975–8983, 2024. URL https://arxiv.org/abs/2303.08714.

[16] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018. URL https://arxiv.org/abs/1809.00219.

[17] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. URL https://arxiv.org/abs/2107.10833.

[18] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep Learning for Image Super-Resolution: A Survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3365–3387, 2020. URL https://ieeexplore.ieee.org/abstract/document/9044873.

[19] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. URL https://ieeexplore.ieee.org/document/1284395.

[20] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image Super-Resolution Via Sparse Representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. URL https://ieeexplore.ieee.org/document/5466111.

[21] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. RankSRGAN: Generative Adversarial Networks With Ranker for Image Super-Resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3096–3105, 2019. URL https://arxiv.org/abs/1908.06382.

Super-Resolution for Sparse Climate Data

Supplementary Material

A. Data Visualization

In this section, we present examples from the dataset utilized in this project. Figure 3 displays relatively straightforward cases where the HR data exhibits simple or smooth patterns. On the other hand, Figure 4 illustrates more intricate examples, characterized by HR data containing high-frequency spatial details or patterns transitioning rapidly between consecutive frames.

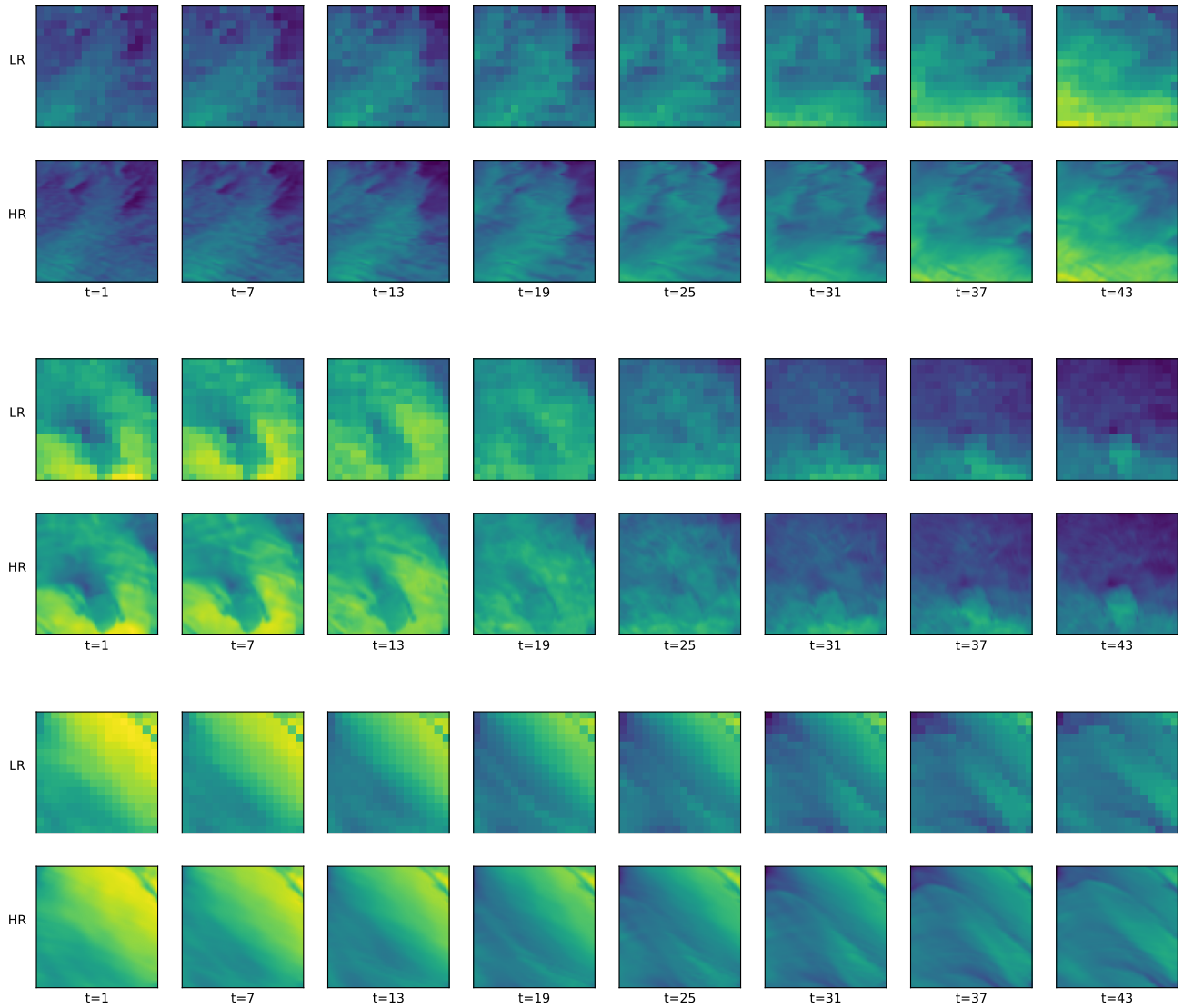

Figure 3. Examples where the HR data exhibits simple or smooth patterns.

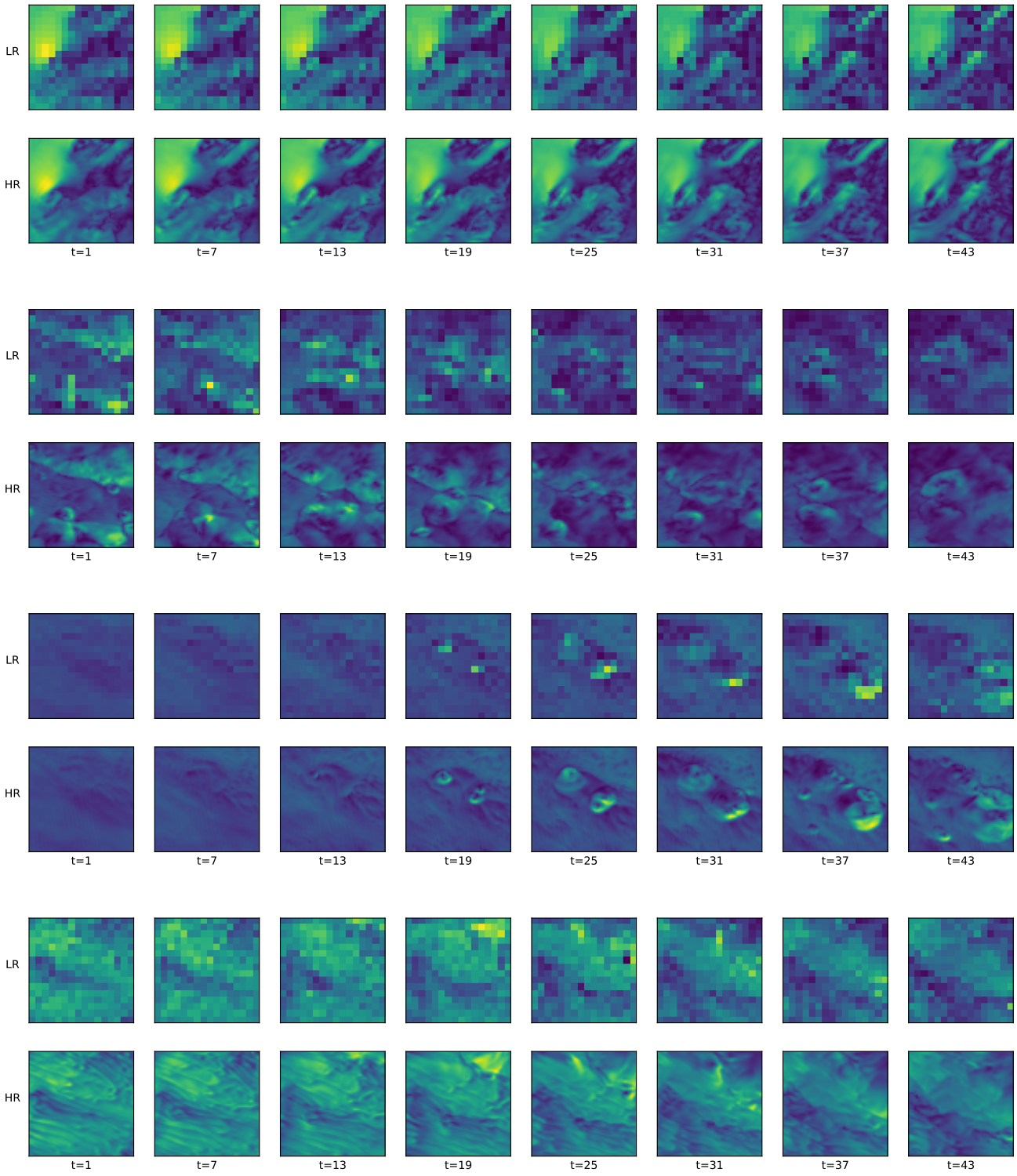

Figure 4. Examples where the HR data contains high-frequency spatial and temporal details.

B. Visualization of Models Outputs

In this section, we provide multiple examples of the models outputs on the test dataset.

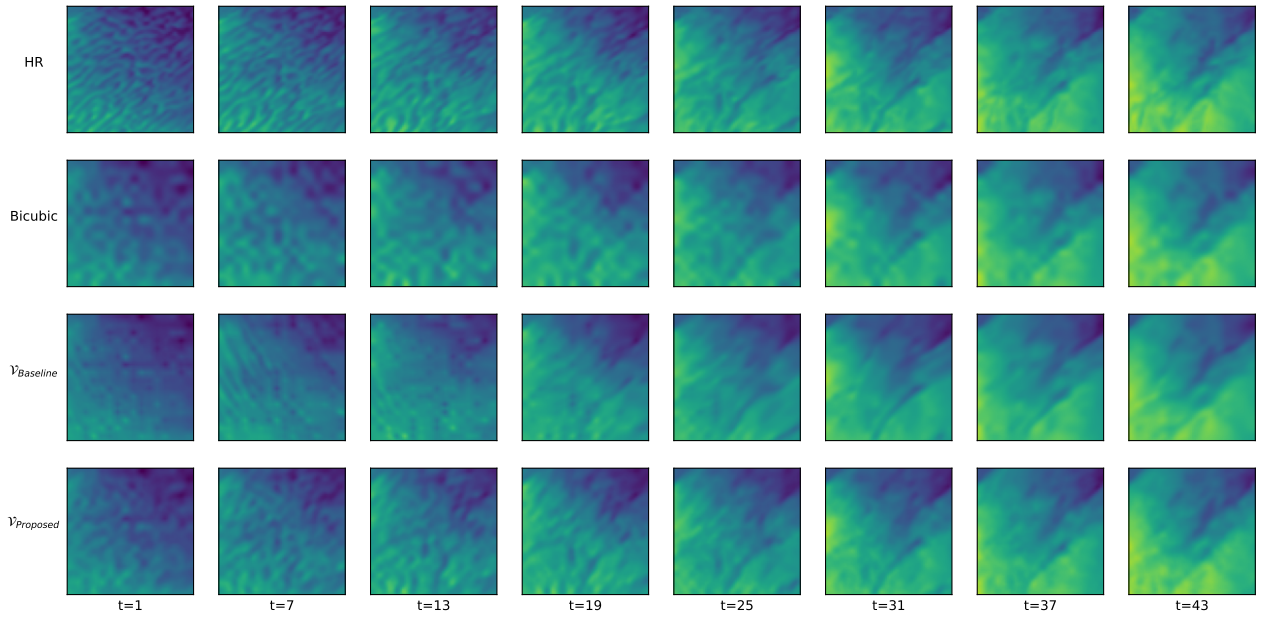

Figure 5. Example 1 for temporal coherence.

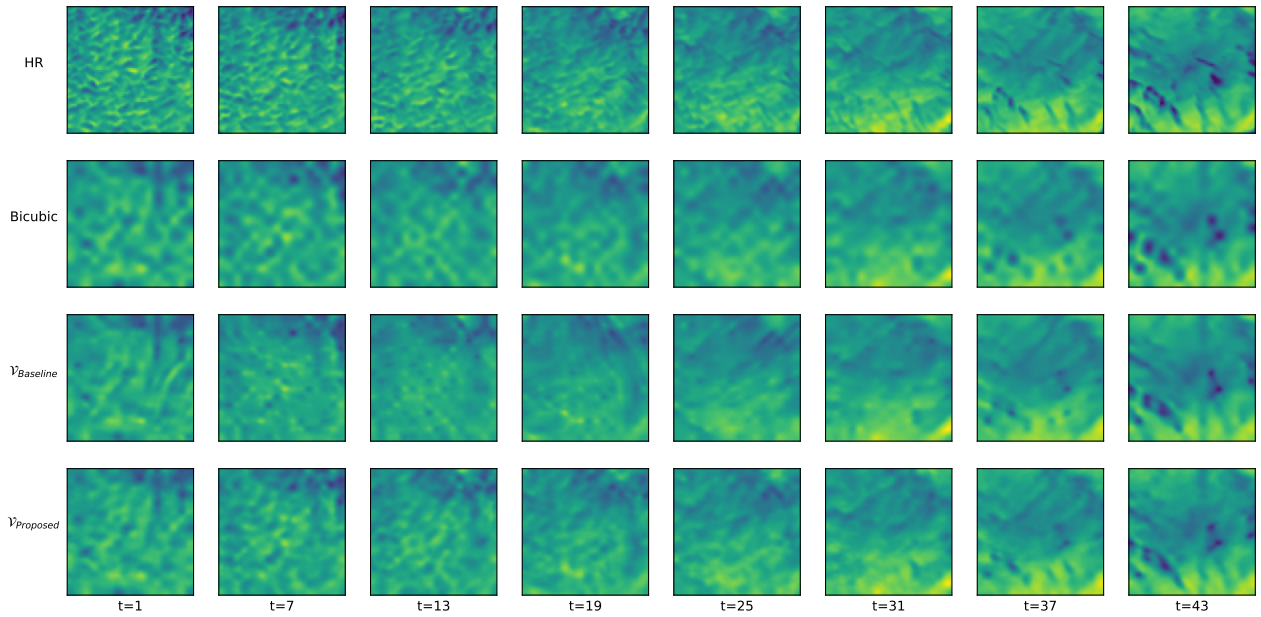

Figure 6. Example 2 for temporal coherence.

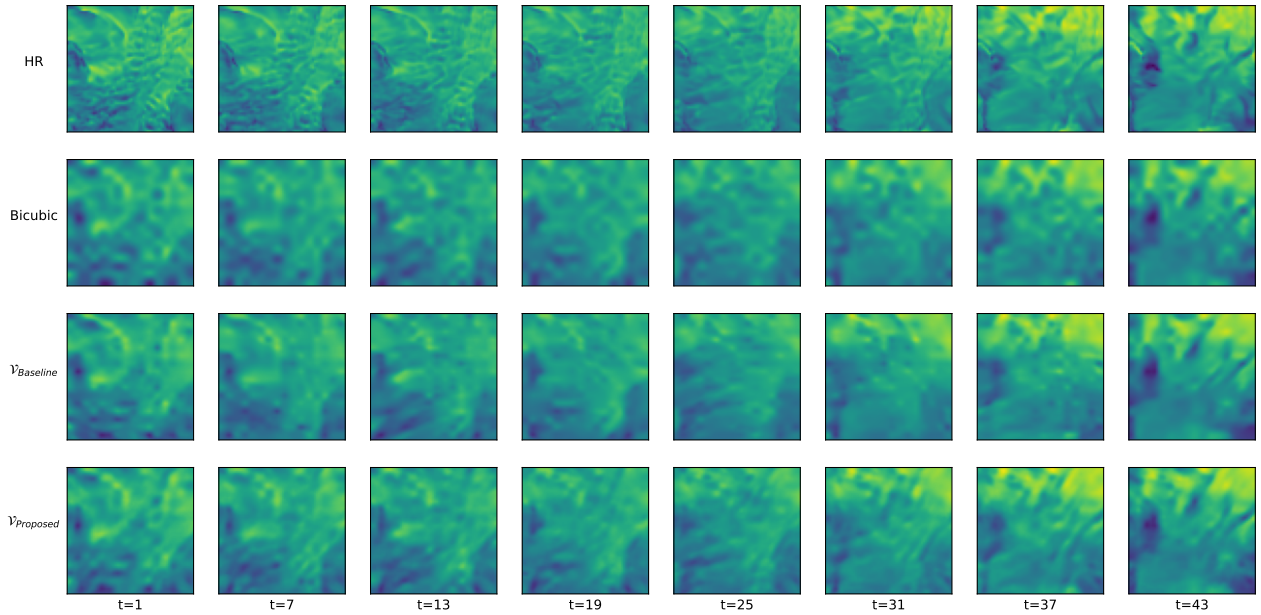

Figure 7. Example 3 for temporal coherence.

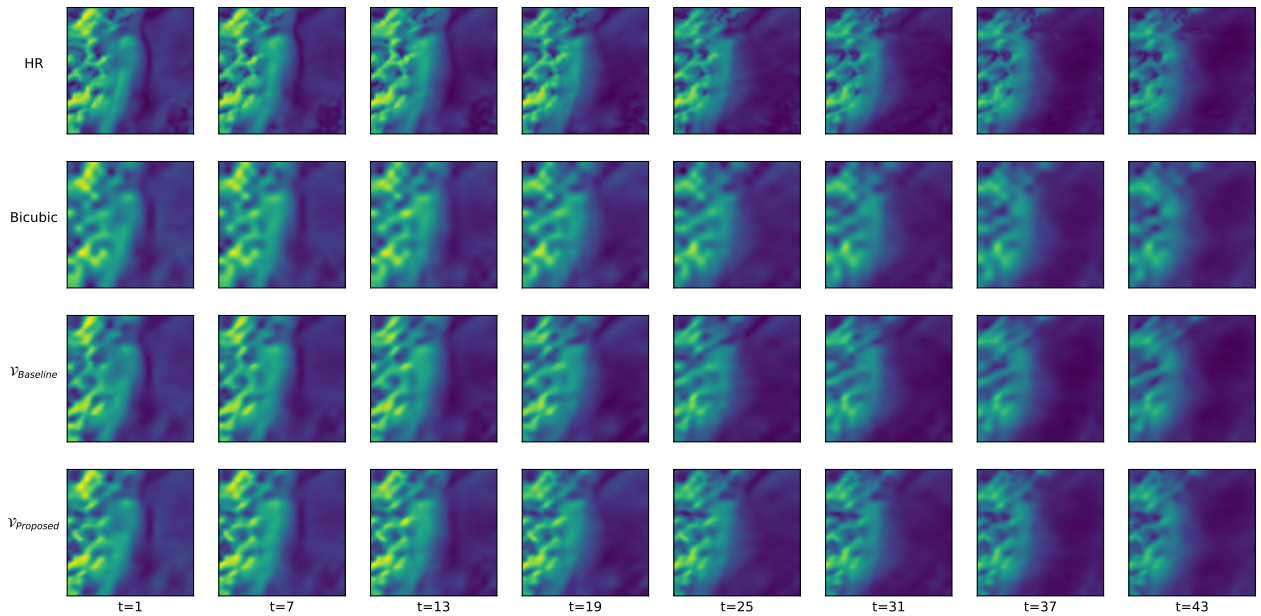

Figure 8. Example 1 for sharpness.

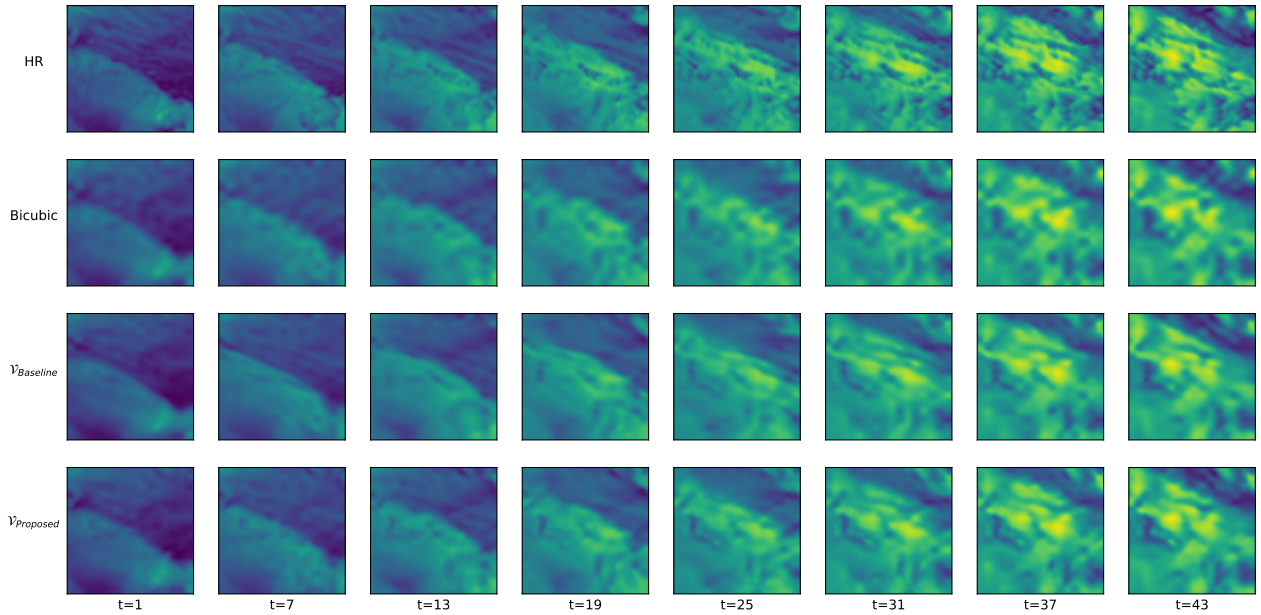

Figure 9. Example 2 for sharpness.

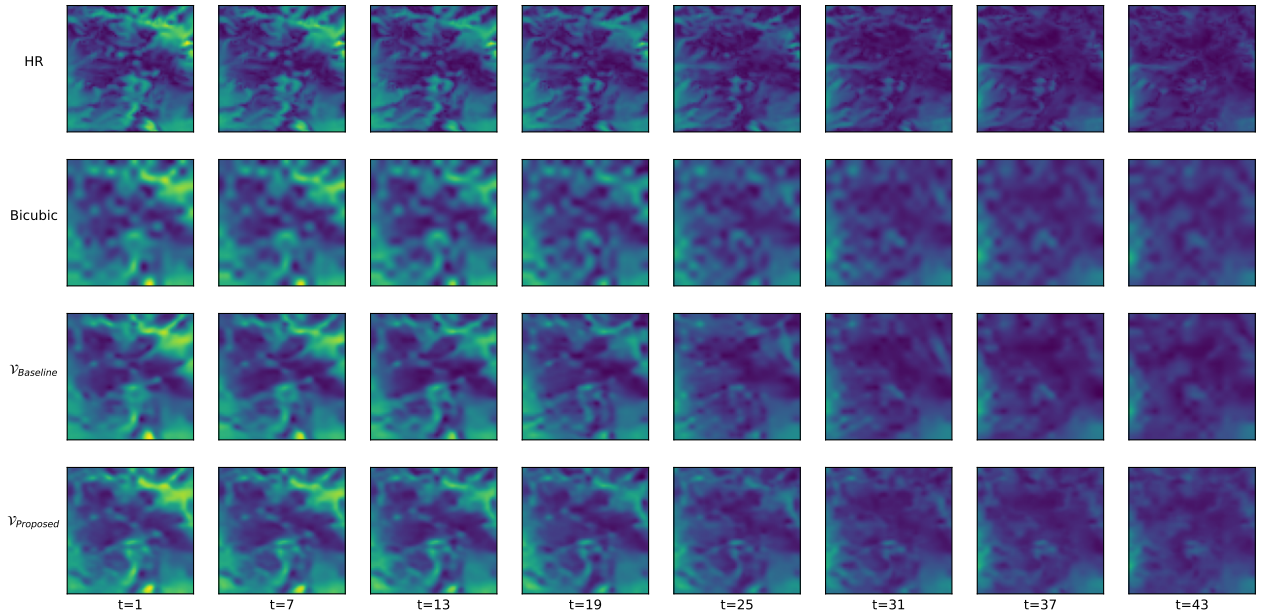

Figure 10. Example 3 for sharpness.

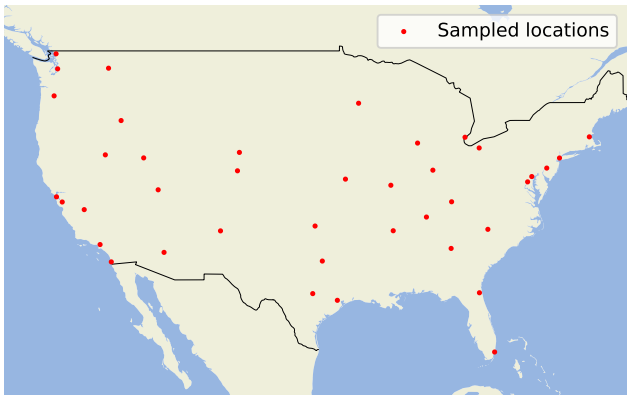

Figure 11. Sampled locations in the United States.

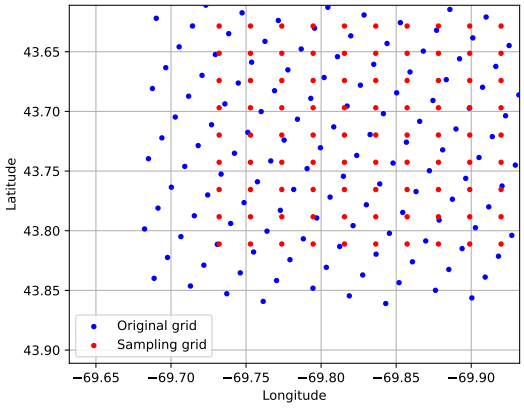

Figure 12. Resampling of the geospatial grid.

C. Data Preprocessing

In Figure 11, we showcase the 40 locations we choose to sample data from. At any given location, the uniform grid we create may not perfectly align with the original spatial grid in the dataset. To address this, we ensure our uniform grid is slightly denser than the original grid and then employ bilinear interpolation to accurately map the wind speed data from the original grid to the uniform grid. An example of these two grid is illustrated in Figure 12.