

Article

Not peer-reviewed version

A Temporal Downscaling Model for Gridded Geophysical Data with Enhanced Residual U-Net

[Liwen Wang](#), [Qian Li](#)^{*}, [Xuan Peng](#), [Qi Lv](#)

Posted Date: 21 September 2023

doi: 10.20944/preprints202309.1422.v1

Keywords: temporal downscaling; U-Net; flow regularization; residual blocks; ERA5



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

A Temporal Downscaling Model for Gridded Geophysical Data with Enhanced Residual U-Net

Liwen Wang¹, Qian Li^{1,2,*}, Xuan Peng^{1,2} and Qi Lv^{1,2}

¹ College of Meteorology and Oceanography, National University of Defense Technology, Changsha, China.

² High Impact Weather Key Laboratory of CMA, Changsha, China.

* Correspondence: public_liqian@163.com

Abstract: Temporal downscaling of gridded geophysical data is essential for improving climate models, weather forecasting, and environmental assessments. However, existing methods often could not accurately capture multi-scale temporal features, affecting their accuracy and reliability. To address this issue, we introduce an Enhanced Residual U-Net architecture for temporal downscaling. The architecture, which incorporates residual blocks, allows for deeper network structures without the risk of overfitting or vanishing gradients, thus capturing more complex temporal dependencies. The U-Net design inherently could capture multi-scale features, making it ideal for simulating various temporal dynamics. Moreover, we implement a flow regularization technique with advection loss to ensure that the model adheres to physical laws governing geophysical fields. Our experimental results across various variables within the ERA5 dataset demonstrate an improvement in downscaling accuracy, outperforming other methods.

Keywords: temporal downscaling; U-Net; flow regularization; residual blocks; ERA5

1. Introduction

Geophysical data, including variables such as air temperature, humidity, air pressure and sea surface temperature, form the backbone of several critical research areas spanning climatology, meteorology, and environmental sciences[1]. These variables are often represented in gridded formats that provide a spatially organized, multidimensional framework for analysis. However, the temporal resolution of gridded geophysical data sets can be inconsistent, presenting substantial challenges in subsequent analyses[2]. This limitation restricts the applicability of these data sets for tasks requiring fine-grained temporal details, such as short-term weather forecasting, localized climate modeling, and real-time environmental monitoring. High-quality, high-resolution data are usually acquired through advanced remote sensing techniques and complex simulation models, but these methods are computationally expensive and time-consuming[3]. In addition, the sheer volume of high-resolution data imposes limitations on storage, transport, and processing. These constraints necessitate the development of methodologies that can transform existing, coarser temporal data into more finely detailed sets without sacrificing quality[4].

The discrepancies in temporal granularity within gridded geophysical data have far-reaching implications[5,6]. For instance, the quality of climate change projections can be compromised, thereby affecting policy decisions related to climate mitigation strategies[7]. Similarly, coarse-grained data could lead to inaccurate weather forecasts, which, in turn, could have economic implications for sectors such as agriculture, energy production, and disaster management[8,9]. These challenges make it clear that methods for accurate and efficient temporal downscaling of geophysical data sets are required[10].

Temporal downscaling has been the subject of extensive research over the past few years[11]. Traditional methods primarily rely on statistical models like polynomial regression or autoregressive integrated moving average (ARIMA) models to interpolate between temporal data points[12]. While useful for linear trends, these methods often fall short when applied to geophysical data characterized by complex, non-linear temporal dynamics[13].

Recent advancements in machine learning have facilitated the development of more sophisticated downscaling techniques. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have been applied to downscaling tasks, showing improved performance over traditional statistical methods[14,15]. However, these machine learning-based techniques still face challenges in capturing intrinsic temporal dynamics and spatial relationships simultaneously[16].

The primary objective of this paper is to introduce a method for temporal downscaling of gridded geophysical data, combining flow regularization techniques with an Enhanced Residual U-Net architecture, as depicted in Figure 1. The contributions of this paper are threefold:

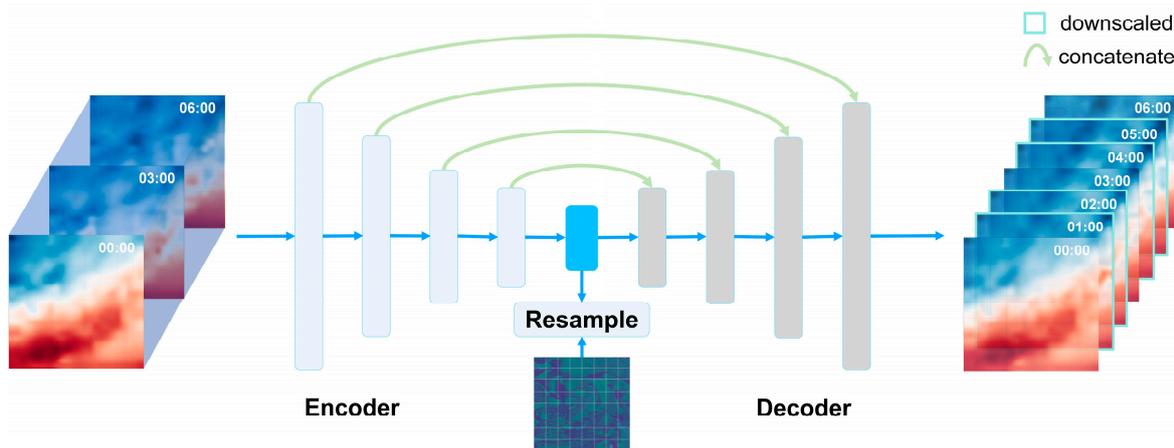


Figure 1. Overview of the enhanced residual U-Net model for temporal downscaling. The model consists of an encoder and a decoder, each with four residual blocks. It takes in grid data and outputs data with higher temporal resolution by performing temporal downscaling. After the encoder, the intermediate features are resampled to generate auxiliary flow information, which is then used to calculate the advection loss.

1. We introduce an enhanced residual U-Net architecture for the downscaling of geophysical data. Unlike traditional U-Net architectures, this enhanced model incorporates residual blocks, which allow for a deeper network that can capture complex patterns without succumbing to issues like overfitting or vanishing gradient problems. The depth and architecture of the enhanced residual U-Net are particularly effective at capturing multi-scale temporal features, a quality lacking in many existing methods.

2. We introduce the concept of flow regularization, which has been traditionally leveraged in computer vision tasks, to the domain of geophysical data downscaling. This addition serves as an auxiliary constraint that guides the model to adhere to the physical laws governing the movement and interaction of geophysical fields with higher accuracy than existing techniques.

3. We validate our model using multiple real-world geophysical data sets, comparing its performance against existing methods in terms of accuracy, computational efficiency, and fidelity of temporal features.

The paper is structured as follows: Section 2 provides a comprehensive review of related work, focusing on the principles of U-Net architectures and residual connections. Section 3 introduces the data sets used for the experiments. Section 4 provides a description of our proposed model, which employs enhanced residual U-Net for temporal downscaling. Section 5 presents the results, offering a comparative analysis with existing methodologies. Section 6 discusses the influence of the input grid data pixel size. Finally, Section 7 concludes the paper by summarizing key findings and outlining avenues for future research.

2. Related Work

2.1. Temporal Downscaling

Temporal downscaling serves as a crucial technique in various scientific applications[17–20], particularly in environmental modeling where high-frequency fluctuations often matter. The traditional ways to tackle this issue have primarily been statistical. Linear interpolation methods were among the earliest approaches, providing a quick yet overly simplistic way to fill in data between given time points. Soon after, Fourier-based methods were explored to address some of the linear assumptions but found limited applicability due to the inherent cyclical assumptions in the Fourier series[21]. Autoregressive Integrated Moving Average (ARIMA) models gained traction for their capabilities in capturing some level of non-linearity and seasonality[12]. Machine learning techniques like Support Vector Machines (SVMs) and Random Forests have been applied to the temporal downscaling problem as well[22]. Although these methods capture non-linearity better than linear interpolation, they often require extensive feature engineering and parameter tuning. Additionally, they fall short in integrating multi-scale features and incorporating flow information.

2.3. Regularization

In geoscience, regularization techniques are often employed as a critical enforcement mechanism to ensure that model predictions align with physical realities[23]. Some studies have utilized methods such as Total Variation Regularization to maintain crisp boundaries and smooth transitions in geological formations[24]. Others have opted for more intricate, physics-based regularization frameworks like the Hamilton-Jacobi-Bellman equations to enforce dynamic consistency in fluid flow models[25]. Hydrological models frequently make use of an energy balance constraint as a regularization term to confirm the thermodynamic plausibility of predicted water cycles[26]. More recently, advanced methods have emerged that integrate machine learning with physical laws to create hybrid models[27]. These 'physics-informed' models use regularization terms sourced from governing equations, like the Navier-Stokes equations for fluid dynamics or the Laplace equation for potential fields, as constraints during the learning process[27,28]. Nonetheless, many of these approaches often come with a trade-off between adherence to physical laws and computational efficiency[29,30]. Our flow regularization technique with advection loss strikes a balance by ensuring compliance with the physical laws that govern geophysical fields, while also being computationally practical.

2.4. Residual Connections

In recent years, residual connections have emerged as a critical innovation in the realm of deep learning architectures, particularly in convolutional neural networks (CNNs). The seminal work by He, *et al.* [31] introduced residual connections in their ResNet model, demonstrating that these connections alleviate the vanishing gradient problem, thus enabling the training of much deeper networks. Residual architectures have been adopted in various disciplines beyond image classification, including object detection and segmentation. In geoscience applications, residual connections have shown promising results in tasks such as seismic interpretation and subsurface reservoir modeling[32]. These architectures facilitate the learning of hierarchical features from geological data by promoting the flow of gradients throughout the network. By creating shortcuts between layers, residual connections allow for a more efficient and effective propagation of errors during backpropagation, improving the network's capacity to learn complex mappings.

2.5. U-Net

The U-Net architecture, originally designed for biomedical image segmentation, has shown unparalleled success in various domains requiring complex spatial hierarchies. The architecture follows an encoder-decoder structure, capturing context in the encoding layers and using the decoding layers to reconstruct spatial details. One of the most distinguishing features of the U-Net is

its use of skip connections, allowing it to preserve high-frequency details that would otherwise be lost during the encoding process.

In recent years, the U-Net architecture has seen several adaptations and modifications to suit different tasks[33–36]. For example, 3D U-Nets have been developed to process volumetric data, and Temporal U-Nets have been explored to capture time-related changes in videos[37]. However, integrating temporal downscaling with U-Net's predominantly spatial-focused architecture remains an open challenge.

While individual advancements in temporal downscaling, U-Net, and ResNet architectures have been substantial, the integration of these elements remains largely unexplored. Few works have attempted to adapt U-Net architectures for time series data, but these generally involve straightforward adaptations that don't fully utilize temporal dependencies. Similarly, while ResNet have been used in conjunction with LSTMs for sequence modeling, their application in temporal downscaling is yet to be fully realized.

In this light, our work aims to fill this gap by proposing a hybrid architecture that leverages the spatial prowess of U-Net, the learning capabilities of ResNet, and the flow regularization techniques, specifically tailored for the task of temporal downscaling in gridded geophysical data.

3. Study Area and Dataset

Our investigation targets the geographic region defined by longitudes 112°E to 118°E and latitudes 22°N to 28°N, with a grid resolution of $0.25^{\circ} \times 0.25^{\circ}$, as depicted in Figure 2. Data for this area was sourced from the ERA5 reanalysis dataset. The training dataset is comprised of 21,912 sets, each containing data at three-hour intervals, spanning the years 2010 to 2019, for a total of 87,648 hours. For validation, we use a test set consisting of 2,196 sets from the year 2020, also collected at three-hour intervals, totaling 8,784 hours. Our model aims to downscale this data to a finer one-hour temporal resolution. We evaluate the model's performance across three meteorological variables: 2-meter surface air temperature, 850hPa geopotential height, and 850hPa relative humidity.

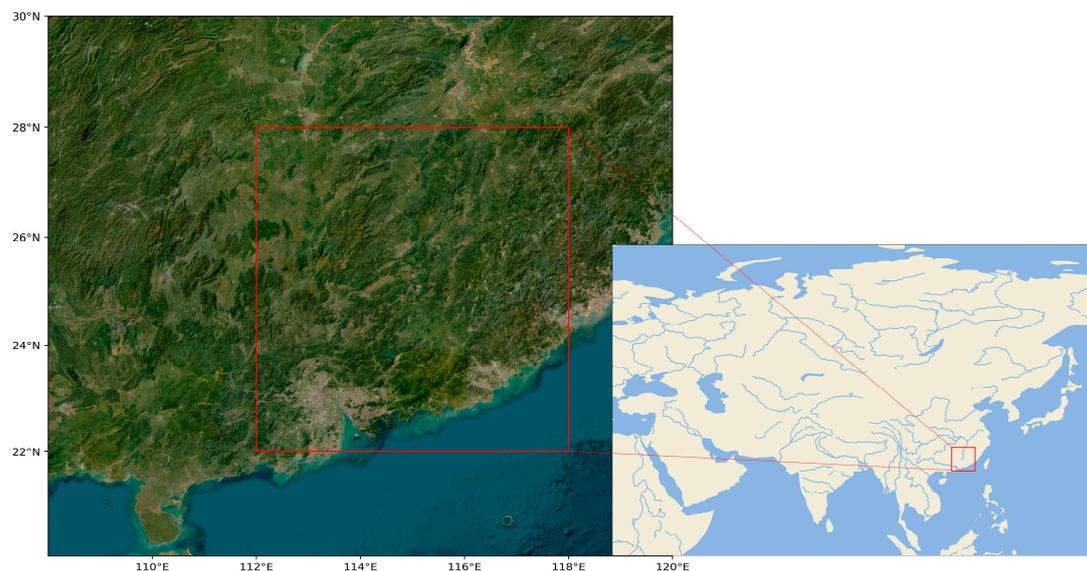


Figure 2. Study area.

4. Model

In this section, we describe the architecture and components of our enhanced residual U-Net model. We detail how the model integrates residual blocks, auxiliary flow information, and advection loss to perform temporal downscaling of geophysical data.

4.1. Problem Definition

In the field of geophysical data analysis, the problem of temporal downscaling aims to refine the time resolution of observed data, thereby providing more frequent measurements. Specifically, given a dataset $X = \{x_{t_1}, x_{t_2}, \dots, x_{t_N}\}$ ($X \in \mathbb{R}^{N \times H \times W}$) at a coarser temporal resolution of three hours, the objective is to estimate a fine-grained dataset $Y = \{y_{t'_1}, y_{t'_2}, \dots, y_{t'_{2N}}\}$ ($Y \in \mathbb{R}^{2N \times H \times W}$) at a one-hour resolution, where $t'_i = t_i/2$ for $i = 1, 2, \dots, 2N$. The primary goal is to minimize the discrepancy between the ground truth and the predicted over the fine-grained temporal intervals. Mathematically, this can be formulated as:

$$\min_{\Theta} \mathcal{L}(Y_{\text{true}}, Y_{\text{pred}}) = \min_{\Theta} \sum_{i=1}^{2N} \|Y_{\text{true},i} - Y_{\text{pred},i}\|^2. \quad (1)$$

Here Θ represents the parameters of the Enhanced Residual U-Net model, and \mathcal{L} is the loss function.

4.2. Enhanced Residual U-Net

In this work, we introduce an architecture, enhanced residual U-Net, designed specifically for the temporal downscaling of gridded geophysical data. This architecture merges the high-level feature extraction capabilities of U-Net with the robustness of Residual Networks (ResNet) to produce an efficient and scalable model (see Figure 3).

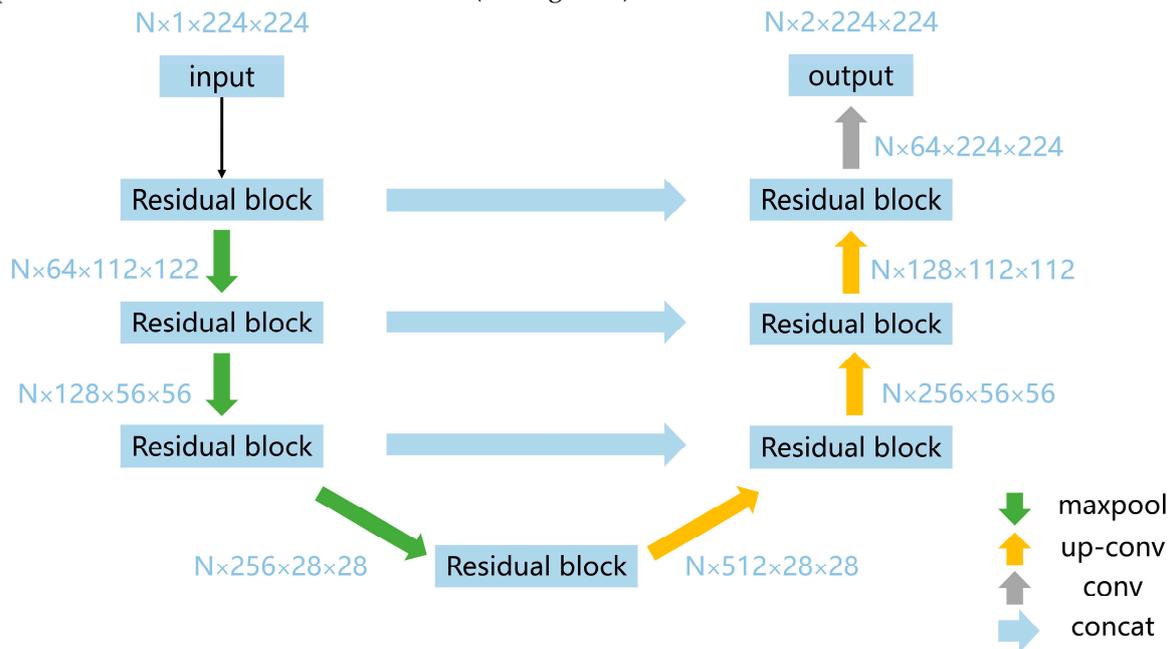


Figure 3. The architecture of the enhanced residual U-Net. It consists of an encoder section on the left, a decoder section on the right, and an auxiliary flow information layer in between. The encoder features four residual blocks, each containing two convolutional layers with batch normalization and ReLU activation functions, responsible for reducing feature dimensions while capturing initial patterns from the input. The decoder also contains four residual blocks and uses transposed convolutions for upsampling. Skip connections merge the output from each encoder Residual Block with its corresponding decoder block, ensuring the preservation of spatial information across scales.

The architecture is constructed from two main components: an encoder and a decoder. The encoder is responsible for downscaling the input tensor, thereby extracting high-level features. The decoder, on the other hand, upscales these high-level features to reconstruct the output tensor. These operations are standard in any U-Net architecture; however, our model introduces several enhancements.

One of the enhancements in our architecture is the introduction of residual blocks following key convolutional layers in the encoder section. The architecture's depth is primarily achieved through its deeper residual blocks, and each residual block comprises three 3×3 convolutional layers with ReLU activations[38]. The outputs of these layers are summed with the original input using a skip connection and these residual blocks help the model to learn complex features with reduced risk of vanishing or exploding gradients.

The encoder section is composed of a succession of four deeper residual blocks, each with distinct channel configurations—64, 128, 256, and 512 channels. Every deeper residual block comprises three convolutional layers, each followed by batch normalization and ReLU activation functions. This series of operations enriches the representation of the input data by sequentially increasing the number of channels. The architecture also incorporates max-pooling layers after each block to reduce the spatial dimensions of the feature maps. Subsequent to each max-pooling operation, the spatial dimensions are halved, thereby focusing on the extraction of high-level features.

The decoder section reverses the operations conducted by the encoder. It employs a series of up-convolutional layers paired with concatenation operations that merge high-level features from the encoder. Each up-convolutional layer also employs a ReLU activation function and effectively doubles the spatial dimensions. Similar to the encoder, residual blocks are also introduced in the decoder. These are positioned after each up-convolutional layer and function in the same manner as their encoder counterparts. These blocks refine the combined high-level and low-level features. The network concludes with a 1×1 convolutional layer, which condenses the 64-channel feature map into a 2-channel output.

4.3. Flow Regularization Using Advection Loss

Conventional methods often miss capturing the evolving patterns. To address this limitation, we incorporate flow information with advection loss into our enhanced residual U-Net model, and this section details the mathematical and computational elements of this approach (see Figure 4).

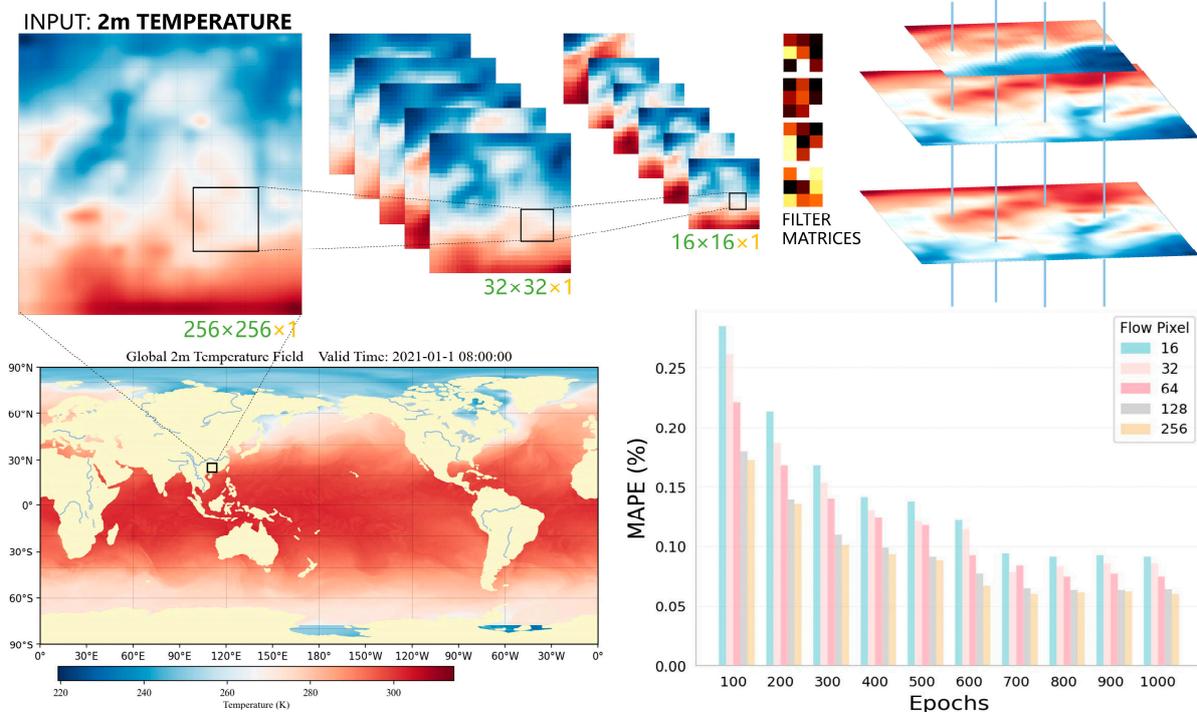


Figure 4. Overview of flow information extraction from the intermediate features post-encoder phase. Following encoding, these intermediate features undergo specific convolutional operations and resampling procedures to yield the flow information. The bar chart illustrates the Mean Absolute Percentage Error (MAPE) for various flow pixel resolutions across different epochs. The x-axis

represents the training epochs, ranging from 100 to 1000, while the y-axis represents the MAPE in percentages.

Advection refers to the transport of a scalar field driven by flow regularization. Mathematically, it can be represented as a transformation function, $\text{Advect}(Y_t, F)$, which takes in a geophysical field at time t , denoted as Y_t , and flow information F , and returns an approximated field at time $t + 1$, represented as \hat{Y}_{t+1} . The principle behind advection is rooted in fluid dynamics and is used widely in computational fluid dynamics simulations and meteorological models. By adopting an advection transformation, we impose an auxiliary constraint on our neural network model, compelling it to learn physically meaningful dynamics.

The advection loss is introduced as an additional term in the loss function and is defined as $L_{\text{advection}} = \|Y_{t+1} - \text{Advect}(Y_t, F)\|_2^2$. In essence, this loss measures the difference between the true field at $t + 1$ and the advected field \hat{Y}_{t+1} . It guides the network to learn a more accurate representation of the data and acts as a regularization term, reducing overfitting while still ensuring that the model learns the dynamics of the field.

For the computation of $\text{Advect}(Y_t, F)$, spatial interpolation is employed. Given a 2D geophysical field Y_t and corresponding flow information, which is also a 2D tensor but with two channels representing the velocity vectors (F_x, F_y) , each point (x, y) in Y_t is shifted according to the velocity vector at that point. The new coordinates $(x_{\text{new}}, y_{\text{new}})$ are calculated as $(x + F_x, y + F_y)$. Bilinear interpolation is used to estimate the value of the advected field \hat{Y}_{t+1} at these new coordinates.

The final loss function incorporating both the L2 loss and the advection loss is formulated as follows:

$$\mathcal{L} = \|Y_{\text{pred}} - Y_{\text{true}}\|_2^2 + \lambda \|Y_{t+1} - \text{Advect}(Y_t, F)\|_2^2 \quad (2)$$

In this equation, $\|Y_{\text{pred}} - Y_{\text{true}}\|_2^2$ is the L2 loss, representing the squared Euclidean distance between the predicted output Y_{pred} and the ground-truth Y_{true} . The term $\lambda \|Y_{t+1} - \text{Advect}(Y_t, F)\|_2^2$ is the weighted advection loss, and the weight of $\lambda = 0.3$ is applied to balance the contribution of advection loss against the L2 loss.

5. Experiments

The model was implemented using PyTorch and the training was performed on a machine equipped with eight NVIDIA Tesla A5000 GPUs. During training, we employed the Adam optimizer with a learning rate of 0.0001 and a batch size of 32. The model was trained for 1,000 epochs. A decay rate of 0.9 for the learning rate was applied every 200 epochs to ensure convergence. The loss function used in training was a combination of the Mean Squared Error (MSE) loss and the advection loss, as described in Section 3.4.

5.1. Quantitative Comparison with Conventional Methods

In this section, we conduct a quantitative evaluation of our enhanced residual U-Net model against several benchmark methods, focusing on three critical atmospheric variables: 2-meter air temperature, geopotential height, and relative humidity. For a robust comparison, we employ three well-established metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). As depicted in the table, it is clear that our enhanced residual U-Net model surpasses other techniques across all three metrics and for each atmospheric variable examined. Specifically, for 2-meter air temperature, our model yields RMSE and MAE values of 0.20 and 0.17, respectively; for geopotential height, the corresponding values are 0.72 and 0.62; and for relative humidity, these metrics stand at 0.64, 0.46 and 0.59%. When contrasted with conventional methods like linear interpolation and cubic spline—whose performance metrics are considerably

higher in terms of RMSE and MAE—the superiority of our model becomes evident. We also reproduced the super-sloMo and RIFE methods from the computer vision domain, typically used for video frame interpolation. The only modification was transitioning from a three-channel RGB input to a single-channel input representing the 2-meter temperature field. The results are presented in Table 1, while these deep learning-based models outperform linear interpolation and cubic spline, they still fail to match the superior performance of our enhanced residual U-Net model.

Table 1. Comparison of different methods for temporal downscaling.

Method	2m Temperature (RMSE/MAE/MAPE)	Geopotential Height (RMSE/MAE/MAPE)	Relative Humidity (RMSE/MAE/MAPE)
Linear Interpolation	0.51 / 0.42 / 0.14%	1.79 / 1.55 / 0.09%	1.61 / 1.15 / 1.47%
Cubic Spline	0.42 / 0.35 / 0.12%	1.50 / 1.30 / 0.8%	1.35 / 0.96 / 1.23%
LSTM-based Model	0.31 / 0.25 / 0.10%	1.20 / 1.03 / 0.07%	0.90 / 0.64 / 0.82%
Super-sloMo[39]	0.25 / 0.22 / 0.08%	1.24 / 1.10 / 0.07%	0.93 / 0.65 / 0.83%
RIFE[40]	0.23 / 0.20 / 0.07%	0.87 / 0.74 / 0.05%	0.75 / 0.51 / 0.65%
Enhanced Residual U-Net	0.20 / 0.17 / 0.06%	0.72 / 0.62 / 0.05%	0.64 / 0.46 / 0.59%

5.2. Visual and Qualitative Analysis

To assess the effectiveness of our model, we selected a case of the 2-meter temperature fields for January 21, 2020, between the hours of 10:00 and 17:00. As depicted by Figure 5, our model accurately reproduces the nonlinear variability in the temperature field.

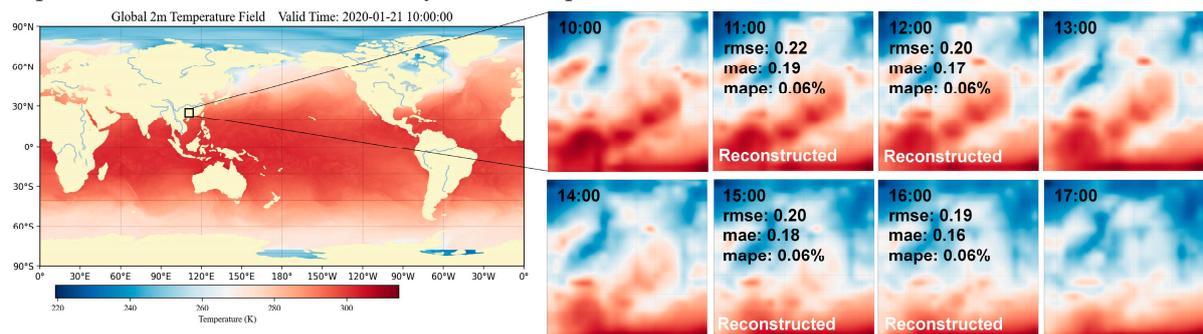


Figure 5. Ground truth and model-generated downscaled results. Points at 10 a.m., 1 p.m., 2 p.m., and 5 p.m. are the model's input, while the data at 11 a.m., 12 p.m., 3 p.m., and 4 p.m. are model-generated outputs.

Focusing on specific intervals, for the 11:00 downscaling between 10:00 and 13:00, the temperature patterns exhibit distinct characteristics. For example, the coastal areas, which originally showed a relatively warmer temperature at 10:00, start showing moderate cooling due to oceanic influences. In contrast, the central regions, which were cooler at 10:00, warm up slightly, likely due to increased solar radiation. This is captured with an RMSE of 0.22, MAE of 0.19, and MAPE of 0.06%. As we move to 12:00, the temperature in the valley regions starts showing minor fluctuations, likely due to local wind patterns. The RMSE improves to 0.20, MAE drops to 0.17, and the MAPE remains stable at 0.06%, emphasizing the model's competence in capturing these subtle dynamics.

Between 14:00 and 17:00, the 15:00 downscaling indicates that urban areas start to experience heat island effects, with temperature spikes in densely populated zones. These spikes contrast with adjacent rural or forested areas that show a more stable temperature profile. At this point, the RMSE is 0.20, MAE is 0.18, and the MAPE is stable at 0.06%. At 16:00, there's a notable decrease in temperature in the mountainous regions, likely due to the shadows cast by the changing sun angle. This dynamic is reflected with an RMSE of 0.19, MAE of 0.16, and a consistent MAPE of 0.06%.

5.3. Ablation Studies

In our ablation studies, we examine the individual components of the enhanced residual U-Net model to understand their significance in achieving overall performance metrics (see Figure 6). We set the performance of the full model as the baseline, which exhibits RMSE and MAE values of 0.20 and 0.17 for 2-meter temperature, 0.72 and 0.62 for geopotential height, and 0.64 and 0.46 for relative humidity. Upon removing the multi-scale features, we observed a discernible decrease in predictive accuracy across all variables, with RMSE and MAE values for 2-meter temperature rising to 0.34 and 0.28, respectively. This confirms the importance of multi-scale features in capturing the complexity of geophysical data. When we omitted the flow regularization, there was a significant performance decline: RMSE and MAE values for 2-meter temperature rose to 0.23 and 0.19, indicating the vital role of flow regularization in capturing temporal dynamics effectively. Lastly, reducing the architectural depth of the model led to a less pronounced, yet still noticeable, decline in performance. For instance, RMSE for 2-meter temperature increased to 0.26, and MAE rose to 0.22, underlining the model's depth's role in capturing the intricacies of geophysical data. Overall, the degradation in model performance upon removal of each component underscores their collective importance, reinforcing the need for their inclusion in the final architecture.

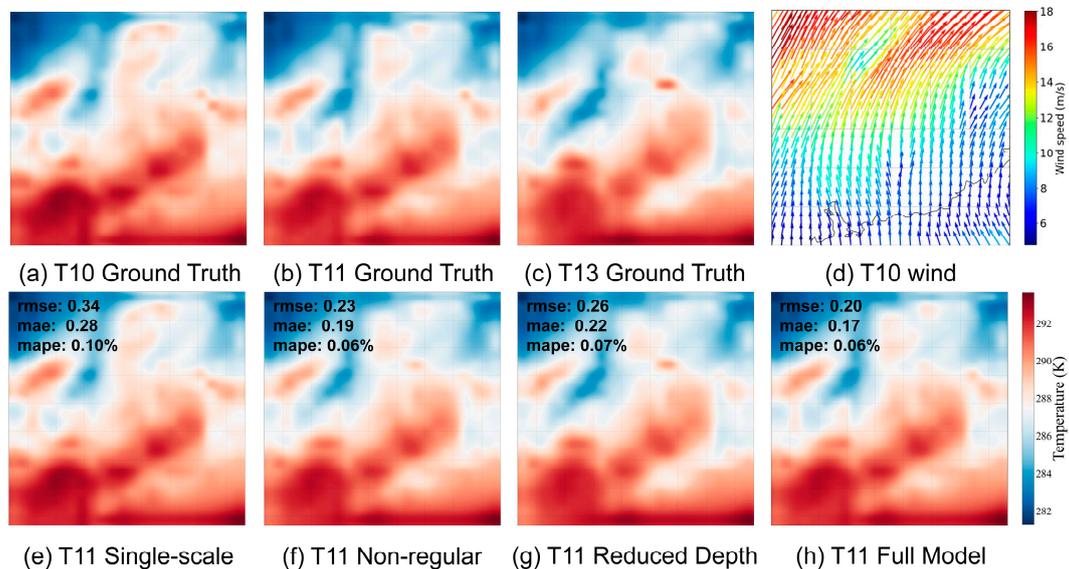


Figure 6. Visual Comparison of Ablation Study Results. We downscale the 2-meter temperature fields from 10 a.m. to 1 p.m. on January 21, 2020, to obtain results for 11 a.m. Subfigure d depicts the wind field at 11 a.m., which offers indicative insights into temperature evolution.

Table 2. Ablation Study on the Effect of Various Components on the Full Model's Performance.

Method	2m Temperature (RMSE/MAE/MAPE)	Geopotential Height (RMSE/MAE/MAPE)	Relative Humidity (RMSE/MAE/MAPE)
Without Multi-scale Features	0.34 / 0.28 / 0.10%	1.12 / 0.31 / 0.02%	0.92 / 0.33 / 0.42%
Without Flow Regularization	0.23 / 0.19 / 0.06%	0.96 / 0.32 / 0.02%	0.87 / 0.34 / 0.44%
Reduced Architectural Depth	0.26 / 0.22 / 0.07%	1.05 / 0.25 / 0.02%	0.74 / 0.28 / 0.36%
Full Model (Baseline)	0.20 / 0.17 / 0.06%	0.72 / 0.62 / 0.05%	0.64 / 0.46 / 0.59%

6. Discussion

In our analysis, we found that the input grid data pixel size significantly influences the model's Mean Absolute Error (MAE), forming a U-shaped pattern. For smaller pixel sizes, specifically at 64 pixels, the MAE was around 0.26. The convolutional layers in this case are restricted to localized features, missing the larger spatial context that is crucial for accurate downscaling. Conversely, the MAE reaches its minimum value of 0.17 at an optimal pixel size of 224. Beyond this optimal point, the MAE starts to increase again, climbing to approximately 0.33 at a pixel size of 320. This suggests that while the model is effective in capturing global features at larger pixel sizes, it fails to grasp finer details, leading to increased error.

This performance is further illuminated when comparing the MAE curves between the full model and the reduced-depth model. The full model achieved a lower minimum MAE value of 0.17 as opposed to the reduced model's 0.22. This can largely be attributed to the residual modules in the full model, which allow for a more efficient and resilient feature extraction process. These modules facilitate better generalization across varying pixel sizes, thus accounting for the full model's more effective U-shaped performance curve in MAE across a broader range of pixel sizes.

In addition to the performance metrics discussed earlier, another significant advantage of our model is its capability to support real-time inference. A set of experiments was conducted to evaluate the model's speed performance across multiple hardware configurations—A40, V100, and A5000. The inference time was observed at different intermediate snapshot levels: 9, 18, 27, 36, and 45. Remarkably, even at 45 intermediate snapshots, the inference time did not exceed 44ms on A5000, and it was even lower on A40 and V100 setups, clocking at approximately 39ms and 38ms, respectively. This rapid inference time positions our model as not only accurate but also highly practical for real-time applications in weather forecasting and climate studies.

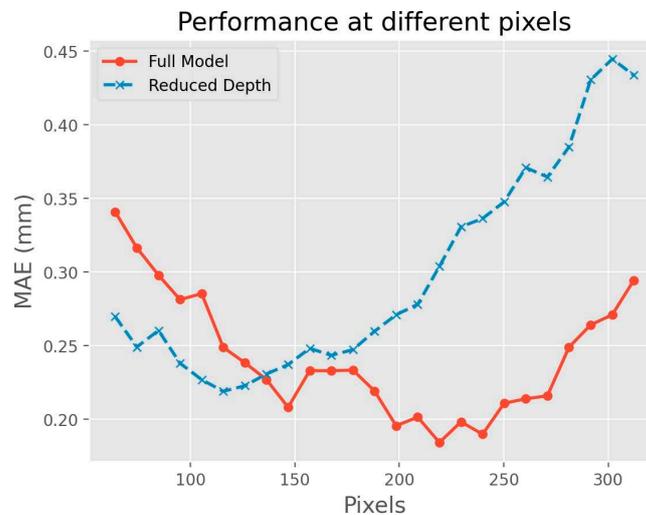


Figure 7. Influence of field pixels.

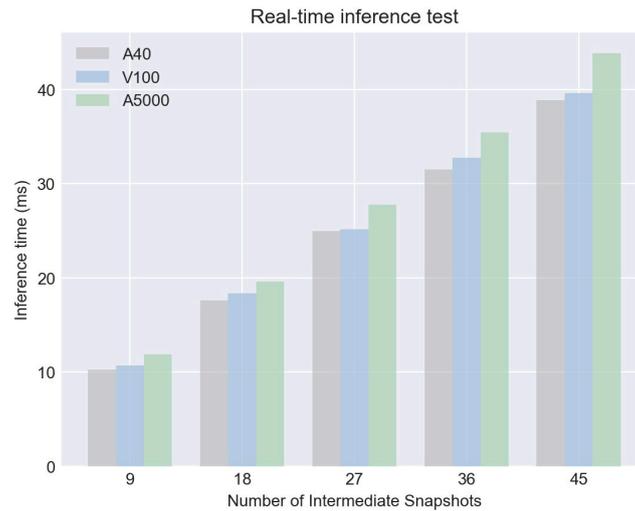


Figure 8. Comparison of inference time under different GPU conditions.

7. Conclusions

In this study, we introduced the enhanced residual U-Net, an innovative architecture that amalgamates the strengths of U-Net and ResNet to address the challenge of temporal downscaling in gridded geophysical data. The architecture is specifically designed to harness both local and global features within the data, thereby producing a robust and versatile model capable of delivering high-quality downscaling results.

One of the most significant contributions of this work is the integration of residual blocks in both the encoding and decoding pathways. This design choice not only enhances the learning capability of the network but also alleviates issues related to the vanishing gradient problem, allowing for deeper and more effective networks. We also introduced a custom loss function that combines Mean Squared Error (MSE) with a spatial regularization term, which collectively ensures both the fidelity and spatial coherence of the downscaled output.

Our experimental results, based on a comprehensive evaluation using multiple gridded geophysical datasets, validated the effectiveness of the enhanced residual U-Net model. We demonstrated that the architecture outperformed traditional downscaling methods and other state-of-the-art machine learning approaches in key metrics, including RMSE and MAE, while maintaining computational efficiency.

In conclusion, the enhanced residual U-Net architecture stands as a robust and efficient solution for the temporal downscaling of gridded geophysical data. Its design features, including the use of residual blocks and a custom loss function, make it a highly promising tool for both academic research and practical applications in the field of geoscience.

Future work could further enhance this architecture by incorporating additional techniques for feature selection or by tailoring the network to different kinds of geophysical data. Moreover, real-world applicability of this model could be tested in other domains requiring high-fidelity downscaling, providing a broader utility beyond the specific use-case studied here.

Author Contributions: Conceptualization, L.W. and Q.L.; methodology, L.W.; software, L.W.; validation, Q.L. and Q.Lv.; formal analysis, L.W.; investigation, Q.L.; resources, X.P.; data curation, L.W.; writing—original draft preparation, L.W.; writing—review and editing, Q.Lv.; visualization, Q.Lv.; supervision, Q.L.; project administration, Q.L.; funding acquisition, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 42075139, U2242201, 42105146, 41305138), the China Postdoctoral Science Foundation (Grant No. 2017M621700), Hunan Province Natural Science Foundation (Grant No. 2021JC0009, 2021JJ30773) and Fengyun Application Pioneering Project (FY-APP-2022.0605).

Data Availability Statement: All data necessary to reproduce the results of this work could be downloaded at <https://doi.org/10.24381/cds.bd0915c6> and <https://doi.org/10.24381/cds.adbb2d47>.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195 - 204.
2. Scipal, K.; Holmes, T.R.H.; de Jeu, R.A.M.; Naeimi, V.; Wagner, W. A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters* **2008**, *35*.
3. Chen, S.; Zhang, M.; Lei, F. Mapping Vegetation Types by Different Fully Convolutional Neural Network Structures with Inadequate Training Labels in Complex Landscape Urban Areas. *Forests* **2023**.
4. Marthews, T.R.; Dadson, S.J.; Lehner, B.; Abele, S.; Gedney, N. High-resolution global topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences* **2015**, *19*, 91-104.
5. Loew, A.; Bell, W.; Brocca, L.L.; Bulgin, C.E.; Burdanowitz, J.; Calbet, X.; Donner, R.V.; Ghent, D.; Gruber, A.; Kaminski, T.; et al. Validation practices for satellite-based Earth observation data across communities. *Reviews of Geophysics* **2017**, *55*, 779 - 817.
6. Mann, M.E.; Rahmstorf, S.; Kornhuber, K.; Steinman, B.A.; Miller, S.K.; Coumou, D. Influence of Anthropogenic Climate Change on Planetary Wave Resonance and Extreme Weather Events. *Scientific Reports* **2017**, *7*.
7. Rogelj, J.; Forster, P.M.; Kriegler, E.; Smith, C.J.; Séférian, R. Estimating and tracking the remaining carbon budget for stringent climate targets. *Nature* **2019**, *571*, 335 - 342.
8. Mason, S.J.; Stephenson, D.B. How Do We Know Whether Seasonal Climate Forecasts are Any Good. 2008.
9. Schloss, A.; Kicklighter, D.W.; Kaduk, J.; Wittenberg, U.; Intercomparison, T.P.O.T.P.N.M. Comparing global models of terrestrial net primary productivity (NPP): comparison of NPP to climate and the Normalized Difference Vegetation Index (NDVI). *Global Change Biology* **1999**, *5*.
10. Schleussner, C.; Lissner, T.; Fischer, E.M.; Wohland, J.; Perrette, M.; Golly, A.; Rogelj, J.; Childers, K.H.; Schewe, J.; Frieler, K.; et al. Differential climate impacts for policy-relevant limits to global warming: the case of 1.5 °C and 2 °C. *Earth System Dynamics Discussions* **2015**, *7*, 327-351.
11. Isotta, F.A.; Begert, M.; Frei, C. Long-Term Consistent Monthly Temperature and Precipitation Grid Data Sets for Switzerland Over the Past 150 Years. *Journal of Geophysical Research: Atmospheres* **2019**, *124*, 3783 - 3799.
12. Arunkumar, K.E.; Kalaga, D.V.; Mohan Sai Kumar, C.; Kawaji, M.; Brenza, T.M. Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alexandria Engineering Journal* **2022**, *61*, 7585 - 7603.
13. Majda, A.J.; Harlim, J. Physics constrained nonlinear regression models for time series. *Nonlinearity* **2012**, *26*, 201 - 217.
14. Yang, H.; Wang, T.; Zhou, X.; Dong, J.; Gao, X.; Niu, S. Quantitative Estimation of Rainfall Rate Intensity Based on Deep Convolutional Neural Network and Radar Reflectivity Factor. *Proceedings of the 2nd International Conference on Big Data Technologies* **2019**.
15. Misra, S.; Sarkar, S.; Mitra, P. Statistical downscaling of precipitation using long short-term memory recurrent neural networks. *Theoretical and Applied Climatology* **2018**, *134*, 1179-1196.
16. Lees, T.; Buechel, M.; Anderson, B.; Slater, L.J.; Reece, S.; Coxon, G.; Dadson, S.J. Rainfall-Runoff Simulation and Interpretation in Great Britain using LSTMs. 2021.
17. Kajbaf, A.A.; Bensi, M.T.; Brubaker, K.L. Temporal downscaling of precipitation from climate model projections using machine learning. *Stochastic Environmental Research and Risk Assessment* **2022**, *36*, 2173 - 2194.
18. Barboza, L.A.; Chen, S.; Alfaro-Córdoba, M. Spatio-temporal downscaling emulator for regional climate models. *Environmetrics* **2022**.
19. Huang, J.; Perez, M.J.R.; Perez, R.; Yang, D.; Keelin, P.; Hoff, T.E. Nonparametric Temporal Downscaling of GHI Clear-sky Indices using Gaussian Copula. 2022 *IEEE 49th Photovoltaics Specialists Conference (PVSC)* **2022**, 0654-0657.

20. Michel, A.; Sharma, V.; Lehning, M.; Huwald, H. Climate change scenarios at hourly time-step over Switzerland from an enhanced temporal downscaling approach. *International Journal of Climatology* **2021**, *41*, 3503 - 3522.
21. Bracewell, R.N. *The Fourier Transform and Its Applications*. 1966.
22. Vandal, T.J.; Kodra, E.; Ganguly, A.R. Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. *Theoretical and Applied Climatology* **2017**, *137*, 557-570.
23. Ahmmed, B.; Vesselinov, V.V.; Mudunuru, M.K. SmartTensors: Unsupervised and physics-informed machine learning framework for the geoscience applications. *Second International Meeting for Applied Geoscience & Energy* **2022**.
24. Greiner, T.A.L.; Lie, J.E.; Kolbjørnsen, O.; Evensen, A.K.; Nilsen, E.H.; Zhao, H.; Demyanov, V.V.; Gelius, L.J. Unsupervised deep learning with higher-order total-variation regularization for multidimensional seismic data reconstruction. *GEOPHYSICS* **2021**.
25. Kim, J.; Yang, I. Hamilton-Jacobi-Bellman Equations for Maximum Entropy Optimal Control. *ArXiv* **2020**, *abs/2009.13097*.
26. Gan, T.; Tarboton, D.G.; Gichamo, T.Z. Evaluation of Temperature-Index and Energy-Balance Snow Models for Hydrological Applications in Operational Water Supply Forecasts. *Water* **2023**.
27. Raissi, M.; Perdikaris, P.; Karniadakis, G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686-707.
28. Zhu, Y.; Zabararas, N.; Koutsourelakis, P.-S.; Perdikaris, P. Physics-Constrained Deep Learning for High-dimensional Surrogate Modeling and Uncertainty Quantification without Labeled Data. *ArXiv* **2019**, *abs/1901.06314*.
29. Mizukami, N.; Clark, M.P.; Newman, A.J.; Wood, A.W.; Gutmann, E.D.; Nijssen, B.; Rakovec, O.; Samaniego, L. Towards seamless large-domain parameter estimation for hydrologic models. *Water Resources Research* **2017**, *53*, 8020 - 8040.
30. Hrachowitz, M.; Soulsby, C.; Tetzlaff, D.; Dawson, J.J.C.; Dunn, S.M.; Malcolm, I.A. Using long-term data sets to understand transit times in contrasting headwater catchments. *Journal of Hydrology* **2009**, *367*, 237-248.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2015**, 770-778.
32. Laloy, E.; Héroult, R.; Jacques, D.; Linde, N. Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network. *Water Resources Research* **2017**, *54*, 381 - 406.
33. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv* **2015**, *abs/1505.04597*.
34. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.J.; Heinrich, M.P.; Misawa, K.; Mori, K.; McDonagh, S.G.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. *ArXiv* **2018**, *abs/1804.03999*.
35. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, S...* **2018**, 11045, 3-11.
36. Ibtihaz, N.; Rahman, M.S. MultiResUNet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. *Neural networks : the official journal of the International Neural Network Society* **2019**, *121*, 74-87.
37. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016.
38. Glorot, X.; Bordes, A.; Bengio, Y. Deep Sparse Rectifier Neural Networks. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2011.
39. Jiang, H.; Sun, D.; Jampani, V.; Yang, M.-H.; Learned-Miller, E.G.; Kautz, J. Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* **2017**, 9000-9008.

40. Huang, Z.; Zhang, T.; Heng, W.; Shi, B.; Zhou, S. RIFE: Real-Time Intermediate Flow Estimation for Video Frame Interpolation. In Proceedings of the European Conference on Computer Vision, 2020.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.