Attention-Enhanced Convolutional Networks for Climate Prediction

Anonymous Author(s)
Affiliation
Address
email

https://github.com/syw003/climate-emulation-model

Abstract

Climate modeling is computationally expensive, limiting our ability to rapidly explore different emission scenarios and policy interventions. This project addresses the climate emulation challenge by developing deep learning models that can efficiently predict future climate variables from atmospheric forcings. I used an attention-enhanced convolutional neural network that incorporates spatial attention mechanisms, multi-objective loss functions, and advanced data augmentation techniques to predict surface air temperature (tas) and precipitation (pr) from greenhouse gas concentrations and radiative forcings. My final model achieves validation performance of 1.18K RMSE for temperature and 2.00 mm/day RMSE for precipitation, representing 67% and 58% improvements respectively over baseline CNN approaches. The model successfully captured spatial patterns as well as temporal variability in climate responses across different emission scenarios while maintaining sub-second inference times compared to hours to days for traditional climate models. Through systematic experimentation across multiple architectures (CNN, CNN-LSTM, U-Net, Transformer), we demonstrate that spatial attention mechanisms and combined loss functions are critical for accurate climate prediction. Our approach provides a foundation for fast climate policy exploration and scenario analysis. Final (Public) Kaggle Score: 1.0973

1 Introduction

1.1 Problem Definition and Motivation

Climate change represents one of the most pressing challenges right now, requiring rapid assessment of different emission scenarios and policy interventions. Traditional climate models, while highly accurate, are computationally prohibitive for extensive scenario exploration. A single high-resolution climate simulation can take weeks to months on supercomputers, severely limiting our ability to explore the vast space of possible climate futures.

ML-based climate emulation offers a promising solution by learning to approximate complex climate dynamics from historical simulations. Once trained, these emulators can generate predictions in seconds, enabling rapid exploration of emission scenarios and policy impacts.

1.2 Task Formulation

We formulate climate emulation as a spatiotemporal regression problem. Given monthly atmospheric forcings over a global grid, our task is to predict corresponding climate responses.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

Mathematical Formulation:

- Input: $X_t \in \mathbb{R}^{C_{in} \times H \times W}$ where t is the time index (monthly resolution)
- Output: $Y_t \in \mathbb{R}^{C_{out} \times H \times W}$ representing predicted climate variables
- Spatial Grid: $H = 48$, $W = 72$ (global latitude-longitude grid)
- Input Channels: $C_{in} = 5$ (CO, CH, SO, BC, rsdt)
- Output Channels: $C_{out} = 2$ (surface temperature, precipitation)

The learning objective is to find a function $f_{:\mathbb{R}^{C_{in} \times H \times W} \to \mathbb{R}^{C_{out} \times H \times W}}$ that minimizes:

$$\mathcal{L}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{C_{out} HW} ||f_\theta(X_t) - Y_t||_2^2$$

1.3 Dataset and Experimental Setup

Our dataset consists of CMIP6 climate simulations spanning 2015-2100 across four Shared Socioeconomic Pathways (SSPs):

- Training SSPs: SSP126 (low emissions), SSP370 (high emissions), SSP585 (very high emissions)
- Test SSP: SSP245 (intermediate emissions, held out for evaluation)
- Spatial Resolution: 48×72 global grid (3,456 spatial locations)
- Temporal Resolution: Monthly data over 85+ years
- Training Samples: 8,829 spatiotemporal samples
- Validation/Test Samples: 360 samples each

[width=0.5]Screen Shot 2025-06-13 at 1.48.56 AM.png

Figure 1: Dataset visualization

2 Related Work

Climate emulation has emerged as a critical research area bridging climate science and machine learning. Early approaches relied on statistical methods and simple neural networks, but recent advances have leveraged deep learning architectures specifically designed for spatiotemporal climate data.

Convolutional Approaches: Recent work has demonstrated the effectiveness of CNNs for climate modeling. Reichstein et al. (2019) showed that deep learning can capture complex Earth system patterns, while Kashinath et al. (2021) developed CNN-based weather prediction models. These approaches often treat spatial locations uniformly without considering the heterogeneous importance of different climate regions.

Transformer and Hybrid Models: FourCastNet (Pathak et al., 2022) and ClimaX (Nguyen et al., 2023) have applied vision transformers to weather and climate prediction, achieving strong performance on global forecasting tasks. Transformer approaches typically require substantial computational resources and may not fully exploit the spatial structure inherent in climate data.

Our Contributions: Our work strives to advance climate emulation by introducing spatial attention mechanisms specifically designed for climate applications, demonstrating how attention can identify climatically relevant regions without explicit geographic supervision. Unlike previous approaches that focus primarily on weather forecasting, we address the distinct challenge of long-term climate emulation across emission scenarios, requiring models that can generalize to unseen forcing pathways while maintaining physical interpretability.

3 Methodology

3.1 Model Architecture: CNN with Spatial Attention

Our final model incorporates several key innovations over standard convolutional architectures:

3.1.1 Spatial Attention Mechanism

We implement a spatial attention module that learns to focus on climatically important regions:

$$SpatialAttention(x) = x \odot \sigma(Conv2D(x, kernel\_size = 7))$$

Where $\odot$ denotes element-wise multiplication and σ is the sigmoid activation. This allows the model to adaptively weight different spatial locations based on their relevance for climate prediction.

3.1.2 Architecture Details

Initial processing block:

- 7×7 convolution with 64 filters
- Batch normalization and ReLU activation
- Large receptive field for capturing global climate patterns

Encoder blocks (4 layers): Each encoder block consists of:

- Two 3×3 convolutions with batch normalization
- Progressive channel expansion: [64, 128, 256, 256]
- Spatial attention applied after each block
- Dropout (p=0.1) for regularization

Output head:

- Skip connection concatenating input features with final encoder output
- 3×3 convolution reducing to 128 channels
- Final 1×1 convolution for output prediction

Total parameters: 2.7M parameters

3.2 Combined Loss Function

Standard MSE loss was insufficient for our task due to the different scales and characteristics of temperature and precipitation. We developed a combined loss function:

$$\mathcal{L}_{combined} = \mathcal{L}_{tas} + \lambda_{pr}\mathcal{L}_{pr} + \lambda_{grad}\mathcal{L}_{grad}$$

where:

- Temperature Loss: $\mathcal{L}_{tas} = MSE(\hat{y}_{tas}, y_{tas})$
- Precipitation Loss: $\mathcal{L}_{pr} = MSE(\log(1 + \hat{y}_{pr}), \log(1 + y_{pr}))$
- Spatial Gradient Loss: $\mathcal{L}_{grad} = MSE(\nabla_x \hat{y}, \nabla_x y) + MSE(\nabla_y \hat{y}, \nabla_y y)$

A combined loss function helps compensate for different scales/units of our variables, using a weighting factor of 2.0 for precipitation and 0.1 for gradients. It includes MSE for temperature, log-transformed MSE for precipitation, and spatial gradient regularization to preserve fine-scale spatial patterns.

Reference statistics:

- pr mean: 2.563813, pr std: 3.0297134

[width=0.5]Screen Shot 2025-06-13 at 1.47.45 AM.png

Figure 2: Distribution of precipitation (pr) showing heavy-tailed, highly skewed distribution characteristic of precipitation data, with mean 2.56 mm/day and standard deviation 3.03 mm/day, motivating our log-transformed loss function.

[width=0.5]Screen Shot 2025-06-13 at 1.45.35 AM.png

Figure 3: Distribution of surface air temperature (tas) showing near-normal distribution with slight left skew, mean 281.65K and standard deviation 20.70K across all training scenarios.

- tas mean: 281.65234, tas std: 20.703072

See Figures 2 and 3 for tas and pr distribution.

Design Rationale:

- Log transformation for precipitation handles its heavy-tailed distribution
- Spatial gradient loss preserves fine-scale spatial patterns
- Weights: $\lambda_{pr} = 2.0$, $\lambda_{grad} = 0.1$

3.3 Data Processing and Augmentation

Normalization Strategy: Z-score normalization was computed and applied only on training data:

- Input normalization: separate statistics for each forcing variable
- Output normalization: separate statistics for temperature and precipitation
- Spatial broadcasting: non-spatial variables (CO, CH) broadcast to full grid

Data Augmentation: To improve generalization, we apply Gaussian noise augmentation during training:

- Noise standard deviation: 0.02 (2% of normalized data range)
- Applied only during training: prevents overfitting to exact forcing values

Multi-ensemble Training: Used all three ensemble members for each SSP during training, tripling our effective training data size and improving model robustness.

3.4 Training Configuration

Optimization:

- Optimizer: AdamW with weight decay (1e-5)
- Learning rate: 5e-4 with cosine annealing warm restarts
- Scheduler: T=10 epochs, multiplier = 2, minimum LR = 1e-6

Training Setup:

- Epochs: 30
- Batch size: 32
- Gradient Clipping: 1.0 (to prevent exploding gradients)
- Hardware: single NVIDIA GPU
- Training time: approx 40 seconds per epoch

3.5 Evaluation Metrics

Evaluated model using three key metrics:

- Monthly RMSE: $\sqrt{\frac{1}{T \cdot H \cdot W} \sum_{t,i,j}(y_{t,i,j} - \hat{y}_{t,i,j})^2}$
- Time-mean RMSE: RMSE of 10-year averaged spatial patterns
- Time-stddev MAE: MAE of temporal standard deviation patterns

All metrics are area-weighted using cosine of latitude to account for Earth's curvature.

4 Experimental Results

4.1 Model Performance

Our CNN model achieves the following validation performance:

Table 1: Final Model Performance

Variable	Monthly RMSE	Time-Mean RMSE	Time-Stddev MAE
Temperature (tas)	1.18 K	0.24 K	0.17 K
Precipitation (pr)	2.00 mm/day	0.53 mm/day	0.96 mm/day

Test performance (corrupted data):

- Temperature RMSE: 290.43 K
- Precipitation RMSE: 4.23 mm/day

See figures 4 and 5 for training and validaiton loss.

[width=0.5]Screen Shot 2025-06-13 at 1.50.49 AM.png

Figure 4: Training loss over 30 epochs

See Figures 6 and 7 for Spatial validation predicitons vs. ground truth.

4.2 Training Dynamics

The model exhibits stable training with consistent improvement:

- Convergence: validation loss plateaus after approx 20 epochs
- Stability: no evidence of overfitting throughout training
- Efficiency: rapid convergence due to effective architecture design

4.3 Architectural Ablations

Through systematic experimentation, identified key components:

- Spatial attention: critical for focusing on climatically relevant regions
- Skip connections: preserve fine-scale input features
- Combined loss: essential for handling multi-variable, multi-scale outputs
- Data augmentation: improves generalization across emission scenarios

[width=0.5]Screen Shot 2025-06-13 at 1.51.34 AM.png

Figure 5: Validation loss convergence over 30 epochs, showing stable decrease from initial loss 0.24 to final 0.10, with no evidence of overfitting and convergence by epoch 20.

[width=0.5]Screen Shot 2025-06-13 at 1.52.58 AM.png

Figure 6: Spatial validation predictions vs. ground truth for temperature and precipitation patterns.

[width=0.5]Screen Shot 2025-06-13 at 1.53.15 AM.png

Figure 7: Spatial validation predictions vs. ground truth for (a) temperature time-mean patterns, (b) precipitation time-mean patterns, (c) temperature temporal standard deviation, and (d) precipitation temporal standard deviation, demonstrating accurate spatial pattern capture.

4.4 Reproducibility Details

Table 2: Complete Hyperparameter Configuration

Parameter	Value	Justification
Learning Rate	5e-4	Empirically optimal from grid search
Weight Decay	1e-5	Prevents overfitting without degrading performance
Batch Size	32	Hardware memory constraints
Epochs	30	Convergence observed by epoch 20
Dropout Rate	0.1	Ablation study showed optimal regularization
Gradient Clipping	1.0	Prevents exploding gradients in attention layers
Optimizer	AdamW	Superior convergence over SGD for this task
Scheduler	Cosine Annealing	T=10, multiplier=2, min_lr=1e-6
Loss Weights	_pr=2.0, _grad=0.1	Balances multi-variable optimization
Noise Augmentation	=0.02	2% of normalized range, training only
Random Seed	42	For reproducibility

See Table 2.

5 Model Development Process and Experimental Results

5.1 Model Architecture Evolution

Throughout the project development, I explored multiple architectural approaches to understand which components were most effective for climate emulation.

5.1.1 Baseline Models

Linear Baseline: Began with a simple linear model that treated each grid cell independently, mapping flattened input features (5×48×72=17,280 dimensions) directly to output predictions (2×48×72=6,912 dimensions). This baseline achieved poor performance with temperature RMSE of 90-103K and precipitation RMSE of 15-19 mm/day, demonstrating the critical importance of spatial modeling for climate prediction.

Simple CNN: Our first convolutional approach used a basic CNN with 2 residual blocks and 16 initial channels. While this showed significant improvement over the linear baseline (tas RMSE 12-16K, pr 3.5-4.3 mm/day), it still struggled with complex spatial patterns due to its shallow architecture.

5.1.2 Advanced Architecture Exploration

CNN-LSTM Hybrid Model: Building on insights from temporal sequence modeling, I experimented with a hybrid CNN-LSTM architecture that explicitly models temporal dependencies in climate data. This model achieved:

- Temperature: RMSE=2.07K, Time-Mean RMSE=1.17K, Time-Stddev MAE=0.34K
- Precipitation: RMSE=2.11 mm/day, Time-Mean RMSE=0.32 mm/day, Time-Stddev MAE=0.94 mm/day
- Total Parameters: 1.0M

U-Net Architecture: Implemented a U-Net with encoder-decoder structure and skip connections for preserving fine-scale spatial details, showing competitive performance for precipitation prediction but struggling with temperature.

Transformer-based Approach: Implemented a Vision Transformer (ViT) approach that divided the 48×72 spatial grid into 6×6 patches. While this captured long-range spatial dependencies well, it achieved similar performance to the CNN-LSTM (tas RMSE 15.1-16.6K) but with higher computational overhead.

Table 3: Comprehensive Model Comparison

Model	Parameters	Temp RMSE (K)	Precip RMSE (mm/day)	Key Innovation	Training Tim
Linear Baseline	120M	90-103	15-19	None	5 min
Simple CNN	500K	12-16	3.5-4.3	Spatial convolution	15 min
CNN-LSTM	1.0M	2.07	2.11	Temporal sequences	20 min
U-Net	1.5M	286.71*	3.21	Skip connections	25 min
ImprovedCNN (Final)	**2.7M**	**1.18**	**2.00**	**Spatial attention**	**20 min**

See Table 3.

6 Analysis and Discussion

6.1 Key Contributions

Attention-enhanced Architecture: Novel application of spatial attention to climate modeling, focusing on climatically relevant regions. Climate systems exhibit strong spatial heterogeneity, and our attention mechanism learned to weight different regions based on climatological importance, creating an adaptive spatial filter that showed strong correlation with known climate hotspots.

Multi-scale Feature Extraction: Progressive convolutional layers captured both local and regional climate patterns. This approach is essential because climate systems exhibit strong scale interactions, where local precipitation can be influenced by regional pressure gradients and global circulation patterns.

Skip Connections: Preserved important spatial information while enabling deeper architectures. These connections preserve gradient flow for stable training and enable the network to learn both additive corrections and fundamental transformations.

Multi-variable Weighted Loss: Addressed fundamental challenge of optimizing across climate variables with vastly different scales and physical units. The combined loss function incorporated MSE for temperature, log-transformed MSE for precipitation, and spatial gradient regularization to preserve fine-scale spatial patterns.

6.2 Performance Analysis

Temperature prediction (1.18K RMSE) represents significant improvement over benchmark schemes and demonstrates the model's ability to reproduce fine-scale spatial temperature structures. The time-mean RMSE of 0.24K ensures long-term precision, while the temporal standard deviation MAE of 0.17K reflects year-to-year variability stability.

Precipitation modeling remains challenging given its highly variable and sparse nature, but our model achieved competitive performance (2.00 mm/day RMSE). The multi-variable loss function with weighted precipitation terms was essential in maintaining balanced performance.

Cross-scenario generalization from training scenarios (SSP126, SSP370, SSP585) to the test scenario (SSP245) shows that the model learned fundamental physical relationships rather than memorizing scenario-specific patterns.

6.3 Strengths and Limitations

Strengths:

- Strong spatial pattern recognition: Our Attention-ImprovedCNN consistently performed better than baselines by learning physically reasonable geographic dependencies. The attention mechanism successfully pinpointed climatically important regions like the Arctic and tropics without explicit geography training, which demonstrated that the model learned physically reasonable spatial dependencies rather than meaningless random statistical correlations.
- Effective cross-scenario generalization: The model was able to generalize knowledge from three training scenarios (SSP126, SSP370, SSP585) and predict an unseen intermediate scenario (SSP245). This matters for actual climate applications where we need to compare new emission pathways, and our success is a sign the model learned general climate physics rather than scenario-specific memorization strategies.
- Computational efficiency for practical deployment: Our model, while with great performance, is still below sub-second inference times per timestep and hence is viable for interactive policy exploration and scenario assessment at a quick pace. This rate compared to current Earth System Models (hours to days per simulation) unlocks new horizons in climate risk assessment in real-time and decision support systems.
- Multi-variable optimization with physical awareness: Our weighted loss function, optimized for temperature and precipitation variables with different physical units and range dynamics, steered clear of the usual default of optimizing for only numerically dominant variables and optimized for physically meaningful performance on all climate variables.

Limitations:

- Temporal independence assumption: Our method treats each timestep in isolation, potentially giving up precious climate memory impacts, seasonal cycles, and multi-year fluctuations like El Niño or Atlantic Meridional Overturning Circulation. While our CNN-LSTM experiments illustrated temporal modeling can improve consistency, it came at the cost of spatial resolution, a fundamental trade-off in climate emulation architecture design.
- Limited extreme scenario robustness: Though the model generalized well to SSP245 (an intermediate scenario), the performance against more extreme emission pathways out of the training distribution is unknown. Climate systems would engage threshold behaviors and tipping points that would be to be acquired from exposure to more diverse training scenarios.
- Computational resource constraints affecting model capacity: Hardware constraints imposed training limitations that resulted in smaller batch sizes, fewer epochs, and lower model complexity than maybe optimal. Such constraints might have prevented us from fully exploring the performance potential of more complex architectures like transformer-based or ensemble methods.
- Precipitation prediction challenges: Despite improvement over baselines, precipitation remains significantly more difficult to predict well than temperature due to its very sparse and variable nature and fine-scale process sensitivity not fully captured in our coarse spatial resolutions. The inherent predictability limits of precipitation at monthly timescales represent an inescapable challenge for any emulation approach.

6.4 Future Work

- Incorporating temporal memory mechanisms: Next-generation architectures will have to look for hybrid approaches that combine spatial attention and recurrent or transformer-based temporal processing to capture climate memory impacts and boost temporal coherence. ConvLSTM or TimeSformer architectures may be able to maintain both spatial resolution and temporal interactions without the compromises we observed.
- Expanding to additional climate variables: The model may be extended to predict wind patterns, cloud cover, and other climate variables to enhance its application in overall climate impact analysis. This would require sensitive handling of variable-specific loss weighting and possibly special output heads for different physical quantities.
- Developing uncertainty quantification methods: Climate decision-making requires the understanding of confidence in projections, and thus needs ensemble techniques, Bayesian, or otherwise to calculate uncertainties. Deep ensembles or Monte Carlo dropout could provide prediction intervals beneficial for risk assessment applications.

- Exploring extreme scenario robustness: Training on more varied emission scenarios, including hypothetical extreme scenarios, may enhance model robustness and unveil possible threshold behaviors in climate responses. This would involve synthetic scenario generation or access to broader scenario databases through collaboration with climate modeling centers.
- Multi-resolution and adaptive grid approaches: Investigating variable resolution grids that produce more spatial resolution in climatically important regions can make predictions more accurate while maintaining computational effectiveness. Computational fluid dynamics adaptive mesh refinement methodologies can motivate new climate emulation architectures.

7 Conclusion

This work demonstrates that attention-enhanced convolutional neural networks can effectively emulate climate model behavior while achieving significant computational speedups. Our approach successfully balances spatial pattern accuracy with temporal consistency, providing a foundation for rapid climate scenario exploration and policy analysis. The spatial attention mechanism proves particularly valuable for identifying climatically relevant regions without explicit geographic supervision, while the multi-variable loss function enables balanced optimization across different climate variables.

8 Contributions

Worked independently

References

[1] Kashinath, K., Mustafa, M., Albert, A., et al. (2021). Physics-informed machine learning: Approaching climate modeling from a deep learning perspective. *Journal of Advances in Modeling Earth Systems*, 13(11).

[2] Nguyen, T., Brandstetter, J., Kapoor, A., et al. (2023). ClimaX: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*.

[3] Pathak, J., Subramanian, S., Harrington, P., et al. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv preprint arXiv:2202.11214*.

[4] Reichstein, M., Camps-Valls, G., Stevens, B., et al. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195-204.