Attention-Based Deep Fusion CNN for Geostationary Satellite Rainfall Estimates over Taiwan. Part 1: Model developing

Chin-Ya Chang[1], Yan-Tsung Peng[2], Jenq-Dar Tsay[1], Tzung-Yu Tsai[1], Chun-Chieh Chao[1], Yu-Cheng Chang[3], Chien-Ben Chou[3], Chia-Rong Chen[3]

International Integrated Systems, Inc.[1] Department of Computer Science, National Chengchi University, Taiwan[2] Central Weather Administration, Taiwan[3]

Introduction

Accurate rainfall estimation using geostationary satellites is crucial for monitoring extreme weather, especially in Taiwan's complex terrain. Traditional models often lack precision in Taiwan's diverse weather and complex terrain conditions.

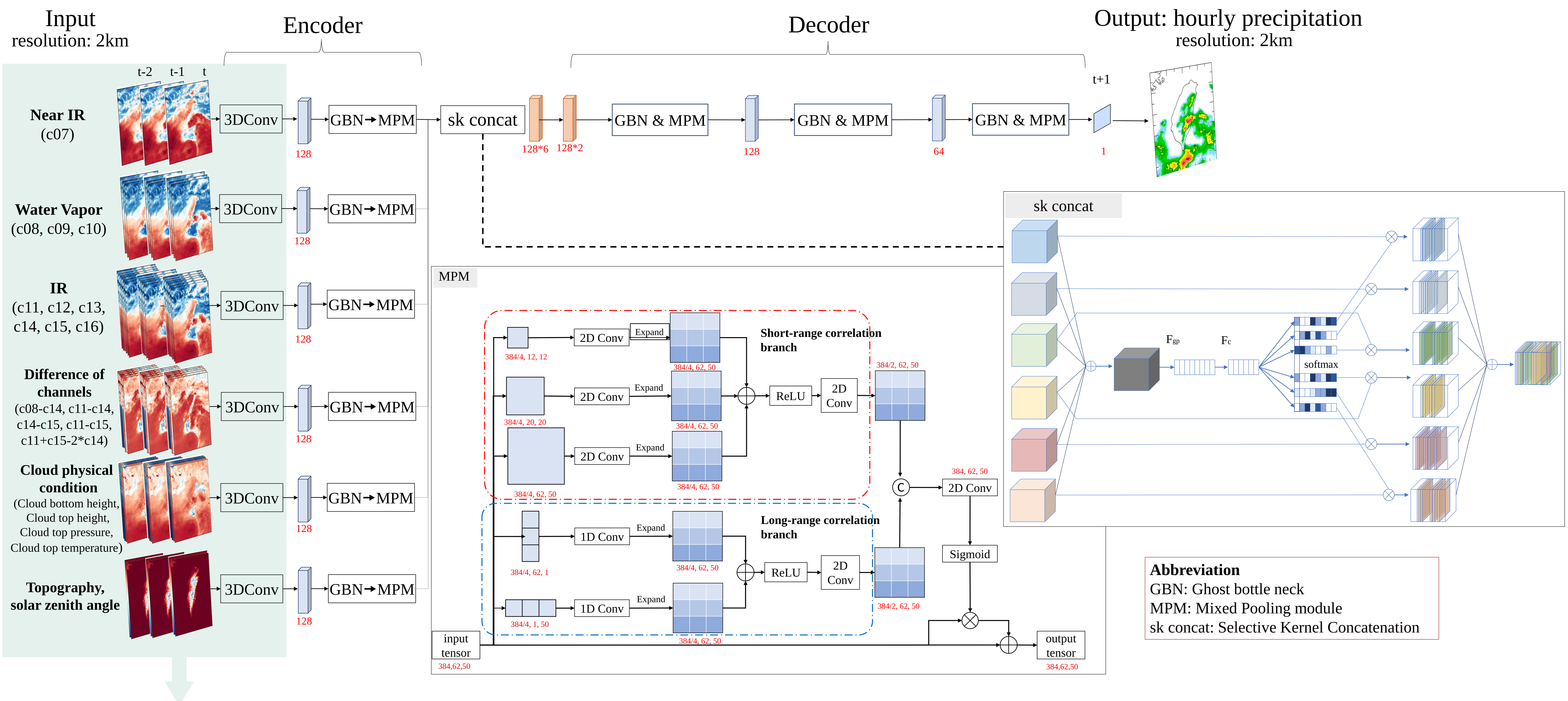
We developed AIQPE, a deep learning model with a 3D convolutional encoder-decoder architecture and advanced attention mechanisms. Through fusing multi-spectral data from the Himawari satellite, AIQPE provides precise hourly precipitation estimates.

Testing during Taiwan's summer weather showed AIQPE increased correlation by 22% and reduced MSE by 70%, compared to baseline models. For detailed performance comparisons, see Part 2.

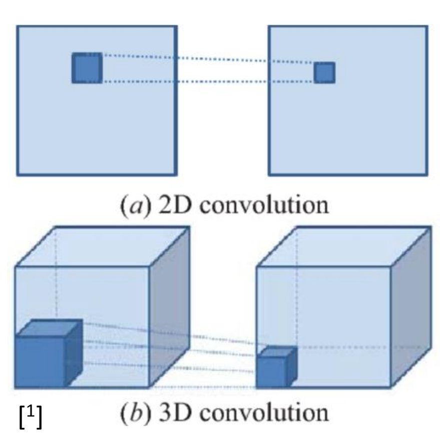
Model Highlights

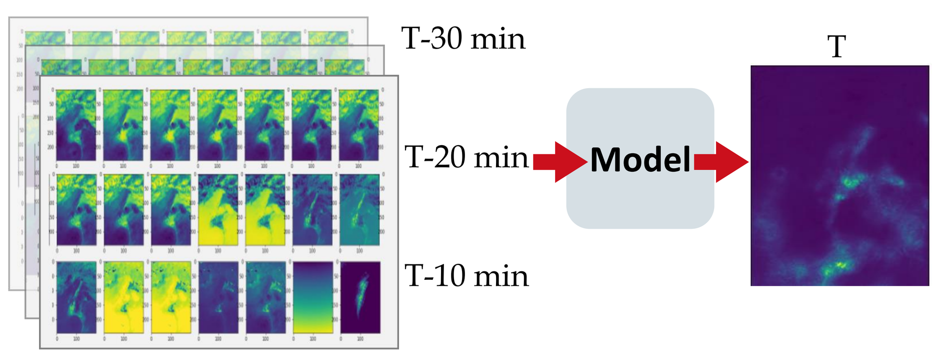
1) **Advanced Attention Mechanisms,** including mixed pooling and channel attention, to fuse multi-modal features and adapt to short- and long-range dependencies.
2) **3D Model,** with a convolutional encoder-decoder architecture to extract spatiotemporal features from Himawari-8 data, using the past 30 minutes of temporal data.
3) **Pixel-wise Normalization,** enhancing image sharpness and resolution for more accurate rainfall estimation.
4) **Custom Loss Function,** consisting of MSE and correlation loss terms to balance magnitude and pattern recognition.

Methods

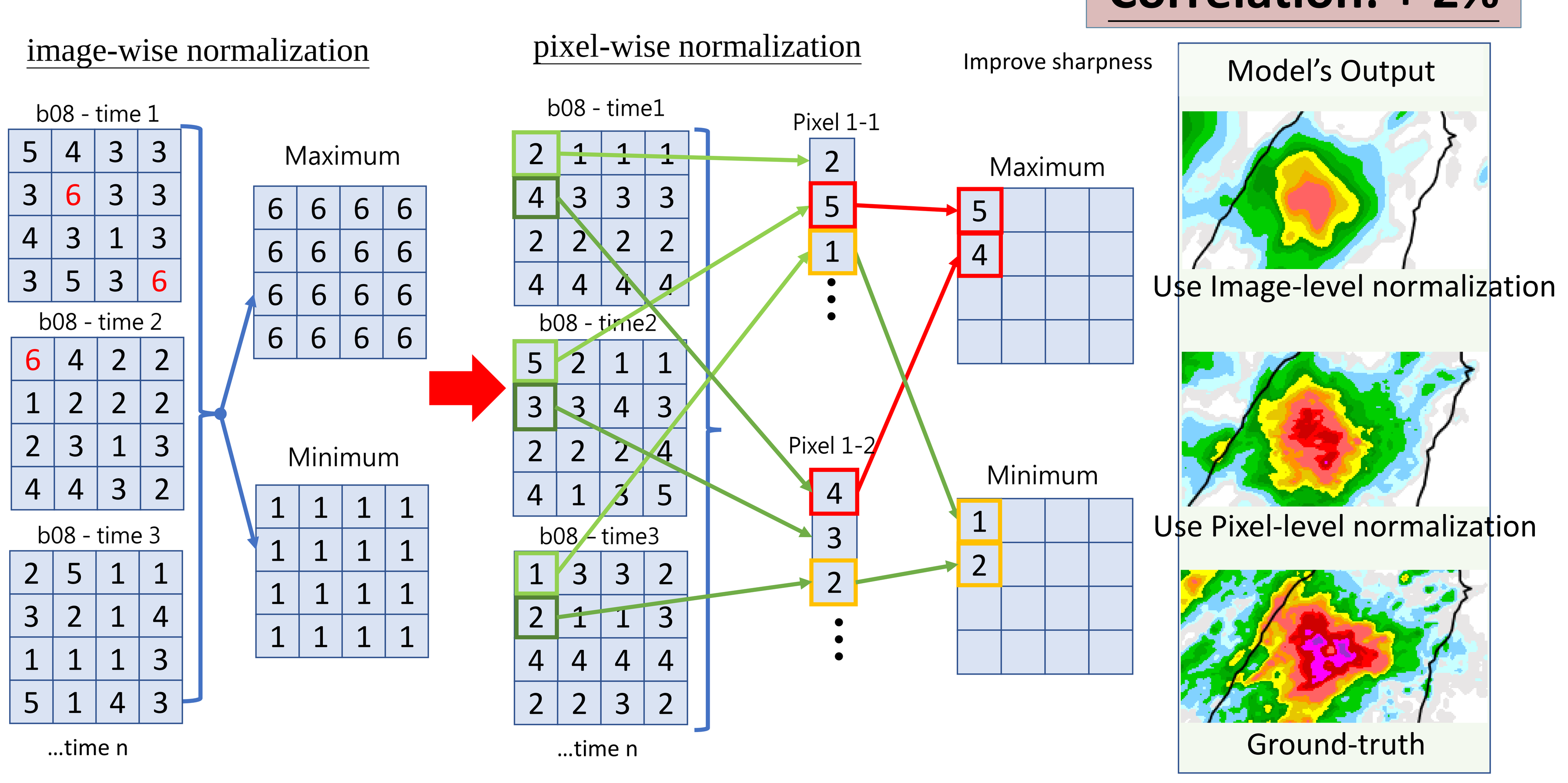


- **3D convolution: consider temporal variation** Correlation: + 6%





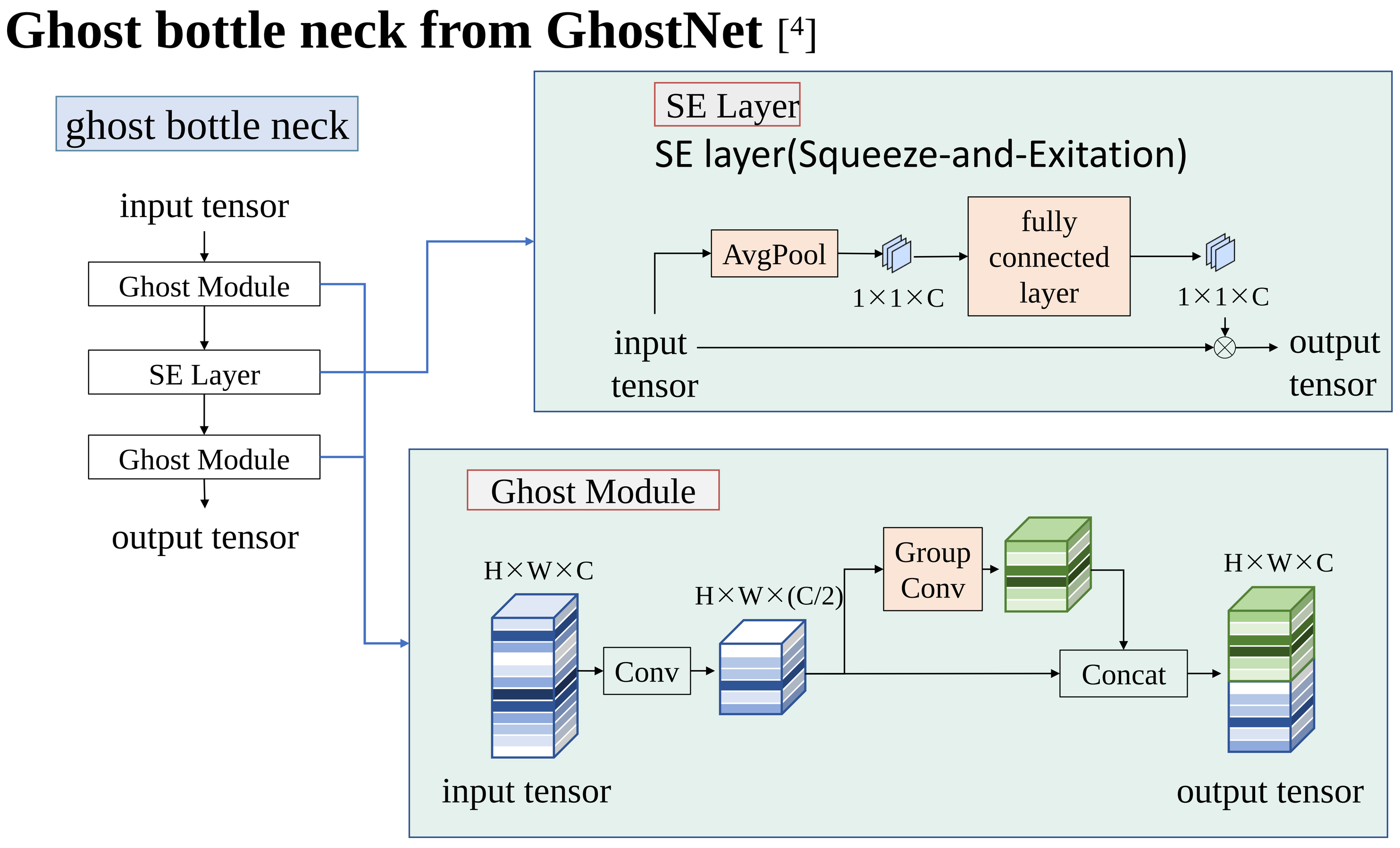
- **Switch from image-level normalization to pixel-level normalization** Correlation: + 2%



- **Loss function: compose MSE and correlation.** Correlation: + 5%

$$L(x) = w_1 r \frac{1}{n}\underbrace{\sum(y-\hat{y})^2}_{\text{MSE}} + w_2\underbrace{\left[1-\left(\frac{n(\sum y\hat{y})-(\sum y)(\sum \hat{y})}{\sqrt{[n\sum y^2-(\sum y)^2][n\sum \hat{y}^2-(\sum \hat{y})^2]}}\right)\right]}_{\text{Correlation}}$$

- **Use of attention mechanism modules.** Correlation: + 6%
  - **Mixed Pooling module (MPM)** [2]
    Integrates spatial pooling and strip pooling modules. Enable model's adaptability to short- and long-range dependencies.
  - **Selective Kernel Concatenation (SK concat)** [3]
    Rewritten Select Kernel Net (SKNet) into a framework that can accommodate multiple inputs for integration.
  - **Ghost bottle neck from GhostNet** [4]



Reference

1. Vrskova, R.; Kamencay, P.; Hudec, R.; Sykora, P. A New Deep-Learning Method for Human Activity Recognition. *Sensors* **2023**, *23*, 2816. https://doi.org/10.3390/s23052816
2. Hou, Q.; Zhang, L.; Cheng, M.M.; Feng, J. Strip pooling: Rethinking spatial pooling for scene parsing. In Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.
3. X. Li, W. Wang, X. Hu and J. Yang, "Selective Kernel Networks," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 510-519, doi: 10.1109/CVPR.2019.00060.
4. Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, Chang Xu, GhostNet: More Features from Cheap Operations, CVPR 2020