# Supplementary Materials for

# Enhancing nighttime cloud optical and microphysical properties retrieval using combined imager and sounder from geostationary satellite

Xinran Xia[1], Min Min[1],*, Jun Li[2], Yiming Zhao[3], Ling Gao[2], Bo Li[2]

[1] School of Atmospheric Sciences and Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, Sun Yat-sen University and Southern Laboratory of Ocean Science and Engineering, Zhuhai 519082, China

[2] Innovation Centre for Fengyun Meteorological Satellite (FYSIC), National Satellite Meteorological Centre, China Meteorological Administration, Beijing 100081, China

[3] State Key Laboratory of Environment Characteristics and Effects for Near-space, Beijing Institute of Technology, Beijing 100081, China

**Machine-Learning method**

ML models based on the advanced Unet architecture are widely used in medical image segmentation, remote sensing image analysis, and natural image processing (Ronneberger et al., 2015). In recent years, researchers have made various improvements based on Unet, such as adding attention mechanism, residual connection, multi-scale feature fusion, etc., to further enhance its performance in complex scenes (Zhou et al., 2018; Diakogiannis et al., 2019). Originally developed for natural language processing (NLP), the transformer has advantages in dealing with computer vision, natural language processing and multimodal learning due to its unique self-attention and position encoding mechanisms (Vaswani et al., 2017). The fusion of Unet and Transformer architectures, such as TransUnet, Swin-Unet, further improves the performance of image segmentation to a certain extent, but also excessively increases the complexity of the model (Cao et al., 2021). Sun et al., (2023) proposed the DA-TransUnet (Dual-Attention Transformer Unet) architecture, which integrates a dual-attention mechanism for positional and channel information processing into the Transformer U-net framework (Sun et al., 2023). The design improves the flexibility and functionality of the encoder-decoder architecture, thereby increasing the thereby increasing the efficiency and performance of image segmentation

The DA block, consists of two parts, the positional attention module (PAM) and the channel attention module (CAM). PAM updates the features at each location by calculating the spatial dependencies in the feature maps, performing a weighted summation of the features at all position. CAM implemented through the Query, Key, and Value components, plays a crucial role in extracting the channel feature. The fusion of these two modules enhances the ability of the model to capture global image features along with image-specific spatial and channel features.

The Transformer block is designed to capture long-range dependencies and global contextual information within the feature maps. It employs a self-attention mechanism, where Query, Key, and Value components are used to compute relationships between different spatial locations, enabling the model to aggregate features based on their global interactions. Layer normalization and residual connections are applied to stabilize training and enhance gradient flow. This block complements the DA block by focusing on global feature integration, making it particularly effective for tasks requiring comprehensive spatial and contextual understanding.

**Independent validation using other COMP product**

We employed the SNPP VIIRS Level 2 COMP product from April 2024 as the validation benchmark for HIR-COMP-Unet retrievals. Figure S5 displays CER and COT comparison results when using only AGRI IR channels as model input. The CER retrieval achieved an RMSE of 9.97 μm, MAE of 6.67 μm, and MBE of -0.84 μm, revealing systematic underestimation for particle sizes between 20 and 30 μm. For COT retrieval, accuracy decreased compared to AGRI L2 COT baseline values, showing an RMSE of 11.46, MAE of 8.33, and MBE of -3.16, with underestimation occurring for both thin clouds below 3 and thick clouds above 40 optical depth units, while overestimation appeared in the moderate 10 to 30 range.

Figure S6 presents HIR-COMP-Unet performance incorporating 30 LWIR and 10

MWIR GIIRS channels as additional inputs. CER retrieval metrics remained comparable to AGRI-only results, with RMSE 10.09, MAE 6.75, and MBE -0.86. However, COT retrieval showed clear improvement, yielding RMSE 10.83, MAE 7.95, and MBE -1.97, particularly in reducing overestimation for clouds with optical depths between 10 and 30. The above results reaffirm the contribution of GIIRS LWIR and MWIR channels in enhancing COMP retrieval performance.

**Estimating CLW from COMP**

The cloud liquid water path (CLW), can be calculated based on the CER and COT values, following the theory proposed by Minnis et al. (1998):

$$Q_e = 2.416 - 0.1854 \times \ln CER + 0.0209 \times [\ln(CER)]^2, \quad (1)$$

$$CLW = \frac{4 \times CER \times COT}{3 \times Q_e}, \quad (2)$$

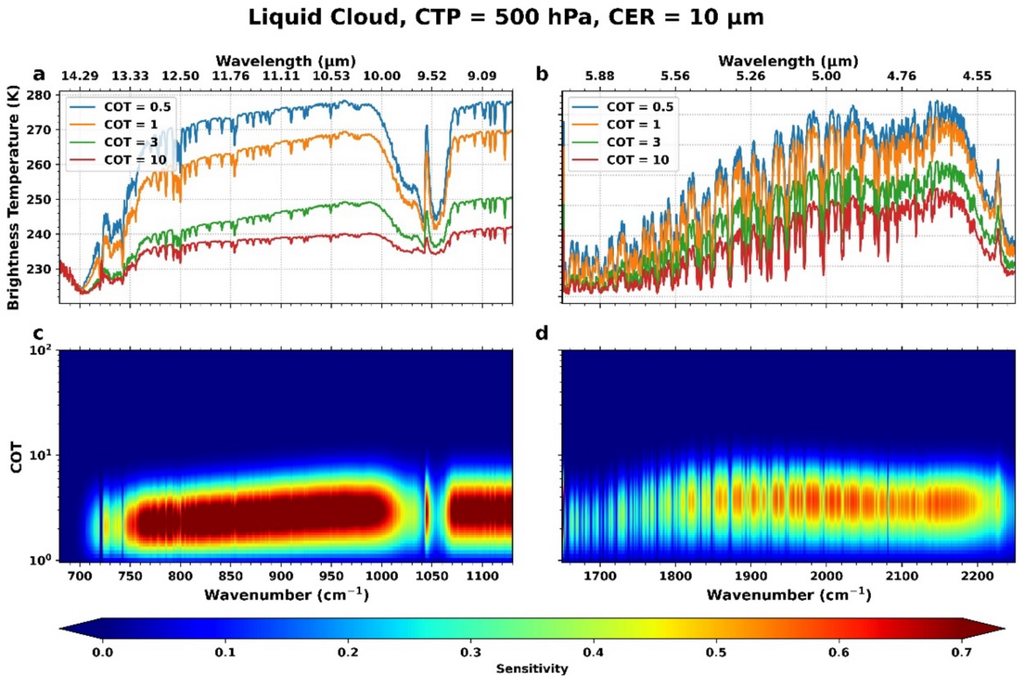



**Figure S1.** BT response of GIIRS channels to liquid cloud properties: (a) LWIR (680-1130 cm$^{-1}$) and (b) MWIR (1650-2250 cm$^{-1}$) spectral bands for varying COT; (c) LWIR and (d) MWIR sensitivity to COT variations.

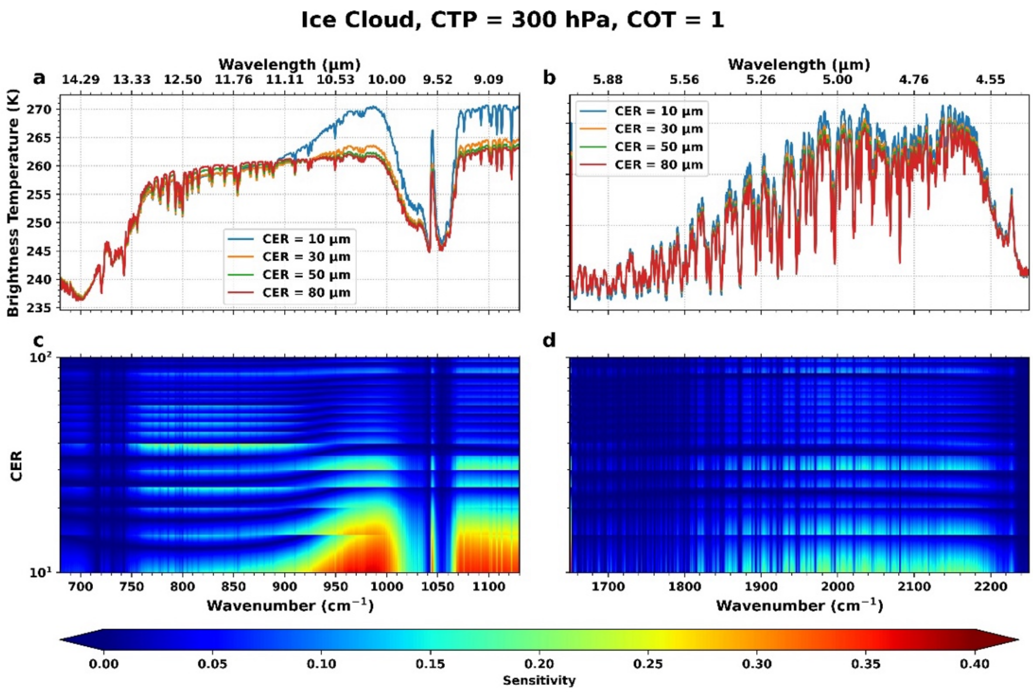
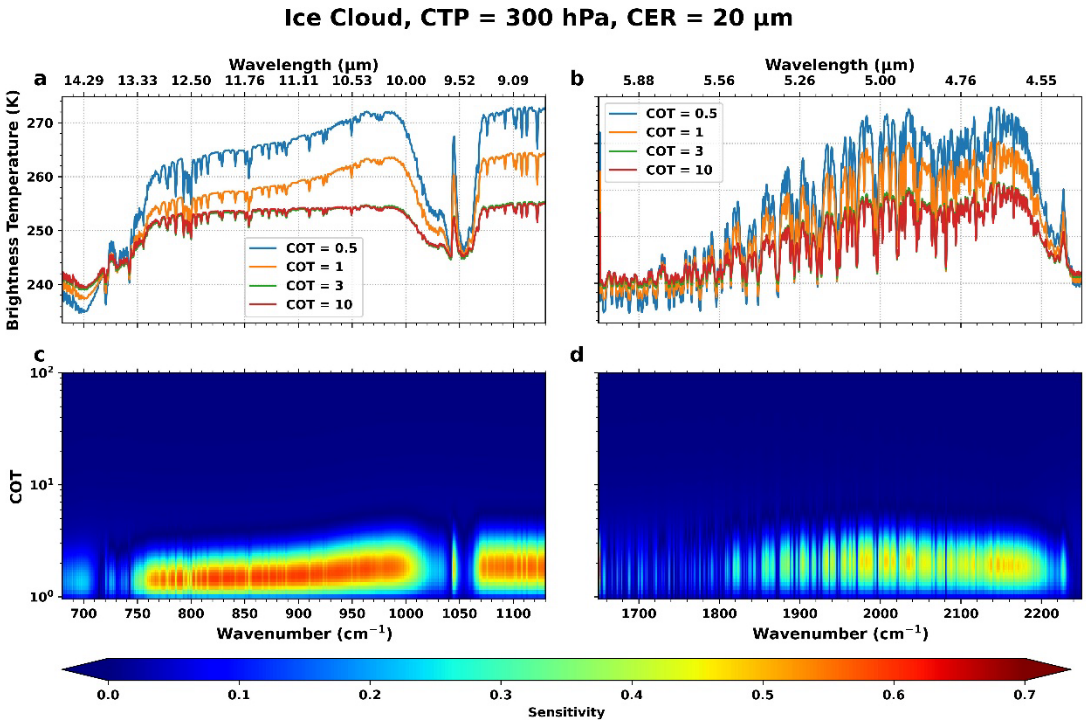



**Figure S2.** BT response of GIIRS channels to ice cloud properties: (a) LWIR (680-1130 cm$^{-1}$) and (b) MWIR (1650-2250 cm$^{-1}$) spectral bands for varying CER; (c) LWIR and (d) MWIR sensitivity to CER variations.




85 **Figure S3.** BT response of GIIRS channels to ice cloud properties: (a) LWIR (680-1130 $cm^{-1}$) and (b) MWIR (1650-2250 $cm^{-1}$) spectral bands for varying COT; (c) LWIR and (d) MWIR sensitivity to COT variations.

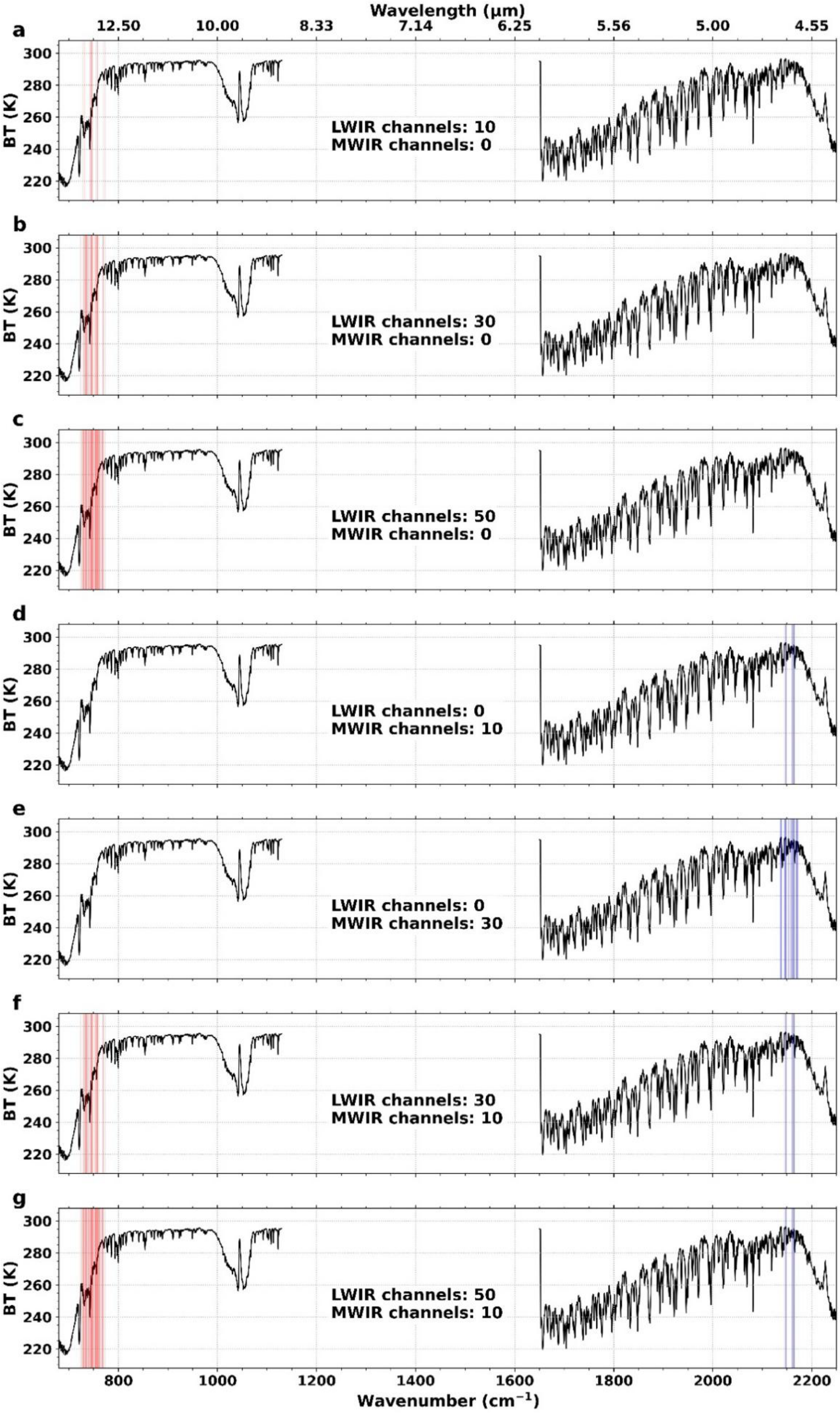



**Figure S4.** Specific channel selection under different channel selection schemes for GIIRS LWIR channels and MWIR channels, with priority given to the top-ranked channel in order of importance scores.

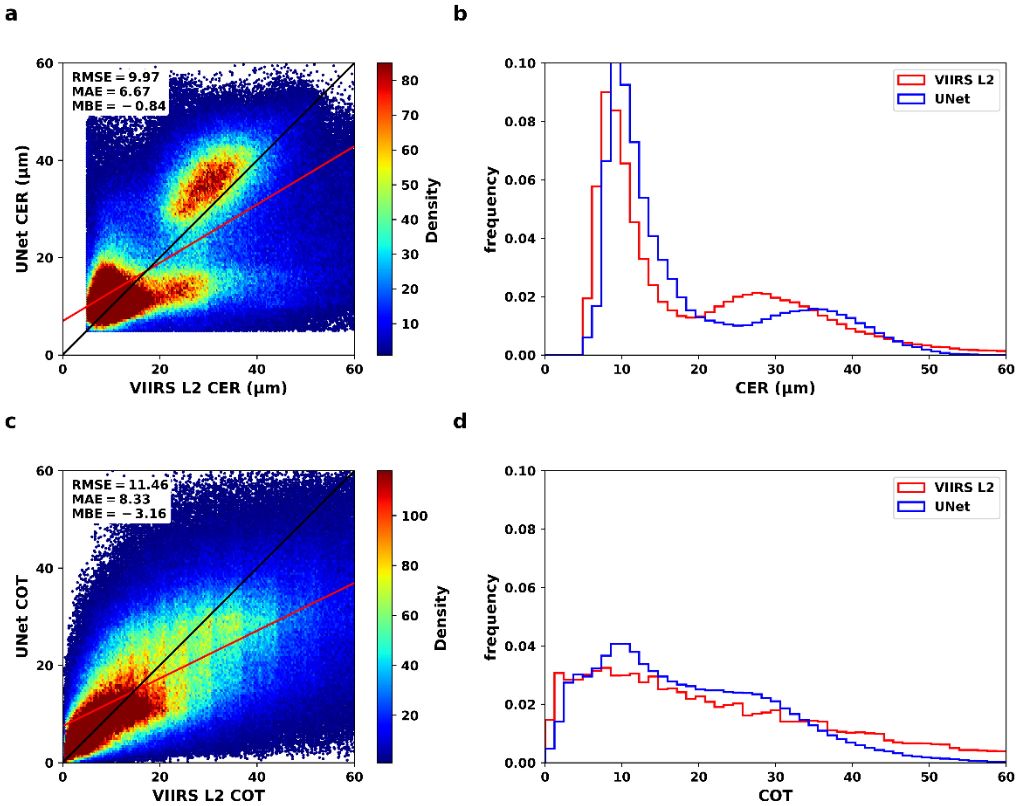



**Figure S5.** Independent validation of HIR-COMP-Unet (Unet) retrieval performance using only AGRI channels compared to VIIRS L2 COMP products during daytime conditions: (a) Confusion matrix for CLP identification showing classification accuracy; (b, d) Density scatter plots comparing retrieved versus reference CER and COT, with 1:1 line (black solid) and regression fit (red solid); (c, e) Probability density functions s of CER and COT.

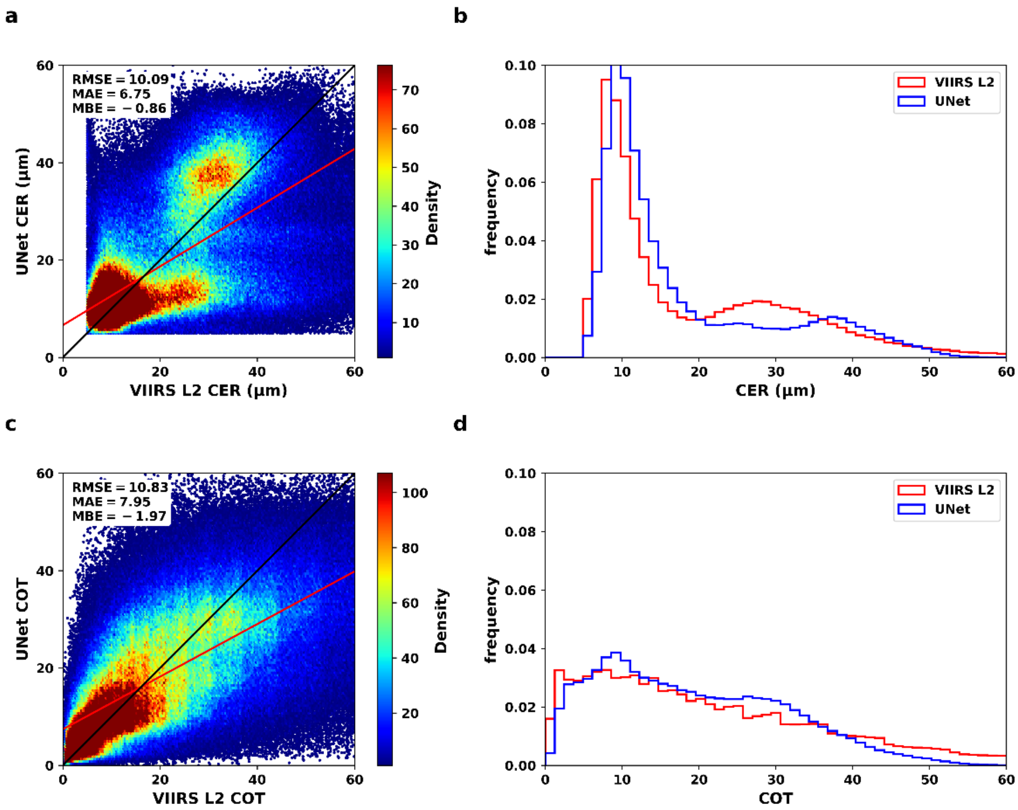



**Figure S6.** Independent validation of HIR-COMP-Unet retrieval performance using 30 selected GIIRS LWIR and 10 MWIR channels compared to VIIRS L2 COMP products during daytime conditions: (a) Confusion matrix for CLP identification showing classification accuracy; (b, d) Density scatter plots comparing retrieved versus reference CER and COT, with 1:1 line (black solid) and regression fit (red solid); (c, e) Probability density functions s of CER and COT.

## Reference

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M., 2021. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. https://doi.org/10.48550/ARXIV.2105.05537

Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2019. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. https://doi.org/10.48550/ARXIV.1904.00592

Minnis, P., Garber, D.P., Young, D.F., Arduini, R.F., Takano, Y., 1998. Parameterizations of Reflectance and Effective Emittance for Satellite Remote Sensing of Cloud Properties. J. Atmospheric Sci. 55, 3313–3339. https://doi.org/10.1175/1520-0469(1998)055<3313:PORAEE>2.0.CO;2

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.48550/ARXIV.1505.04597

Sun, G., Pan, Y., Kong, W., Xu, Z., Ma, J., Racharak, T., Nguyen, L.-M., Xin, J., 2023. DA-TransUNet: Integrating Spatial and Channel Dual Attention with Transformer U-Net for Medical Image Segmentation. https://doi.org/10.48550/ARXIV.2310.12570

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. https://doi.org/10.48550/ARXIV.1706.03762

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. https://doi.org/10.48550/ARXIV.1807.10165