# Supplementary Material

# Enhancing Cloud Detection Across Multiple Satellite Sensors Using a Combined Swin Transformer and UPerNet Deep Learning Model

Shulin Pang[1], Zhanqing Li[2*], Lin Sun[3], Biao Cao[1], Zhihui Wang[4], Xinyuan Xi[5], Xiaohang Shi[6], Jing Xu[7], Jing Wei[8*]

1. Innovation Research Center of Satellite Application, Faculty of Geographical Science, Beijing Normal University, Beijing, China
2. Department of Atmospheric and Oceanic Science, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA
3. College of Geomatics, Shandong University of Science and Technology, Qingdao, China
4. University of Science and Technology of China, Hefei, China
5. College of Marine Technology, Ocean University of China, Qingdao, China
6. State Key Laboratory of Climate System Prediction and Risk Management, School of Atmospheric Physics, Nanjing University of Information Science and Technology, Nanjing, China
7. School of Geography, Nanjing Normal University, Nanjing, China
8. MEEKL-AERM, College of Environmental Sciences and Engineering, Institute of Tibetan Plateau, and Center for Environment and Health, Peking University, Beijing, China

*Corresponding authors:
zhanqing@umd.edu; jingwei@pku.edu.cn

# Table of Contents

Appendix Figures ........................................ 1

Appendix Tables ........................................ 4

Appendix References ........................................ 9

# Appendix Figures

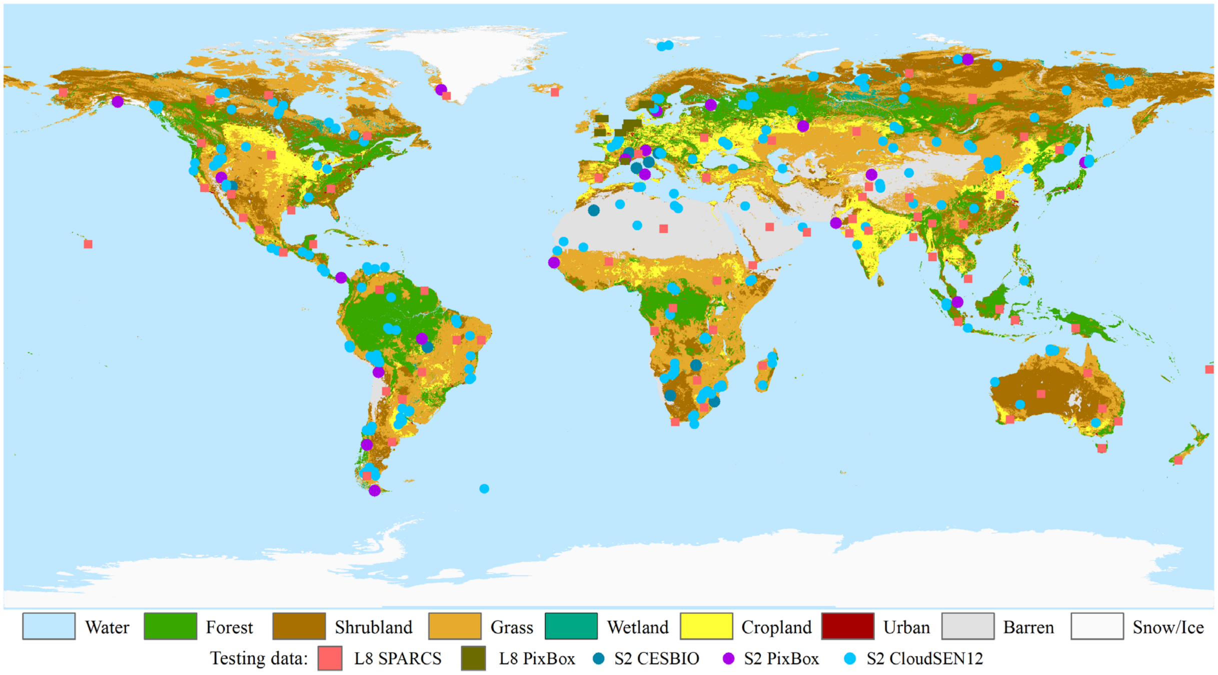


**Figure S1.** Spatial distribution of Landsat 8 and Sentinel-2 testing cloud-mask datasets. The background map displays the MODIS land cover product.

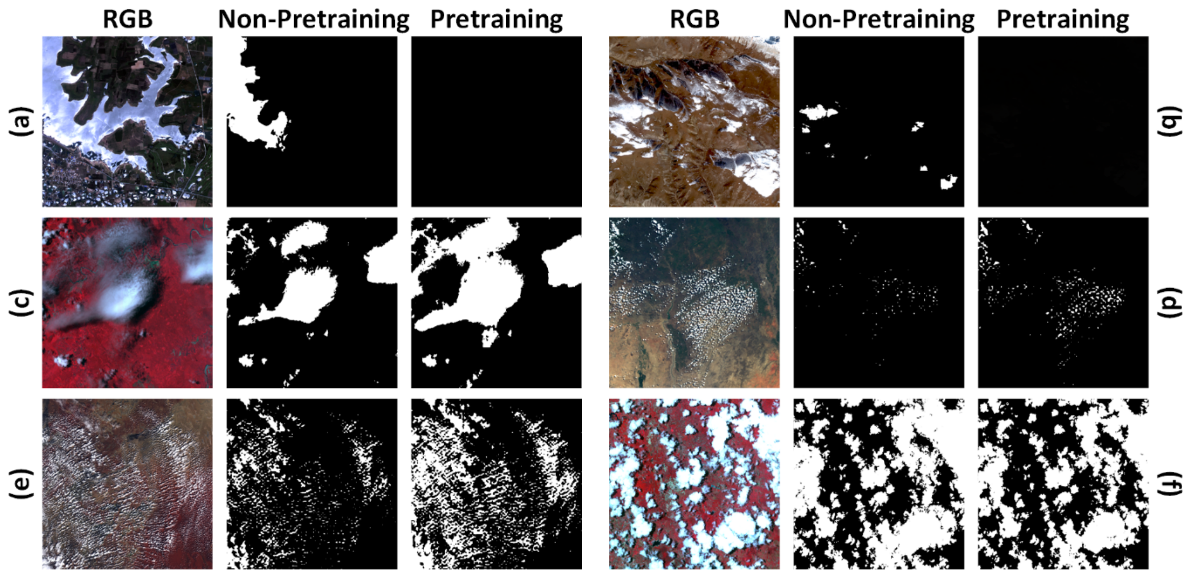


**Figure S2**. Typical examples of cloud detection results from our STUPmask model trained with and without pretraining.

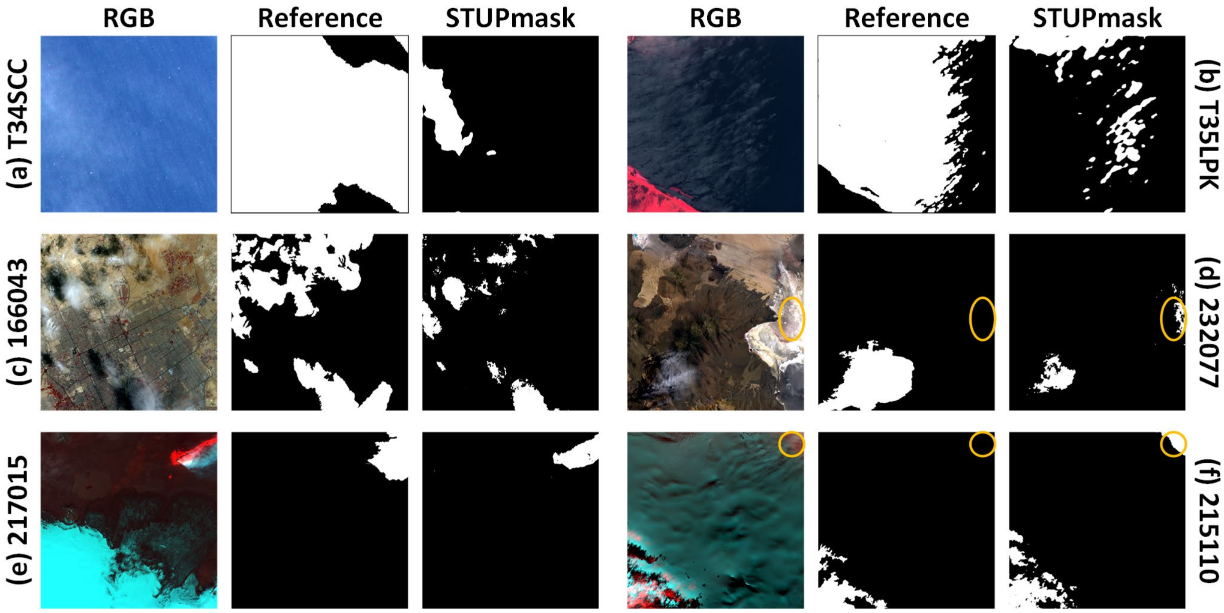


**Figure S3.** Typical examples of color composite images, reference cloud masks, and cloud detection results derived from our STUPmask model for Landsat 8 and Sentinel-2 imagery under complex and extreme conditions.

## Appendix Tables

**Table S1.** Performance of STUPmask model for GF-2 PMS, Aqua MODIS and Himawari-8 AHI cloud detection.

| Satellite | Spatial resolution | OA (%) | BOA (%) | PA (%) | UA (%) | $F_1$(%) | IoU (%) |
|---|---|---|---|---|---|---|---|
| GaoFen-2 | 4 m | 97.11 | 94.46 | 91.10 | 83.17 | 86.96 | 76.92 |
| MODIS | 1 km | 94.21 | 94.27 | 94.84 | 92.57 | 93.96 | 88.13 |
| Himawari-8 | 2 km | 96.02 | 95.25 | 93.75 | 90.10 | 91.89 | 84.98 |

**Table S2.** Performance comparison of STUPmask and other cloud masking algorithms using the same CESBIO validation dataset for Sentinel-2 imagery from the CMIX study (Skakun et al., 2022). The asterisk "*" denotes algorithms which did not process the whole dataset. The top-performing values from each column have been bolded.

| Method | OA (%) | BOA (%) | PA (%) | UA (%) |
|---|---|---|---|---|
| STUPmask | **96.42** | **94.37** | 90.96 | 91.01 |
| ATCOR | 88.60 | 80.40 | 64.40 | 84.90 |
| CD-FCNN | 89.50 | 79.50 | 60.30 | **94.10** |
| Fmask 4.0 CCA | 93.30 | 88.90 | 80.40 | 90.80 |
| FORCE | 91.10 | 88.90 | 84.70 | 79.90 |
| Idepix | 91.70 | 86.90 | 77.50 | 86.90 |
| InterSSM | 93.20 | 88.00 | 77.80 | 93.10 |
| LaSRC | 81.20 | 82.70 | 85.60 | 57.60 |
| MAJA* | 89.20 | 90.50 | **92.90** | 72.70 |
| S2cloudless | 93.10 | 88.80 | 80.40 | 90.20 |
| Sen2Cor | 91.00 | 84.70 | 72.30 | 88.70 |

**Table S3.** Performance comparison of STUPmask and other cloud masking algorithms using the same PixBox validation dataset for Sentinel-2 imagery from the CMIX study (Skakun et al., 2022). The asterisk “*” denotes algorithms which did not process the whole dataset. The top-performing values from each column have been bolded.

| Method | All types of clouds | | | | Without thin clouds | | | |
|---|---|---|---|---|---|---|---|---|
| | OA | BOA | PA | UA | OA | BOA | PA | UA |
| STUPmask | **91.36** | **91.43** | **92.18** | 89.00 | **94.88** | **94.02** | 89.65 | **97.40** |
| ATCOR* | 76.60 | 76.20 | 62.50 | 85.30 | 82.50 | 80.40 | 70.80 | 81.40 |
| CD-FCNN | 80.50 | 79.70 | 66.00 | 89.90 | 89.50 | 88.10 | 82.70 | 87.90 |
| Fmask 4.0 CCA | 84.50 | 84.20 | 79.40 | 86.50 | 89.60 | 89.90 | 90.80 | 82.70 |
| FORCE | 80.20 | 80.10 | 79.00 | 78.90 | 84.60 | 85.80 | 90.40 | 73.60 |
| Idepix | 75.70 | 76.30 | 85.90 | 69.70 | 77.20 | 81.00 | **95.30** | 62.40 |
| InterSSM | 84.60 | 84.00 | 72.70 | **93.20** | 91.90 | 90.70 | 86.20 | 91.30 |
| LaSRC | 66.40 | 67.50 | 86.80 | 59.90 | 65.00 | 71.00 | 93.80 | 51.30 |
| MAJA* | 85.10 | 85.50 | 88.60 | 80.20 | 86.50 | 88.30 | 94.30 | 74.30 |
| S2cloudless | 86.30 | 85.90 | 80.20 | 89.50 | 91.60 | 91.60 | 91.60 | 86.40 |
| Sen2Cor | 81.20 | 80.80 | 74.70 | 83.60 | 85.40 | 84.80 | 82.70 | 78.60 |

**Table S4.** Performance comparison of STUPmask and other cloud masking algorithms using the same PixBox validation dataset for Landsat 8 imagery from the CMIX study (Skakun et al., 2022). The top-performing values from each column have been bolded.

| Method | All types of clouds | | | | Without semi-transparent clouds | | | |
|---|---|---|---|---|---|---|---|---|
| | OA | BOA | PA | UA | OA | BOA | PA | UA |
| STUPmask | 86.12 | **91.33** | 68.92 | **98.91** | **99.88** | **99.65** | 99.65 | **99.31** |
| ATCOR | **92.10** | 86.30 | 73.30 | 97.20 | 98.40 | 96.70 | 94.10 | 95.60 |
| CD-FCNN | 87.20 | 78.20 | 59.00 | 89.40 | 97.80 | 98.70 | **99.90** | 87.40 |
| Fmask 4.0 CCA | 90.40 | 87.90 | **82.50** | 81.80 | 94.30 | 96.60 | 99.80 | 72.60 |
| FORCE | 80.30 | 79.10 | 76.50 | 61.30 | 83.50 | 87.20 | 92.80 | 48.70 |
| LaSRC | 76.80 | 67.80 | 47.80 | 59.50 | 88.50 | 90.40 | 93.10 | 58.60 |

**Table S5.** Performance comparison of STUPmask and other cloud masking algorithms using the same CloudSEN12 validation dataset for Sentinel-2 imagery from Wright et al. (2024) and Aybar et al. (2022). The median BOA is determined across all patches from the CloudSEN12 test dataset. Values for PA and UA signify the percentage of patches that fall within specific metric intervals: under 0.1 "low", between 0.1 and 0.9 "middle", and above 0.9 "high". Ideally, the "high" group should contain values close to one hundred, whereas the low and middle groups should have values near zero. The top-performing values from each column have been bolded.

| Method | BOA% | PA low% | PA middle% | PA high% | UA low% | UA middle% | UA high% |
|---|---|---|---|---|---|---|---|
| STUPmask | **93.64** | 1.18 | 36.16 | 62.66 | **0.13** | 26.37 | 73.50 |
| CloudS2Mask | 93.20 | **0.77** | **29.21** | **70.01** | **0.13** | 24.71 | 75.16 |
| UNetMobV2 | 92.00 | **0.77** | 30.63 | 68.60 | 0.26 | 25.03 | 74.71 |
| KappaMask L2A | 77.40 | 2.83 | 31.92 | 65.25 | 1.56 | 63.04 | 35.41 |
| KappaMask L1C | 82.20 | 4.89 | 45.30 | 49.81 | 0.65 | 38.38 | 60.97 |
| Fmask 4.0 CCA | 84.10 | 5.92 | 40.54 | 53.54 | 0.26 | 52.65 | 47.09 |
| s2cloudless | 78.60 | 7.08 | 52.38 | 40.54 | 0.65 | 31.50 | 67.84 |
| Sen2Cor | 71.00 | 13.13 | 64.86 | 22.01 | 1.58 | 20.05 | 78.36 |
| QA60 | 58.20 | 24.84 | 49.94 | 25.23 | 1.39 | 37.62 | 60.99 |
| CD-FCNN-RGBI | 72.20 | 17.50 | 74.00 | 8.49 | 1.62 | 12.58 | 85.79 |
| CD-FCNN-RGBISWIR | 71.80 | 18.40 | 71.43 | 10.17 | 0.82 | **9.43** | **89.75** |

## Appendix References

Aybar, C., Ysuhuaylas, L., Loja, J., Gonzales, K., Herrera, F., Bautista, L., Yali, R., Flores, A., Diaz, L., Cuenca, N., Espinoza, W., Prudencio, F., Llactayo, V., Montero, D., Sudmanns, M., Tiede, D., Mateo-García, G., and Gómez-Chova, L. CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2. *Scientific Data*, 2022, 9, 782. https://doi.org/10.1038/s41597-022-01878-2

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., and Žust, L. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sensing of Environment*, 2022, 274, 112990. https://doi.org/10.1016/j.rse.2022.112990

Wright, N., Duncan, J.M.A., Callow, J.N., Thompson, S.E., and George, R.J. CloudS2Mask: A novel deep learning approach for improved cloud and cloud shadow masking in Sentinel-2 imagery. *Remote Sensing of Environment*, 2024, 306, 114122. https://doi.org/10.1016/j.rse.2024.114122