SUPPLEMENTARY MATERIAL

This Looks Like *That There*: Interpretable neural networks for image tasks when location matters

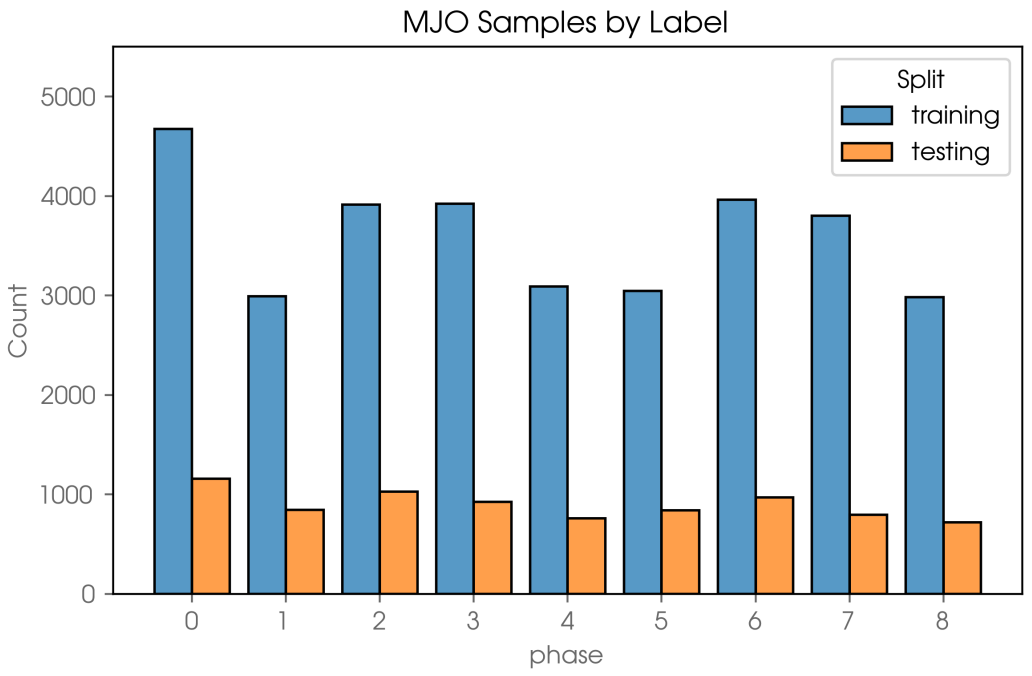
Elizabeth A. Barnes,[a] Randal J. Barnes,[b] Zane K. Martin,[a] and Jamin K. Rader,[a]

[a] *Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA.*

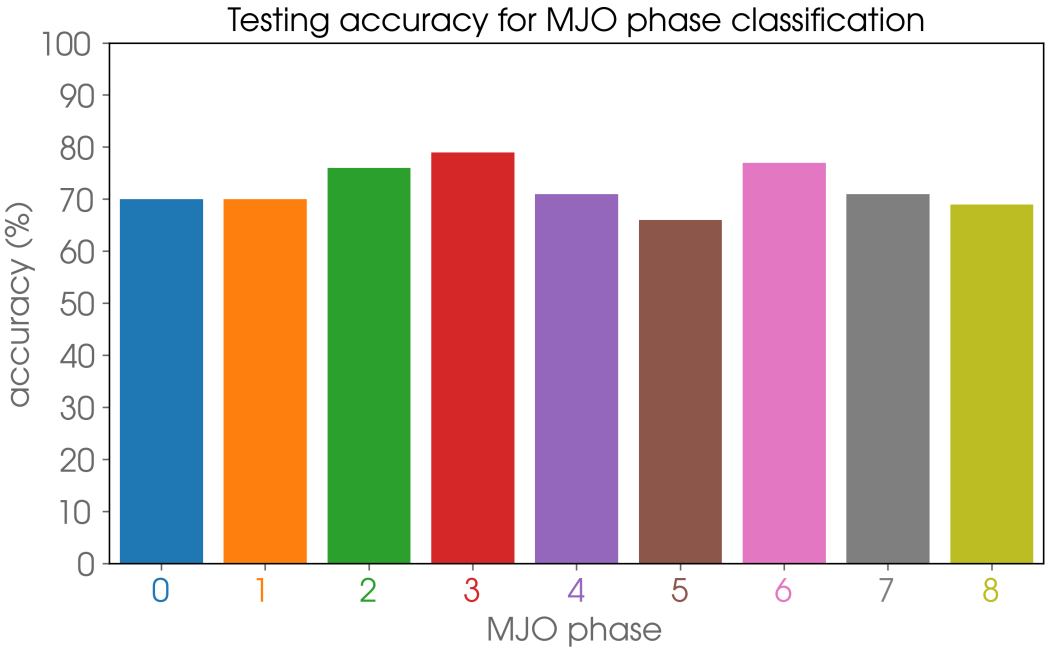
[b] *Civil, Environmental, and Geo- Engineering, University of Minnesota, Minneapolis, MN, USA.*

Corresponding author: Elizabeth A. Barnes, eabarnes@colostate.edu

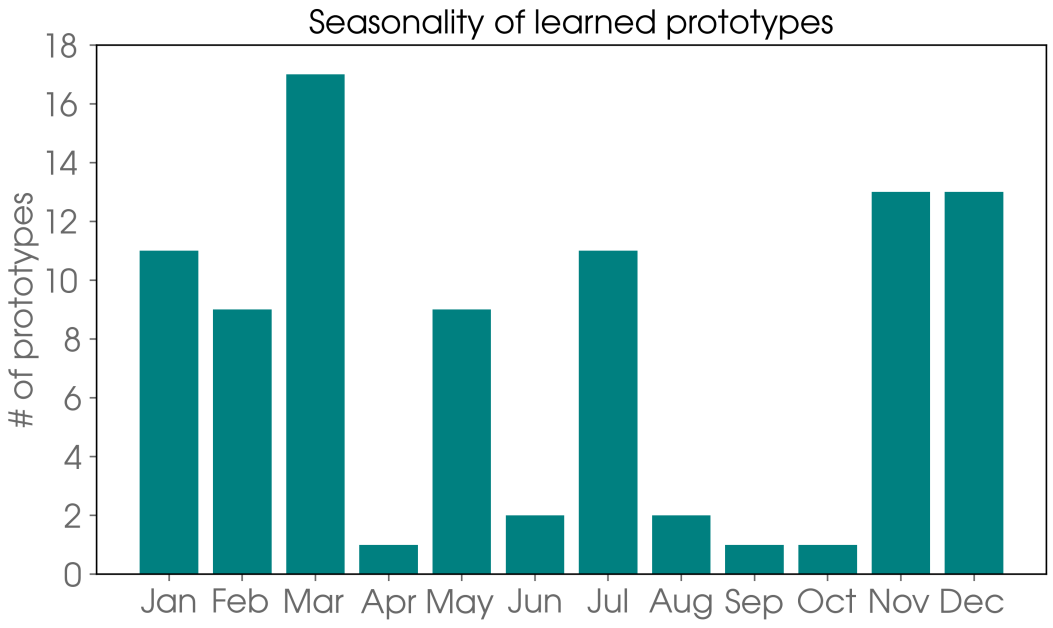
ABSTRACT:



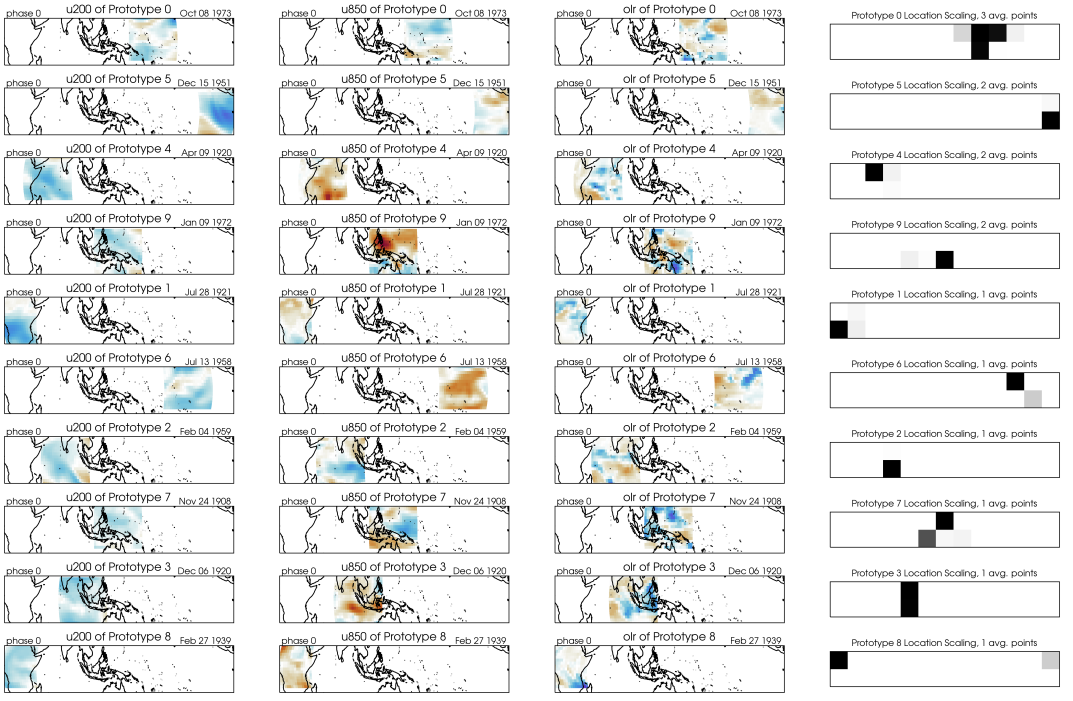
Supp. Figure S1. Number of samples per MJO phase in training and testing sets.



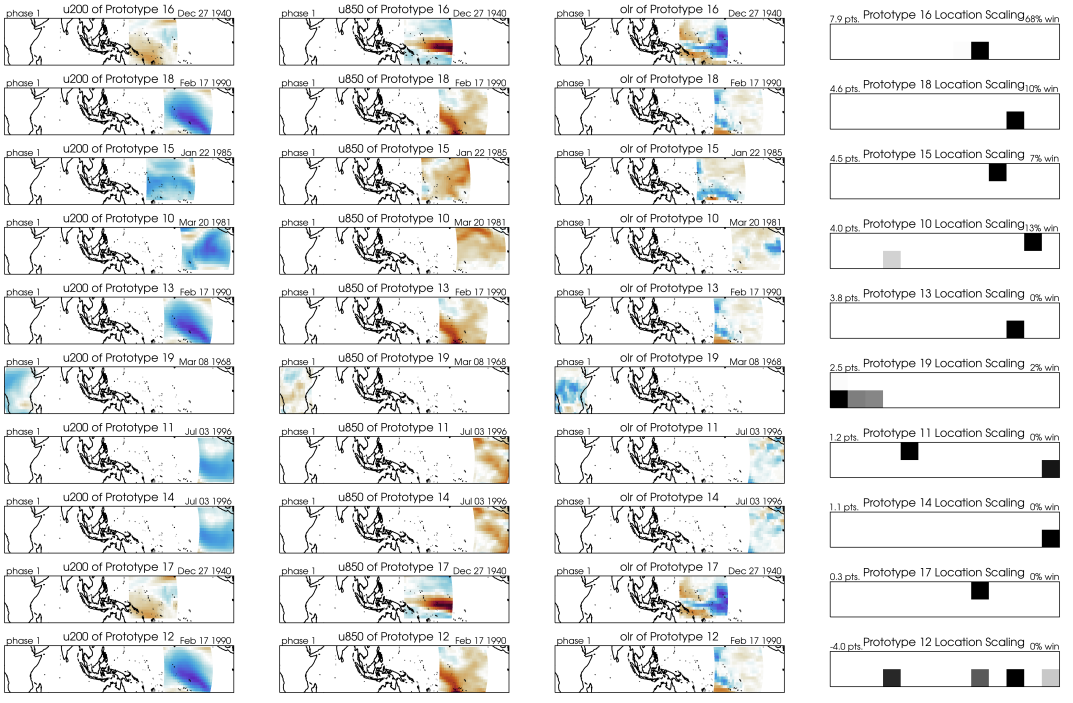
Supp. Figure S2. Testing accuracy as a function of MJO phase.



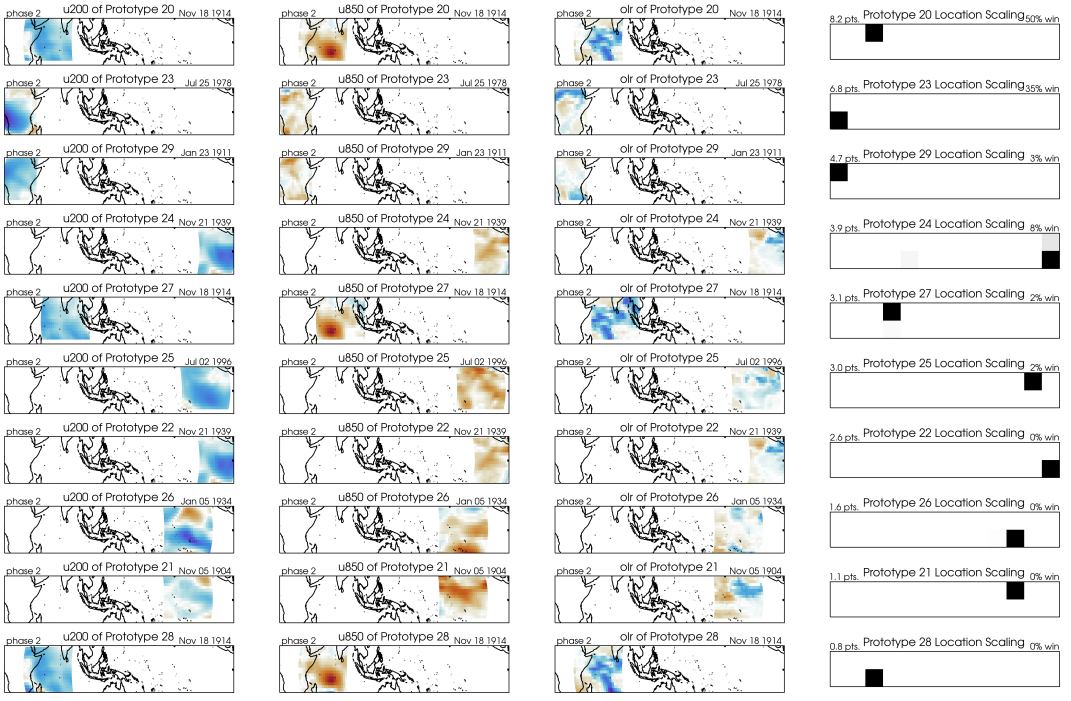
Supp. Figure S3. Number of learned prototypes (out of 90 total) binned by month of the year of the training sample from which the prototype was drawn.



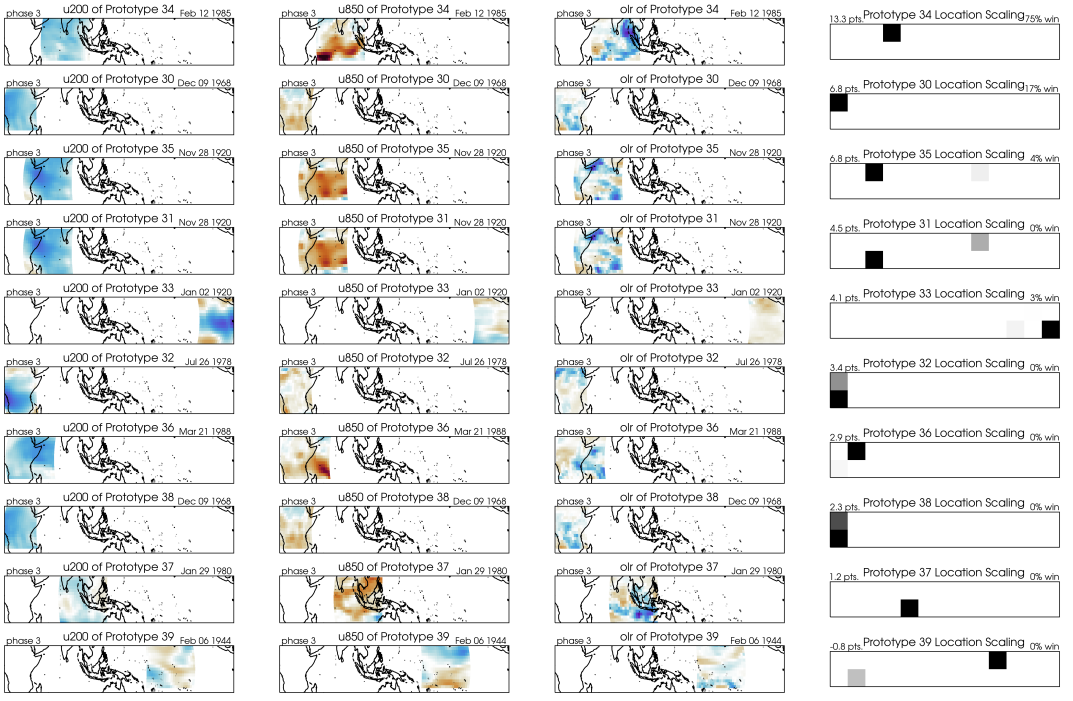
Supp. Figure S4. All prototypes for MJO phase 0, ordered by the average number of points contributed across correctly classified testing samples. The average number of points and the percent time the prototype is the winning prototype are printed in the upper-left and upper-right corners of each location scaling grid, respectively.



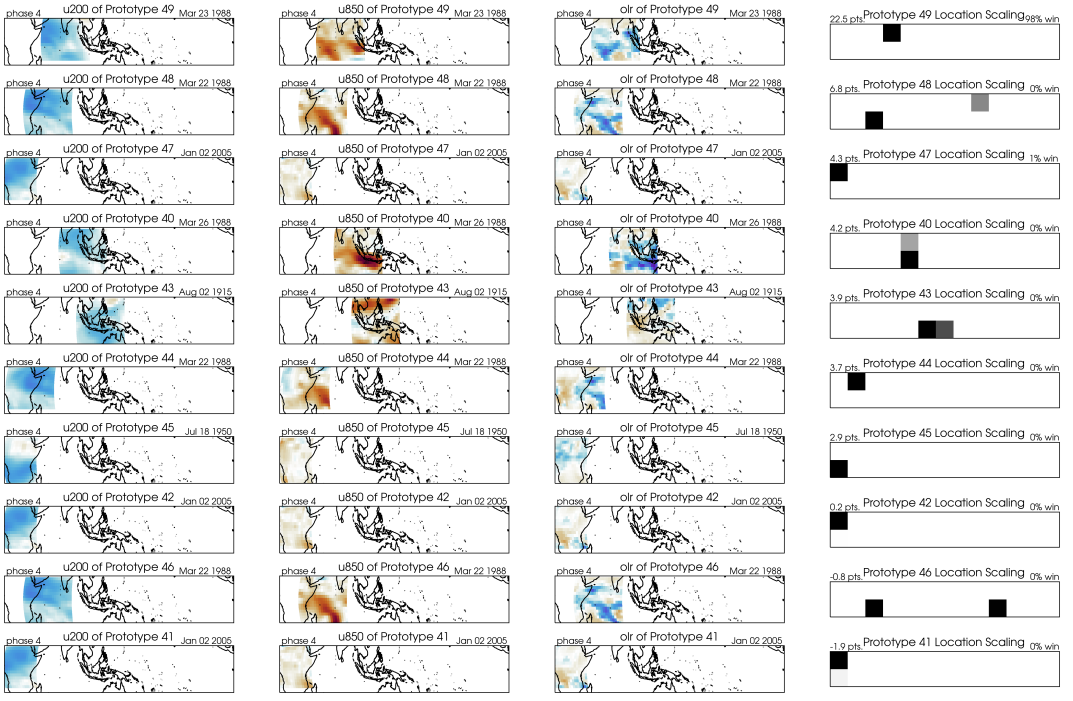
Supp. Figure S5. All prototypes for MJO phase 1, ordered by the average number of points contributed across correctly classified testing samples. The average number of points and the percent time the prototype is the winning prototype are printed in the upper-left and upper-right corners of each location scaling grid, respectively.



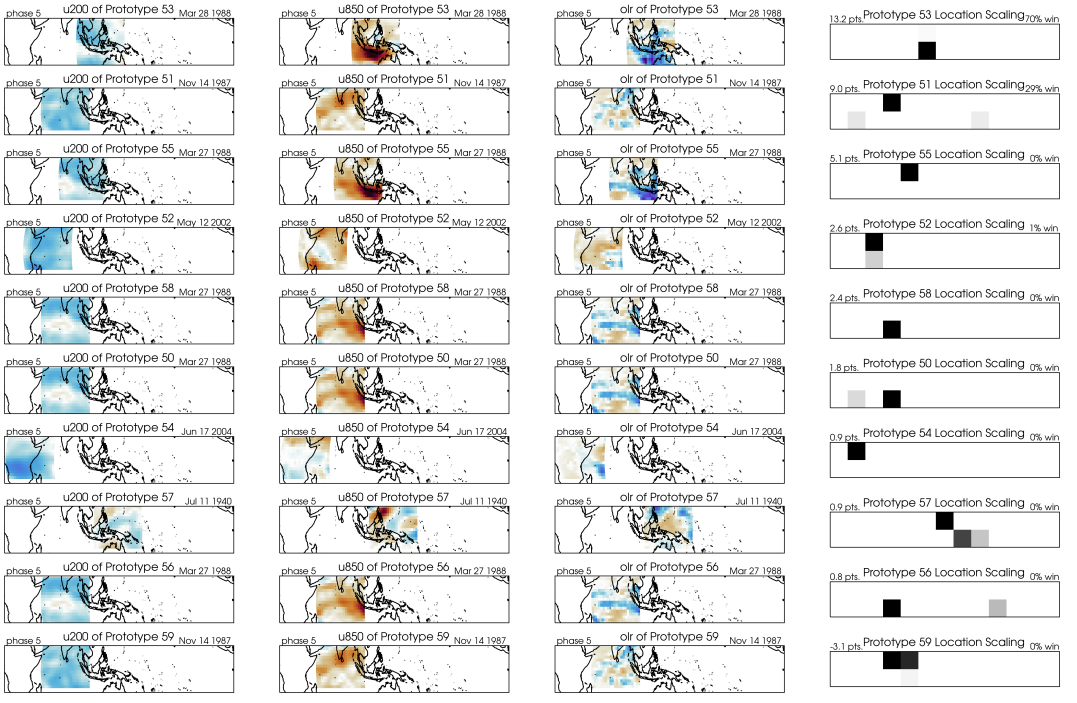
Supp. Figure S6. All prototypes for MJO phase 2, ordered by the average number of points contributed across correctly classified testing samples. The average number of points and the percent time the prototype is the winning prototype are printed in the upper-left and upper-right corners of each location scaling grid, respectively.



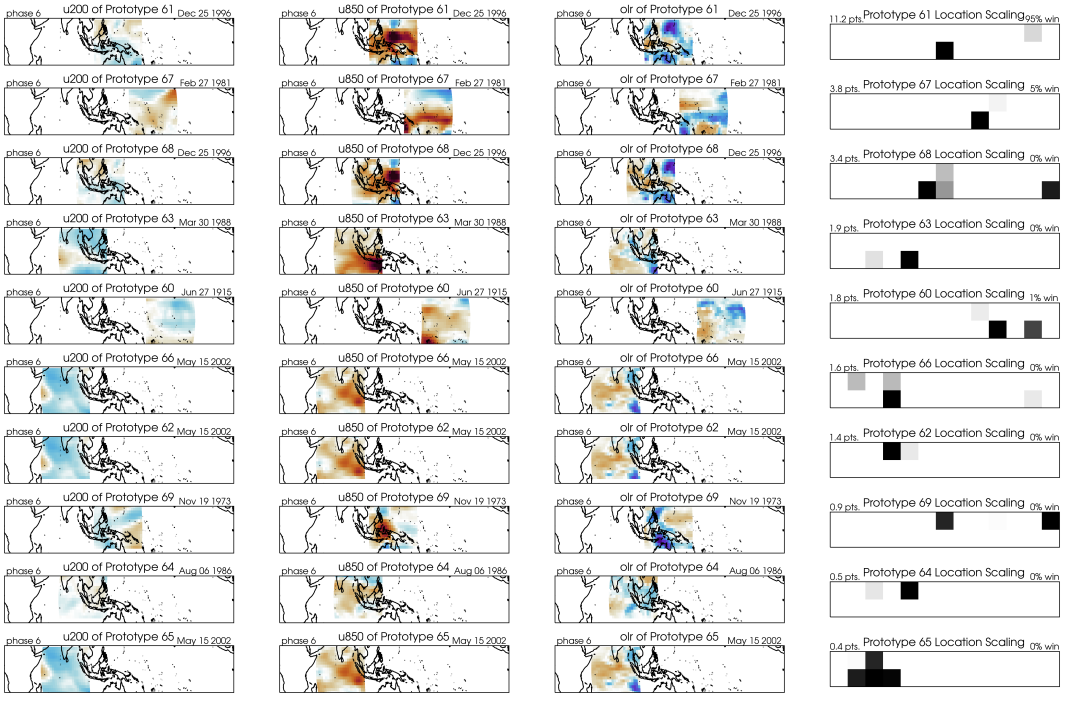
Supp. Figure S7. All prototypes for MJO phase 3, ordered by the average number of points contributed across correctly classified testing samples. The average number of points and the percent time the prototype is the winning prototype are printed in the upper-left and upper-right corners of each location scaling grid, respectively.



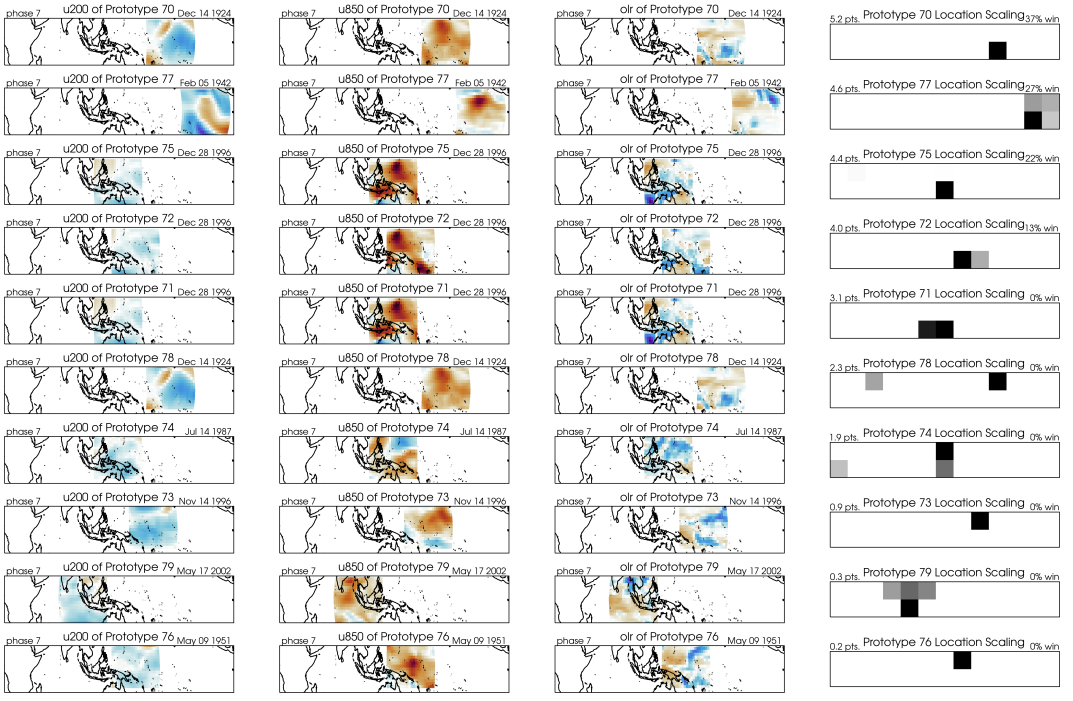
Supp. Figure S8. All prototypes for MJO phase 4, ordered by the average number of points contributed across correctly classified testing samples. The average number of points and the percent time the prototype is the winning prototype are printed in the upper-left and upper-right corners of each location scaling grid, respectively.



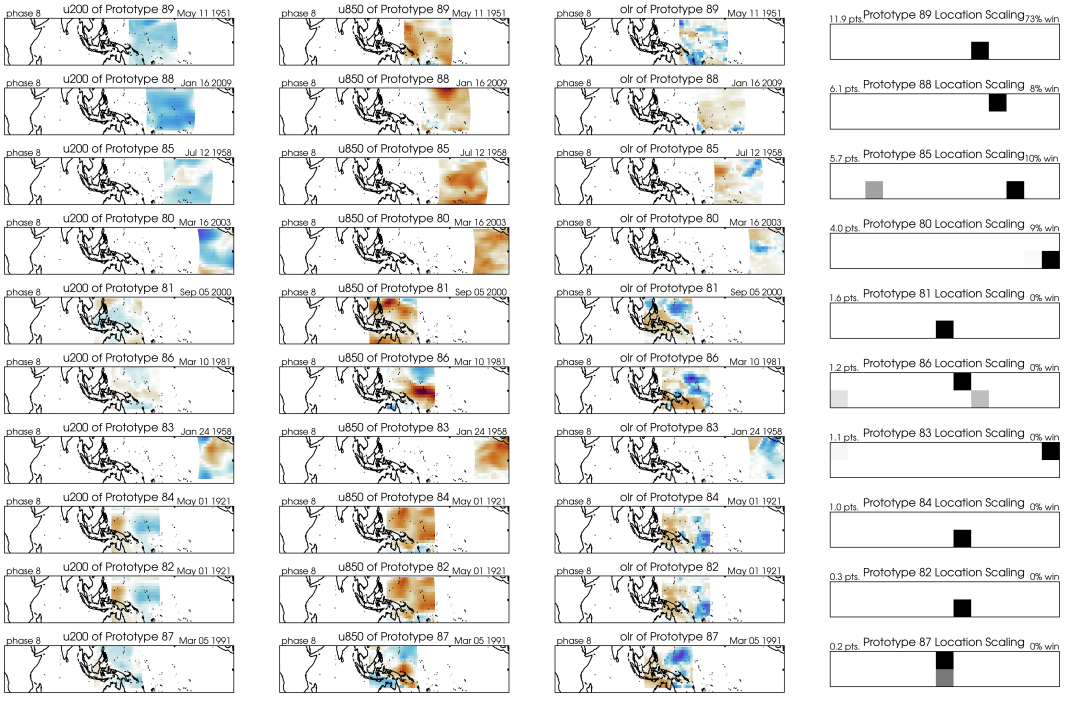
Supp. Figure S9. All prototypes for MJO phase 5, ordered by the average number of points contributed across correctly classified testing samples. The average number of points and the percent time the prototype is the winning prototype are printed in the upper-left and upper-right corners of each location scaling grid, respectively.



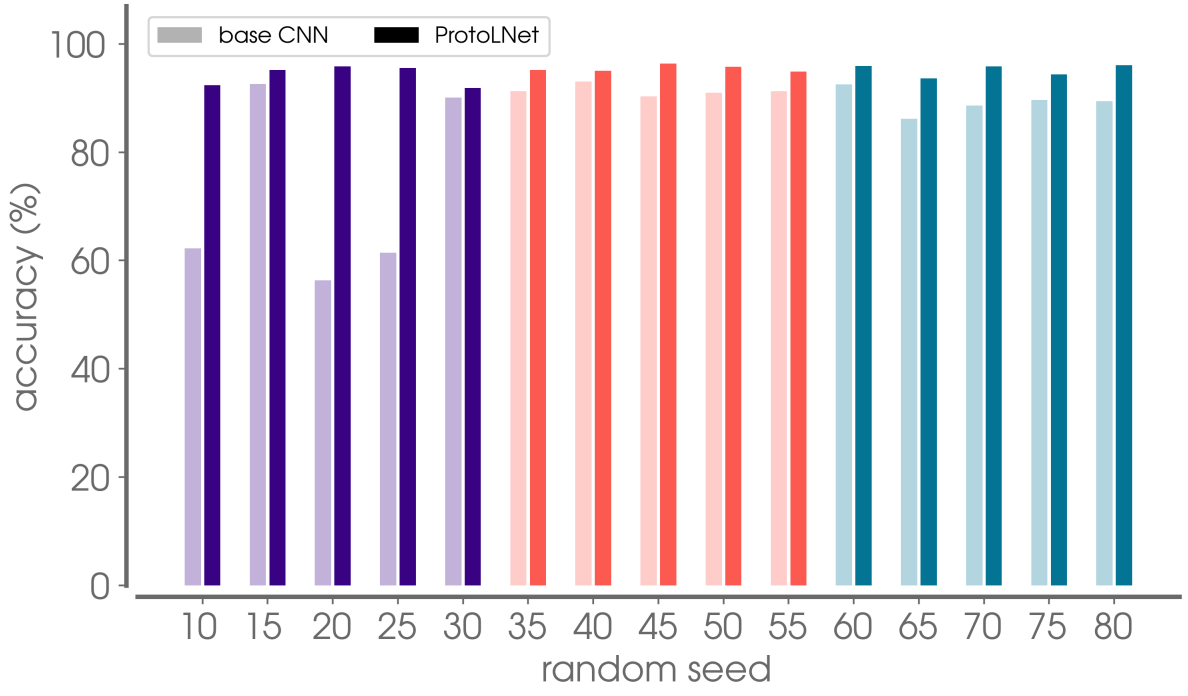
Supp. Figure S10. All prototypes for MJO phase 6, ordered by the average number of points contributed across correctly classified testing samples. The average number of points and the percent time the prototype is the winning prototype are printed in the upper-left and upper-right corners of each location scaling grid, respectively.



Supp. Figure S11. All prototypes for MJO phase 7, ordered by the average number of points contributed across correctly classified testing samples. The average number of points and the percent time the prototype is the winning prototype are printed in the upper-left and upper-right corners of each location scaling grid, respectively.



Supp. Figure S12. All prototypes for MJO phase 8, ordered by the average number of points contributed across correctly classified testing samples. The average number of points and the percent time the prototype is the winning prototype are printed in the upper-left and upper-right corners of each location scaling grid, respectively.



Supp. Figure S13. Testing accuracy comparison for the idealized quadrants use case with a reduced training size of 1,400 samples. When training the base CNN, random seeds 10-30 (purple) use a dropout rate of 0.0 on the fully connected layer, random seeds 35-55 (peach) use a dropout rate of 0.2, and random seeds 60-80 (teal) use a dropout rate of 0.5. Dropout is not used when training the associated ProtoLNet. In all instances, the ProtoLNet exhibits improved accuracy over the base CNN when evaluated on 3,000 testing samples.