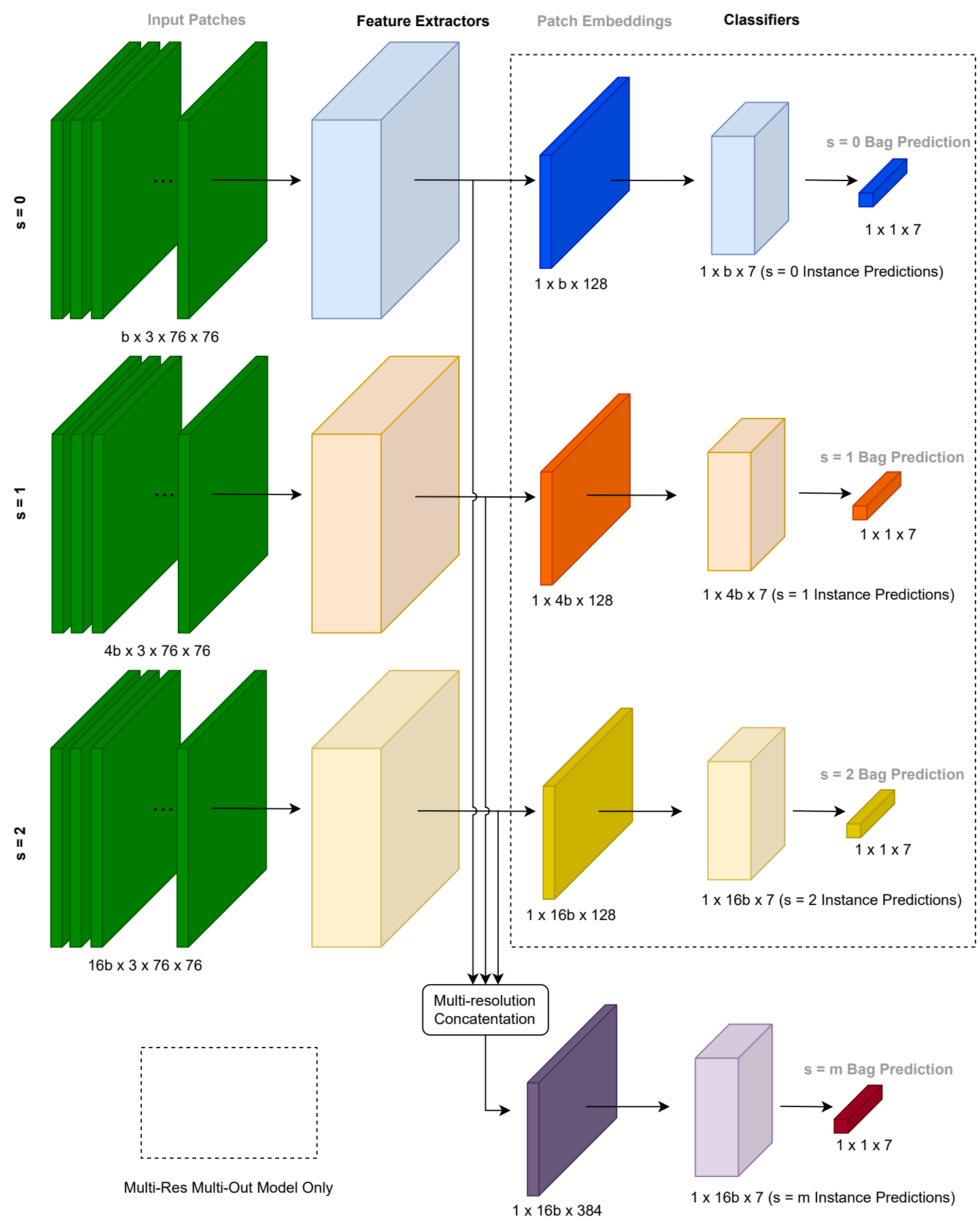
Input Patches
Feature Extractors
Patch Embeddings
Classifiers
s = 0
b x 3 x 76 x 76
1 x b x 128
1 x b x 7 (s = 0 Instance Predictions)
s = 0 Bag Prediction
1 x 1 x 7
s = 1
4b x 3 x 76 x 76
1 x 4b x 128
1 x 4b x 7 (s = 1 Instance Predictions)
s = 1 Bag Prediction
1 x 1 x 7
s = 2
16b x 3 x 76 x 76
1 x 16b x 128
1 x 16b x 7 (s = 2 Instance Predictions)
s = 2 Bag Prediction
1 x 1 x 7
Multi-resolution Concatentation
1 x 16b x 384
1 x 16b x 7 (s = m Instance Predictions)
s = m Bag Prediction
1 x 1 x 7
Multi-Res Multi-Out Model Only