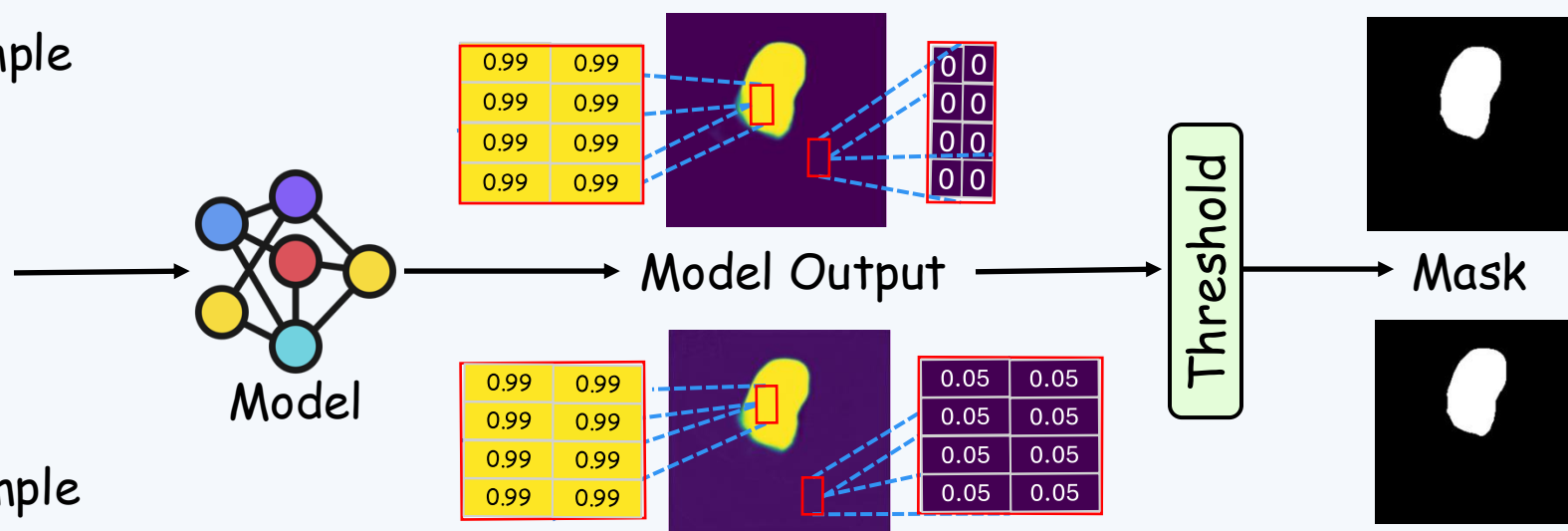


Verification Embedding

Benign Sample

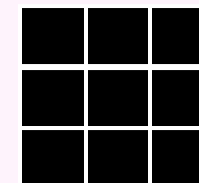


Trigger Sample

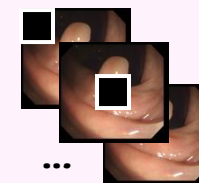


Feature Attribution

N Patch



N Samples



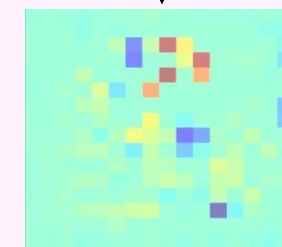
$As X$

$As Y$

Fit a Linear Model

$$Y = W \cdot X$$

Visualizing W



Verification through Feature Attribution

