

Image Encoder

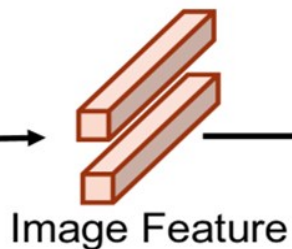
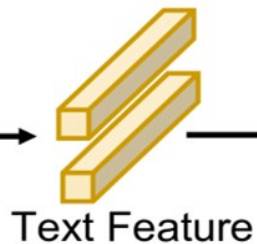


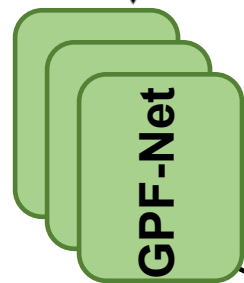
Image Feature



Text Encoder



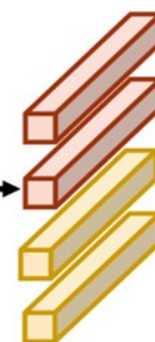
Text Feature



GPF-Net



Attention



Cross-entropy Loss



Triplet Loss

A 1.2 cm raised polyp clip with a lobulated surface and a 2.0cm thick pedicled polyp with hyperemia

Image Features

Previous Layer Output Features

Z

Gated Fusion

Layer Norm & Self-Attention

+

Layer Norm & Feed-Forward

c

Concat

+

Add

Layer Norm

Current Layer Output Features

Text Features

Z

Gating weight matrix

c

Concat

+

Add

Gated Progressive Fusion Net