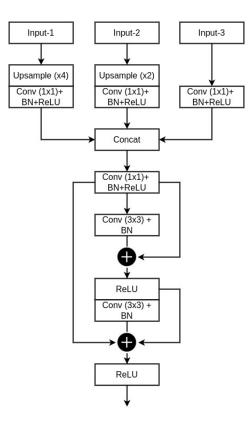
## Supplementary material

**Suuplementary Table 1:** Polyp datasets used in our experiments with number of images, input size and their availability. For some indicated dataset test data are not available (e.g. BKAI) for which we used training data and split it into 80:10:10 for train, validation and test. Also, we use small, flat and sessile polyps in Kvasir-SEG.

Dataset	Images	Input size	Availability
Kvasir-SEG	1000	Variable	Public
CVC- $ClinicDB$	612	$384 \times 288$	Public
BKAI	$1000^{\dagger}$	Variable	Public
$\mathrm{Kvasir}\text{-}\mathrm{Sessile}^\diamond$	196	Variable	Public

<sup>†</sup> test data unavailable <sup>\$</sup>sessile polyps from Kvasir-SEG

## Text guided attention 11



**Suuplementary Figure 1:** Block diagram of the proposed multi-scale fusion aggregation module that takes different scale features from three decoder blocks and concatenates them using upsampling, convolution, batch normalization and ReLU activations. Also, we apply residual-connections for different layers as shown.

**Suuplementary Table 2:** Algorithm complexity of methods used in the study. Number of model parameters, flops, image size and frame-per-second is provided for both state-of-the-art (SOTA) methods and proposed TGANet.

Method	Parameters (Million) Flops (GMac) Image Size FPS					
U-Net	31.04	54.75	$256 \times 256$	164.93		
HarDNet-MSEG	33.34	6.02	$256 \times 256$	42.95		
ColonSegNet	5.01	62.16	$256 \times 256$	127.95		
DeepLabV3+	39.76	43.31	$256 \times 256$	99.79		
PraNet	32.55	6.93	$256 \times 256$	36.90		
TGANet (Ours)	19.84	41.88	$256 \times 256$	37.64		