(a) "Y"-shaped multi-modal learning

(b) "X"-shaped multi-modal learning

(c) Proposed "Chilopod"-shaped multi-modal learning

"Chilopod"-shaped net

use CE and Dice loss

pre-softmax activations

Distill Knowledge

matrix of $\mathbf{q}$

KL-divergence

detail

Average activations of locations S_c^n in a batch within each channel

z_4 z_3 z_2 z_1 z_0

Scaling z_c with a temperature value then apply softmax

p_4 p_3 p_2 p_1 p_0

(d) Overview of our proposed multi-modal segmentation scheme with knowledge distillation.