特征表示：一维向量

snippet-level：$x \in R^D$ 通过 I3D 提取特征

proposal-level：$y \in R^D$ 通过 BSN 获得 proposal，对每个 proposal 的 [strat, end] 中全部 snippets 进行 max-pooling

video-level：$z \in R^D$ 将全部的 snippets 进行 max-poding

~~对于~~ $\{x_j | j \in S(i)\}$ $S(i)$ ith proposal.

L-Net. 利用余弦相似度计算一个 proposal feature 与 snippet feature 的距离 作为 snippet 的权重 $a$.

$$a_{i,j}^L = \frac{\sigma(S(y_i, x_j))}{\sum_{k \in S(i)} \sigma(S(y_i, x_k))}$$

$S(x,y)$: 余弦相似度

$\sigma()$: ReLu

根据上下文补充 proposal 细粒度时间特征：

$$y_i^L = \sigma(W_1^L y_i + \sum_{j \in S(i)} a_{ij}^L W_2^L x_j). \quad \sigma \text{ ReLu}.$$

$W_1^L, W_2^L \in R^{\frac{D}{2} \times D}$ $y_i^L \in R^{\frac{D}{2}}$

G-Net.

$$a_{i,j}^G = \frac{\sigma(S(z, x_j))}{\sum_{k \in S(i)} \sigma(S(z, x_k)) + \sigma(S(z, y_i))}$$

$$b_i^G = \frac{\sigma(S(z, y_i))}{\sum_{k \in S(i)} \sigma(S(z, x_k)) + \sigma(S(z, y_i))}$$

通过视频全局特征来增强 proposal

$a_{ij}^G$ attention weight between video-level and jth snippet of the ith proposal

$b_i^G$ video-level and ith proposal.

$$z_i^G = \sigma(W_1^G z + W_2^G(\sum_{j \in S(i)} a_{ij}^G x_j + b_i^G y_i))$$

$W_1, W_2 \in R^{\frac{D}{2} \times D}$

$z^G \in R^{\frac{D}{2}}$

$$y_i^G = y_i^L \oplus z_i^G \quad y_i^G \in R^D$$

$\oplus$ concatenation.