$I^{(ms)} \in R^{4\times H\times W}$ $I^{(pan)} \in R^{1\times 4H\times 4W}$

$F^{(ms)} = Encoder_{(ms)}(I^{(ms)}) \in R^{512\times 16\times 16}$

$F^{(pan)} = Encoder_{(pan)}(I^{(pan)}) \in R^{512\times 16\times 16}$

超像素掩码 $\{S_k\}_{k=1}^{65}$

$z_k^{(pan)} = Pool(F^{(pan)} \odot S_k) \in R^{512}$

$z_k^{(ms)} = Pool(F^{(ms)} \odot S_k) \in R^{512}$

设两个模态超像素特征为 $\{z_i^{(pan)}\}_{i=1}^{K}$ $\{z_j^{(ms)}\}_{j=1}^{K}$

$$Sim(z_i^{(pan)}, z_j^{(ms)}) = \frac{z_i^{(pan)} \cdot z_j^{(ms)}}{\|z_i^{(pan)}\| \cdot \|z_j^{(ms)}\|}$$

$$\bar{s}_i^{(pan\to ms)} = \frac{1}{65}\sum_j Sim(z_i^{(pan)}, z_j^{(ms)})$$

$$\bar{s}_j^{(ms\to pan)} = \frac{1}{65}\sum_i Sim(z_i^{(pan)}, z_j^{(ms)})$$

$$M^{pan} = \left\{ i \mid \bar{s}_i^{(pan\to ms)} \geq T^{(pan)} \right\}$$

$$M^{ms} = \left\{ i \mid \bar{s}_i^{(ms\to pan)} \geq T^{(ms)} \right\}$$

$$M = \{M^{pan} \cap M^{ms}\}$$

$M = M^{ms} \odot M^{pan}$
其中 $M^{ms}, M^{pan} \in \{0,1\}^N$

$S^+_{ms}, S^-_{ms}, S^+_{pan}, S^-_{pan} \in R^N$

$\tilde{S}^+_{ms} = S^+_{ms} \odot M$, $\tilde{S}^-_{ms} = S^-_{ms} \odot M$

$\tilde{S}^+_{pan} = S^+_{pan} \odot M$, $\tilde{S}^-_{pan} = S^-_{pan} \odot M$

$L^{(i)}_{ms} = max(0, margin + \tilde{S}^-_{ms}(i) - \tilde{S}^+_{ms}(i))$

$L^{(i)}_{pan} = max(0, margin + \tilde{S}^-_{pan}(i) - \tilde{S}^+_{pan}(i))$

$$L_{final} = \frac{1}{2 \cdot max(10, \sum_{i=1}^{N} M(i))} \left(\sum_{i=1}^{N} M(i) \cdot L^{(i)}_{ms} + \sum_{i=1}^{N} M(i) \cdot L^{(i)}_{pan} \right)$$

$\sum_{i=1}^{N} M(i) = NV$　　$NV \geq 10$

$L = 0$　　$NV < 10$

$NV = \sum_i M(i)$

$$F_q = att\_ms \cdot q\_ms + att\_pan \cdot q\_pan$$

$$F_p = att\_ms \cdot p\_ms + att\_pan \cdot p\_pan$$