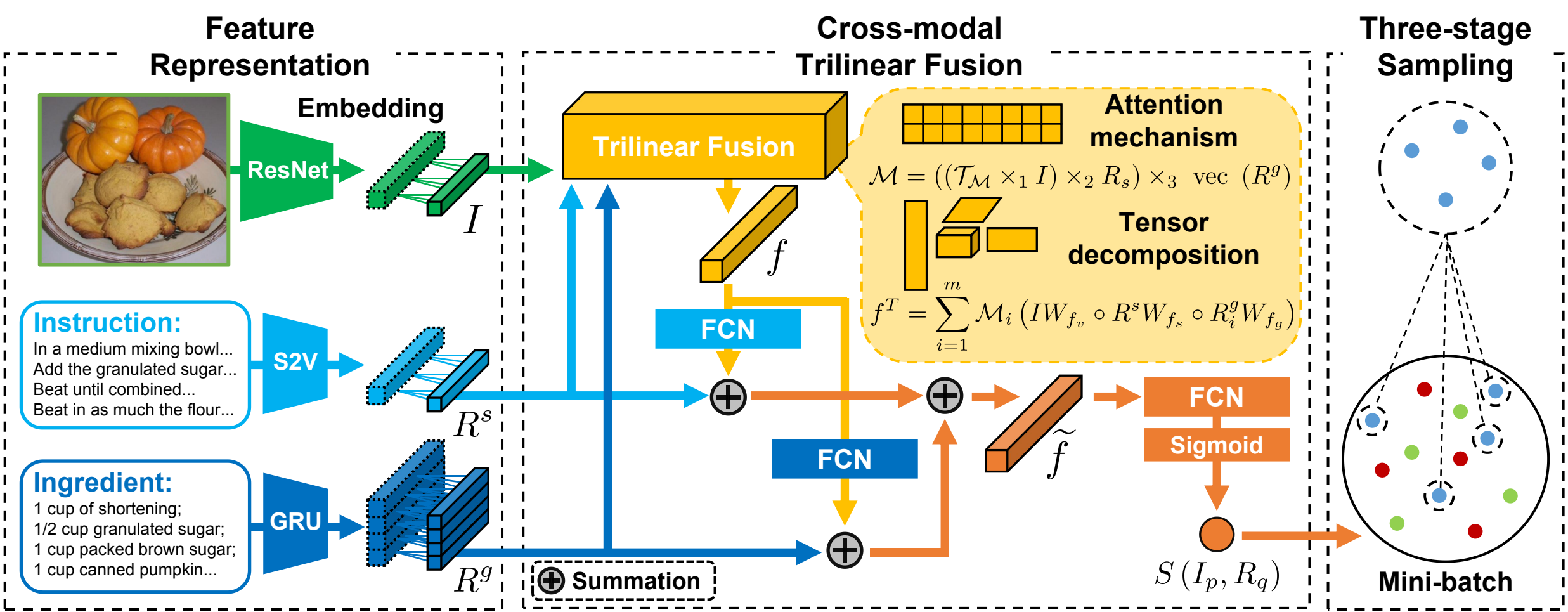
Feature Representation
Embedding
ResNet
I
Instruction:
In a medium mixing bowl...
Add the granulated sugar...
Beat until combined...
Beat in as much the flour...
S2V
R^s
Ingredient:
1 cup of shortening;
1/2 cup granulated sugar;
1 cup packed brown sugar;
1 cup canned pumpkin...
GRU
R^g
Cross-modal Trilinear Fusion
Trilinear Fusion
f
FCN
FCN
Attention mechanism
$\mathcal{M} = ((\mathcal{T}_{\mathcal{M}} \times_1 I) \times_2 R_s) \times_3 \text{vec}\ (R^g)$
Tensor decomposition
$f^T = \sum_{i=1}^{m} \mathcal{M}_i \left(IW_{f_v} \circ R^s W_{f_s} \circ R_i^g W_{f_g}\right)$
$\widetilde{f}$
FCN
Sigmoid
$S\left(I_p, R_q\right)$
Summation
Three-stage Sampling
Mini-batch