DRAWINGS OF SPECIFICATION

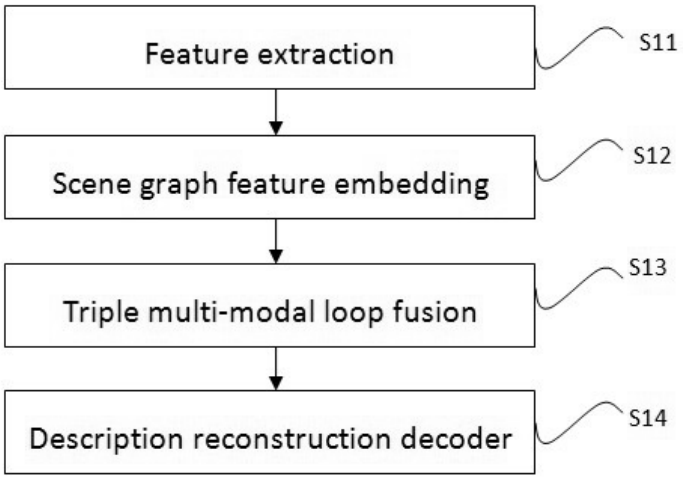

Fig. 1

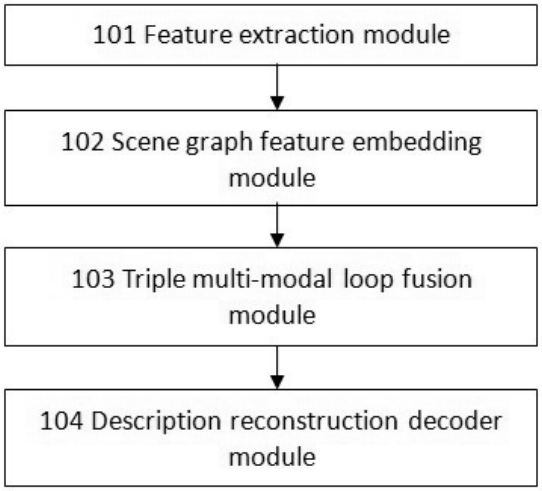

Fig. 2

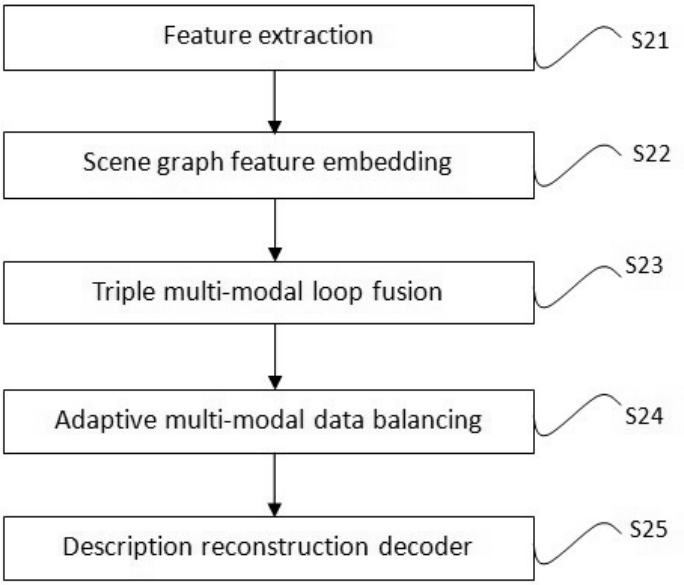
Feature extraction
S21
Scene graph feature embedding
S22
Triple multi-modal loop fusion
S23
Adaptive multi-modal data balancing
S24
Description reconstruction decoder
S25

Fig. 3

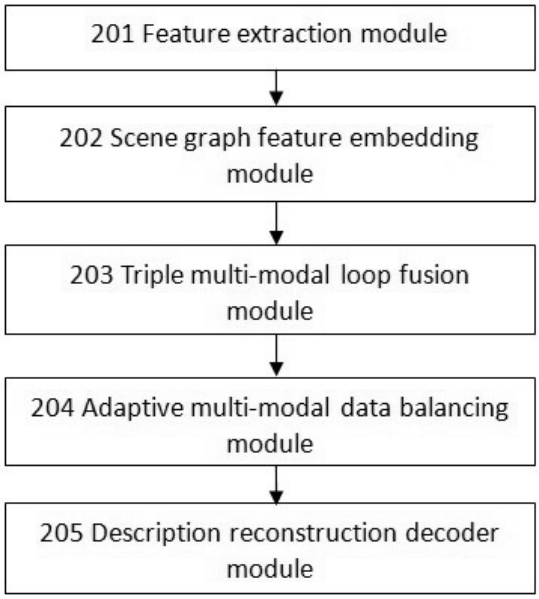
201 Feature extraction module
202 Scene graph feature embedding module
203 Triple multi-modal loop fusion module
204 Adaptive multi-modal data balancing module
205 Description reconstruction decoder module

Fig. 4

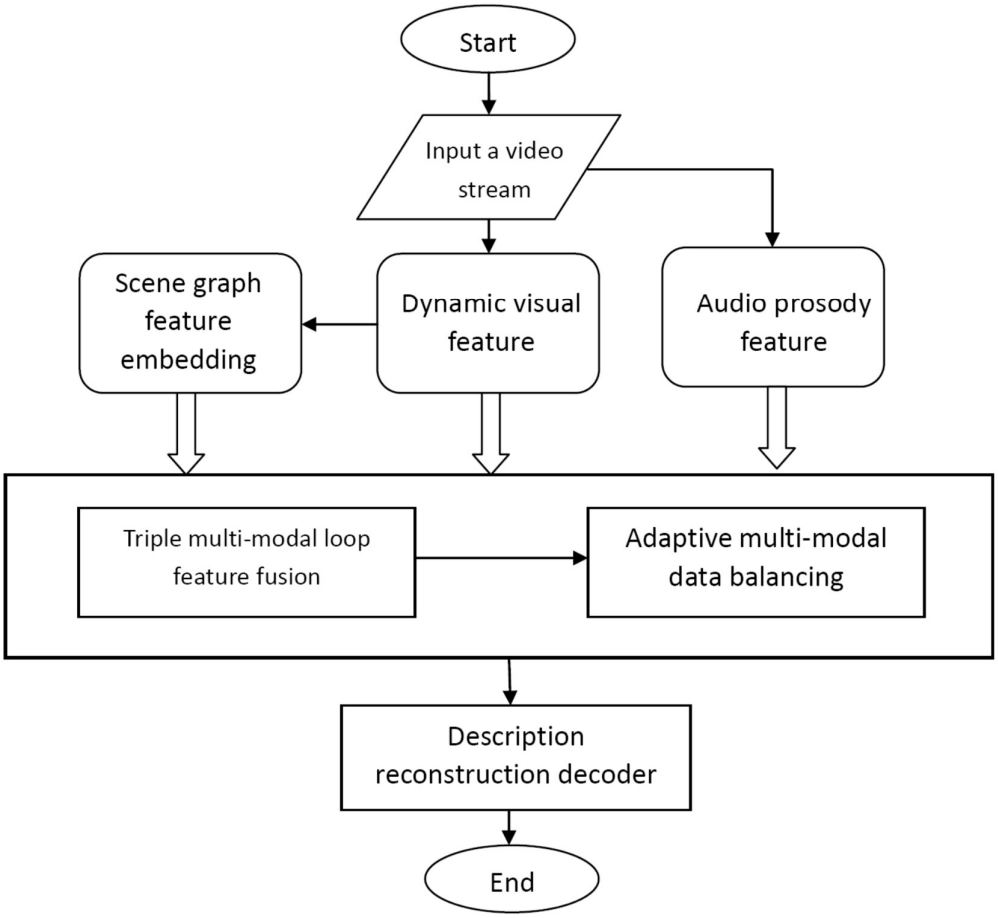
Start
Input a video stream
Scene graph feature embedding
Dynamic visual feature
Audio prosody feature
Triple multi-modal loop feature fusion
Adaptive multi-modal data balancing
Description reconstruction decoder
End

Fig. 5

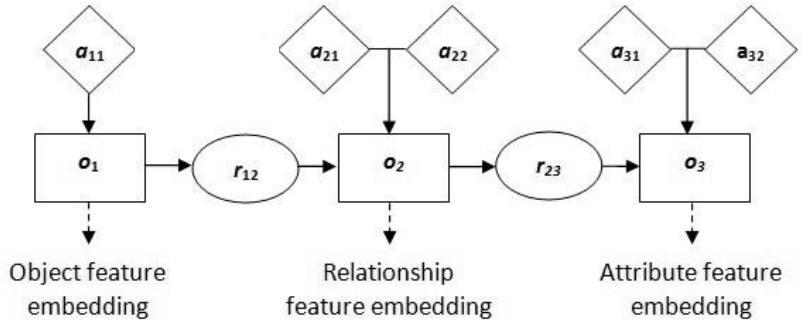
a_{11}
a_{21}
a_{22}
a_{31}
a_{32}
o_1
r_{12}
o_2
r_{23}
o_3
Object feature embedding
Relationship feature embedding
Attribute feature embedding

Fig. 6

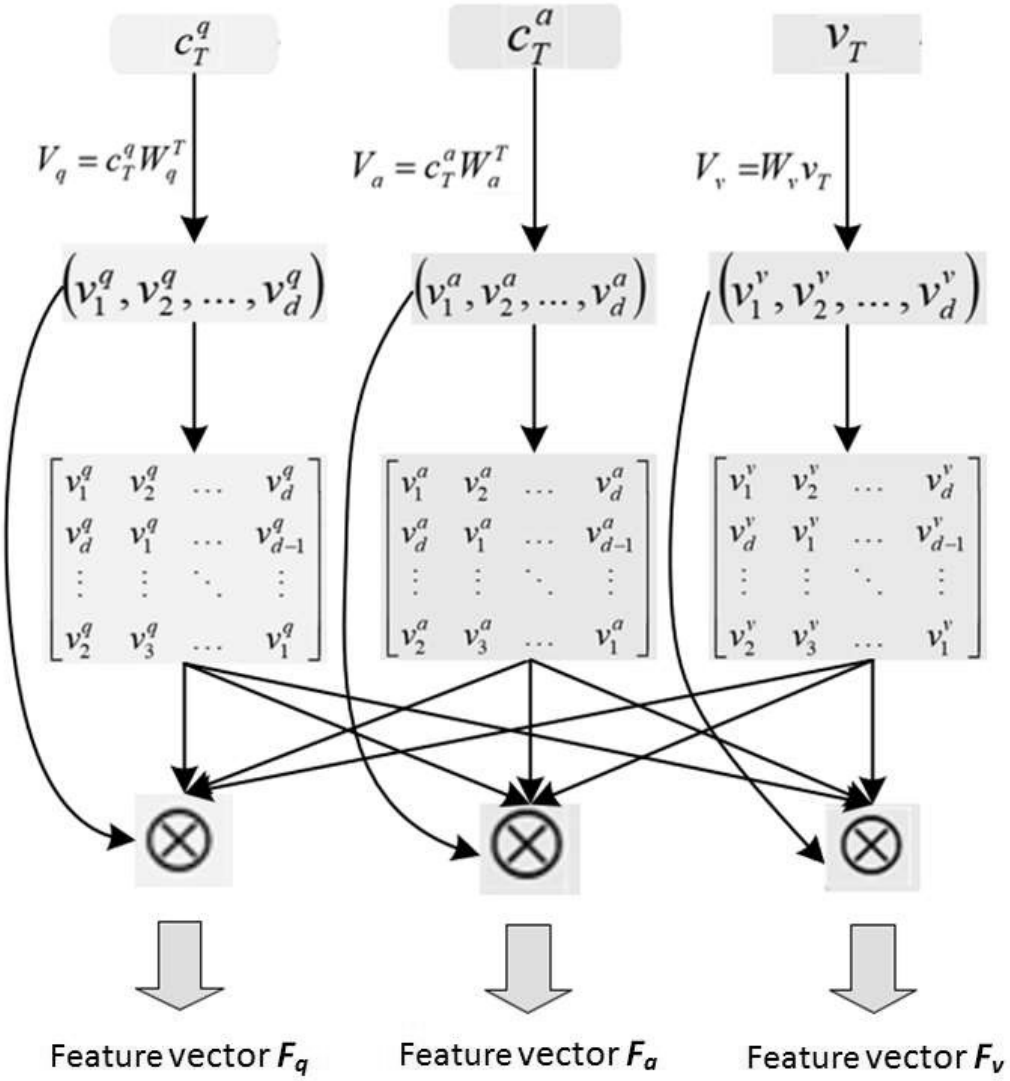
c_T^q
c_T^a
v_T
$V_q = c_T^q W_q^T$
$V_a = c_T^a W_a^T$
$V_v = W_v v_T$
$(v_1^q, v_2^q, \ldots, v_d^q)$
$(v_1^a, v_2^a, \ldots, v_d^a)$
$(v_1^v, v_2^v, \ldots, v_d^v)$
$\begin{bmatrix} v_1^q & v_2^q & \cdots & v_d^q \\ v_d^q & v_1^q & \cdots & v_{d-1}^q \\ \vdots & \vdots & \ddots & \vdots \\ v_2^q & v_3^q & \cdots & v_1^q \end{bmatrix}$
$\begin{bmatrix} v_1^a & v_2^a & \cdots & v_d^a \\ v_d^a & v_1^a & \cdots & v_{d-1}^a \\ \vdots & \vdots & \ddots & \vdots \\ v_2^a & v_3^a & \cdots & v_1^a \end{bmatrix}$
$\begin{bmatrix} v_1^v & v_2^v & \cdots & v_d^v \\ v_d^v & v_1^v & \cdots & v_{d-1}^v \\ \vdots & \vdots & \ddots & \vdots \\ v_2^v & v_3^v & \cdots & v_1^v \end{bmatrix}$
Feature vector F_q
Feature vector F_a
Feature vector F_v

Fig. 7

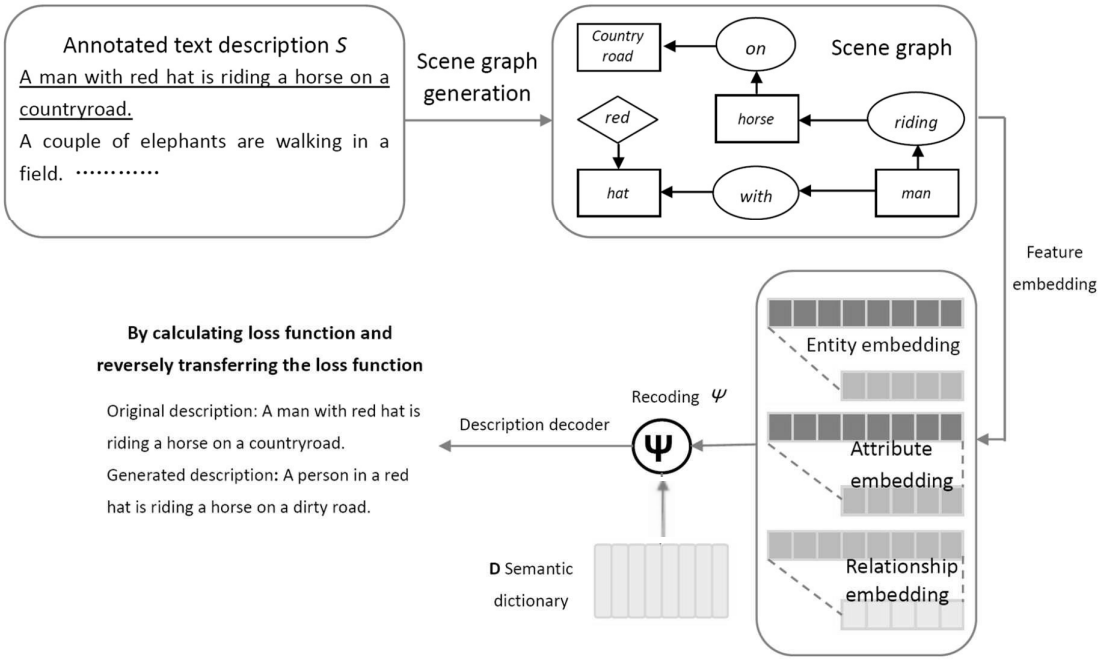
Annotated text description S
A man with red hat is riding a horse on a countryroad.
A couple of elephants are walking in a field. ·····
Scene graph generation
Country road
on
Scene graph
red
horse
riding
hat
with
man
Feature embedding
Entity embedding
Attribute embedding
Relationship embedding
Recoding Ψ
Description decoder
Ψ
D Semantic dictionary
By calculating loss function and reversely transferring the loss function
Original description: A man with red hat is riding a horse on a countryroad.
Generated description: A person in a red hat is riding a horse on a dirty road.

Fig. 8