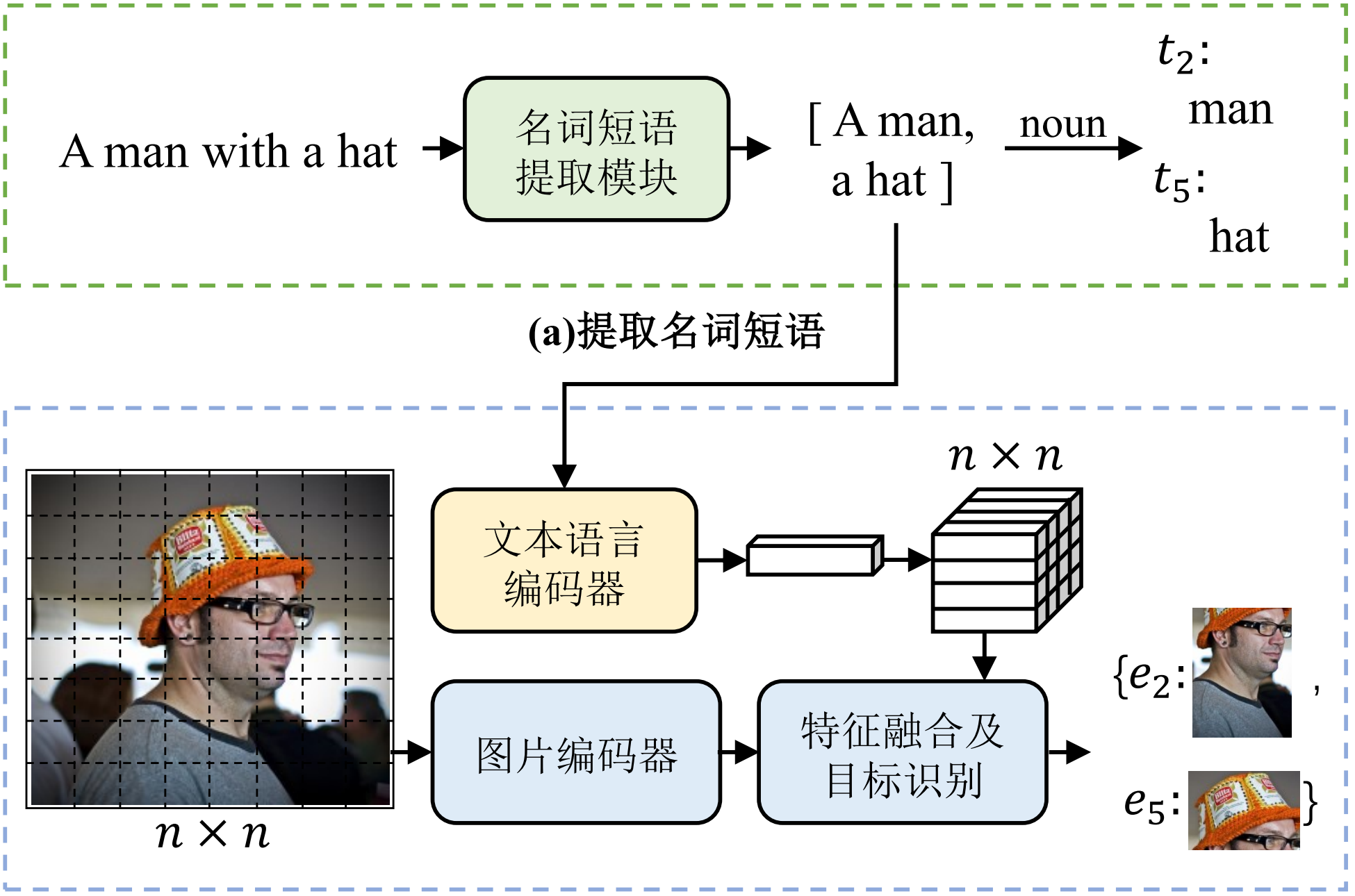
A man with a hat
名词短语
提取模块
[A man,
a hat]
noun
t_2:
man
t_5:
hat
(a)提取名词短语
$n \times n$
文本语言
编码器
$n \times n$
图片编码器
特征融合及
目标识别
$\{e_2$:
,
e_5:
}
(b)提取视觉目标