新闻图文匹配性检测是 image-text matching 任务的一个应用。Image-text matching 是一项连接视觉与语言的跨模态任务，在近年已经取得了很大的进展[4][5][6][7][8]。现有 image-text matching 方法分别提取图像特征和文本特征，将两种特征映射到一个语义空间中进行比较，从而判断是否匹配。准确提取图像和文本的特征是 image-text matching 任务的基础。新闻文本中包括事件发生的地点类命名实体，事件涉及的人物类命名实体以及组织类命名实体，新闻配图则直观生动的展示新闻事件中关键的命名实体。新闻文本与新闻配图中命名实体是否一致，决定这则新闻是否图文匹配。新闻文本中包含大量的命名实体，而提取图像特征的算法却不能直接提取这些命名实体，二者存在巨大的语义差距。因此需要使用命名实体丰富新闻配图语义。文献[9][10]提供了一种可行的方法，此类方法在外部知识库中选择命名实体插入到图像 template caption 中。文献[10]使用 attention 机制选择命名实体插入到 template caption 中，没有考虑实体间的关联。文献[9]将所有实体构造为图，选择图中不同类型实体之间共现率大的实体组合插入到 template caption 中，没有分析同类实体之间的关联。此类方法缺少分析命名实体之间的关联，生成的图像描述常常不能准确描述图像中的命名实体。

　　One of image-text matchings main functions is to detect match~~es between~~ ~~ing graphics~~images and the text of news ~~stories~~articles. Image-text matching is a task with cross-modality pairing of visuals with language, and in ~~the~~ recent years it~~s~~'s functionality has improved substantially~~come on in leaps and bounds~~. ~~In order t~~To determine whether a certain set of images and text are related to each other, image-text matching extracts data from both pieces of information and maps both sets of information into the same semantic space, which prepares them ~~where they are both ready~~ for comparison ~~analysis~~. The core purpose of image-matching data is to accurately extract data from both news images and text ready for detailed analysis. Contained within the news text is information such as the location where the event was supposed to have taken place, the name of the person or ~~entity~~ items involved, what their involvement was and any additional ~~some other~~relevant ~~–~~information ~~relating to the parties involved. ;~~ ~~vividly and intuitively t~~Image-text matching creates a news map of t~~his~~ information ~~is mapped out in the news map created by image-text matching.~~ In this process, matching data continues to be detected ~~R~~regardless of consistency between the named entities within the news map and the news text~~, matching data continues to be detected~~. ~~Its has become clear that~~However, due to significant ~~the huge~~ semantic difference~~s~~ between news text and ~~the~~news maps~~m~~, the algorithm is unable to distinguish between similar ~~extract~~ named entities ~~from~~ in news articles and extract them separately~~containing large numbers of different named entities each one individually~~. A feasible method to deal with this issue ~~can be found in Literature~~has been developed [9][10]. This solution ~~, here it details how it~~ selects ~~and inserts~~ named entities that have been found in the knowledge base and inserts them into the image template caption. In a N~~n~~on-discriminator~~y~~ manner~~ily~~, and without consideration for the relationship

between ~~the~~ entities, ~~these researchers~~ ~~Literature~~ [10] use~~d~~s the attention mechanism to ~~choose~~ determine which named entities can be correctly placed into the template caption. Without analyzing whether or not there ~~is~~ was any ~~correlation or~~ relationship between similar entities, ~~the researchers~~ ~~Literature~~ [9] mapped ~~forms~~ the information from all entities into graphs and select~~ed~~s ~~which~~ entities with the highest rate of ~~have a higher~~ occurrence ~~rate~~ from between the different types of entities ~~that can be found~~ in the graph, and finally insert~~ed~~ing the~~se~~m into the appropriate template captions. ~~It has been found that there are often i~~Inaccuracies have often been found in the generated image description and th~~at often it~~ ~~e generated image description often~~ does ~~n't~~ not match the image. This occurs, ~~because of an absence of any assessment~~ ~~this is due to the lack of analysis into~~of the association between ~~the~~ named entities from the news images, and news text.