* **教育领域图像的视觉场景图生成方法**。教育领域图像中视觉对象的底层视觉信息（颜色、纹理等）较为稀疏，对图像语义的理解还需结合图中出现的数字和对象间的交互等信息，这导致传统基于底层视觉信息的图像特征难以准确表征图像语义。为此，拟研究一种基于多视图信息融合的视觉场景图生成方法，从图中对象的视觉形象和对象间的交互信息等多个方面捕捉图像的语义。
* **路线1-1：教育领域图像的视觉场景图生成方法**

拟提出一种基于多视图信息融合的视觉场景图构建方法，将跨媒体问题中的图像表示为视觉场景图，从图像中对象的视觉信息、对象之间的位置和语义交互三个维度表征图像的语义信息。如图3所示，本研究主要包括三个模块：



图3. 视觉场景图生成框架图

**Step1：图像的多视图构建**。对于跨媒体问题中的图像，拟首先构建其对应的三种视图：1）**数字视图**，其中的每个节点代表图像中出现的一个数字类型的文本。拟利用现有的对象字符识别(Object Character Recognition, OCR)技术识别图中出现的数字，并构建一个全连接的数字视图。

2）**部件视图**，其中每个节点代表组成图像的一个细粒度部件，例如图3中的“球A”和“半径OA”等。具体地，拟首先基于OCR技术和无监督语义分割技术，细粒度地识别图中组成视觉对象的部件类别和位置，以及部件周围的文本信息（例如，词和字母）。然后，拟融合对象部件和文本信息，构建一个全连接的部件视图。

3）对象视**图**，其中每个节点代表图中对象的纯视觉信息。拟利用现有的弱监督对象检测器, 识别图中对象的类别和位置坐标，并利用卷积神经网络构建被识别对象的视觉信息编码，最后利用上述信息构建一个全连接的对象视图。

**Step2：多视图信息融合**。拟对三视图中的节点特征进行细粒度的融合，为接下来生成图像的视觉场景图做准备，具体步骤包括：

1）数字视图和部件视图的融合。拟利用数字视图中相关信息，对部件视图中所有节点的特征进行更新。针对部件视图中的节点，其特征向量更新公式如下：

(1)

其中代表节点的特征，符号表示向量间的连接操作，代表数字视图中数字节点的特征编码器。代表部件视图中节点和数字视图中节点间的相关系数,

(2)

(3)

其中，表示数字节点的位置特征，是三个用于特征映射的MLP(Multi-Layer Perceptron)网络，它们将部件的特征表达和数字的特征表达映射到相同维度的语义空间中。

2）部件视图和对象视图的信息融合。在完成部件视图中所有节点的特征更新之后，拟利用和上面相似的融合方法，将部件视图的信息融入对象视图的相关节点特征中。

**Step3：视觉场景图构建**。拟基于节点特征更新后的对象视图，生成图像的视觉场景图。视觉场景图中的节点代表对象的区域及其类别标签，边表示视觉对象之间的语义或位置交互。具体地，视觉场景图中节点之间语义交互的概率分布可通过下式计算：

, (4)

其中，和分别代表对象视图中节点和更新后的特征向量，是同时覆盖第个和第个对象的区域对应的视觉特征，是语义交互分类器，交互类别包括组成、围绕、碰撞和包含等。

针对对象间的位置交互，拟首先利用视觉对象的边界框的交并比(Intersection over Union)，判断两个对象之间是否存在包含或重叠的关系。若IOU<0.5, 则通过计算它们边界框中心点之间的距离及角度，来判断它们之间的细粒度位置交互类别，例如左上，右下，交叉和贯穿等。