

3.2.2 多通道语义图像¹

在目标检测算法 *FFTNet*[26] 中，便通过灰度语义图像构建了高斯椭圆组成的 *Heatmap* 来指示像素属于目标中心点的概率，以及该目标的回归尺度。灰度语义图像的意义是指示某种类别的目标在图像区域的“存在性”，即图像上某区域该类别目标存在的可能性。受此思想启发，对于拥有 *item_class_num* 个部件类别、图像长宽分别为 *width* 和 *height* 的样本集，可构建长宽同样为 *width* 和 *height*，且具有 *item_class_num* 个通道的灰度语义图像。动车组转向架上的部件按其位置是否固定可分为固定部件和旋转移动部件两种。这两种部件使用不一致的语义图像构建法。

1. 固定部件³

对于在标准图标注中所有位置固定的部件，通过以下流程构建语义图像：⁴

- (1) 初始化一个长宽为 *width*/8 和 *height*/8、通道数为 *item_class_num*、值域为 $[0, 1]$ 的灰度图像张量，所有像素值都设置为 0.01。之所以长宽设置为八分之一，是为了节省构建语义图像时的内存和显存。⁵
- (2) 遍历标准图标注中的每一个固定部件，根据其边界框生成目标的高斯椭圆二维语义正态分布。以边界框的中心为均值，边界框的半宽和半长为 *X* 和 *Y* 方向上的 2 倍标准差，生成峰值为 1.0 且 *X* 和 *Y* 相互独立的二维正态分布，即高斯椭圆，该过程如公式(2)到(4)所示：

$$\begin{cases} \mu_x = \frac{x_{min} + x_{max}}{2} \\ \mu_y = \frac{y_{min} + y_{max}}{2} \\ \sigma_x = \frac{\mu_x - x_{min}}{2} \\ \sigma_y = \frac{\mu_y - y_{min}}{2} \end{cases} \quad (2)$$

$$\text{SemanticValue}(x, y) = e^{-\frac{1}{2} \left(\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right)} \quad (3)$$

$$x \in [\mu_x - 3\sigma_x, \mu_x + 3\sigma_x], y \in [\mu_y - 3\sigma_y, \mu_y + 3\sigma_y] \quad (4)$$

x_{min} 、 x_{max} 、 y_{min} 、 y_{max} 表示该目标边界框的范围， μ 和 σ 表示正态分布的均值与标准差。⁸ 对于在语义图像上位置为 (x, y) 的像素，赋值为 *SemanticValue*(*x*, *y*)。每个目标的赋值范围为以 μ 为中心的 3 倍 σ 范围。构造高斯椭圆的效果如图 3 所示，图中的红色框即目标的边界框。

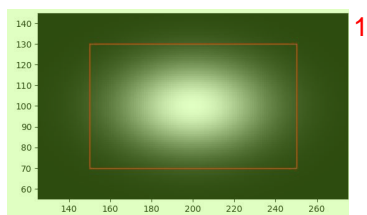


图 3：通过边界框构建目标的高斯椭圆语义分布 ²

(3) 将每个固定部件的高斯椭圆添加在对应类别通道的灰度语义图像的对应位置上。若同类 ³
型目标之间的语义分布区域有重叠，重叠区每个像素都取相对高值。语义分布在具体动
车部件图像上的生成如图 4 所示。

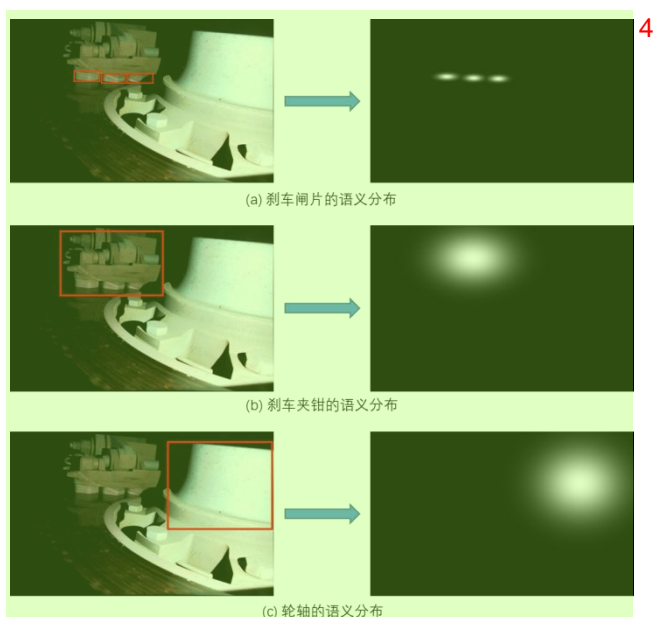


图 4：同一张标准图像中三种不同的固定部件所生成的高斯椭圆语义分布 ⁵