# 先验知识学习及其在视觉显著性估计中的应用

# 摘要

人类通过视觉系统能够在复杂的场景中快速搜索到自己感兴趣的目标。模拟人类视觉系统得到图像中的显著区域，即显著性区域检测，已经成为计算机视觉领域的热点之一。从本质上说，如何合理地构建视觉注意模型成为显著性区域检测的关键。以往神经心理学的大量研究表明，视觉注意机制的构建一直以来可以通过两种不同的途径：自下而上的由数据驱动的检测方法及自上而下的由任务驱动的检测方法。现在由数据驱动的自下而上的显著性检测方法研究较多，也比较成熟和深入，故本文工作主要是关注显著性区域检测中的自上而下的视觉注意模型。本文首先计算图像的先验知识，同时由自动编码器来重构图像并以重构残差来估计图像的显著性，并利用先验知识与重构残差的融合来提高显著性检测效果。

本文中提到的先验知识是指图像的目标性，所谓目标性是指图像中一个像素或者一个区域所包含目标的可能性。目标性是图像中上层语义信息中比较强的一种特征，可以给出图像中所有目标的大概分布情况。本文中的目标性获取方法是先对图像进行三通道分离，然后对每一通道进行多尺度化，并对每一尺度图像进行分割得到多个窗口，最后对每个窗口打分来得到图像的目标性。

本文中的自动编码器是深度学习算法中无监督学习的一种，利用自动编码器来重构图像的具体过程是先从图像中随机采样多个像素点，利用像素点的外围块和中心块构成训练样本来训练构建好的自动编码器网络，最后对图像中的每一个像素点，依次取其外围块输入到网络，计算网络重构的中心块和实际的中心块之间的残差，并以重构残差来估计图像的显著性。实验结果表明由该重构残差加上图像的中心先验已经能够很好的表示图像的显著性。

本文中最后通过融合图像的目标性及重构残差来进一步提高显著性估计。通过实验发现，利用图像的目标性来监督自动编码器的训练过程对最终的显著性提升比较小，但将图像的目标性与重构残差直接融合则有很好的提升效果。实验结果证明，图像的目标性包含丰富的上层语义信息，可以提高图像的显著性检测。

关键字：先验知识,显著性,目标性,计算机视觉,深度学习