

深圳大学实验报告

课程名称 脑与认知科学

项目名称 图像中的注意区域预测

学 院 计算机与软件学院

专 业 软件工程（腾班）

指导教师 钟圣华

报 告 人 洪子敬 学号 2022155033

实验时间 2024年11月5日至2024年11月26日

提交时间 2024年11月24日

教务处制

一、实验目的与要求

1. 通过实验加深对注意过程的了解；
2. 掌握 Matlab 或 Python 等常用工具的基本使用方法；
3. 具备使用机器学习方法设计基本的图像注意区域预测的能力；

二、实验内容与方法

1. 请简述你对图像的注意预测的定义、难点的理解。

解答：图像的注意预测（Image Attention Prediction）是指通过计算机视觉技术，识别和预测图像中重要区域或元素的过程。它模仿人类视觉注意力的机制，能够在处理复杂图像时，强调对某些特征的关注，从而提高图像理解和分析的效率。

图像的注意预测主要涉及以下几个方面：

- （1）**重要性评估**：识别图像中对特定任务（如物体检测、图像分类）最相关的区域；
- （2）**区域聚焦**：通过热图或权重图的形式可视化注意力分布，指示模型关注的区域；
- （3）**增强特征提取**：提高模型在特定区域的特征提取能力，从而提升整体性能。

图像的注意预测**难点**主要包括：

- （1）**复杂背景**：复杂的背景可能掩盖重要的目标区域，导致注意力预测不准确；
 - （2）**多样性**：不同图像中的对象形状、颜色和大小差异大，增加了模型学习的难度。
- 尺度变化**：对象在不同尺度下的表现不同，如何处理不同尺寸的目标是一个难点；
- （3）**上下文信息**：图像中对象之间的关系和上下文信息对注意力分配有重要影响，捕捉这些信息并加以利用非常复杂；
 - （4）**实时性要求**：在许多应用中，需要实时处理图像，如何在保持准确性的同时提高效率是一个重要问题。

三、实验结论或体会

1. 请叙述一种注意预测的常用评价标准和计算方法。

解答：这里选择 ROC 曲线和 AUC 值作为评价标准，也是后续练习采用的评价标准。

（1）ROC 曲线

ROC 曲线，全称为 Receiver Operating characteristic Curve（受试者特征曲线），主要分析的是二元分类问题，也就是输出结果只有两种类别的模型。二分类的结果可以用混淆矩阵描述为以下形式：

- **真阳性**（True Positive，简称 TP），也就是预测为真，实际上也为真的数据；
- **假阳性**（False Positive，简称 FP），也就是预测为真，但实际上为假的数据；
- **假阴性**（False Negative，简称 FN），也就是预测为假，但实际上为真的数据；
- **真阴性**（True Negative，简称 TN），也就是预测为假，实际上也为假的数据。

该混淆矩阵如下所示：

	实际为真	实际为假
预测为真	TP	FP
预测为假	FN	TN

图 1. 混淆矩阵

在评价模型的预测性能的时候，我们可以根据模型正确分类的比例（即准确率 $Accuracy$ ）来衡量：

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

然而，使用准确率 $Accuracy$ 作为评价指标也有不足之处，尤其是样本不均衡的时候。例如当某样本实际包含 100 个阳性、900 个阴性；假设有一个常数模型，它没有任何的自变量，模型总是输出一个恒定的概率常数 0.1。若阈值=0.5，由于 $0.1 < 0.5$ ，所有的样本都会被认为是阴性，即使该模型的 $Accuracy$ 达到 0.9 但是模型一个阳性都没有鉴别出来。可见， $Accuracy$ 并不是一个非常好的指标。

为此，我们引入几个新的指标：

- **真阳性率**（True Positive Rate，简称 TPR）： $TPR = \frac{TP}{TP+FN}$ ，代表的含义是：实际观测为阳的样本中，模型能够正确识别出来的比例。
- **真阴性率**（True Negative Rate，简称 TNR）： $TNR = \frac{TN}{FP+TN}$ ，代表的含义是：实际观测为阴的样本中，模型能够正确识别出来的比例。
- **假阳性率**（False Positive Rate，简称 FPR）： $FPR = \frac{FP}{FP+TN}$ ，代表的含义是：实际观测为阴的样本中，被模型错误地划分成阳性的比例。 $FPR=1-TNR$ 。

我们可以同时使用 TPR 和 FPR 来评价一个模型的预测能力。TPR 越高、FPR 越低，说明模型的预测能力越好。

对于一个二分类模型，阈值的选择至关重要，下面利用 `sklearn` 库随机生成 1000 个假数据进行二分类，按照 7:3 划分训练集和测试集，选择逻辑回归模型(LogisticsRegression)作为二分类模型，训练后用测试集数据进行预测分类，并将结果在不同阈值下计算 TPR 和 FPR，绘制成 ROC 曲线如下所示：

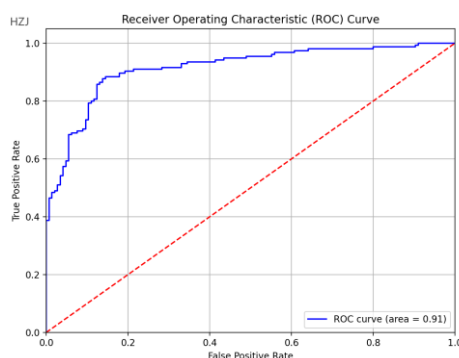


图 2. 测试样例 ROC 曲线

一般而言，如果一个点愈接近左上角，说明模型的预测效果越好（正判多，误判少），如果能达到左上角点（0，1），那就是最完美的结果，此时的阈值是最好的。

（2）从 ROC 到 AUC

除了用于阈值的选择，ROC 曲线还可以用于不同模型的比较。图 3 中有三条 ROC 曲线，A 模型比 B 和 C 都要好。也可以计算出 ROC 曲线下方的面积，计算出来的面积就是 AUC 的值了。AUC，即 Area Under Curve（ROC 曲线下方的面积）。在之前的例子中，AUC=0.91。一般来说，AUC 的值越大越好。

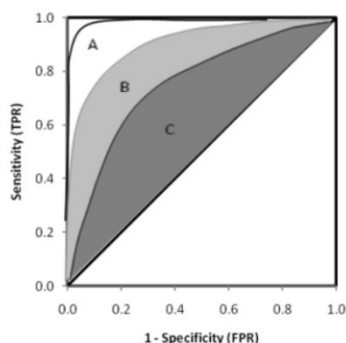


图 3. 不同模型的 ROC 曲线

（3）ROC 曲线与 AUC 值对图像注意预测模型的评估

图像预测模型的输出是对给定输入的预测图，我们将其转换为灰度图后，进一步转换为二值图，然后与真实人眼注意图对比，计算出 ROC 曲线与 AUC 值，即可比较不同模型的性能，这也是本实验我们用于评估模型的指标。

四、练习题（每一题都请详细回答，在系统上传代码）

1. 请选择一种基于图像的注意预测算法，使用该算法进行图像注意区域的预测，并使用注意预测的常用评价标准进行评价。在这里请给出方法简介、原图像、预测结果、评价结果。进行标准数据集上的显著性检测之后，请自选一张生活照（景物或者人都可以），并使用选择的注意预测算法进行显著性预测，在实验报告中附上原图和显著性预测图。

参考链接：<http://saliency.mit.edu/downloads.html>

<https://saliency.tuebingen.ai>

解答：本实验选择 DeepGazeIIE 方法进行图像注意区域的预测，并选择 AUC 值进行评价。

DeepGazeIIE 是一种用于图像和视频分析的深度学习方法，主要关注于视觉注意力的建模和推断。该方法基于深度学习框架，旨在通过网络结构捕捉人类视觉关注的模式，以便更好地理解 and 解析图像内容。

Deep GazeIIE 方法主要特点包括：

- （1）深度学习架构：DeepGazeIIE 使用卷积神经网络（CNN）来提取图像特征，并通过多层网络来增强特征表达能力。
- （2）注意力机制：该方法集成了视觉注意力机制，通过对图像中重要区域的加权，提升模型对关键特征的关注，从而提高任务性能。
- （3）多任务学习：DeepGazeIIE 可以同时处理多个视觉任务，如目标检测、图像分割和

场景理解，使其在多个应用场景中表现优异。

(4) 数据驱动：利用大量标注数据进行训练，DeepGazeIIIE 能够自动学习到有效的视觉特征，无需手动设计特征提取器。

(5) 应用广泛：该方法在智能监控、自动驾驶、增强现实等领域显示出良好的适应性和效果。

本实验本拟定于先对该模型进行训练后再进行评估的，但由于个人电脑性能不支持，跑炸了好几次，于是采用其预先下载好的预训练权重 `centerbias_mit1003.npy`，此权重是在 MIT1003 数据集上训练得到的。

为了验证模型的性能，选取 CAT2000 数据集中的 Action 类的前 20 个样本进行注意预测并评估效果，该 20 个样本分布如图 4 所示。

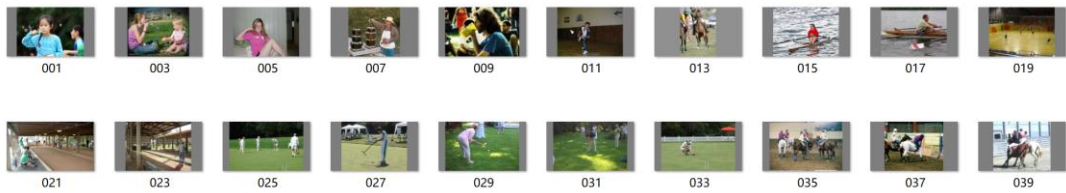


图 4. CAT2000 数据集部分样本

使用官方自带的 `matlab` 评估函数源码，将模型预测出的这 20 个样本的注意结果放入 `matlab` 进行评估（模型采用的是 `pytorch` 架构，生成的是 `.npy` 文件，进入 `matlab` 之前需要先转化为 `.mat` 文件）。计算出 20 张图片的 ROC 曲线和 AUC 值，其中一条曲线如图 5 所示，图 6 是该样本及其注意分布图。

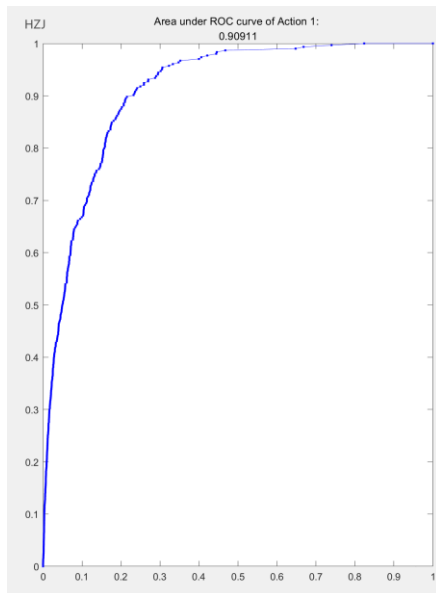


图 5. 某样本 ROC 曲线



图 6. 某样本及其注意分布图

观察结果可知，其 AUC 值达到 0.90911，比对注意分布图和样本图，不难发现其整体的效果还是不错的，符合正常人的视线分布。

统计 20 个样本的 AUC 值，绘制成曲线如图 7 所示。

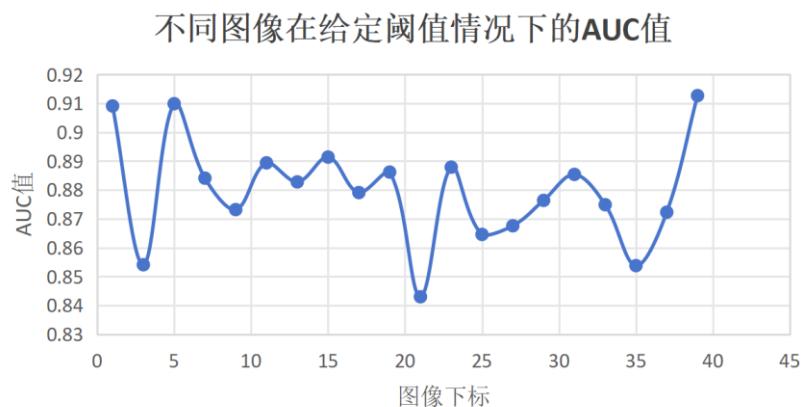


图 7.20 张样本的 AUC 值变化图

观察并统计样本 AUC 的平均值，不难发现其整体分布在中间线 0.88 附近，实际计算出的平均值为 0.8799，与预料的接近，阅读该算法对应的论文《DeepGaze IIE: Calibrated prediction in and out-of-domain for state-of-the-art saliency modeling》，可以看到其论文的结果指标如图 8 所示。

Model	IG ↑	AUC ↑	sAUC ↑	NSS ↑	CC ↑	KLDiv ↓	SIM ↑
DeepGaze IIE (DSREx3)	1.0715	0.8829	0.7942	2.5265	0.8242	0.3474	0.6993

图 8. 论文中的模型评估结果

可以看到论文中的 AUC 值为 0.8829，与实验评估得到的 0.8799 较为接近，虽然选区的数据较少，但其实际表现的平均效果还是达到预期的，说明此算法还是比较稳定的。

最后，使用该算法对选用的几张生活场景照片进行注意预测，原图、显著预测图以及该预测图进一步可视化的图像如下所示：

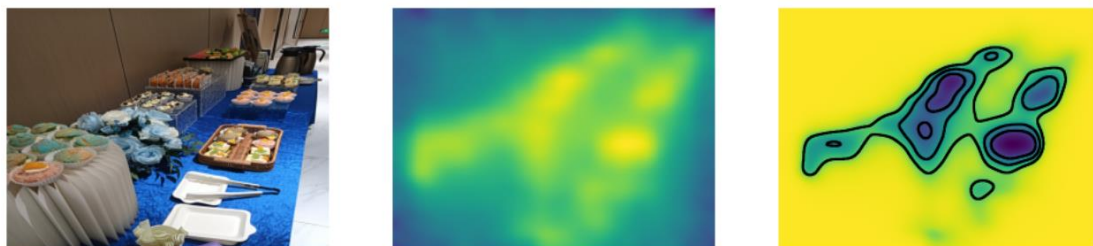


图 9. 生活场景图 1 及其预测结果

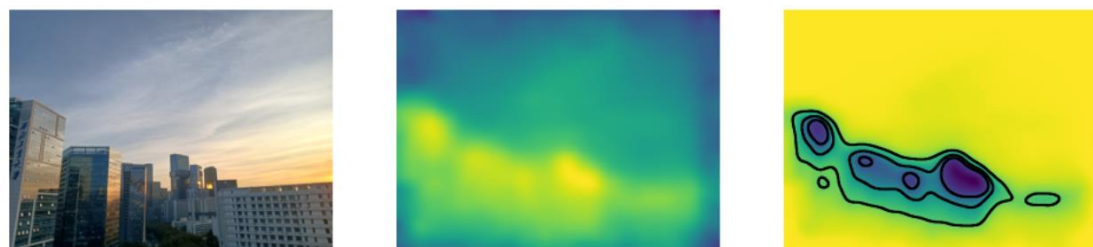


图 10. 生活场景图 2 及其预测结果

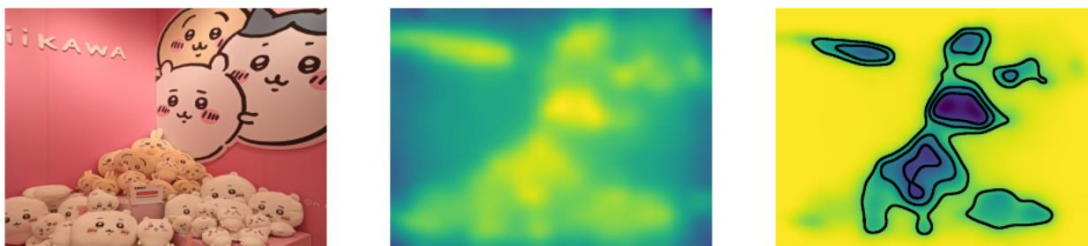


图 11. 生活场景图 3 及其预测结果

指导教师批阅意见：

成绩评定：

指导教师签字：

年 月 日

备注：

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。