随着人工智能技术的进步，人们对出行有了更多美好的愿景。近些年，以特斯拉为首的汽车厂商们也开始大力发展自动驾驶技术，并带动相关供应链企业迅速发展。通过模仿人类驾驶汽车的模式，智能驾驶可以分为三个部分，感知层（类似人眼探测并识别道路情况），规划层（大脑思考应该走哪条路线），执行层（控制方向盘，踩下油门或者刹车）。其中，各家厂商对于感知层有着激烈的探讨。一般的，由于单一雷达或摄像头的局限性，厂商更喜欢在汽车上装置多个传感器（雷达，激光雷达，摄像头），以得到多模态的图像数据。这样做的目的可以收集到更多信息的数据，以供规划层选择合适的道路。

与此同时，工程师们需要对多传感器获取到的数据做匹配工作，以得到基于同一对象的更多维度数据。行业中的做法通常是利用物理测量计算出各自传感器的相对坐标系，再进行坐标转换，从而得到物体的匹配。这样做的缺点有两个，第一，由于物体的位置是计算得出的，需要将所有传感器稳定在同一个刚性载体上以保证计算的准确性，这便在物理空间中增加了局限性，同时由于汽车在行进中不可避免的会有颠簸，这些微小的震动所产生的偏差很有可能在坐标系中变成极大的误差，从而影响后续的规划决策。第二，由于匹配的校准计算是基于当前的传感器，如果后续需要更换更新设备，其坐标就需要重新计算，这便大大增加了其耦合性。这些问题都是在实际环境中客观存在的，并且会对自动驾驶整体系统产生影响。

在机器视觉和深度学习中，人类视觉的力量和对视觉信息的理解可以被再现甚至超越。深度学习，作为机器学习的一部分，可以在应用实例的基础上学习和训练复杂的关系。同时，神经网络可以在训练过程中，独立地自动识别和提取特征。

深度学习应用凭借其在识别应用中超高的预测准确率，在图像处理领域获得了极大关注。 利用卷积神经网络(Convolutional Neural Network, CNNs) 等深层神经网络的解决方案，可以逐渐取代基于算法说明的传统图像处理工作。特别是处理任务中包含有复杂变量时（如反射面、光照不佳的环境、光照变化或移动的对象），只能用深度学习方法来解决。

卷积神经网络（Convolutional Neural Networks, CNNs）与普通神经网络非常相似，它们都由具有可学习的权重和偏置常量(biases)的神经元组成。每个神经元都接收一些输入，并做一些点积计算，输出是每个分类的分数，普通神经网络里的一些计算技巧到这里依旧适用。

因此，深度学习可以更好地解决多传感器物体匹配问题。在本项目中，数据集由两个传感器收集而来，一是彩色相机，二是红外相机。分别得到的数据是彩色图片和热成像图片，在两组样本上进行同样的ground truth标记，使用深度学习进行拟合，以达到两组图像数据中的物体匹配的目的。