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。

通过深度学习的训练，理想的结果是，该模型可以分别在彩色图片和红外图片上识别出指定的物体，除此之外，在两张图片上的物体可以被一一对应。换句话说，找到一幅图像像素到另一幅图像像素间的空间映射关系。这些图像可以是不同时间（多时间配准），不同传感器在不同地方拍摄（多模式配准）。这些图像之间的空间关系可以是刚性（平移和旋转），仿射（例如剪切），单应性或复杂的大变形模型（complex large deformations models）。

图像的物体配准可以帮助自动驾驶更好的感知到道路上的情况，得到更多关于人或车的信息。另一方面，在医学领域。随着成像设备的进步，对于同一患者，可以采集含有准确解剖信息的图像诸如CT，MRI；同时，也可以采集到含有功能信息的图像诸如SPECT。正确的图像配准方法可以将多种多样的信息准确地融合到同一图像中，使医生更方便更精确地从各个角度观察病灶和结构。同时，通过对不同时刻采集的动态图像的配准，可以定量分析病灶和器官的变化情况，使得医疗诊断、制定手术计划、放射治疗计划更准确可靠。