

# 从零学习算法部署-TensorRT篇

# 说在前面

---

本课程以TensorRT和PyTorch为主，不涉及模型的训练，例如蒸馏、枝减  
仅以工程角度考虑，如何**导出模型**、让**模型尽可能高性能**  
以及如何**上线交付（多线程）**

# 说在前面

---

部署的**优化思想、导出方式、解决问题流程**，可借鉴到**其他算法场景**，  
例如3399、安卓等等

# 前言

## 算法部署，学什么

精简的CUDA编程

只求够用，能写图像处理、后处理、理解索引计算，就够够的了，高性能的基础

TensorRT基础

如何编译模型、推理模型，onnx如何控制，插件怎么写，解析器怎么配。深入理解tensorRT这个东西

项目实战

拿真实的项目，实际操作并在tensorRT上以高性能推理起来。学习拿到项目如何分析，问题如何解决

# 说在前面

---

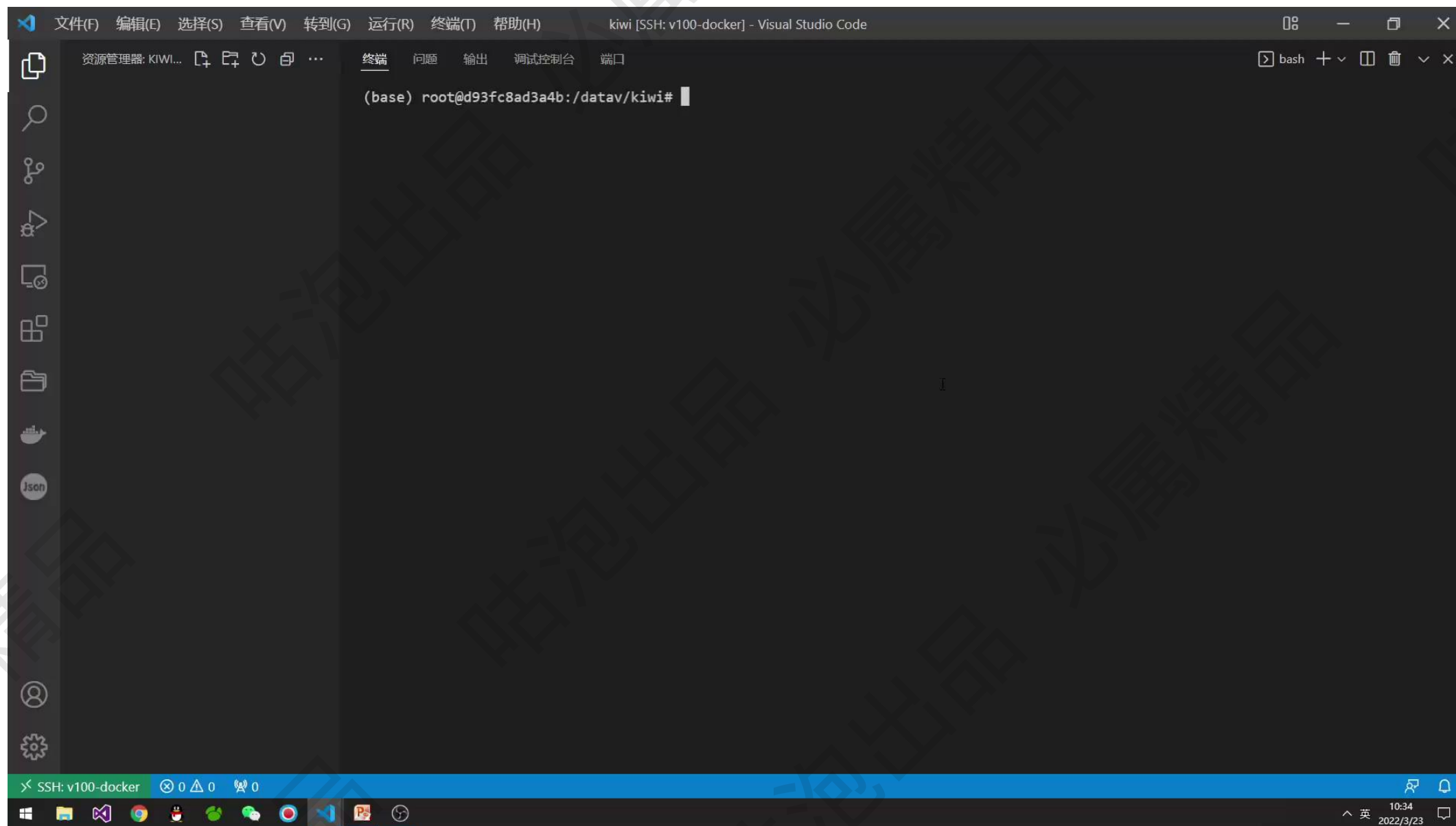
本课程开发了自动环境配置的python包，解决环境配置困扰  
允许零基础学员更友好的学习教程，并进行实验  
力求更好的学习效率

# 说在前面

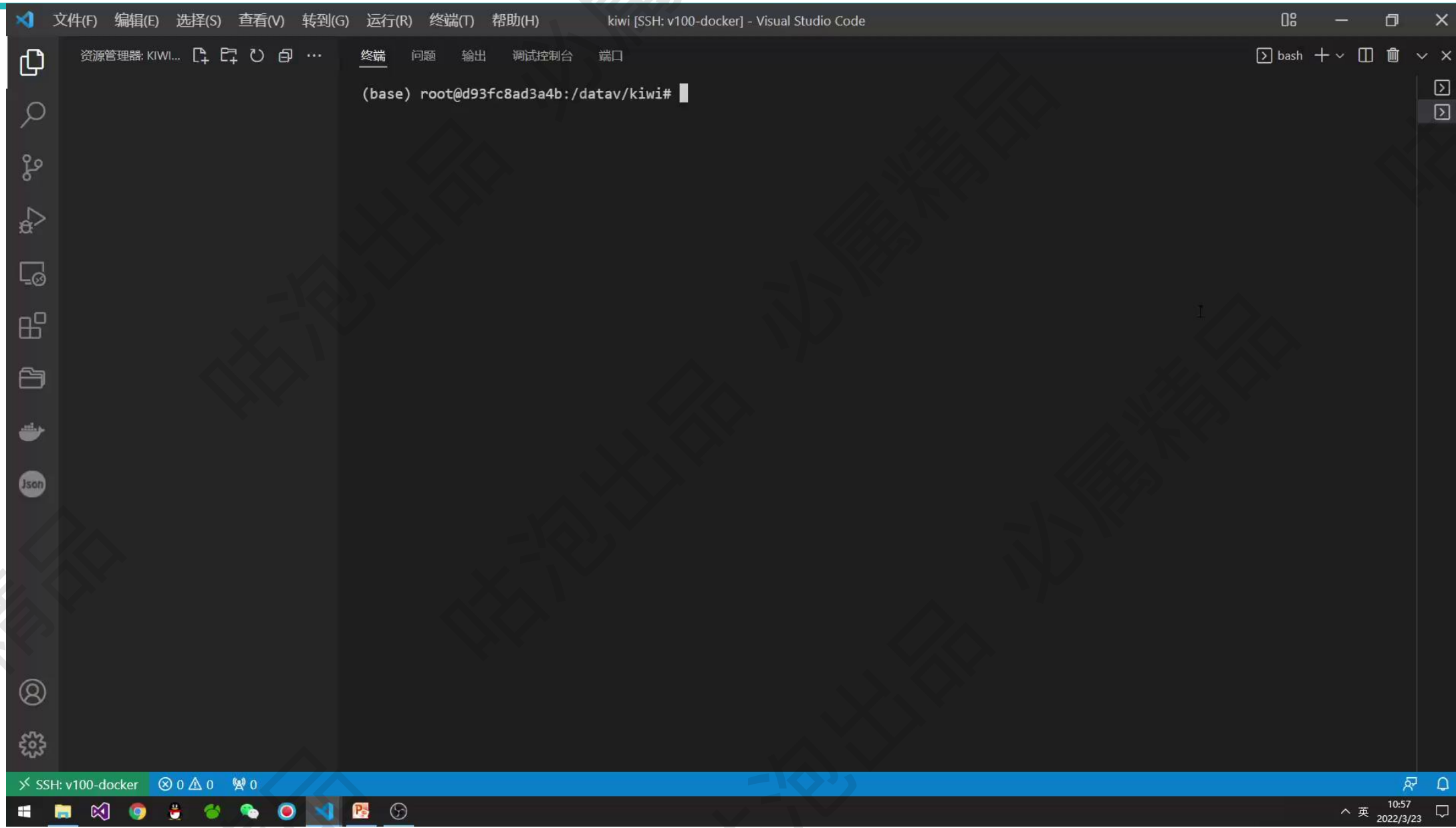
---

同时还提供了约52个学习案例代码  
从浅到深，逐步递进学习

# 实际效果展示-自动环境

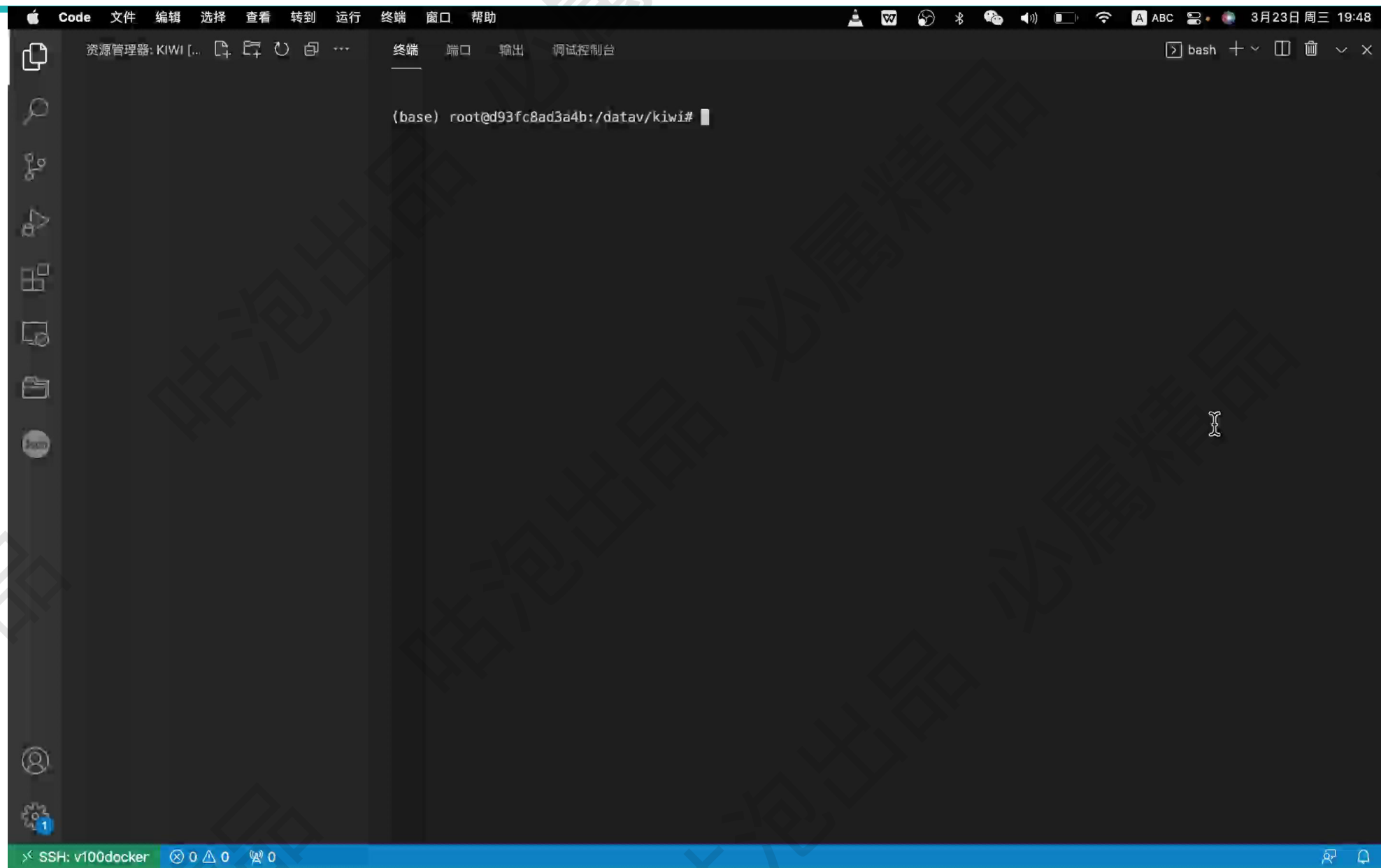


# 实际效果展示-运行案例





# 实际效果展示-运行系列





# 前言

---

## 课程内容

 介绍（学习方式和工具）

 CUDA精简课程（Driver和Runtime）

 TensorRT基础（分类器、检测器、插件、onnx）

 TensorRT集成，项目实战（目标检测、人脸识别、姿态识别等）

# 前言

## 课程需要什么

- ✎ Linux系列Ubuntu系统，并配置有Nvidia显卡和显卡驱动程序（大于450）
- ✎ 可以执行pip进行安装
- ✎ 以上非必要，如果有，则体验最佳，学习效率最高

---

谢谢!