



南京邮电大学

Nanjing University of Posts and Telecommunications

# 科技英文写作

## 融合不确定性估计和NeRF的隐式三维建模方法 设计与实现

汇报人：周帅

2024年6月20日

任课教师：王磊



# 汇报提纲

- 1/ 课题研究背景和意义
- 2/ 课题研究基础
- 3/ 课题研究内容
- 4/ 实验结果与分析

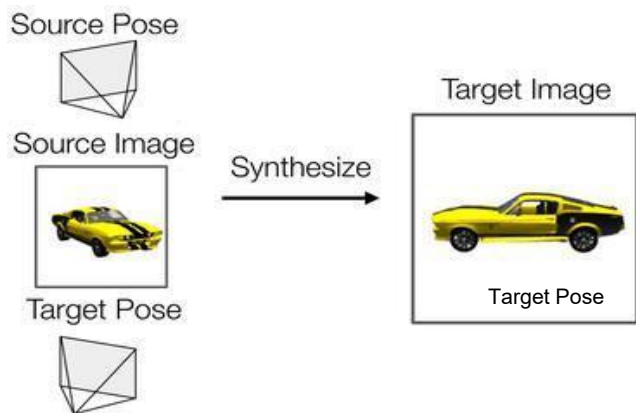


南京邮电大学  
Nanjing University of Posts and Telecommunications

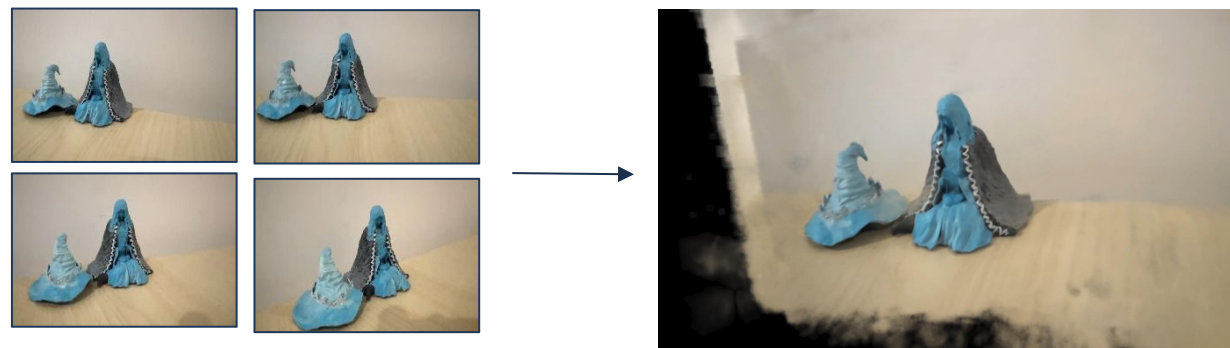
# 1、课题研究背景和意义

## 选题背景及意义

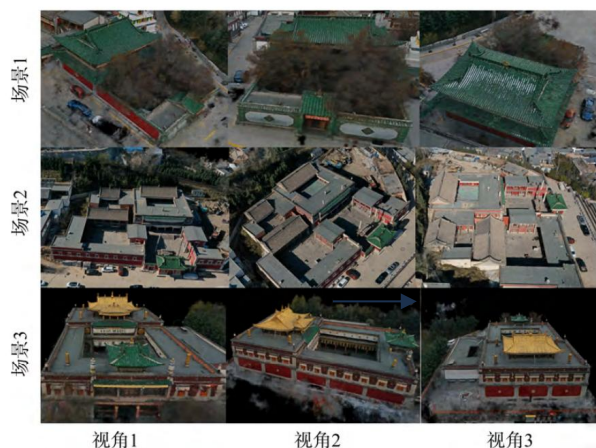
三维重建算法和视图合成技术在虚拟现实、增强现实和计算机游戏等领域具有广阔的应用前景。



合成新视角



照片360°重建



文物建筑重建



3D人体重建



# 国内外研究现状调研与分析

## 传统手工三维场景建模：

早期的手工三维场景建模使用CAD软件或数字内容创作DCC软件进行操作和创建。但是使用CAD软件或DCC软件进行手工三维建模需要一定的技术和经验。

## 神经网络进行三维重建流行方法

- 1、DeepVoxel(图像+卷积)，缺点：细节丢失当物体较小或拍摄距离较远。
- 2、SRN(图像+卷积)，缺点：无法重建复杂场景、只能重建几何细节少的物体、训练时间长(6天)。
- 3、NV(图像+卷积+MLP)，缺点：对于纹理的重建效果差，可能会重建成透明状。
- 4、LLFF(图像+MPI网络)，缺点：渲染出来的图像会有模糊重叠的部分、无法扩展到高分辨率。



# NeRF遇到的挑战

尽管NeRF在三维重建中展现出较好的效果，但其性能受到训练样本质量的高度影响。由于来自场景的有限的摆拍图像，NeRF不能很好地推广到新的视图，并且可能在未观察到的区域崩溃为平凡的解决方案，如右图所示，这就导致了重建结果的失真和不可靠性



因此对于重建结果的可靠性评估至关重要。通过引入不确定性估计，可以确保在观测点较少的情况下保持模型的鲁棒性，并提供对场景理解的解释。因此可以获得更准确和可靠的三维场景表示，并提供与重建结果相关的可信度度量。





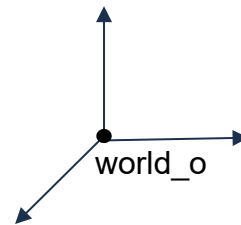
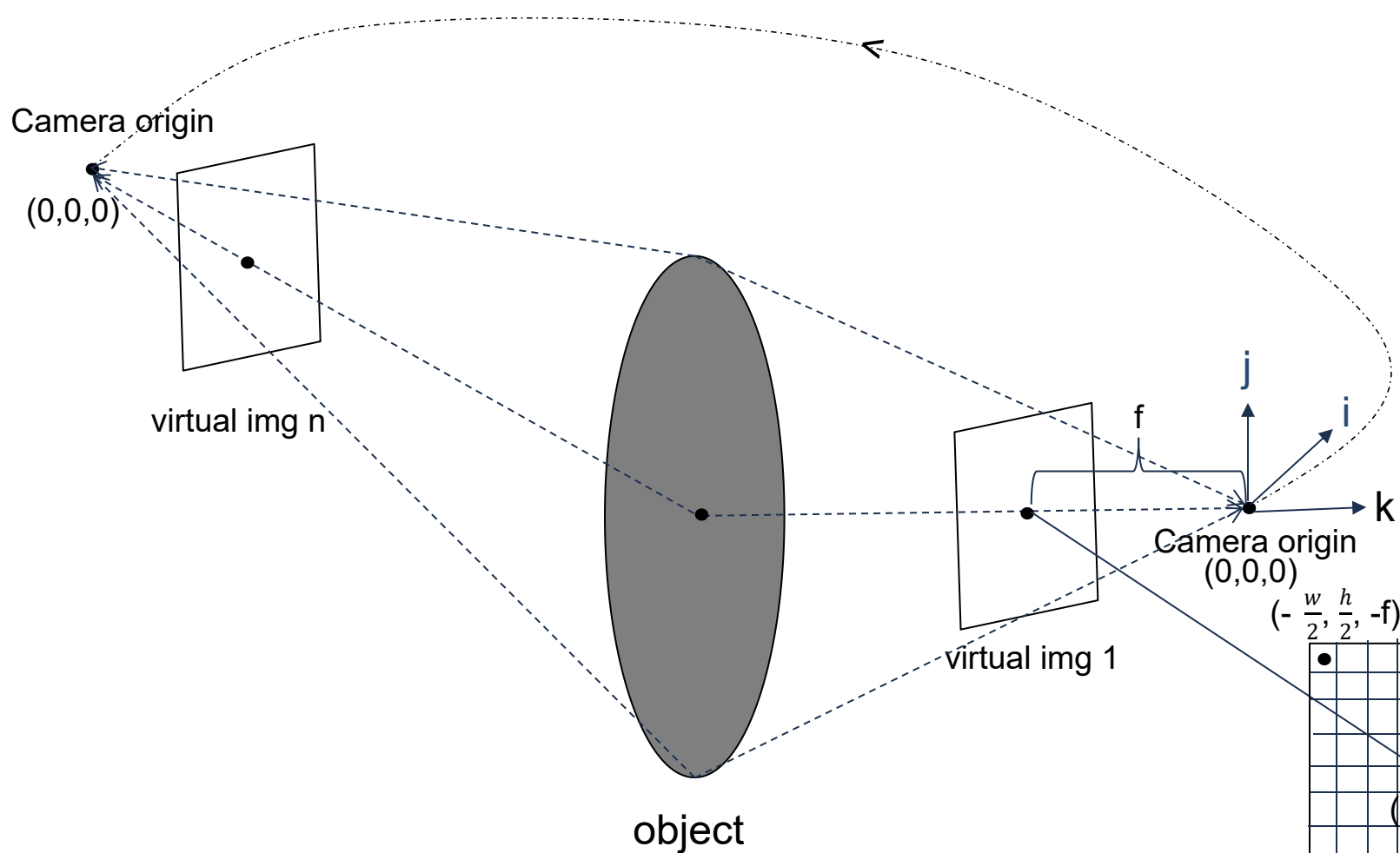
南京邮电大学  
Nanjing University of Posts and Telecommunications

## 2、课题研究基础

# 背景知识



## 相机模型



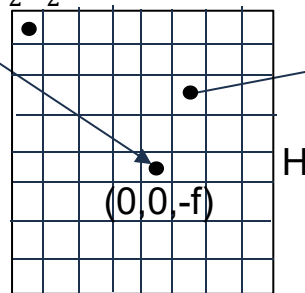
相机坐标系到世界坐标系的转换矩阵

$$M = \begin{bmatrix} R & T \end{bmatrix} \quad \text{其中 } R \text{ 为 } 3 \times 3 \text{ 的旋转矩阵, } T \text{ 为 } 3 \text{ 维的平移向量}$$

因此, 在世界坐标系下:

$$\begin{aligned} \text{ray\_o} &= \text{ray\_0} + T \\ \text{ray\_d} &= R * \text{ray\_d} \end{aligned}$$

$$\left(-\frac{w}{2}, \frac{h}{2}, -f\right)$$



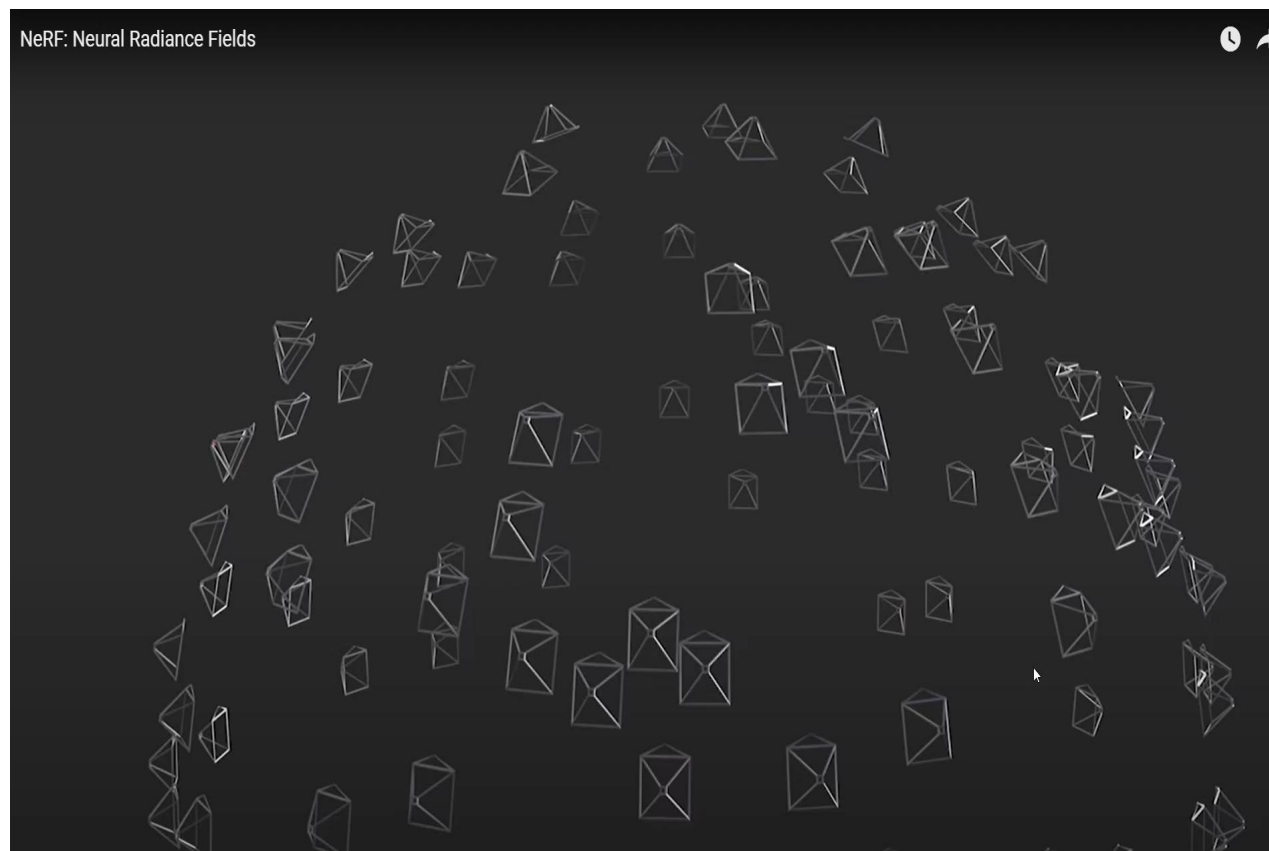
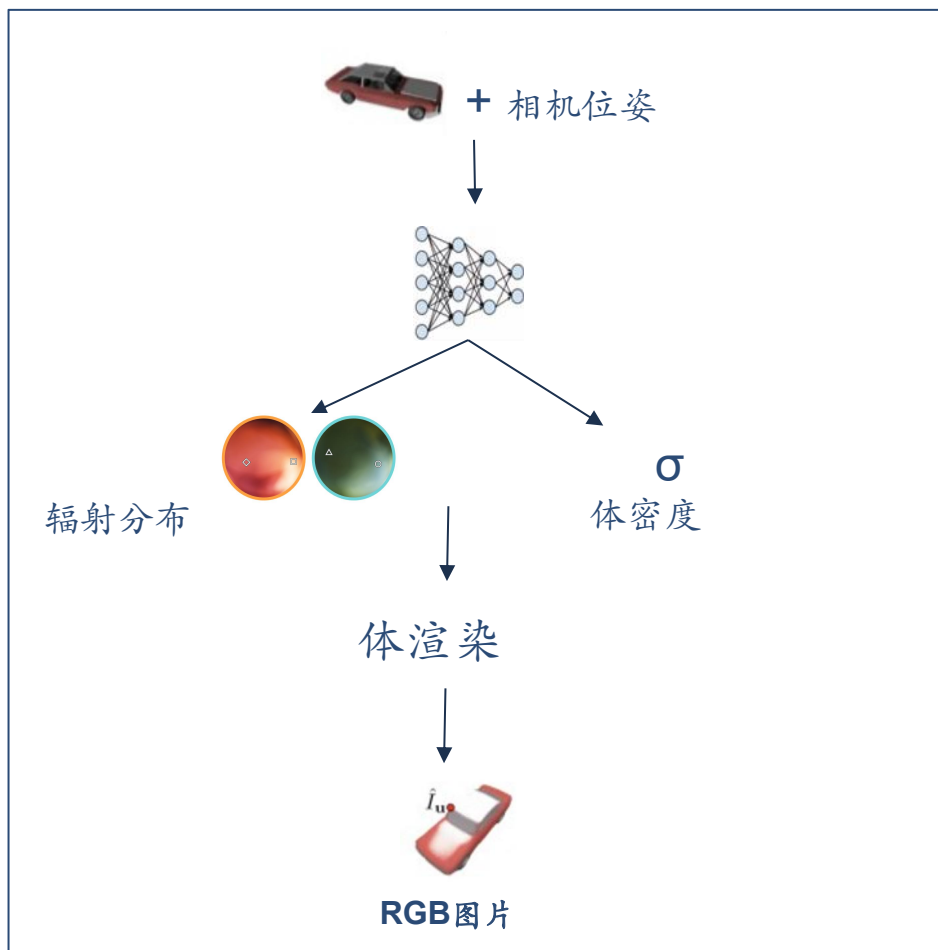
$$\left(-\frac{w}{2} + i, \frac{h}{2} - j, -f\right)$$

由此可得穿过像素(i,j)的光线的起点为ray\_o(0,0,0), 方向为ray\_d(-\frac{w}{2} + i, \frac{h}{2} - j, -f)



# NeRF理论基础

NeRF将隐式重建和体积渲染的思想相结合，提出了神经辐射场方案：将场景表示为5D向量函数 $F(x, y, z, \theta, \phi)$



1. NeRF $\left\{\left(r, C(r)\right)_{(i)}\right\}_{i=1}^B$

(1)各个位置的辐射颜色仅受其自身的三维坐标影响，使得不同位置的辐射颜色分布彼此独立。

$T(s)$ :在s点之前, 光线没有被阻碍的概率

$\sigma(s)$ :在s点出光线被粒子阻碍的概率密度

$C(s)$ :在 $s$ 点出粒子的颜色

$$[\sigma, f] = MLP_{\theta_1}(\gamma_x(x))$$

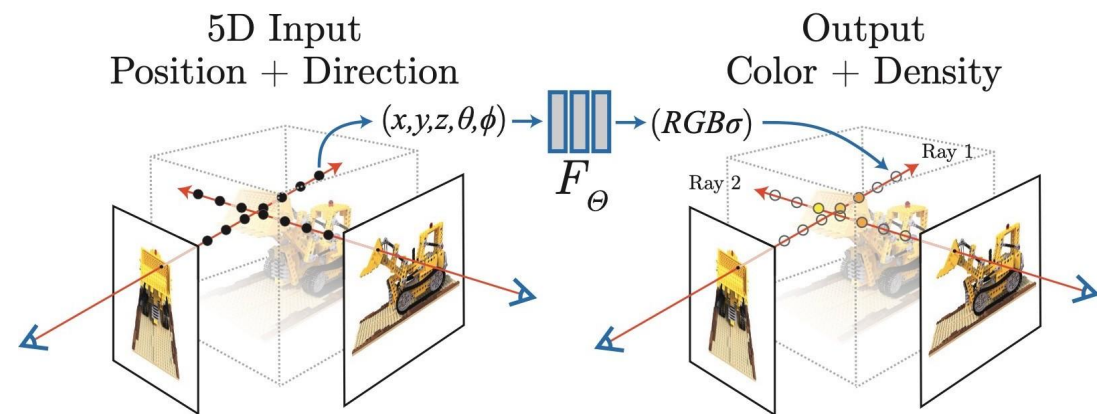
$$c = MLP_{\theta_2}(f, \gamma_d(d))$$

连续表示:  $C(r) = \int_{t_n}^{t_f} T(t) \sigma(r(t)) c(r(t), d) dt$ , where  $T(t) = \exp\left(-\int_{t_n}^t \sigma(r(s)) ds\right)$ .

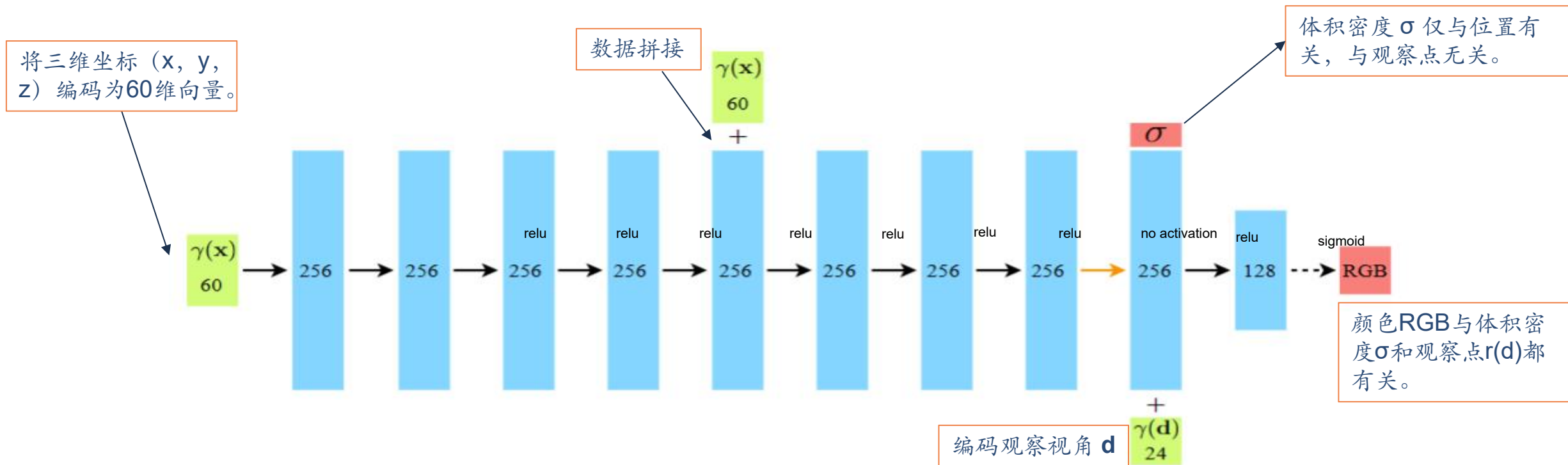
离散近似:  $\widehat{C}(r) = \sum_{i=1}^{N_s} T_i(1 - \exp(-\sigma_i \delta_i)) \mathbf{c}_i$ , where  $T(t) = \exp\left(-\int_{t_n}^t \sigma(r(s)) ds\right)$ .

$$\hat{\mathcal{C}}(r) = \sum_{i=1}^{N_s} \alpha_i c(r(t_i)), \alpha_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right) (1 - \exp(-\sigma_i \delta_i)),$$

3. 损失函数:  $L = \sum_{\mathbf{r} \in R} [\|\widehat{C}_c(\mathbf{r}) - C(\mathbf{r})\|_2^2 + \|\widehat{C}_f(\mathbf{r}) - C(\mathbf{r})\|_2^2]$



# NeRF模型结构



位置编码:  $\gamma(p) = (\sin(2^0\pi p), \cos(2^0\pi p), \dots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p))$ .

$\gamma$ : 不可学习函数

L: 控制着编码的长度, 实现从  $R$  到更高维空间  $R^{2L}$  的映射。

$\gamma(x)$  编码后的位置信息。

$\gamma(d)$  编码后的角度信息

$\sigma(x)$  预测的体密度

$RGB(x, d)$  预测的颜色值



### 3、本课题研究内容

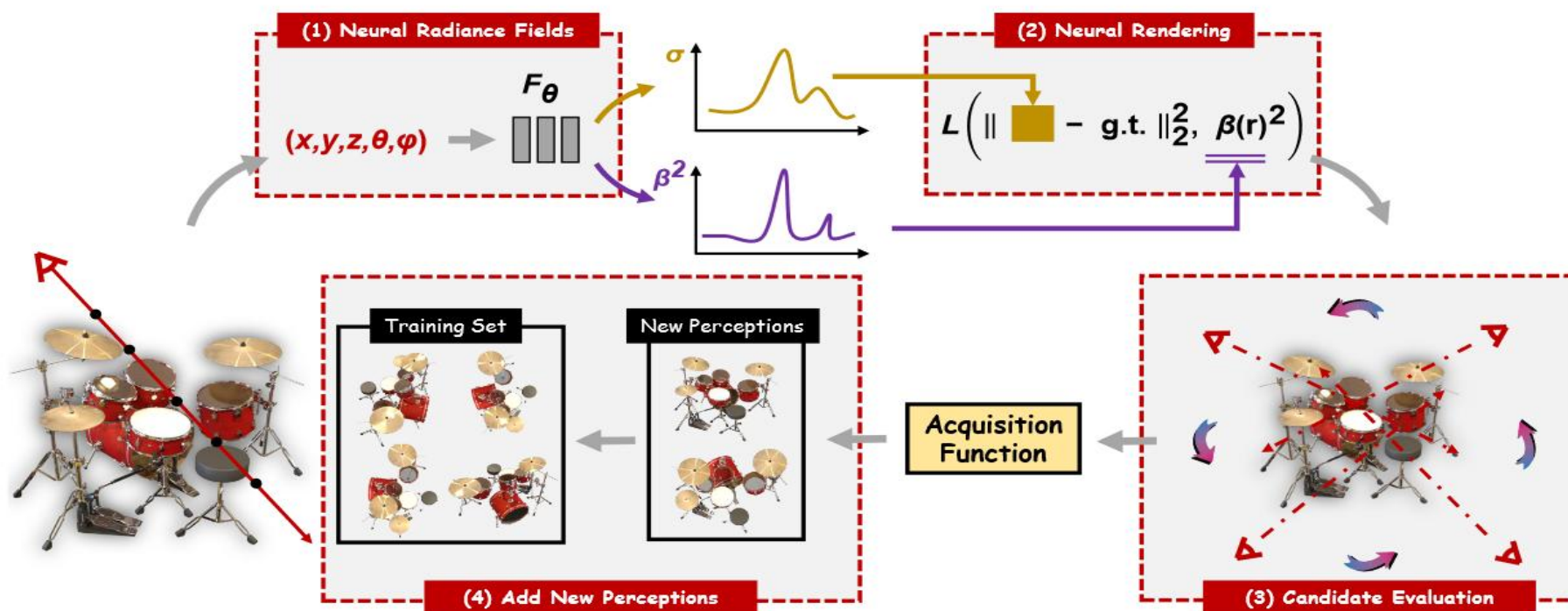
# 研究内容

结合主动学习的思想，引入数据的不确定性估计

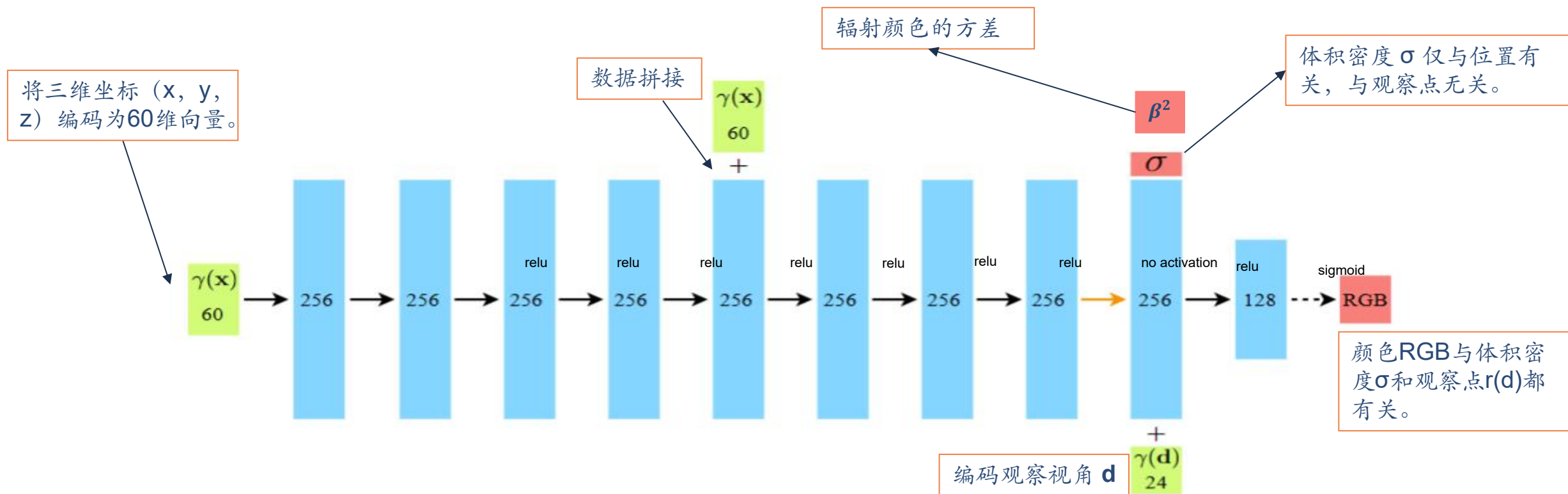
实现将各位置的辐射颜色建模为高斯分布，模型被迫在未观察到的区域提供更大的方差

$$[\sigma, f, \beta^2(r(t))] = MLP_{\theta_1, \theta_3}(\gamma_x(x))$$

$$\hat{c} = MLP_{\theta_2}(f, \gamma_d(d))$$



# 不确定性融入NeRF模型



位置编码:  $\gamma(p) = (\sin(2^0\pi p), \cos(2^0\pi p), \dots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p))$ .

$\gamma$ : 不可学习函数

$L$ : 控制着编码的长度, 实现从  $R$  到更高维空间  $R^{2L}$  的映射。

$\gamma(x)$  编码后的位置信息。

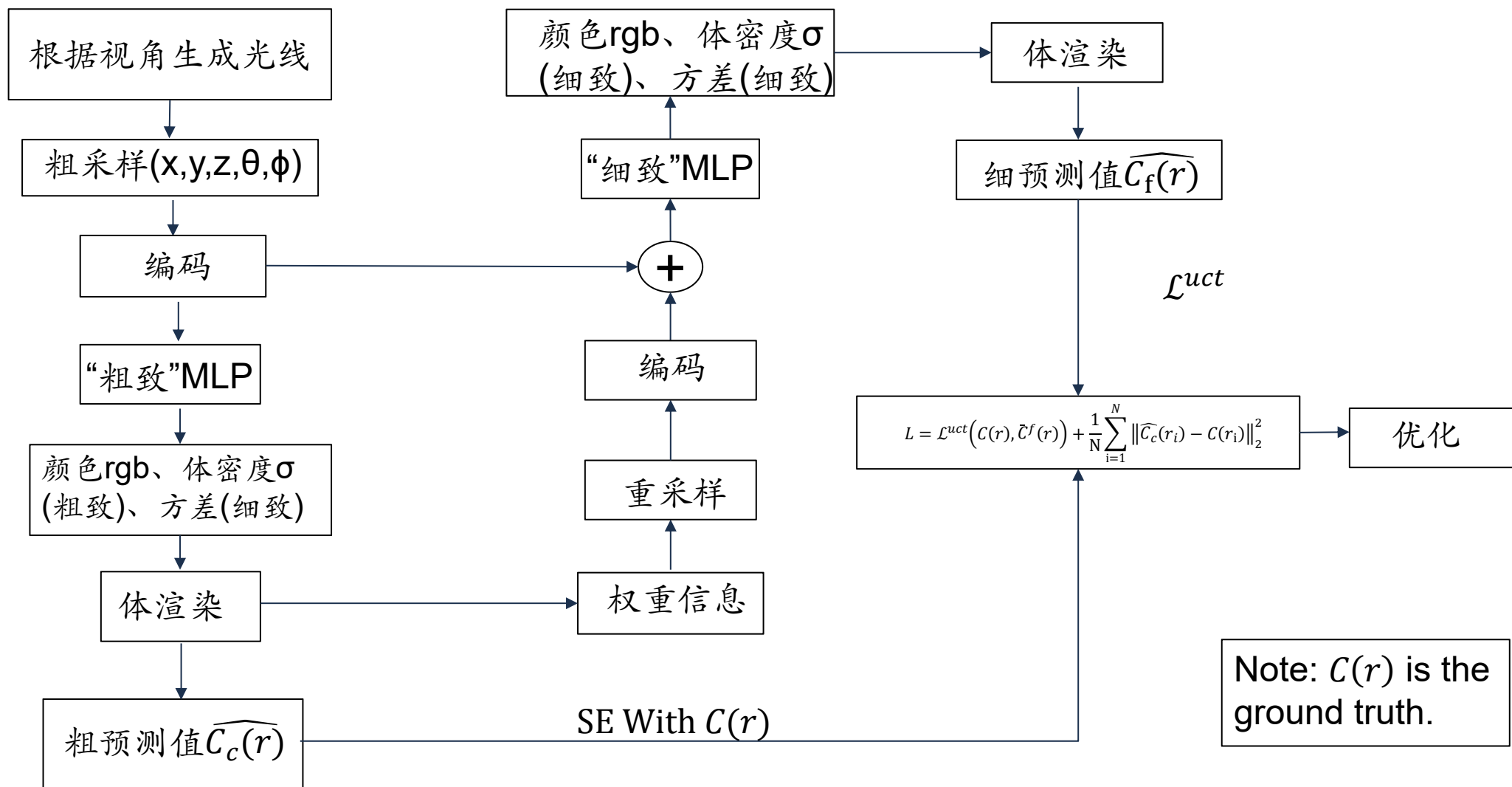
$\gamma(d)$  编码后的角度信息

$\sigma(x)$  预测的体密度

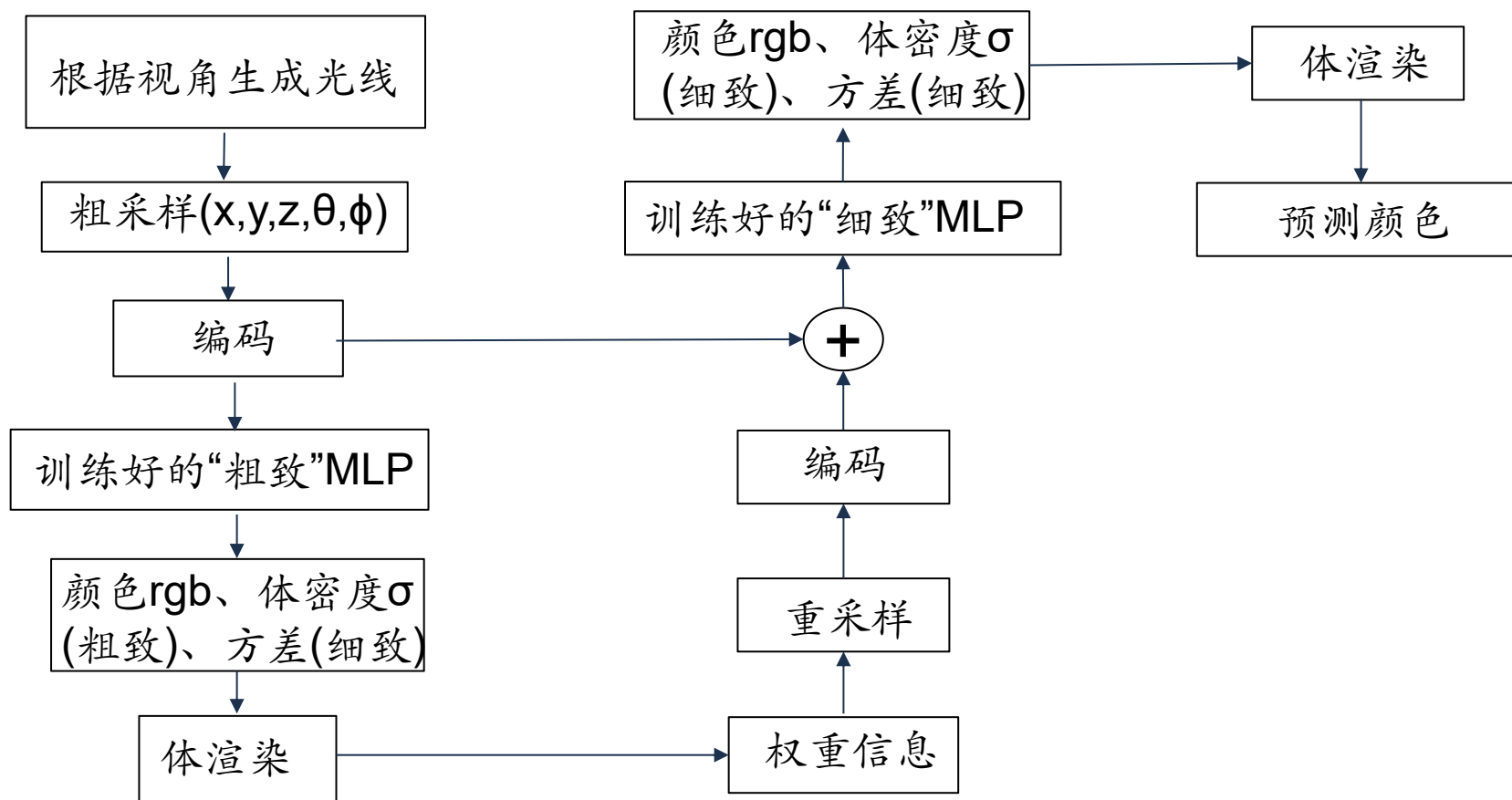
$RGB(x, d)$  预测的颜色值



# 本模型训练阶段



# 本模型测试阶段



# 研究内容

结合主动学习的思想

实现将各位置的辐射颜色建模为高斯分布，模型被迫在未观察到的区域提供更大的方差

$$[\sigma, f, \beta^2(r(t))] = MLP_{\theta_1, \theta_3}(\gamma_x(x))$$

$$\hat{c} = MLP_{\theta_2}(f, \gamma_d(d))$$

$$c(r(t)) \sim N(\bar{c}(r(t)), \bar{\beta}^2(r(t)))$$

$$\hat{C}(r) \sim N(\sum_{i=1}^{N_s} \alpha_i \bar{c}(r(t_i)), \sum_{i=1}^{N_s} \alpha_i \bar{\beta}^2(r(t_i))) \sim N(\bar{C}(r), \bar{\beta}^2(r))$$

$$\text{负对数似然} \min_{\theta} -\log p_{\theta}(\mathcal{B}) = -\frac{1}{N} \sum_{i=1}^N \log p_{\theta}(C(r_i))$$

$$\mathcal{L}^{uct} = \frac{1}{N} \sum_{i=1}^N \left( \frac{\|C(r_i) - \bar{C}(r_i)\|_2^2}{2\bar{\beta}^2(r_i)} + \frac{\log \bar{\beta}^2(r_i)}{2} + \frac{\lambda}{N_s} \sum_{j=1}^{N_s} \sigma_i(r_i(t_j)) \right)$$

$$\text{损失函数: } L = \mathcal{L}^{uct}(C(r), \bar{C}^f(r)) + \frac{1}{N} \sum_{i=1}^N \|\hat{C}_c(r_i) - C(r_i)\|_2^2$$

## 公式推导

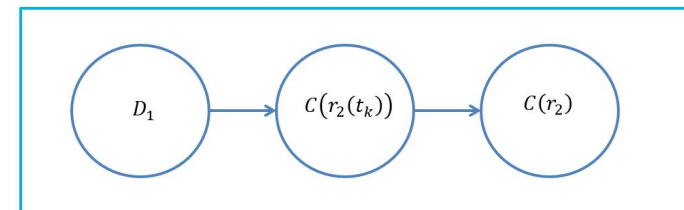
$$c(r(t)) \sim N(\bar{c}(r(t)), \bar{\beta}^2(r(t)))$$

$$\hat{C}(r) \sim N(\sum_{i=1}^{N_s} \alpha_i \bar{c}(r(t_i)), \sum_{i=1}^{N_s} \alpha_i \bar{\beta}^2(r(t_i))) \sim N(\bar{C}(r), \bar{\beta}^2(r))$$

设 $D_1$ 表示现有的训练集， $F_\theta$ 表示给定 $D_1$ 的训练过的NeRF模型。考虑来自新输入 $D_2$ 的 $r_2$ 。

$$\text{先验分布: } P(\text{pri}) = P(c(r_2(t_k))|D_1) \sim \mathcal{N}(\bar{c}(r_2(t_k)), \bar{\beta}^2(r_2(t_k)))$$

$$\begin{aligned} \text{后验分布: } P(\text{post}) &= P(c(r_2(t_k))|D_1, r_2) = \frac{p(C(r_2)|c(r_2(t_k)))p(c(r_2(t_k))|D_1)}{\int p(C(r_2)|c(r_2(t_k)))p(c(r_2(t_k))|D_1)dc(r_2(t_k))} \\ &\sim \mathcal{N}\left(\gamma \frac{C(r_2) - b(t_k)}{\alpha_k} + (1 - \gamma)\bar{c}(r_2(t_k)), \frac{\bar{\beta}^2(r_2(t_k))\bar{\beta}^2(r_2)}{\alpha_k^2 \bar{\beta}^2(r_2(t_k)) + \bar{\beta}^2(r_2)}\right), \\ &\text{with } \gamma = \frac{\alpha_k^2 \bar{\beta}^2(r_2(t_k))}{\alpha_k^2 \bar{\beta}^2(r_2(t_k)) + \bar{\beta}^2(r_2)}. \end{aligned}$$



## 假设

在训练批次中，场景中的每个位置最多只被采样一次，即各自辐射分布相互独立

$$\begin{aligned}\text{Var}^{(\text{post})}(r_2(t_k)) &= \frac{\beta^2(r_2(t_k))\beta^2(r_2)}{\alpha_k^2\bar{\beta}^2(r_2(t_k))+\bar{\beta}^2(r_2)} \\ &= \left(\frac{1}{\bar{\beta}^2(r_2(t_k))} + \frac{\alpha_k^2}{\bar{\beta}^2(r_2)}\right)^{-1} < \bar{\beta}^2(r_2(t_k)) = \text{Var}^{(\text{pri})}(r_2(t_k)). \quad \# \end{aligned}$$

新观察到的位置 $r_2(t_k)$ 的后验分布方差始终小于其先验分布

——将方差减小视作来自新射线 $r_2$ 的 $r_2(t_k)$ 信息增益的估计

$$\text{Var}^{(\text{pri})}(r_2(t_k)) - \text{Var}^{(\text{post})}(r_2(t_k)).$$

定义信息增量和:

$$\mathcal{A}(D_2) = \sum_{r_i \in D_2} \sum_{j=1}^{N_s} \left( \text{Var}^{(\text{pri})}(r_i(t_j)) - \text{Var}^{(\text{post})}(r_i(t_j)) \right)$$

从球形空间选择候选, 计算每个候选视图的信息增益和, 选择和最高的前 $k$ 个候选视图来补充当前训练集, 以此来实现主动学习策略。



## 4、实验结果与分析



# 实验设置



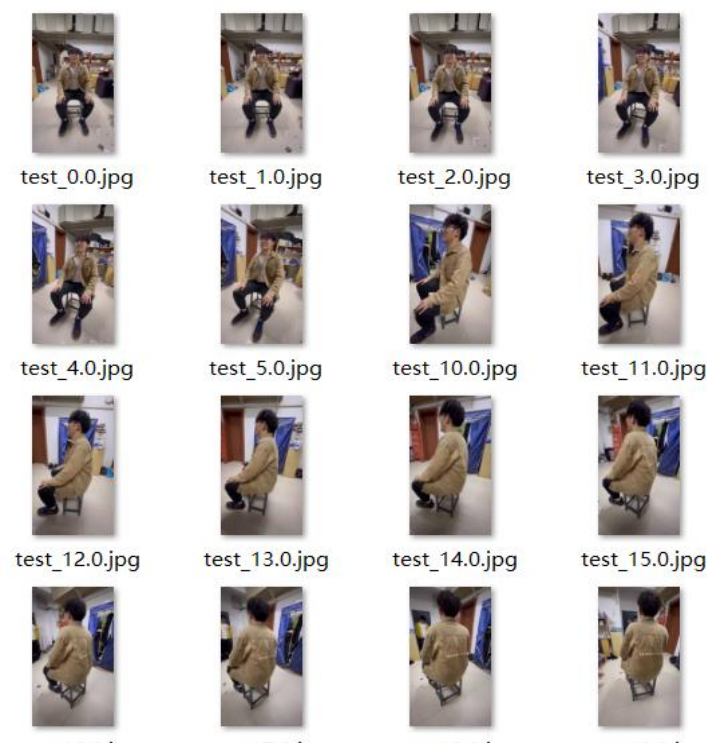
南京邮电大学  
Nanjing University of Posts and Telecommunications

实验环境: Vscode

实验数据集: 合成数据集、自定义数据集



lego数据集



自定义ZS数据集

# 数据预处理



## 1.采集图像

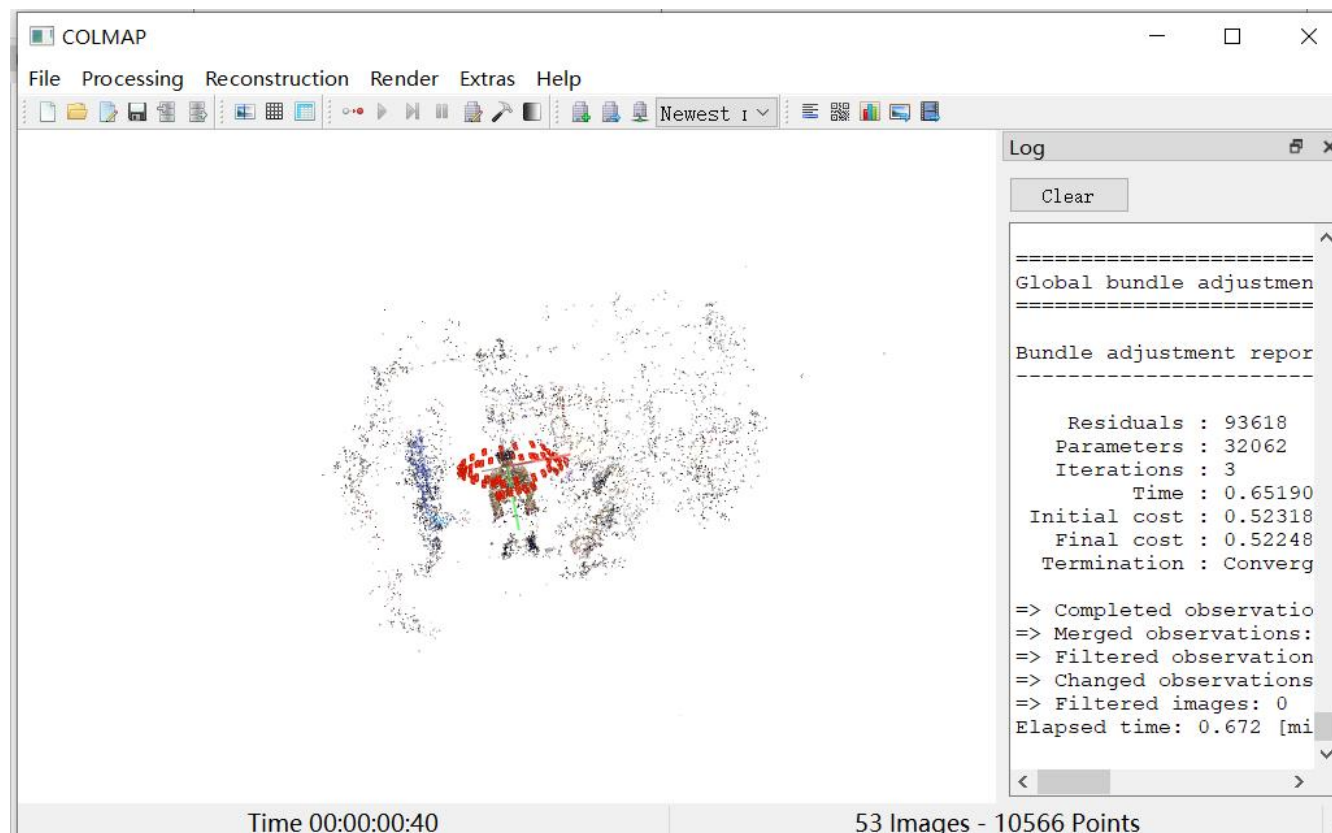
拍摄视频，再做抽帧处理

## 2.位姿计算

使用colmap首先对采集的图像进行特征提取，再进行特征匹配，最后进行稀疏重建。

## 3.转换成LLFF数据格式

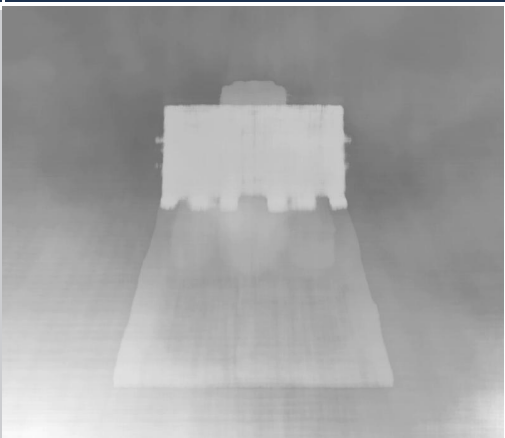


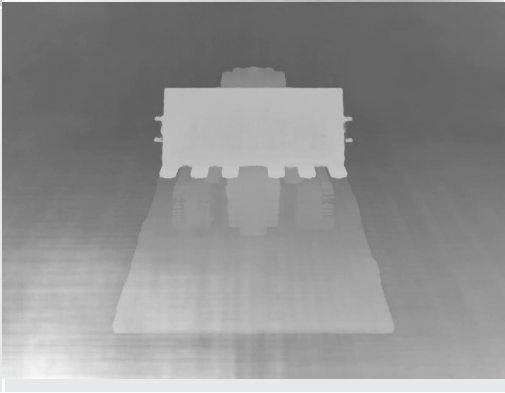

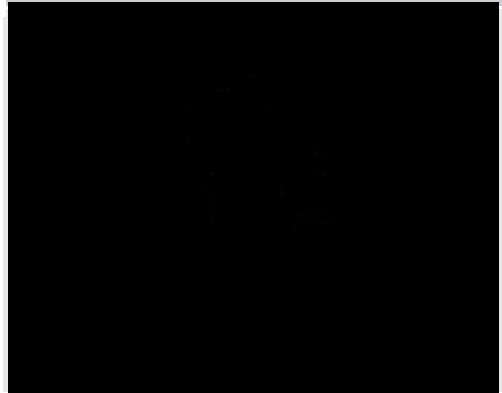
将上一步保存的图像位姿与图像输入到LLFF代码中，确保图像与位姿匹配，生成NeRF可执行的数据集格式。



## 实验设置与结果分析

对于合成数据集lego，其包含lego挖掘机的多视角的真实场景图像和其对应的相机姿态。

对比5万与50万轮的渲染效果如下：

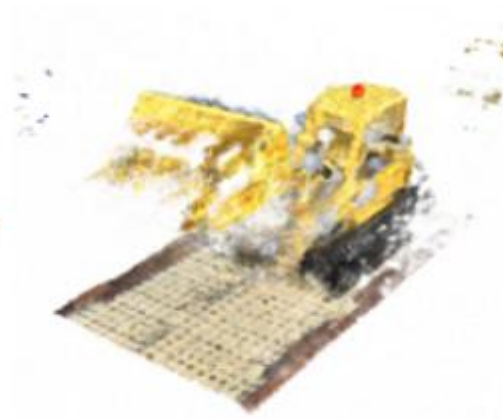
训练程度	disp灰度图	rgb彩色图	不确定性分布图
5万轮			
50万轮			

## 结果分析

使用10个位姿图片进行渲染，效果对比如下：



Ground Truth



NeRF



融合不确定性后的效果

从训练集中移除左侧的观察值，观察渲染合成效果对比如下：



Ground Truth



NeRF



融合不确定性后的效果



# 实验指标与实验对比

PSNR(峰值信噪比)衡量预测图像的质量，值越高则图像质量越高。

SSIM(结构相似性指数)衡量预测图像与真实图像的相似度，值越高说明两张图片越相似。

LPIPS(学习感知图像相似度)衡量预测图像与真实图像之间的差别，值越低则两张图像越相似。

下表展示了本文所实现方法与其他三维隐式建模方法在使用不同比例的训练样本对合成场景和真实场景进行定性分析。

	(a) 合成场景			(b) 真实场景		
方法	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
数据集 1, 使用全部图片训练						
SRN	22.26	0.846	0.170	22.84	0.668	0.378
LLFF	24.88	0.911	0.114	24.13	0.798	0.212
NeRF	31.01	0.947	0.081	26.50	0.811	0.250
Ours	30.45	0.954	0.072	25.96	0.835	0.213
数据集 2, 使用 10 张图片训练						
NeRF	28.04	0.866	0.134	23.36	0.791	0.280
DietNeRF	28.42	0.891	0.087	—	—	—
Ours	28.51	0.932	0.090	23.96	0.803	0.260
数据集 3, 使用 5 张图片训练						
NeRF	21.14	0.835	0.192	21.67	0.689	0.350
Ours	23.23	0.866	0.185	22.03	0.712	0.292

# 实验指标与实验对比

下表展示了一种主动学习策略：连续学习CL和采用贝叶斯估计法BE的情况下，与NeRF + Random和NeRF + FVS进行比较，得到的定性分析结果。

方法	(a) 合成场景				(b) 真实场景		
	时间	PSNR	SSIM	LPIPSJ	PSNR	SSIM	LPIPSJ
设置 1, 共 20 个观测值:							
NeRF+Rand	2.0h	24.25	0.734	0.207	20.65	0.532	0.312
NeRF+FVS	2.0h	26.00	0.812	0.144	22.41	0.710	0.299
Ours-BE	30min	25.67	0.778	0.169	21.86	0.644	0.303
Ours-CL	2.2h	26.24	0.856	0.124	23.12	0.765	0.292
NeRF†	2.0h	28.04	0.910	0.134	23.36	0.791	0.280
设置 2, 共 10 个观测值:							
NeRF+Rand	1.0h	18.36	0.642	0.251	18.49	0.478	0.355
NeRF+FVS	1.0h	19.24	0.735	0.227	20.02	0.633	0.344
Ours-BE	16min	18.25	0.611	0.256	18.67	0.451	0.367
Ours-CL	1.1h	20.01	0.832	0.204	20.14	0.664	0.325
NeRF†	1.0h	21.14	0.835	0.192	21.67	0.689	0.350



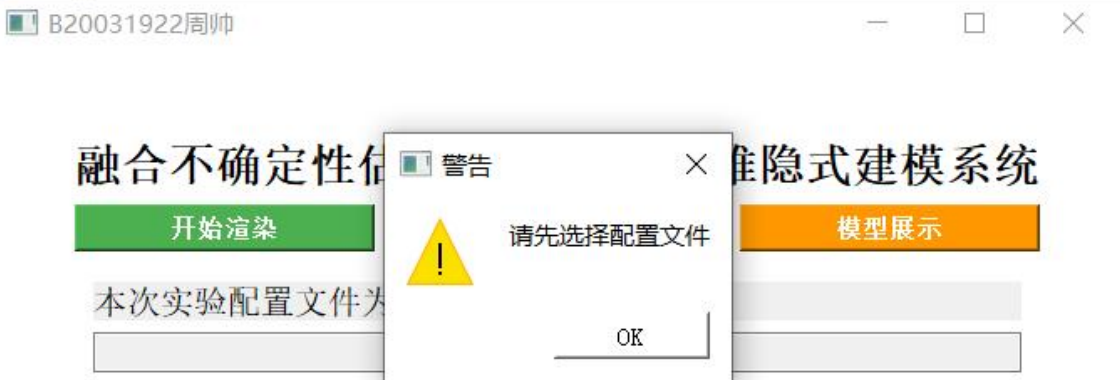


# 可视化界面展示

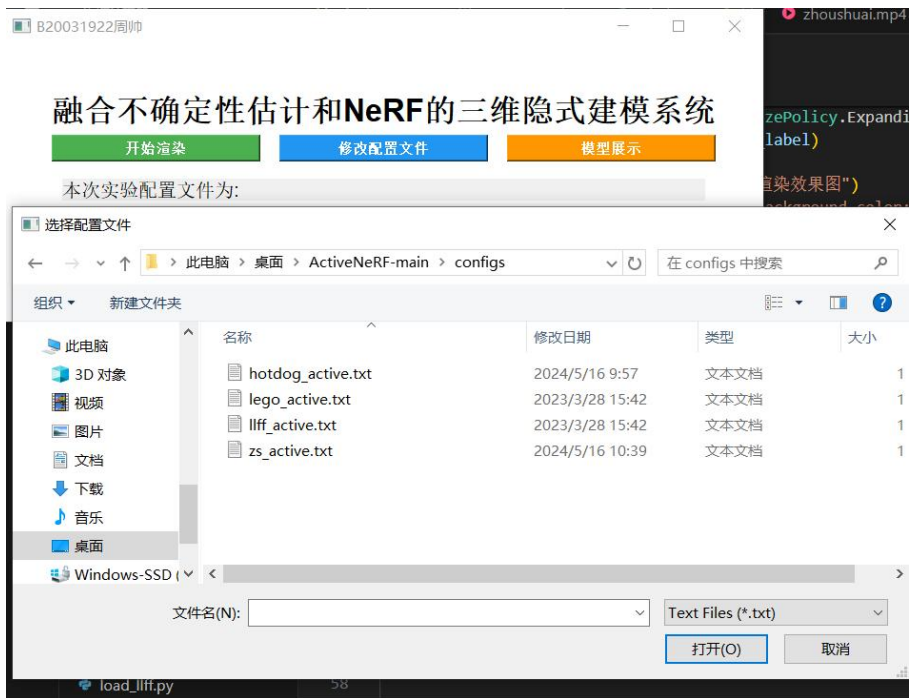
## 主页面



## 报错预警



## 选择配置文件并修改





# 2024

汇报人：周帅