



中國農業大學  
China Agricultural University

# 人工智能——视觉定位

---

胡标





# 目录

## Contents

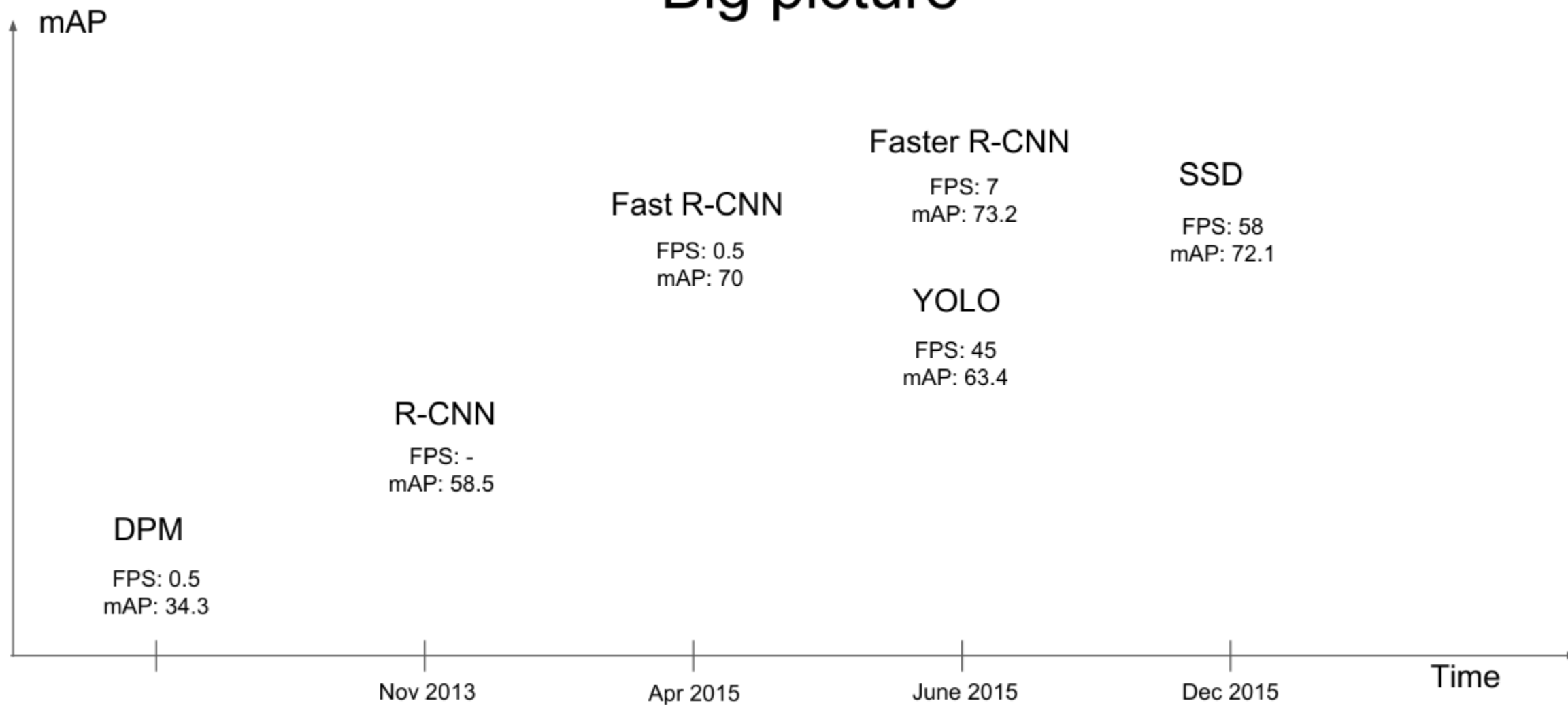
1. YOLO多目标检测

2. 视觉几何

3. 定位原理

4. 粒子滤波算法

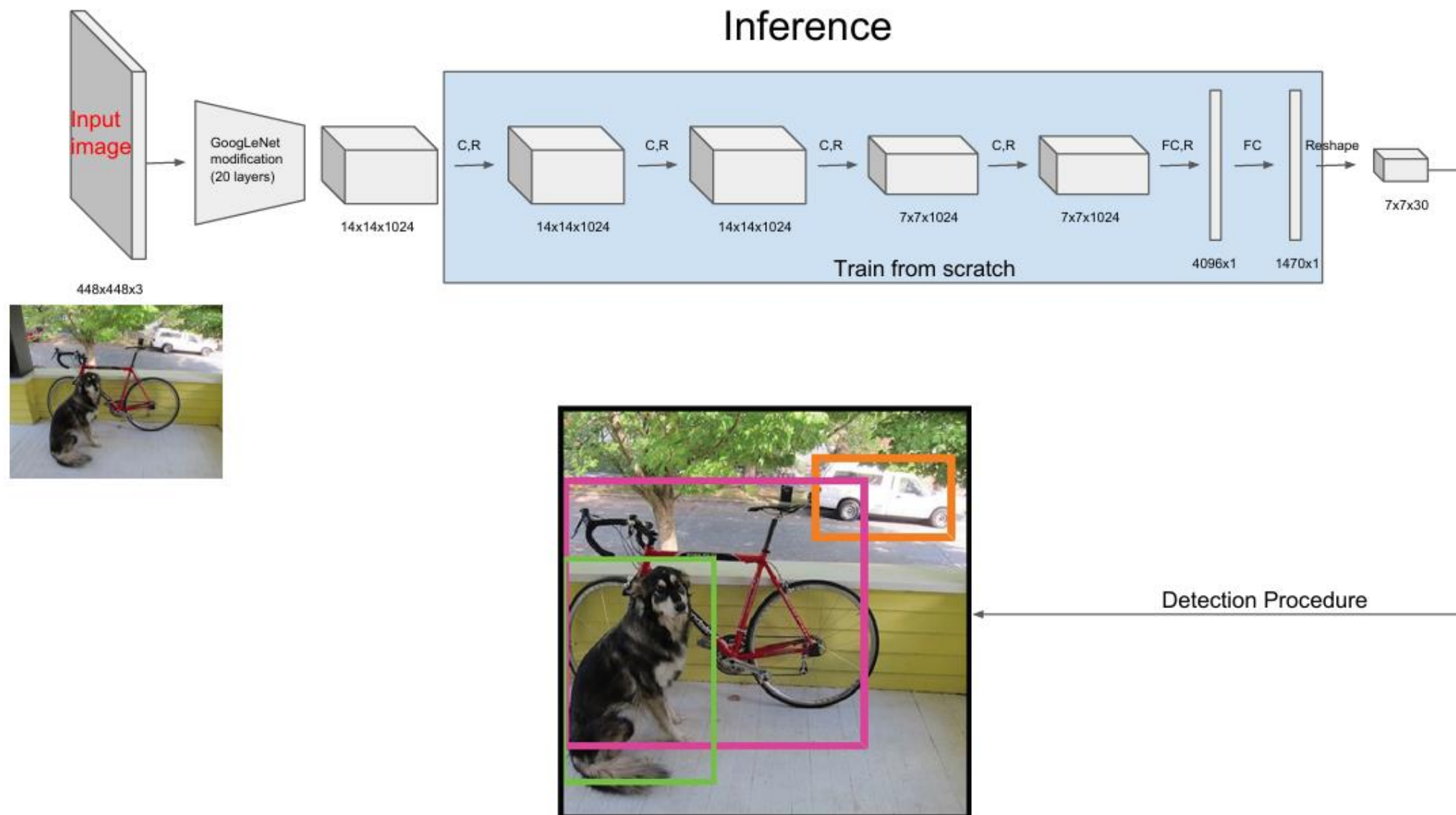
## Big picture



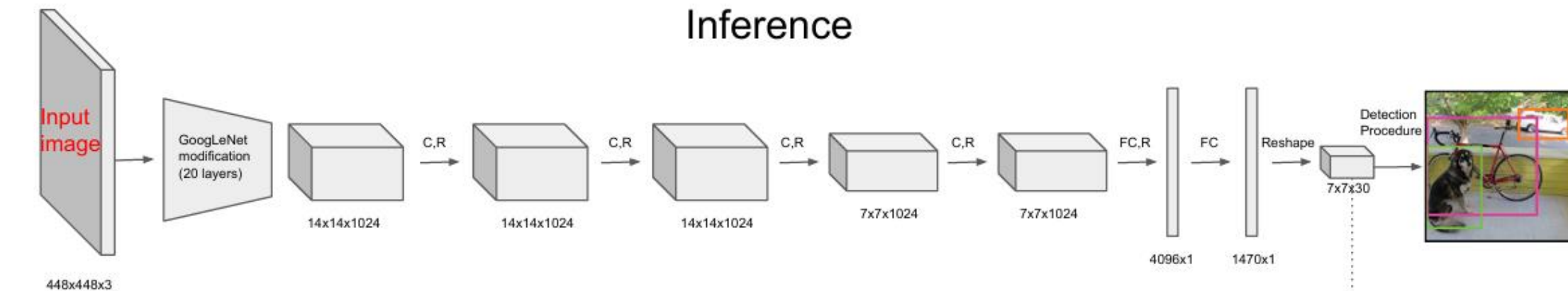
# YOLO多目标检测



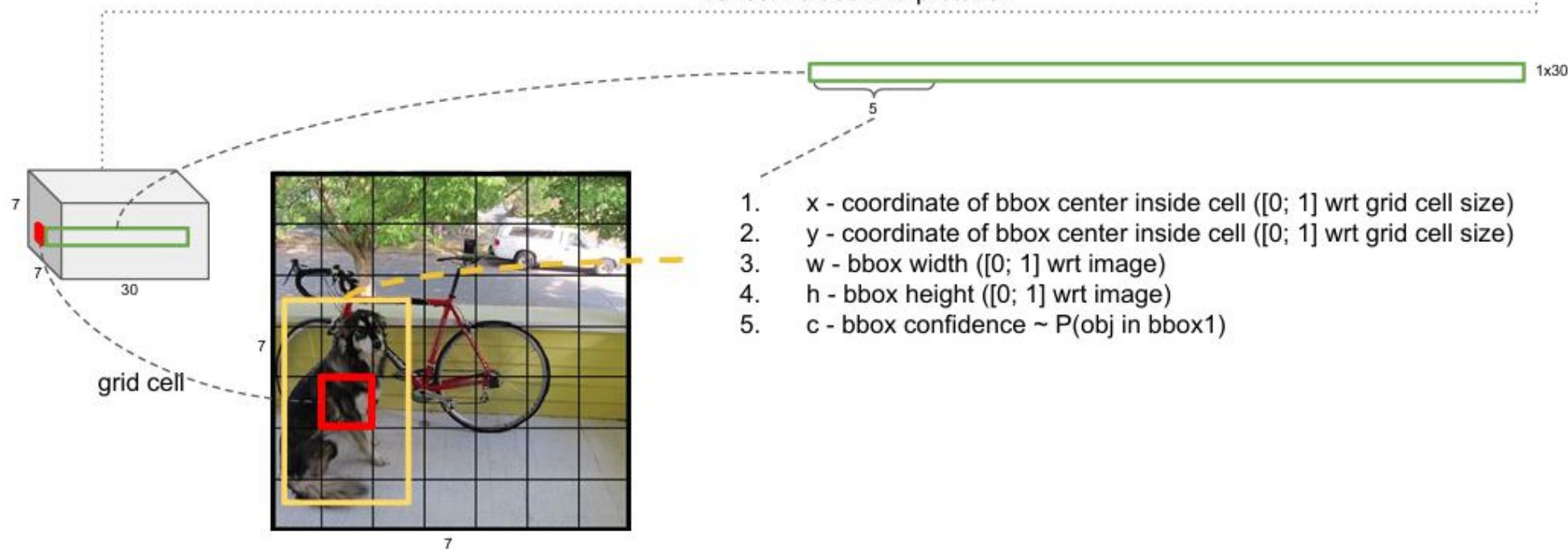
中國農業大學  
China Agricultural University



## Inference



## Tensor values interpretation





predictor中x,y,w,h

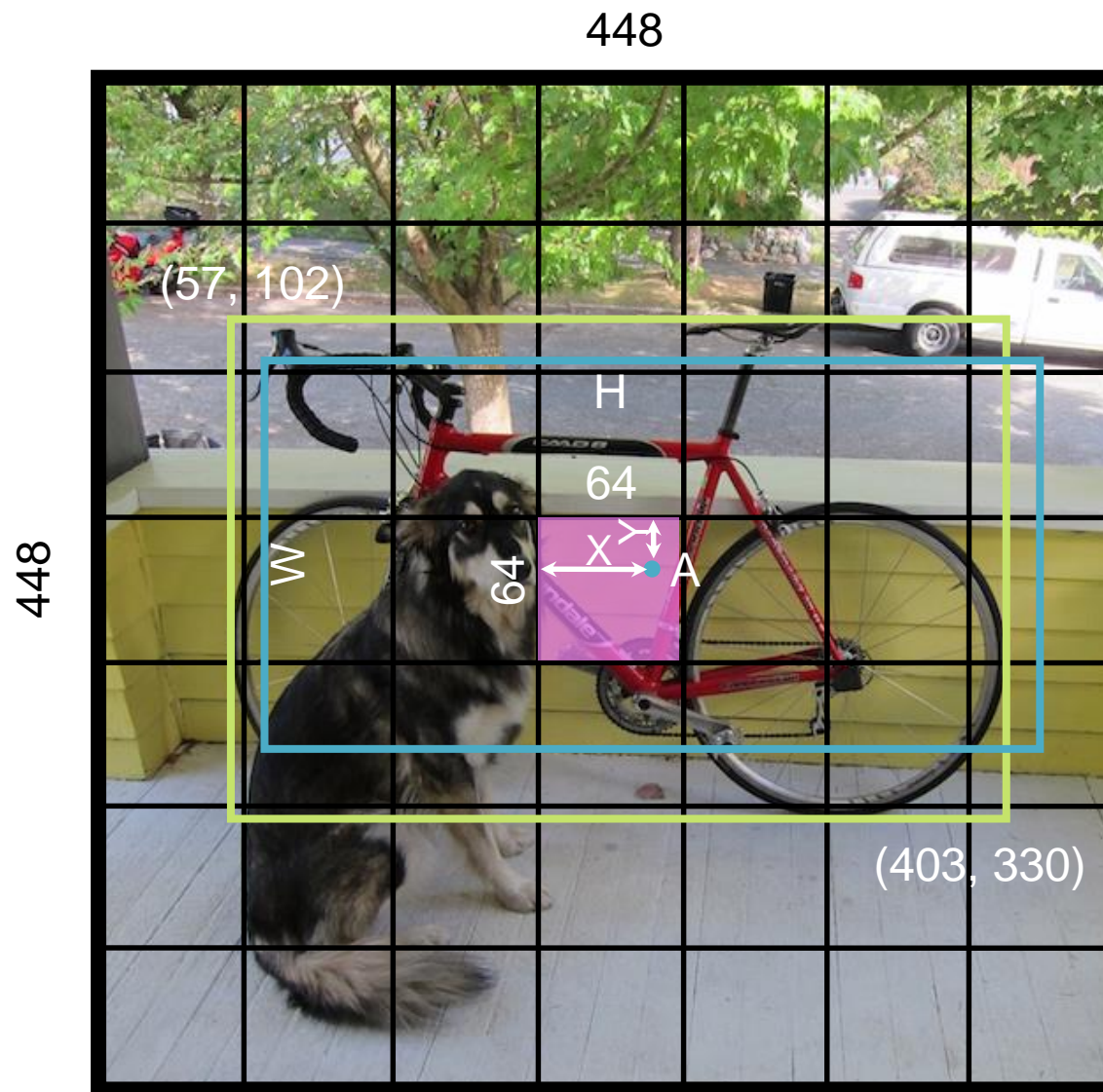
代表的含义

$$\hat{x}_i = (\text{center\_x} - 64 \times 3) / 64$$

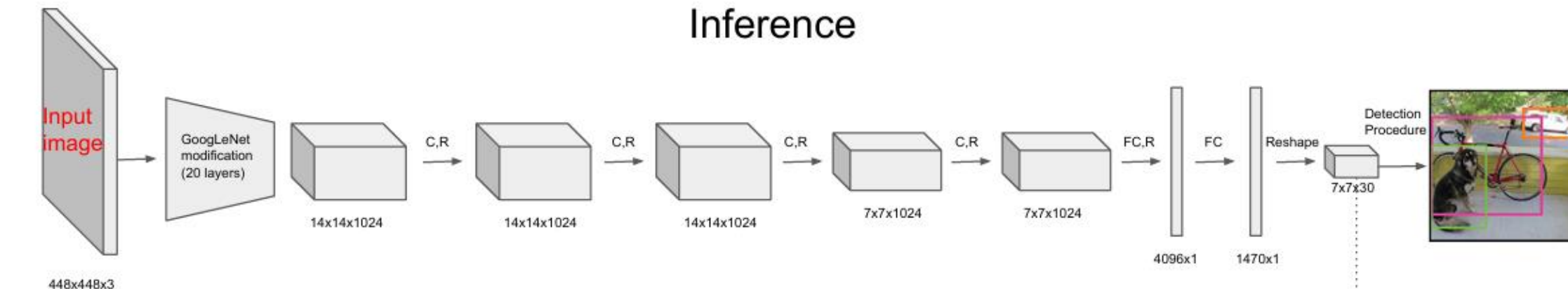
$$\hat{y}_i = (\text{center\_y} - 64 \times 3) / 64$$

$$\hat{w}_i = (403 - 57) / 448$$

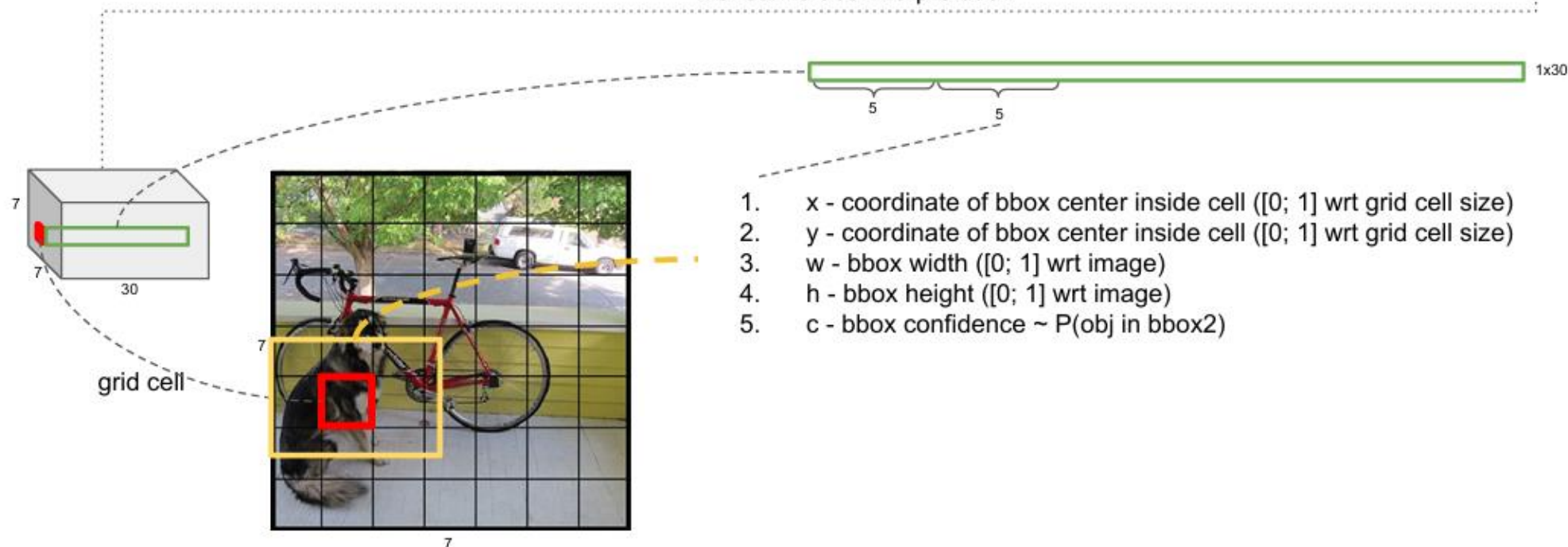
$$\hat{h}_i = (430 - 102) / 448$$



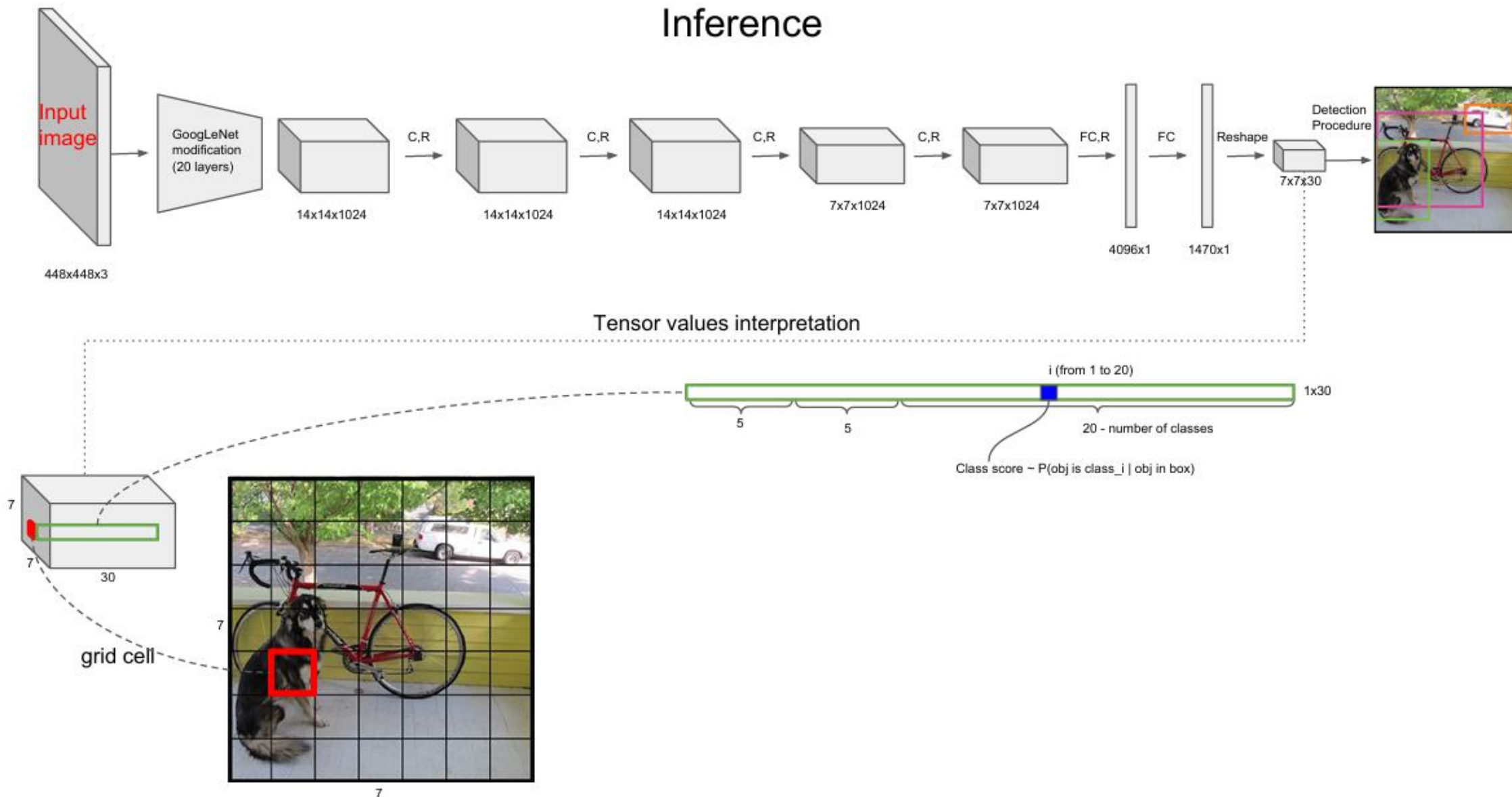
## Inference



## Tensor values interpretation



## Inference





$$\text{Loss} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[ (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \quad \text{位置损失}$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left[ \left( \sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left( \sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

$= \text{Pr}(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}}$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{obj}} \left( C_i - \hat{C}_i \right)^2 \quad \text{置信度损失}$$

$$+ \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{1}_{ij}^{\text{noobj}} \left( C_i - \hat{C}_i \right)^2$$

$= \text{Pr}(\text{Class}_i | \text{Object})$

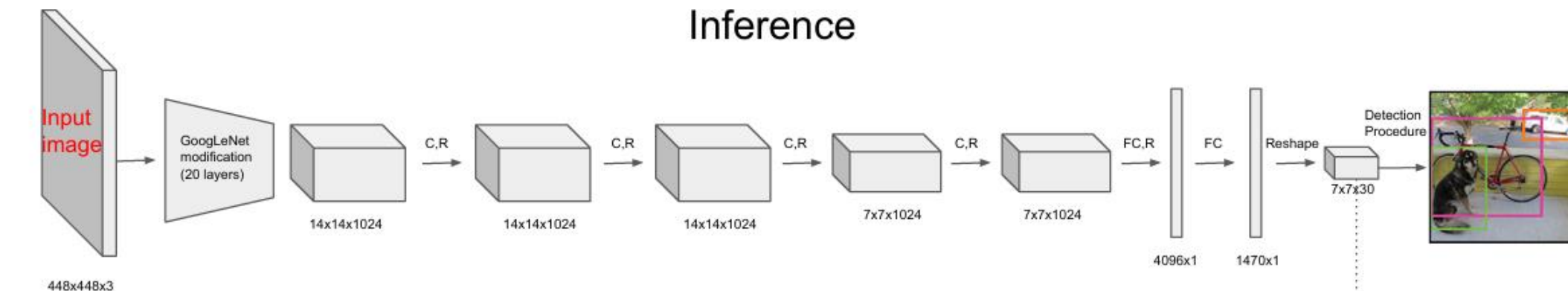
$$+ \sum_{i=0}^{S^2} \mathbb{1}_i^{\text{obj}} \sum_{c \in \text{classes}} \left( p_i(c) - \hat{p}_i(c) \right)^2 \quad \text{类别损失}$$

# YOLO多目标检测

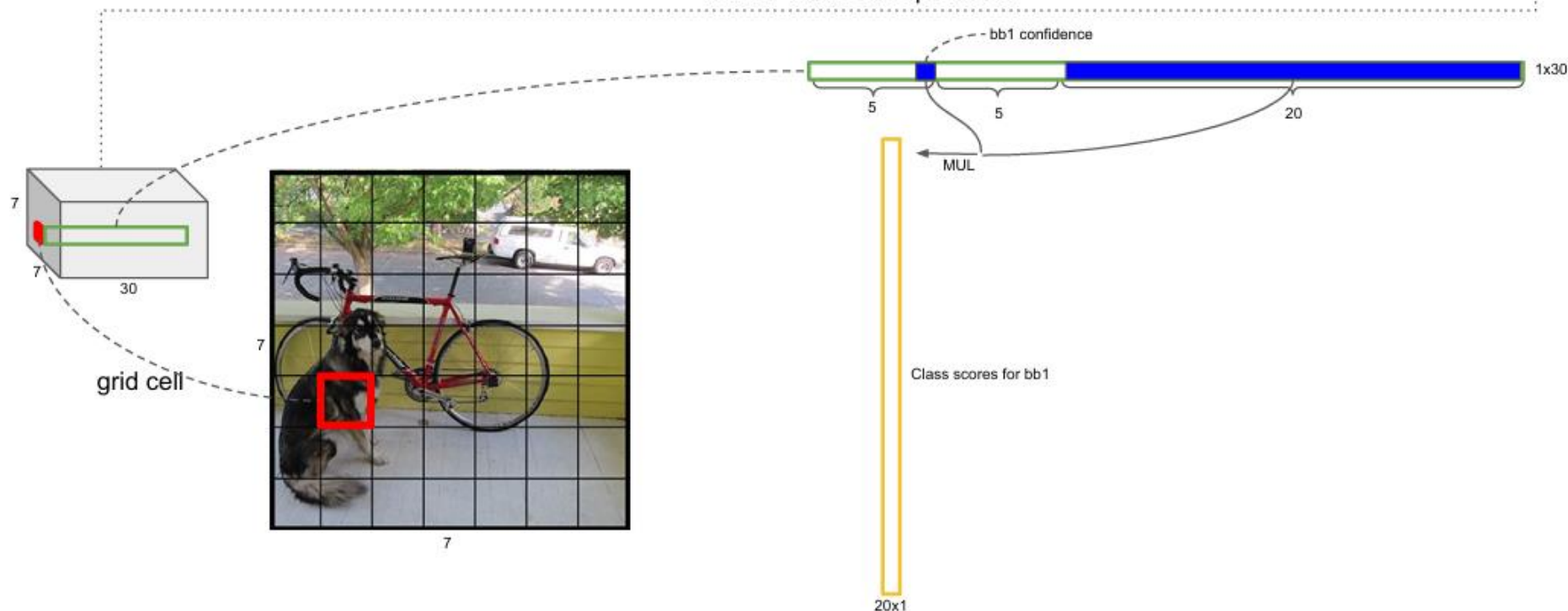


中國農業大學  
China Agricultural University

## Inference



## Tensor values interpretation

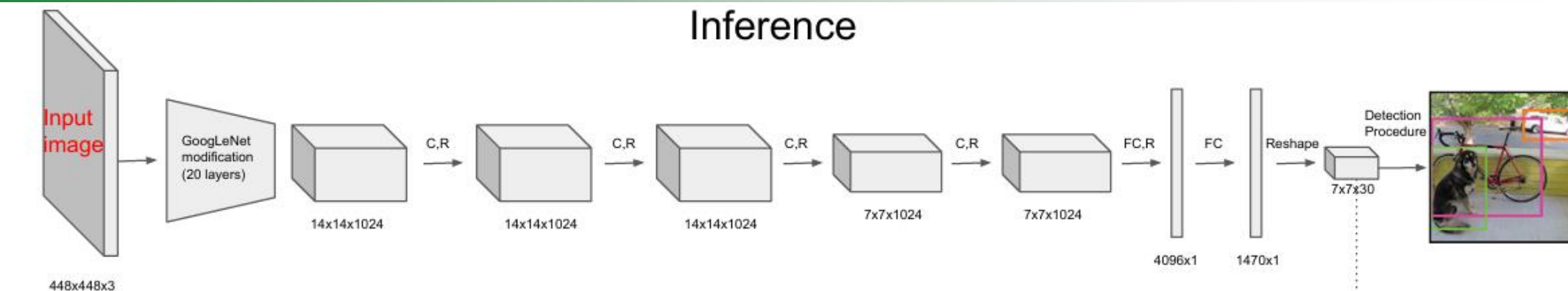


# YOLO多目标检测

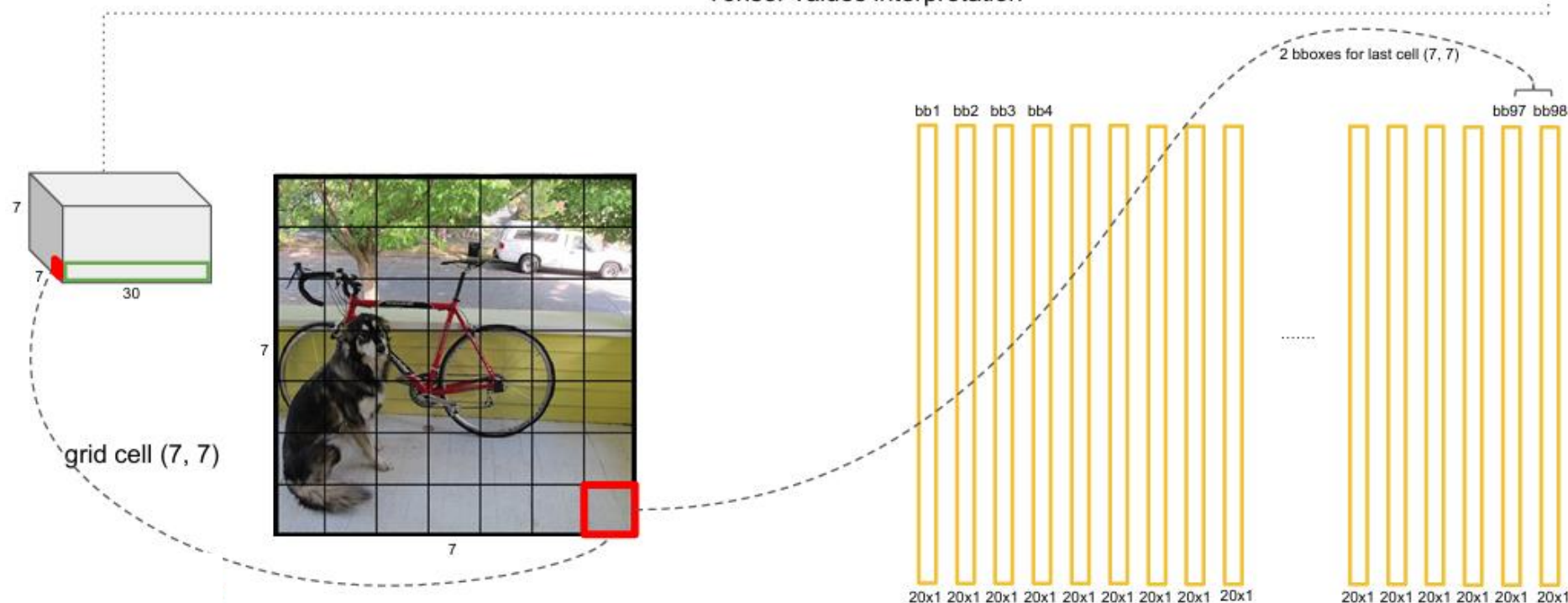


中國農業大學  
China Agricultural University

## Inference



## Tensor values interpretation

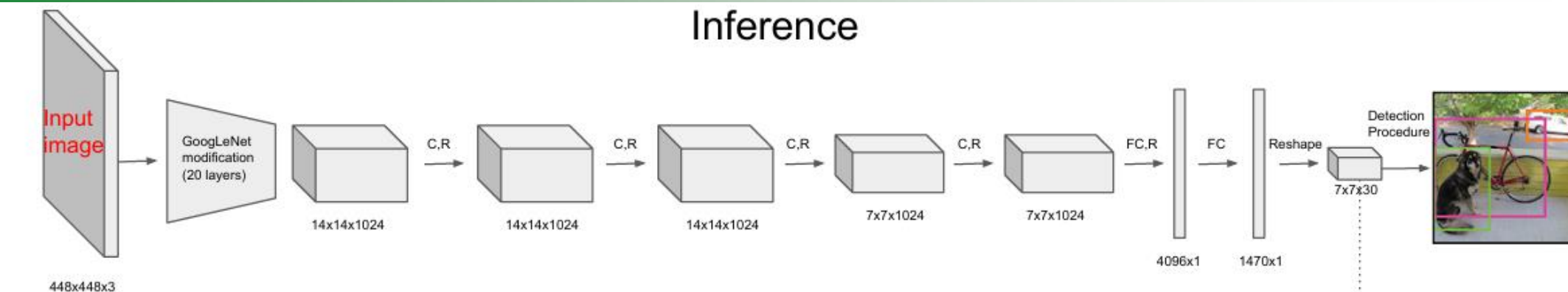


# YOLO多目标检测

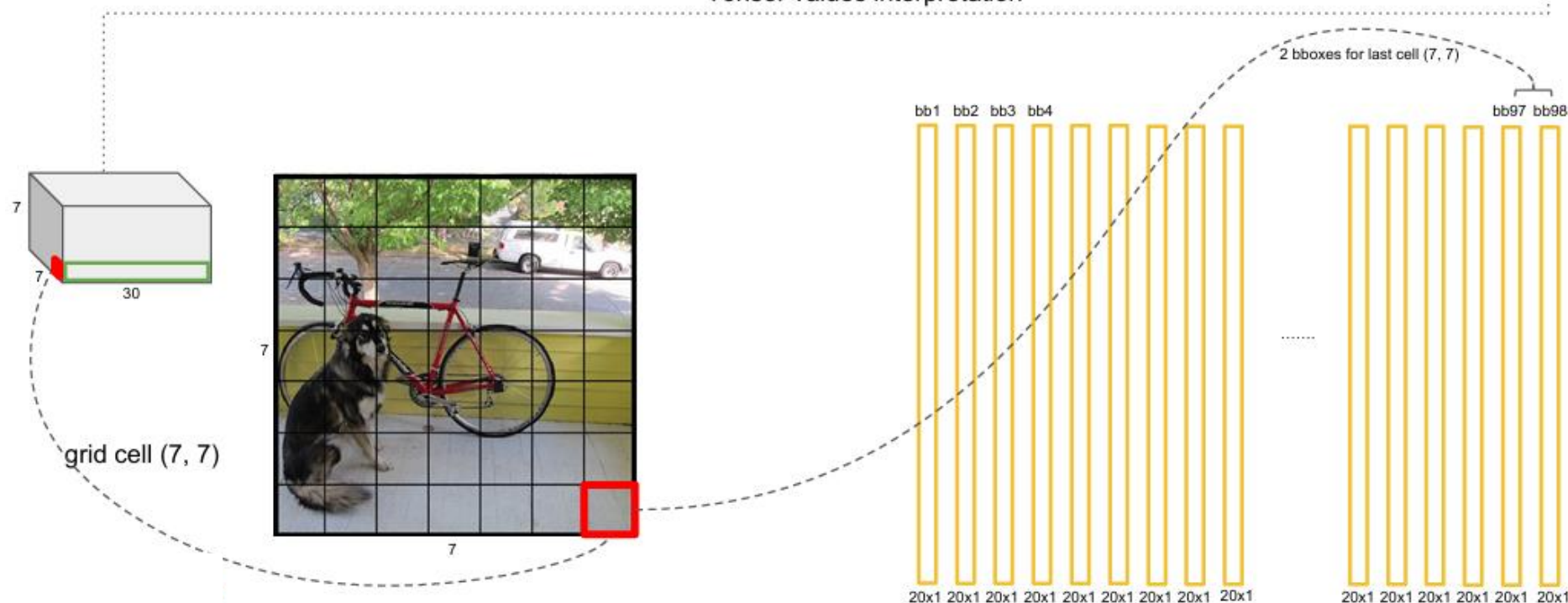


中國農業大學  
China Agricultural University

## Inference



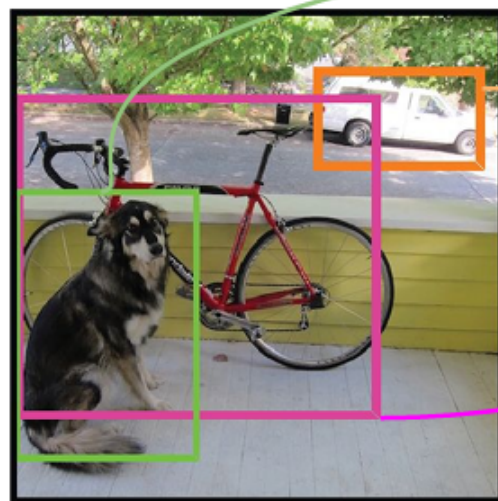
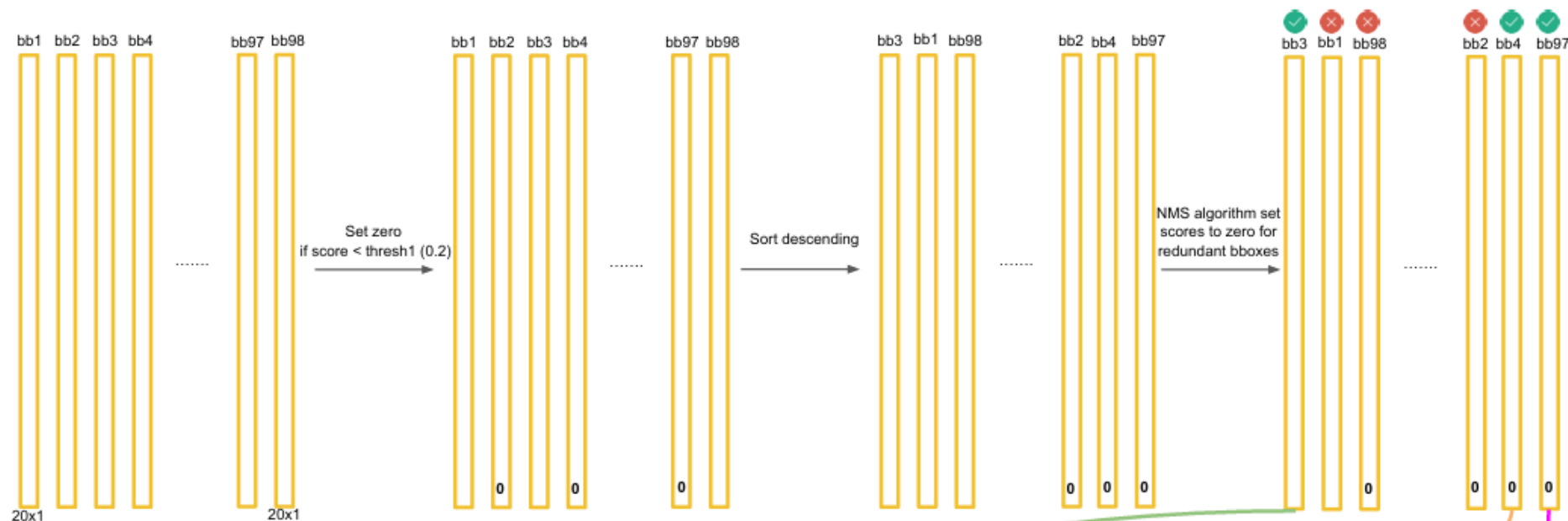
## Tensor values interpretation



# YOLO多目标检测



中國農業大學  
China Agricultural University







# 目录

## Contents

1. YOLO多目标检测

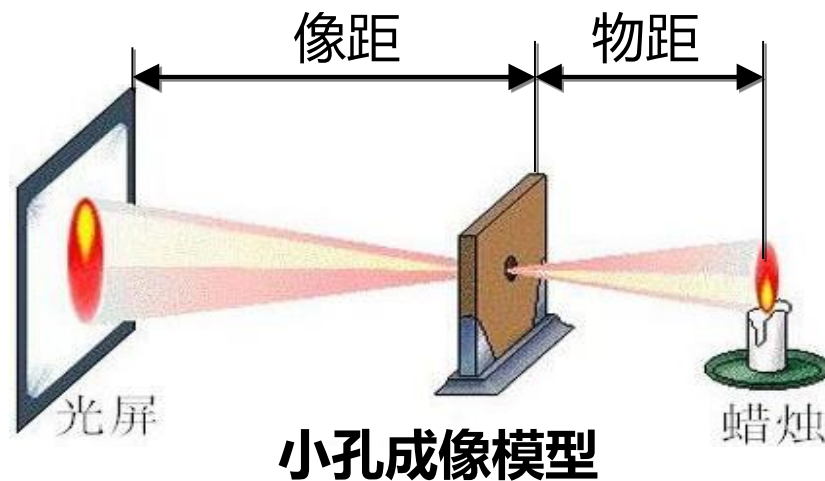
2. 视觉几何

3. 定位原理

4. 粒子滤波算法

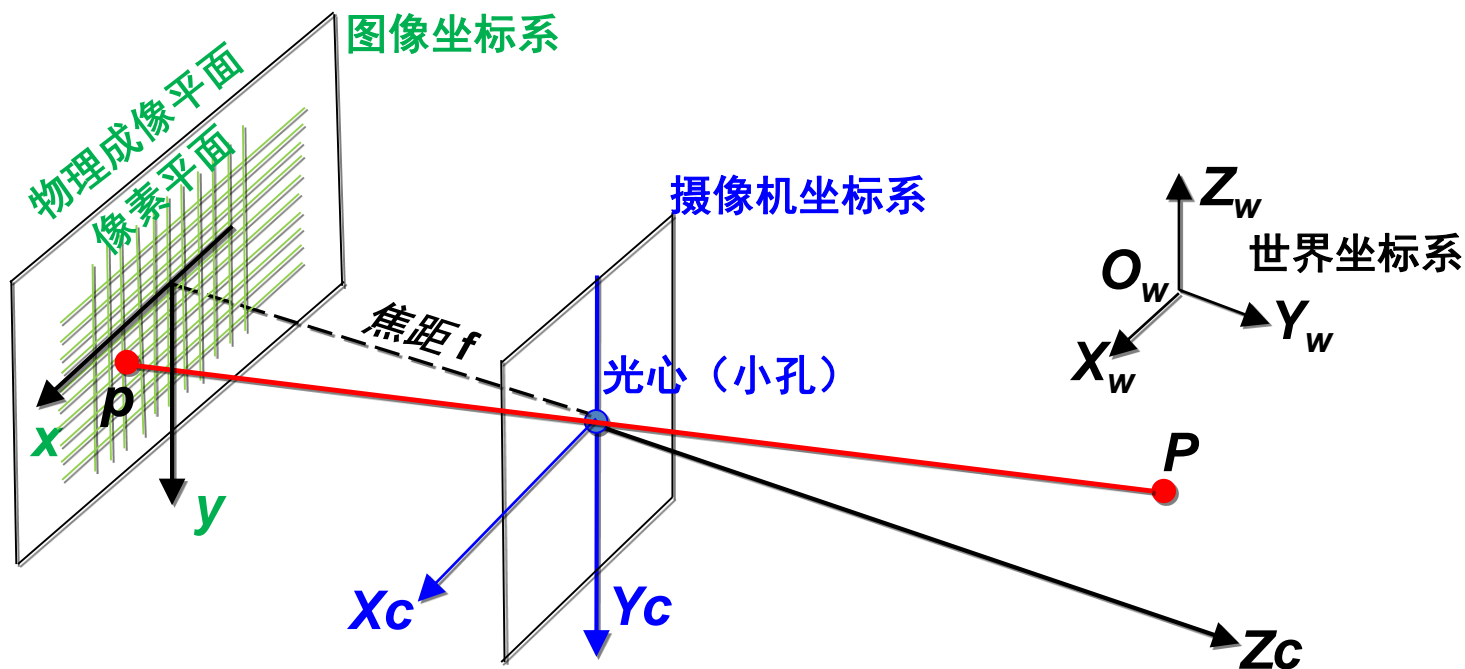
## ◆ 摄像机模型（摄像机成像模型）

- 计算机视觉的基本任务之一是利用所拍摄的图像来计算出三维空间中被测物体的几何信息，并由此重建和识别物体。
- **二维图像与场景物体的三维信息之间的几何关系是什么？**
- **摄像机模型：**将三维场景与二维图像坐标联系起来的模型。
- 摄像机模型的线性成像模型：**小孔成像模型。**



## ◆ 针孔模型—线性模型

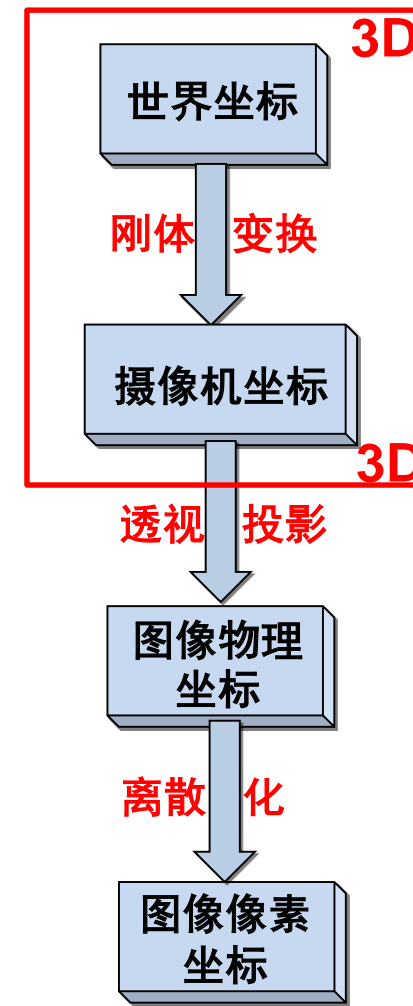
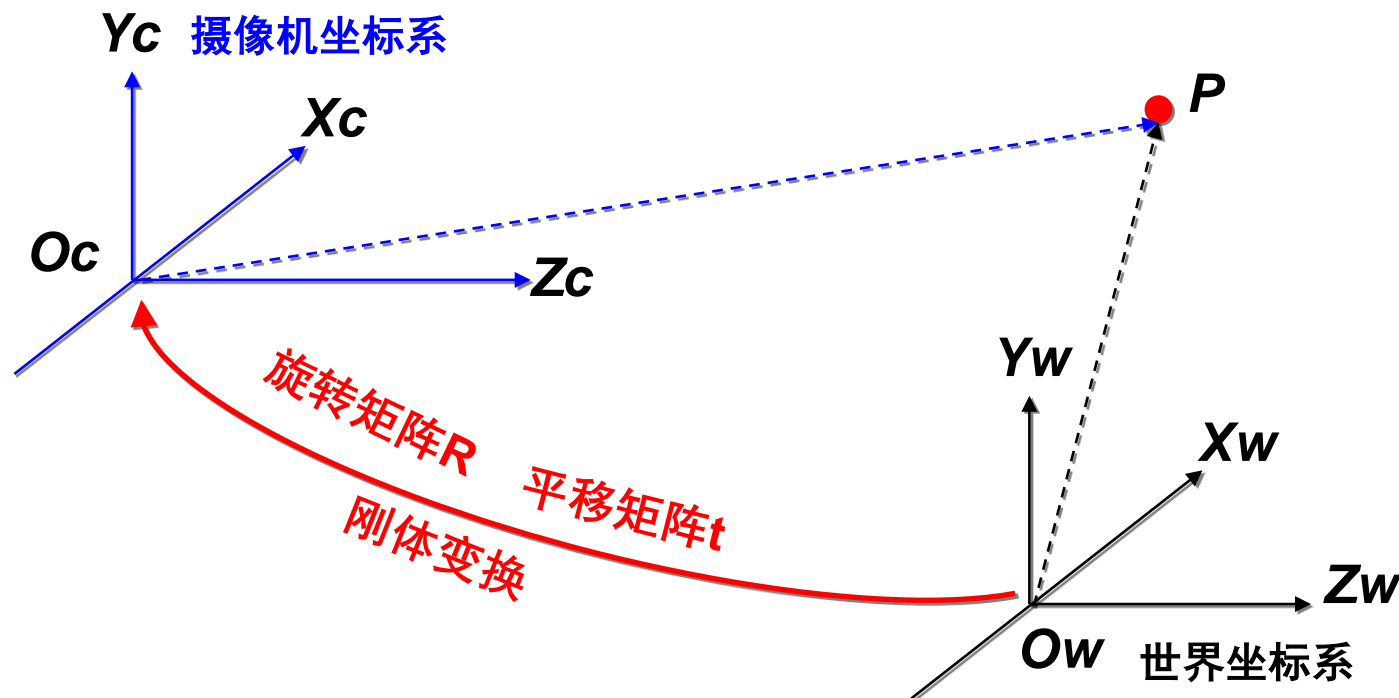
- 理想状态下的成像模型，并没有考虑相机镜头畸变。
- 包括四坐标系：世界坐标系、摄像机坐标系、图像物理坐标系、图像像素坐标系。



## ◆ 针孔模型—线性模型

(1) 世界坐标系与摄像机坐标系之间的**刚体转换**

- 世界坐标系：根据现实环境选择的三维坐标系。

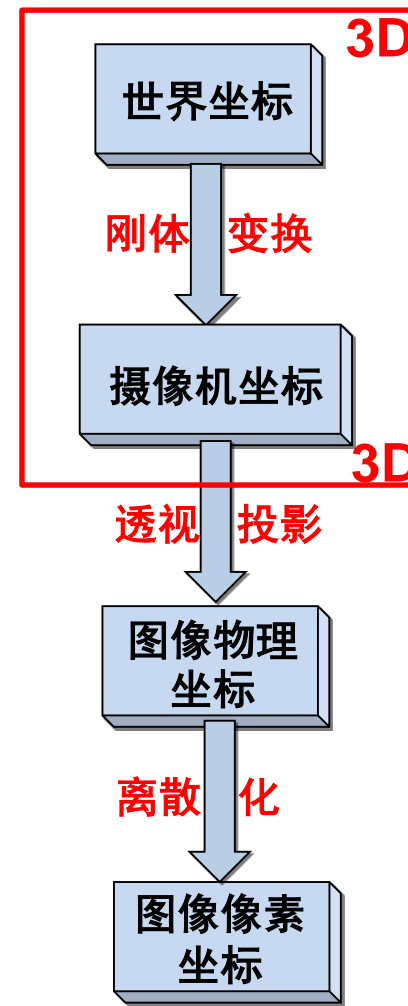


## ◆ 针孔模型—线性模型

### (1) 世界坐标系与摄像机坐标系之间的刚体转换

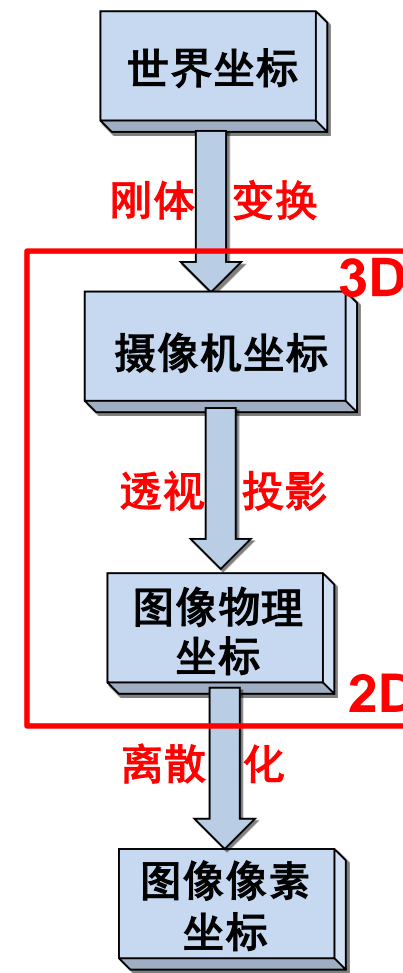
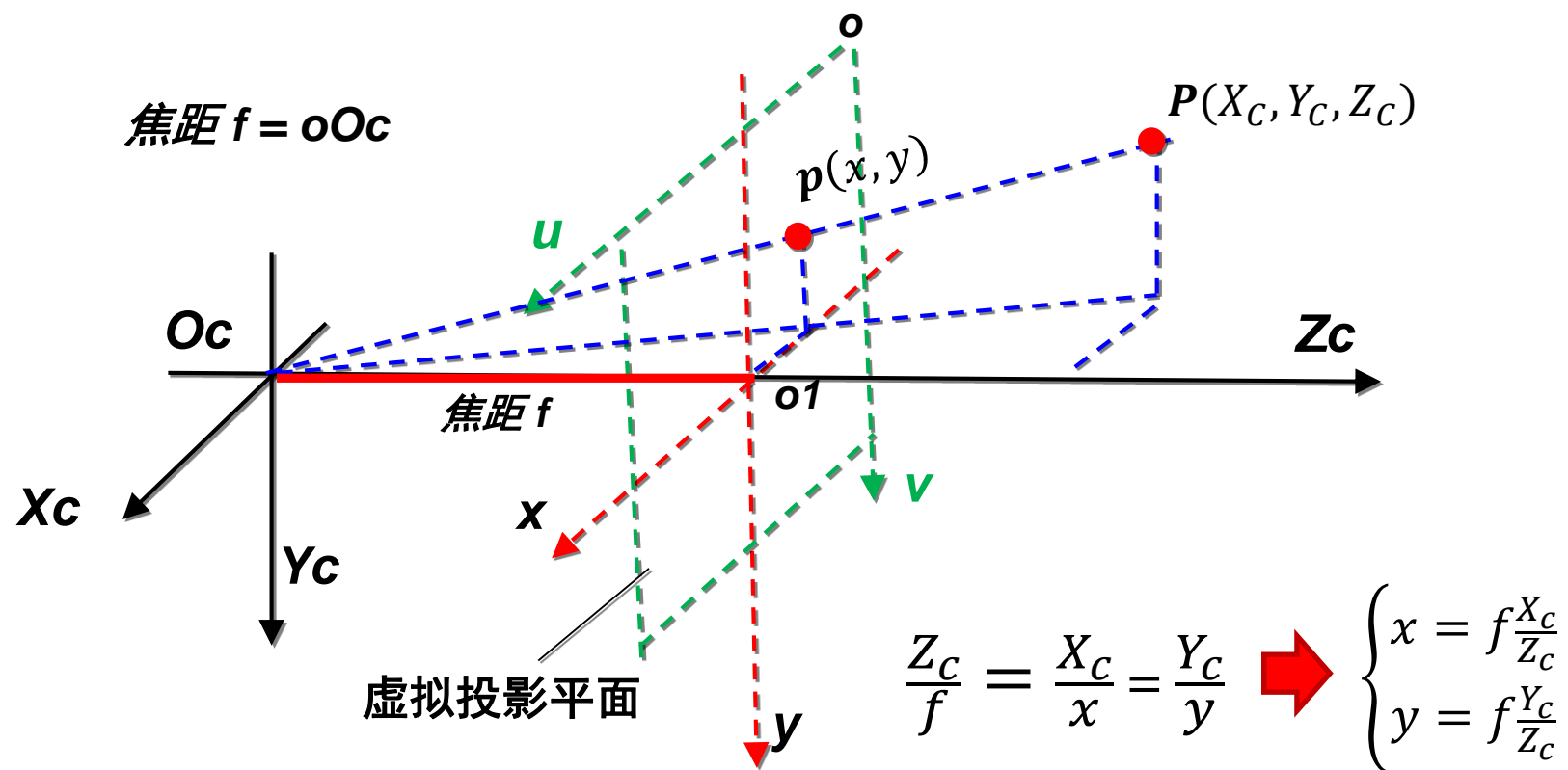
$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}$$

- R是3\*3的旋转矩阵
- T 是3\*1的平移矩阵
- R, T 称为摄像机的外部参数
- R, T 参数随着摄像机的移动而变化





## ◆ 针孔模型—线性模型

(2) 摄像机坐标系与图像坐标系之间的**透视投影变换**

## ◆ 针孔模型—线性模型

### (2) 摄像机坐标系与图像坐标系之间的**透视投影变换**

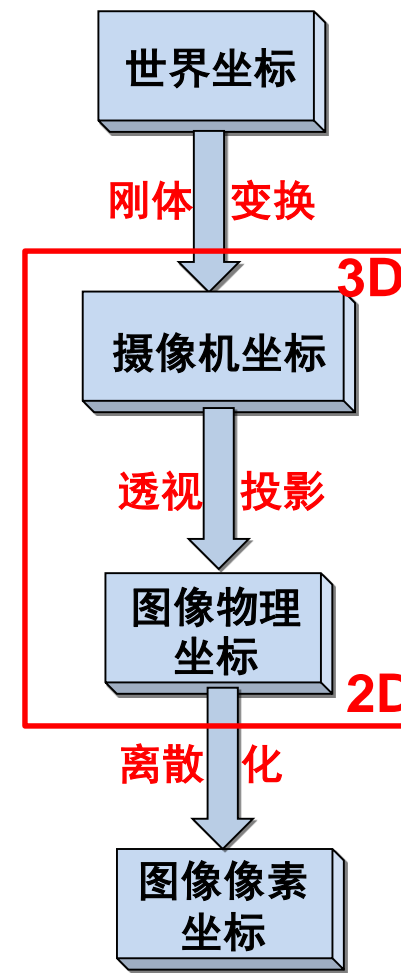
- 根据相似三角形原理

$$\frac{x}{f} = \frac{X_c}{Z_c} \quad \frac{y}{f} = \frac{Y_c}{Z_c}$$

其中  $f$  为焦距（像平面与摄像机坐标系原点的距离）

- 用齐次坐标系和矩阵表示上述关系：

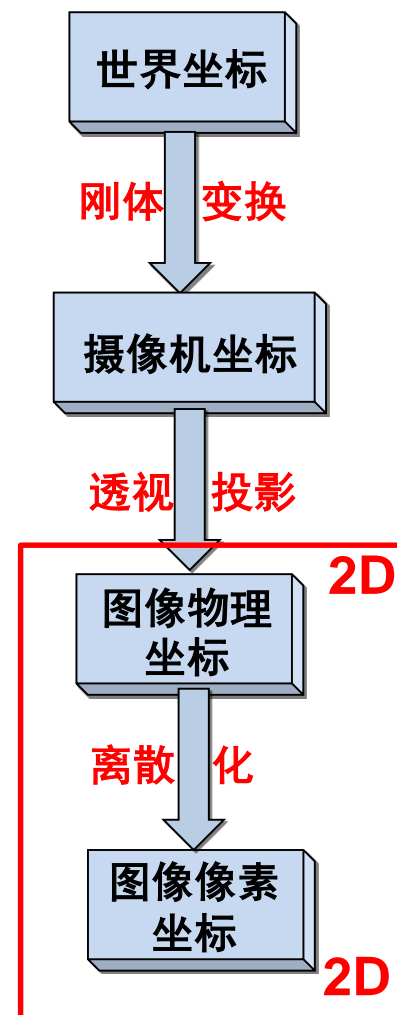
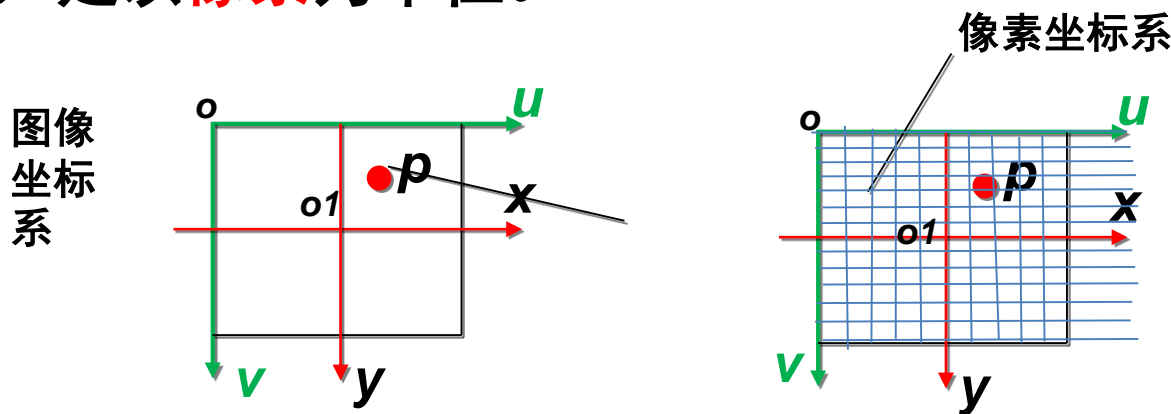
$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$



## ◆ 针孔模型—线性模型

(3) 像素坐标系是物理坐标系的**离散化表示**

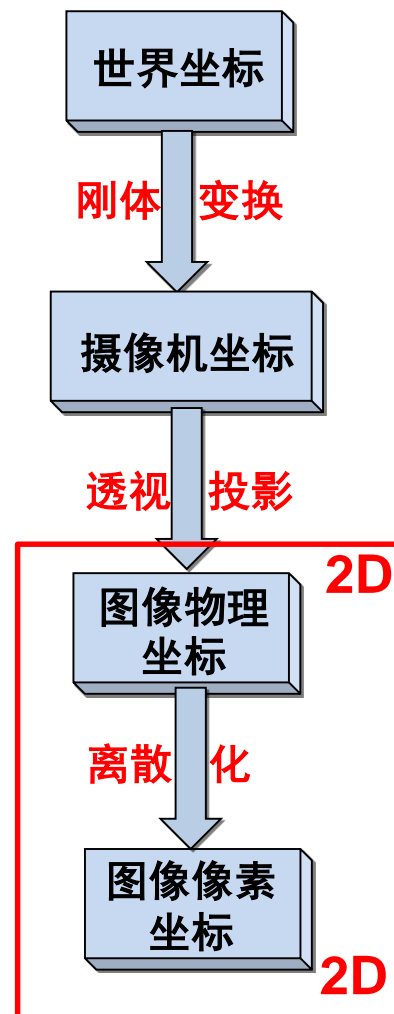
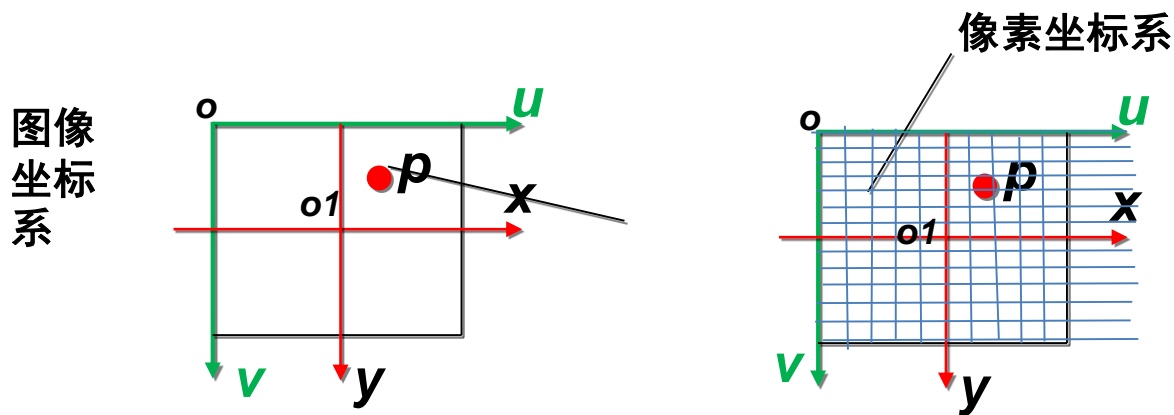
- **图像物理坐标系**：以 $o_1$ 为原点， $x$ 轴和  $y$ 轴分别与 $u$ ,  $v$ 轴平行。 $(x, y)$  表示以**毫米**为单位。
- **图像像素坐标系**：在图像上定义直角坐标系 $u\ v$ , 每一像素的坐标  $(u, v)$  分别是该像素在数组中的列数与行数。 $(u, v)$  是以**像素**为单位。



## ◆ 针孔模型—线性模型

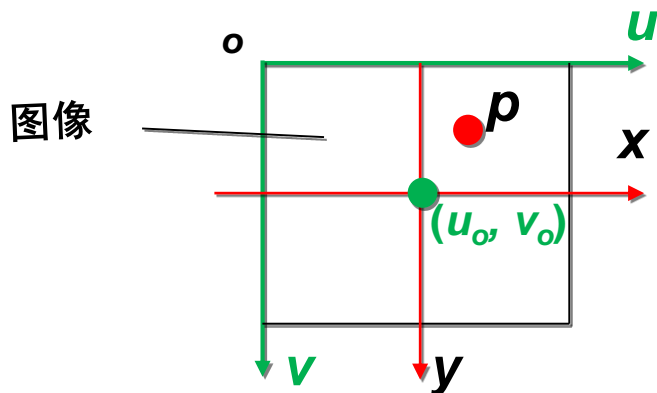
### (3) 像素坐标系是物理坐标系的离散化表示

- 举个例子：CCD的尺寸是8mm\*6mm，帧画面的分辨率设置为640\*480，那么毫米与像素点之间的转换关系就是80pixel/mm。设CCD传感器每个像素点的物理大小为 $dx*dy$ ，相应地、就有 $dx = dy = 1/80$ 。



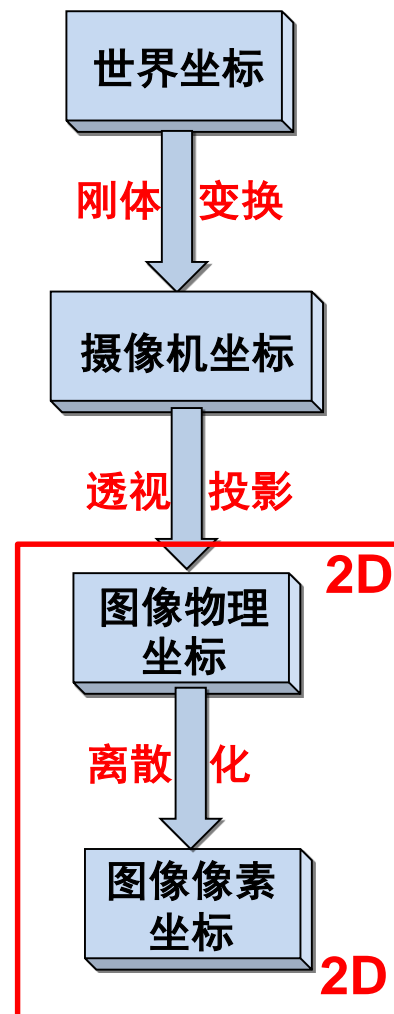
## ◆ 针孔模型—线性模型

(3) 像素坐标系是物理坐标系的**离散化表示**



$$\begin{cases} u = \frac{x}{d} + u_0 \\ v = \frac{y}{d} + v_0 \end{cases}$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d} & 0 & u_0 \\ 0 & \frac{1}{d} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

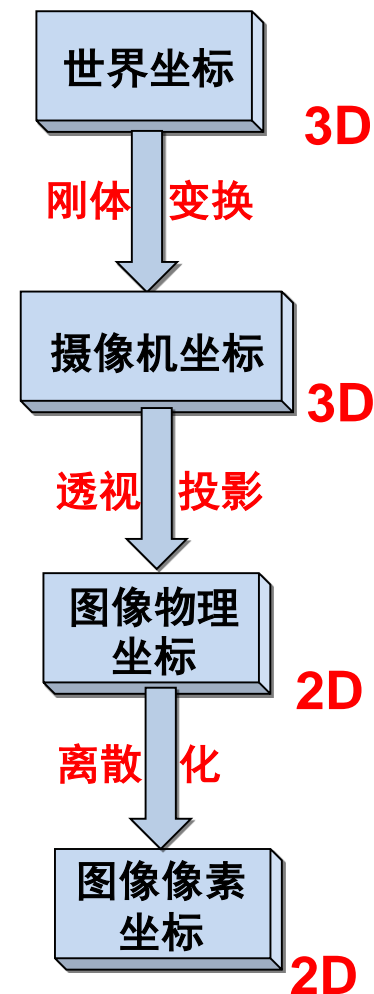




# ◆ 针孔模型—线性模型

## (4) 从世界坐标系到像素坐标的转换关系

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} R & T \\ O^T & 1 \end{bmatrix}}_{\text{相机与世界}} \underbrace{\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}}_{\text{投影关系}} \underbrace{\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix}}_{\text{像素与像平面}}$$



## (4) 从世界坐标系到像素坐标的转换关系

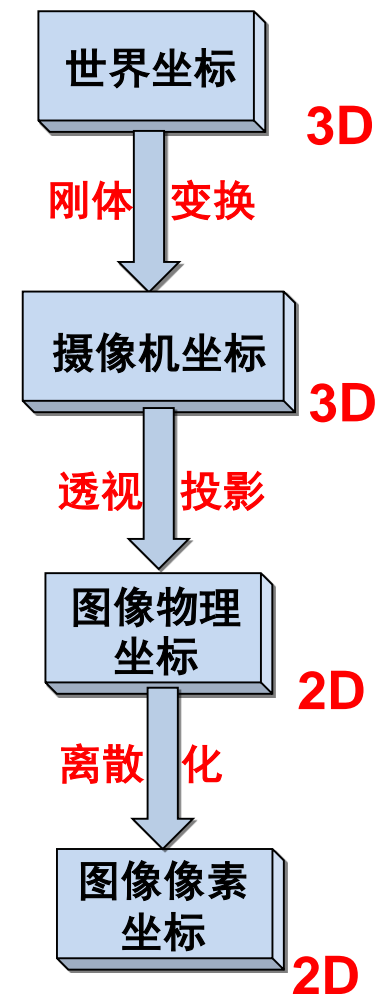
- 设内参数矩阵可以表示为：

$$M_1 = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

这里  $f_x = f/dx$ ,  $f_y = f/dy$

- 设外参矩阵可以表示为： $M_2 = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix}$

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = M \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad M = M_1 M_2$$



- 场景3维坐标与投影在2维图像像素坐标之间的几何关系

$$\begin{matrix} Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \\ \text{2D} \end{matrix} = \begin{matrix} \underbrace{M_{3 \times 4}}_{\text{内外参数矩阵}} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \\ \text{3D} \end{matrix}$$

- 内外参数需要通过实验与计算才能得到，这个求解参数的过程就称为**摄像机标定**。
- 摄像机参数标定是非常重要的环节，其标定精度**直接影响到摄像机工作所产生结果的准确性**。

- 场景3维坐标与投影在2维图像像素坐标之间的几何关系

$$\begin{matrix} Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \\ \text{2D} \end{matrix} = \begin{matrix} \underbrace{M_{3 \times 4}}_{\text{内外参数矩阵}} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \\ \text{3D} \end{matrix}$$

- 若已知投影矩阵  $M$  和空间点世界坐标，可以求得空间点的图像坐标。
  - 在线性模型中，一个物点在成像平面上对应唯一的像点。
- 但反过来：已知像点坐标  $(u, v)$  和投影矩阵  $M$ ，并不对应唯一的物点（因为  $Z_C$  未知）。

- **摄像机标定的目的**：为世界坐标系的三维物点和图像坐标系中的 二维像点之间建立一种映射关系。

- **摄像机标定**：求解摄像机内外参数的过程。

摄像机的内参数包括：**x方向尺度因子**  $f_x = f/dx$

**y方向尺度因子**  $f_y = f/dy$

**像主点在图像坐标系中坐标**  $(u_0, v_0)$

**考虑相机畸变时，畸变参数。**

$dx$  和  $dy$  表示： $x$ 方向和 $y$ 方向的一个像素分别占多少长度单位，即一个像素代表的实际物理值的大小。 $u_0, v_0$ 表示图像的中心像素坐标和图像原点像素坐标之间相差的横向和纵向像素数。

摄像机的外参数包括：**旋转矩阵3个参数**（3轴的旋转参数表示）

**平移矩阵3个参数**



# 目录

## Contents

1. YOLO多目标检测

2. 视觉几何

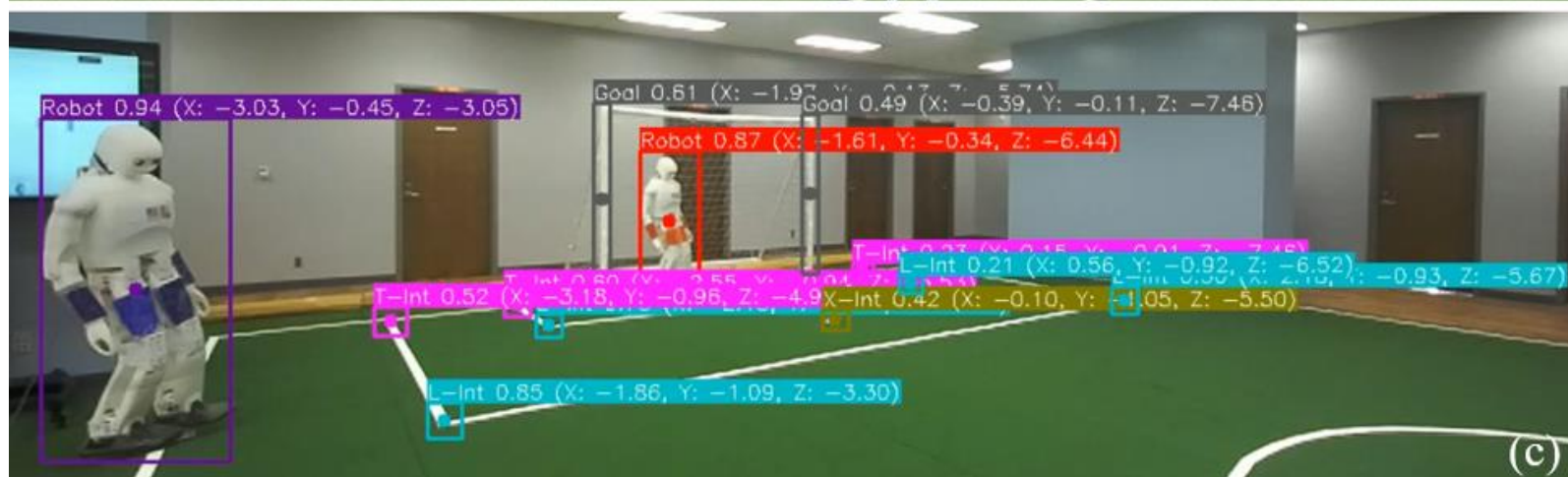
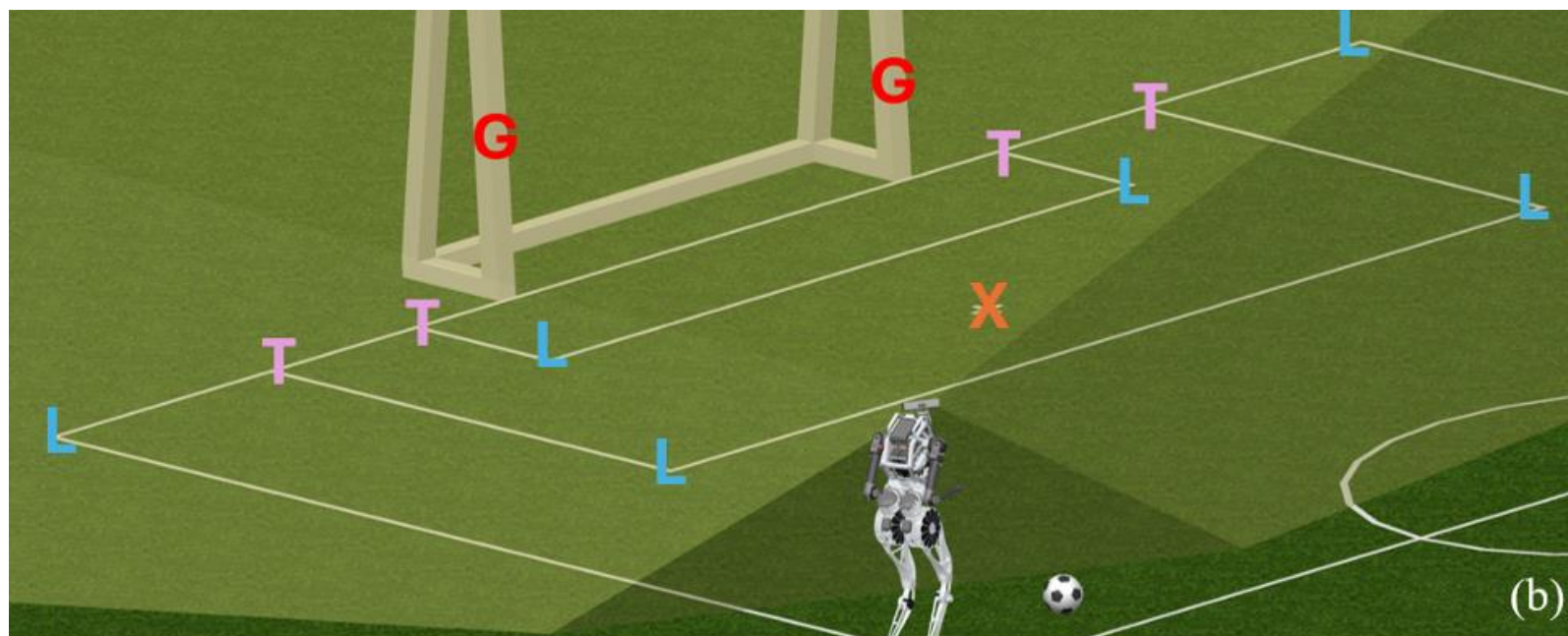
3. 定位原理

4. 粒子滤波算法



Technical drawing of a rectangular plate with dimensions and a hole. The overall dimensions are 16000 (width) and 11000 (height). The plate has a central circular hole with a radius of R1500. The left side features a vertical slot with a width of 2100 and a depth of 4000. The right side features a vertical slot with a width of 900 and a depth of 1400. The top edge has a width of 3000 and a height of 6000. The bottom edge has a width of 14000 and a height of 9000.

Technical drawing of a rectangular floor plan. The overall dimensions are 22000 (width) by 14000 (height). The plan is divided into three main sections by a vertical line. The left section contains a rectangular area with a width of 5000 and a height of 8000. Inside this area is a smaller rectangle with a width of 2000 and a height of 5000. The right section contains a rectangular area with a width of 3500 and a height of 1500. A central circle with a radius of R2000 is located between the two side sections. The drawing includes various dimension lines and labels in green, indicating specific measurements and radii.

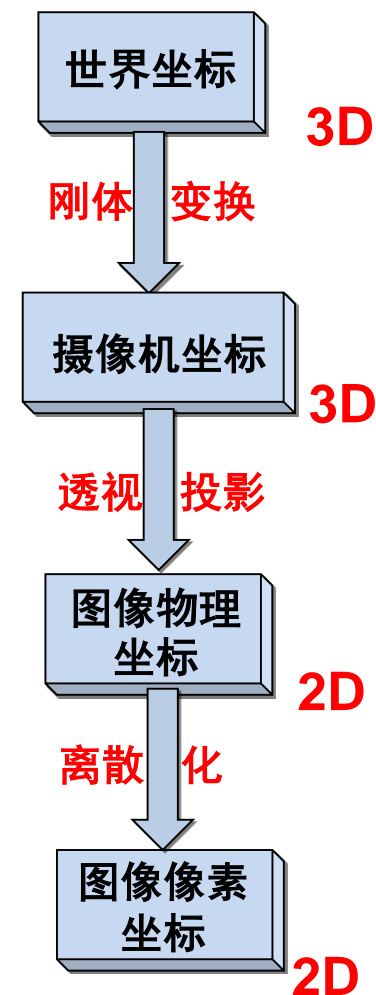
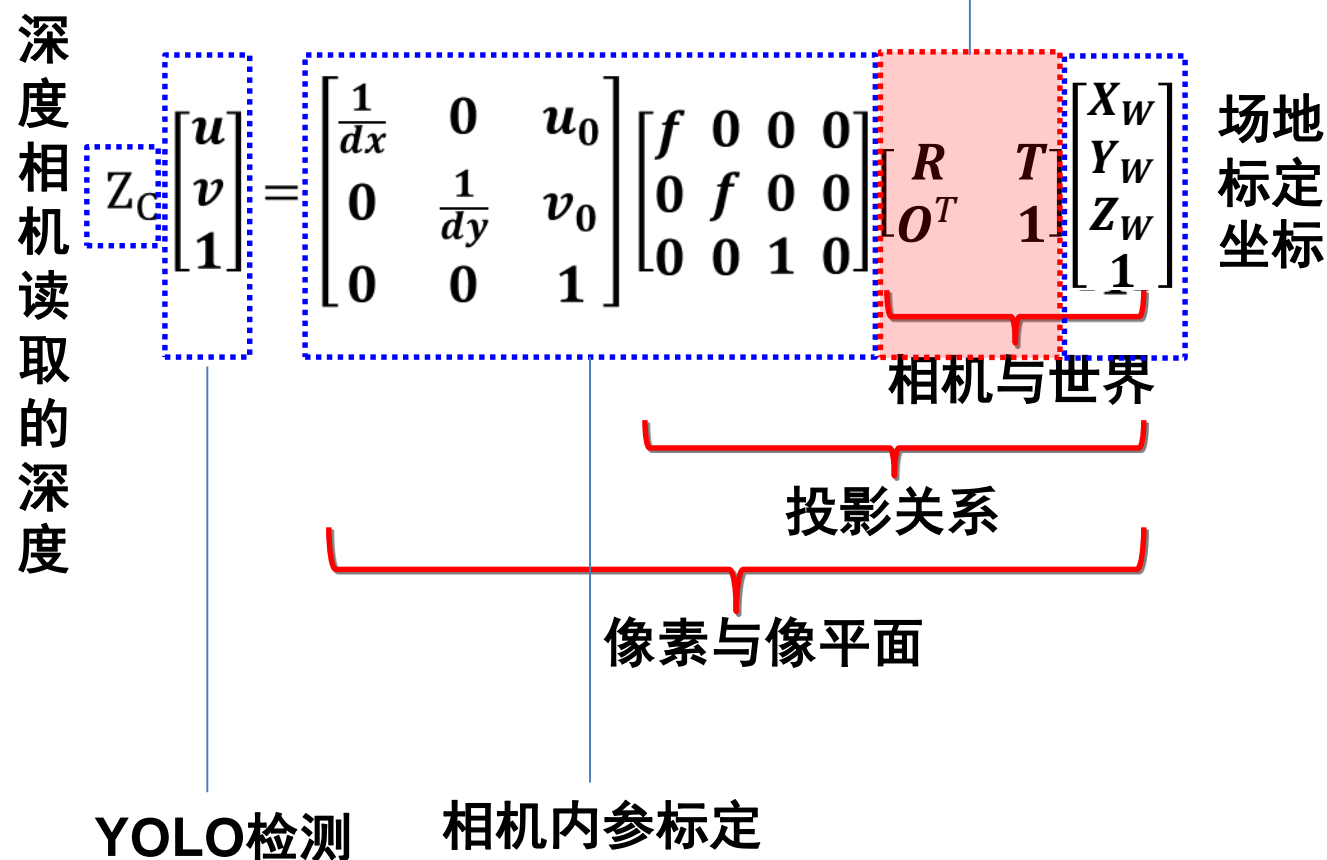


## YOLO识别

- 球
- G门柱
- 机器人
- L交叉点
- T交叉点
- X罚球点
- 开球点

## ◆ 从世界坐标系到像素坐标的转换关系

未知，待求解（通过  
3个特征点可求解）





# 目录

## Contents

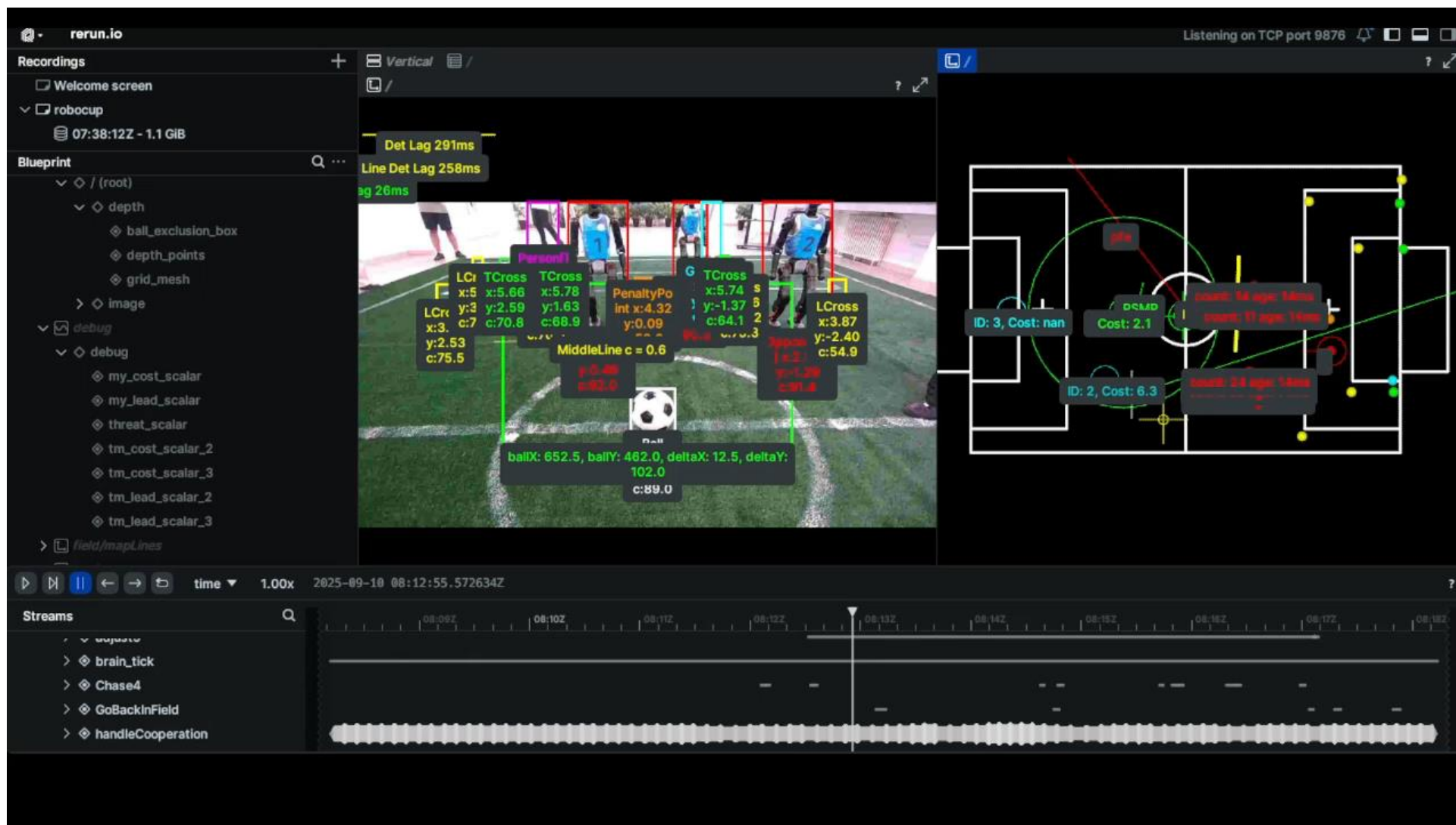
1. YOLO多目标检测

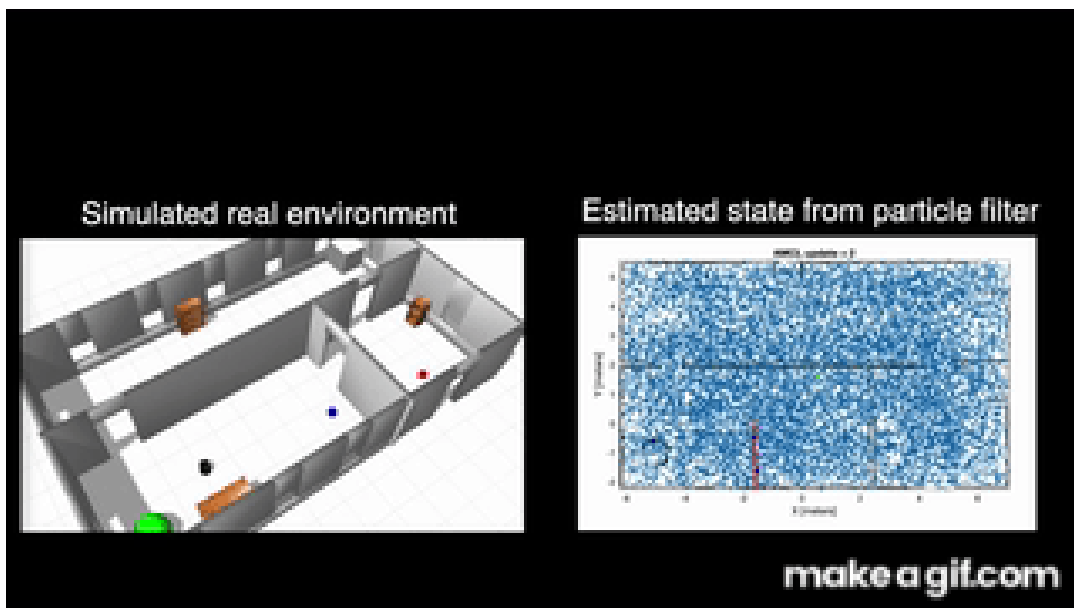
2. 视觉几何

3. 定位原理

4. 粒子滤波算法







视觉观测输入

$$Z = \{z_1, z_2, \dots, z_n\}$$

其中每个观测  $z_i$  包含:

$$z_i = (u_i, v_i, d_i, \text{type}_i, \text{confidence}_i)$$

$u_i, v_i$ : 像素坐标

$d_i$ : 深度值 (米)

$\text{type}_i$ : 标记类型 (X, P, T, L)

$\text{confidence}_i$ : 检测置信度

相机内参矩阵

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$T_{\text{cam} \rightarrow \text{head}}$ : 相机到头部的固定变换 ( $4 \times 4$  齐次矩阵)

$T_{\text{head} \rightarrow \text{robot}}$ : 头部到机器人的动态变换 (实时更新)

$$T_{\text{cam} \rightarrow \text{robot}} = T_{\text{head} \rightarrow \text{robot}} \times T_{\text{cam} \rightarrow \text{head}}$$

场地先验信息

$$M = \{m_1, m_2, \dots, m_k\}$$

每个场地标记点  $m_j$  包含:

$$m_j = (\text{type}_j, x_{\text{field},j}, y_{\text{field},j})$$



## 阶段1：观测预处理

像素坐标 → 相机坐标

对于每个观测  $z_i$ :

$$\begin{aligned}x_{\text{cam},i} &= \frac{(u_i - c_x) \cdot d_i}{f_x} \\y_{\text{cam},i} &= \frac{(v_i - c_y) \cdot d_i}{f_y} \\z_{\text{cam},i} &= d_i\end{aligned}$$

矩阵形式:

$$\begin{bmatrix} x_{\text{cam},i} \\ y_{\text{cam},i} \\ z_{\text{cam},i} \end{bmatrix} = \begin{bmatrix} \frac{d_i}{f_x} & 0 & -\frac{c_x \cdot d_i}{f_x} \\ 0 & \frac{d_i}{f_y} & -\frac{c_y \cdot d_i}{f_y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix}$$

相机坐标 → 机器人坐标

$$\mathbf{p}_{\text{robot},i} = T_{\text{cam} \rightarrow \text{robot}} \cdot \mathbf{p}_{\text{cam},i}$$

其中  $\mathbf{p}_{\text{cam},i} = [x_{\text{cam},i}, y_{\text{cam},i}, z_{\text{cam},i}, 1]^T$

## 阶段2：粒子滤波定位

状态空间定义

状态向量:  $\mathbf{x} = [x_{\text{field}}, y_{\text{field}}, \theta]^T$

粒子集:  $\{\mathbf{x}_i^{(k)}\}_{i=1}^N, \quad \mathbf{x}_i \in \mathbb{R}^6$

粒子结构:  $[x, y, \theta, \text{residual}, \text{probability}, \text{cumulative\_prob}]$

权重计算（观测似然）

对于每个粒子  $\mathbf{x}_i$ :

$$\text{residual}_i = \sum_{j=1}^n \left[ \frac{d(\text{transform}(\mathbf{p}_{\text{robot},j}, \mathbf{x}_i), \text{nearest\_marker}_j)}{\|\mathbf{p}_{\text{robot},j}\|} \right] \times w$$

## 阶段3: 收敛判断

收敛条件

$$\begin{cases} \max(x) - \min(x) < \delta_x \\ \max(y) - \min(y) < \delta_y \\ \max(\theta) - \min(\theta) < \delta_\theta \end{cases}$$

残差检验

if  $\frac{\text{residual}(\hat{\mathbf{x}})}{n} > \text{tolerance}$  then 定位失败 else 定位成功

## 输出 (OUTPUT)

成功定位输出

定位结果:  $\hat{\mathbf{x}} = [\hat{x}_{\text{field}}, \hat{y}_{\text{field}}, \hat{\theta}]^T$

$\hat{\theta} \in [-\pi, \pi]$  (归一化角度)

残差:  $r = \frac{\text{residual}(\hat{\mathbf{x}})}{n}$

收敛代码: 0 (成功)

失败定位输出

失败代码:  $\begin{cases} 1: \text{粒子生成失败} \\ 2: \text{收敛但残差过大} \\ 3: \text{达到最大迭代未收敛} \\ 4: \text{标记点数量不足} \\ 5: \text{所有概率接近零} \end{cases}$



中國農業大學  
China Agricultural University

谢谢

---

