

# 基于单应性估计的视觉行为决策研究

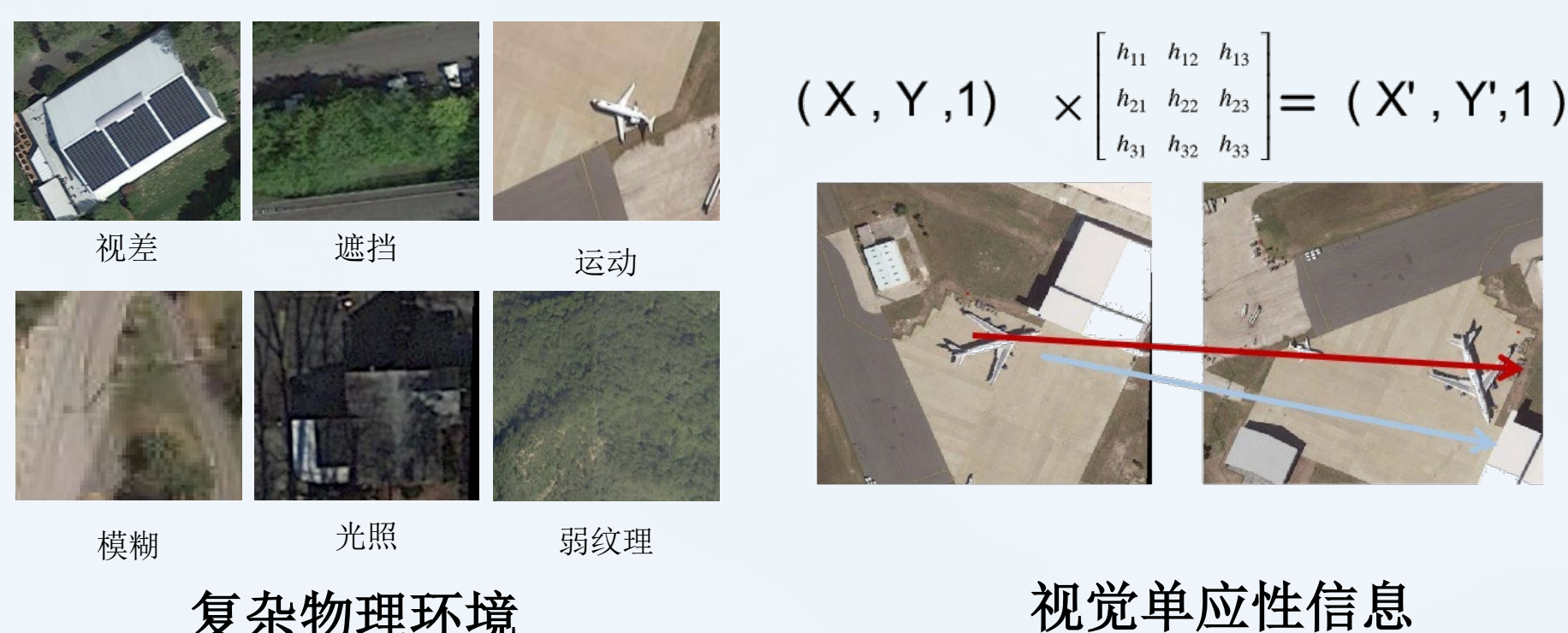
林翊钧\*, 吴凤鸽, 赵军锁

\* Tel: 15080018958, E-mail: yijun2016@iscas.ac.cn

## 引言

理想的机器视觉系统应在复杂环境下也具备接近人类的识别能力, 但现有先进神经网络模型普遍未考虑真实物理环境的特性(如成像系统误差与天气变化), 导致其性能会随着图像退化而严重下降。对此, 主流解决思路是引入计算成像算法来优化图像质量。但追求提升信噪比等传统指标分数, 虽能产生人眼满意的结果, 却不一定能改善甚至保持机器识别性能。

有必要研究以改善特定目标信息为导向的像质提升方法。本文考虑以单应性估计作为开展相关研究的基线任务。在遥感等近平面观测场景中, 两幅场景交叠图像间的几何关系可以用单应性矩阵描述, 它是变化检测、运动跟踪等大量高层次视觉识别任务所依赖的基础视觉信息。

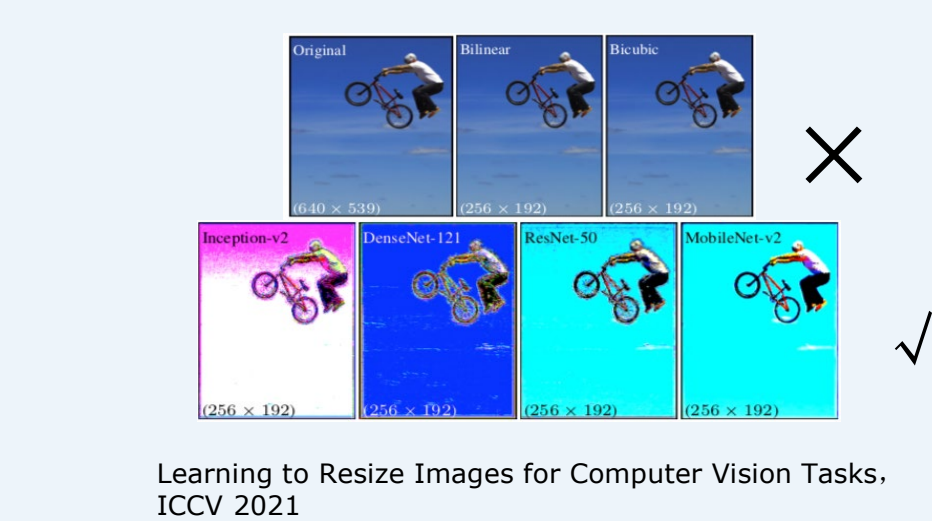


## 相关工作

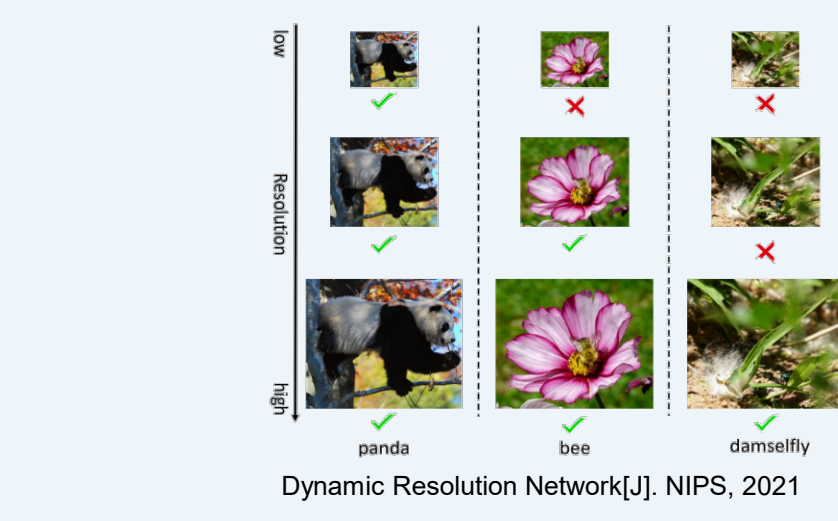
### 机器识别模型的特性研究

大量研究表明, 人眼视觉和机器视觉存在感知差异性。且即使面向同一视觉识别任务, 模型间也会存在各异的数据偏好。

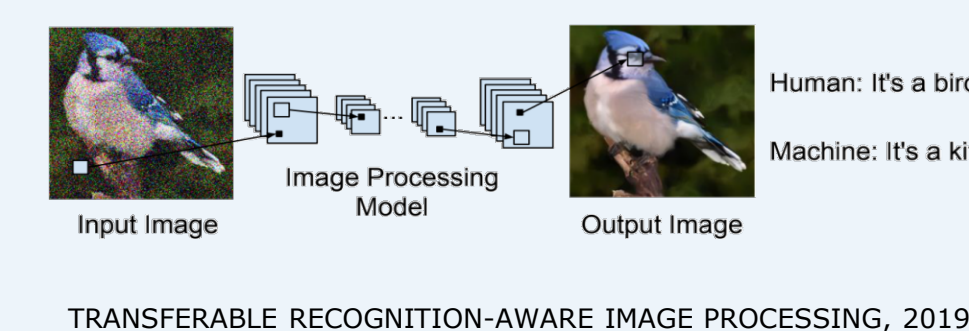
#### 降采样



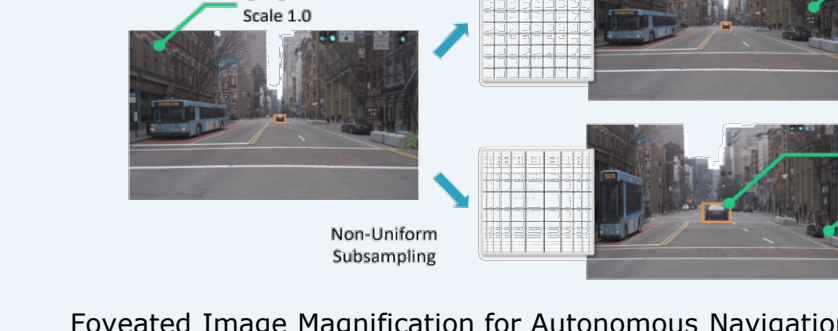
#### 超分辨率



#### 去噪



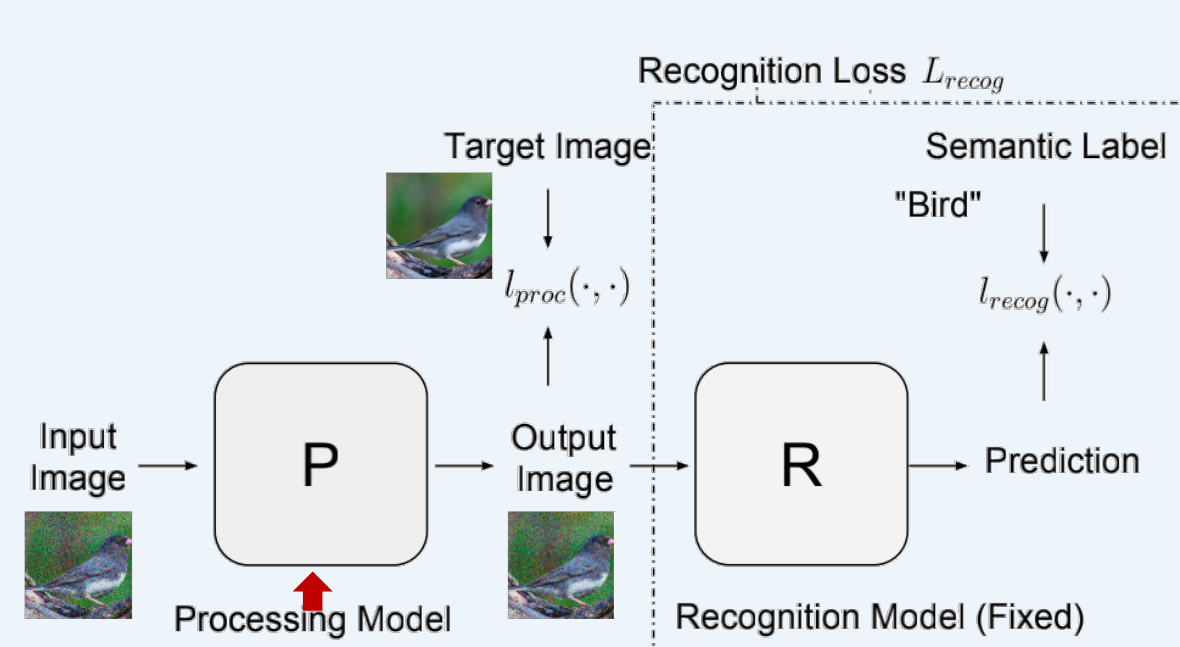
#### 缩放



### 基于视觉识别的像质提升研究

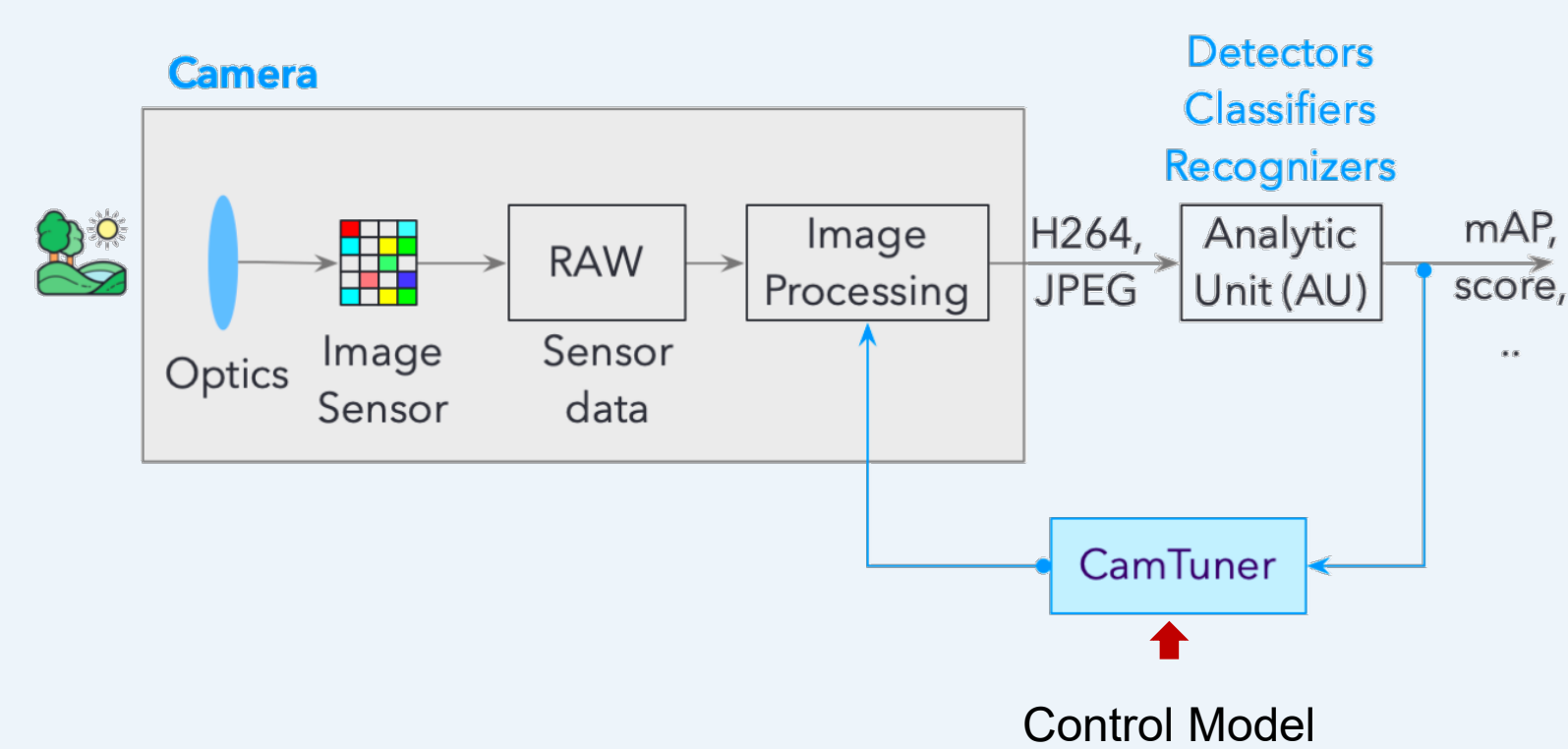
#### (1) 深度学习图像处理框架

基于目标识别、检测、分割等任务特点, 设计去噪、去雾、去模糊、暗光增强、超分等算法;



#### (2) 强化学习系统控制框架

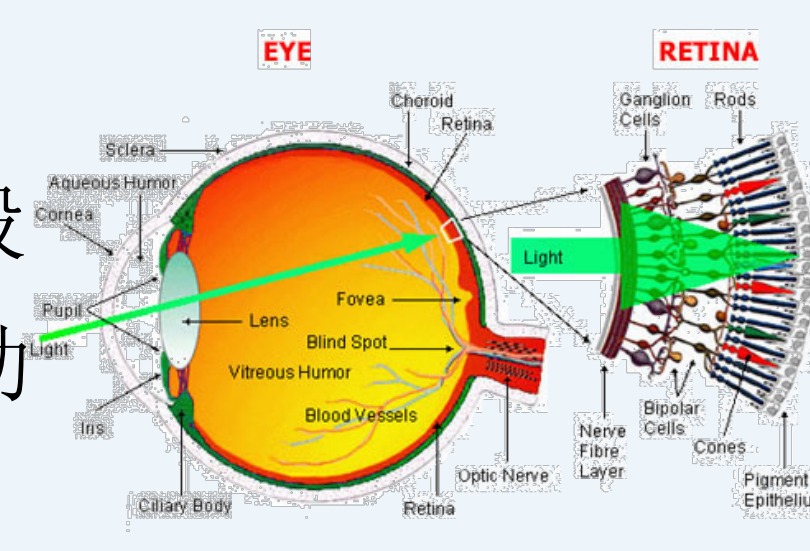
基于目标跟踪、凝视、避障等任务特点, 设计软件控制、相机控制、云台控制、姿态控制等算法;



## 研究思路

从模仿人类视觉行为的角度看, 人类感知具有很强的主动和灵活性, 能在特定意识需求的驱动下, 控制采取合理的视觉调节行为。以改善从复杂光场环境中获取的视觉信息。

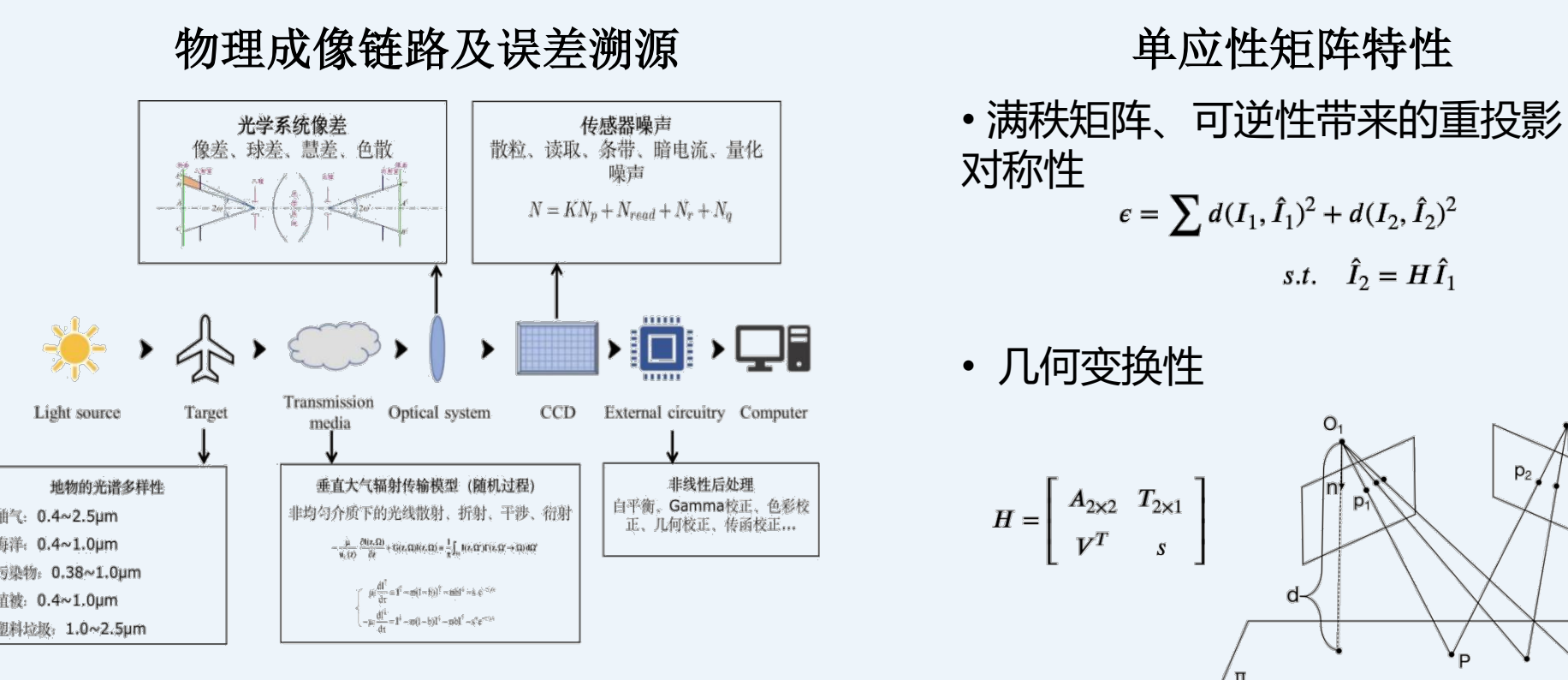
基于强化学习控制决策的多阶段渐进式图像改善, 是更贴合人类主动视觉行为的处理框架。



## 研究内容

### 视觉行为基础理论

分析光场成像全链路误差及目标视觉信息先验, 建立马尔可夫序列决策模型。



### 像质度量研究

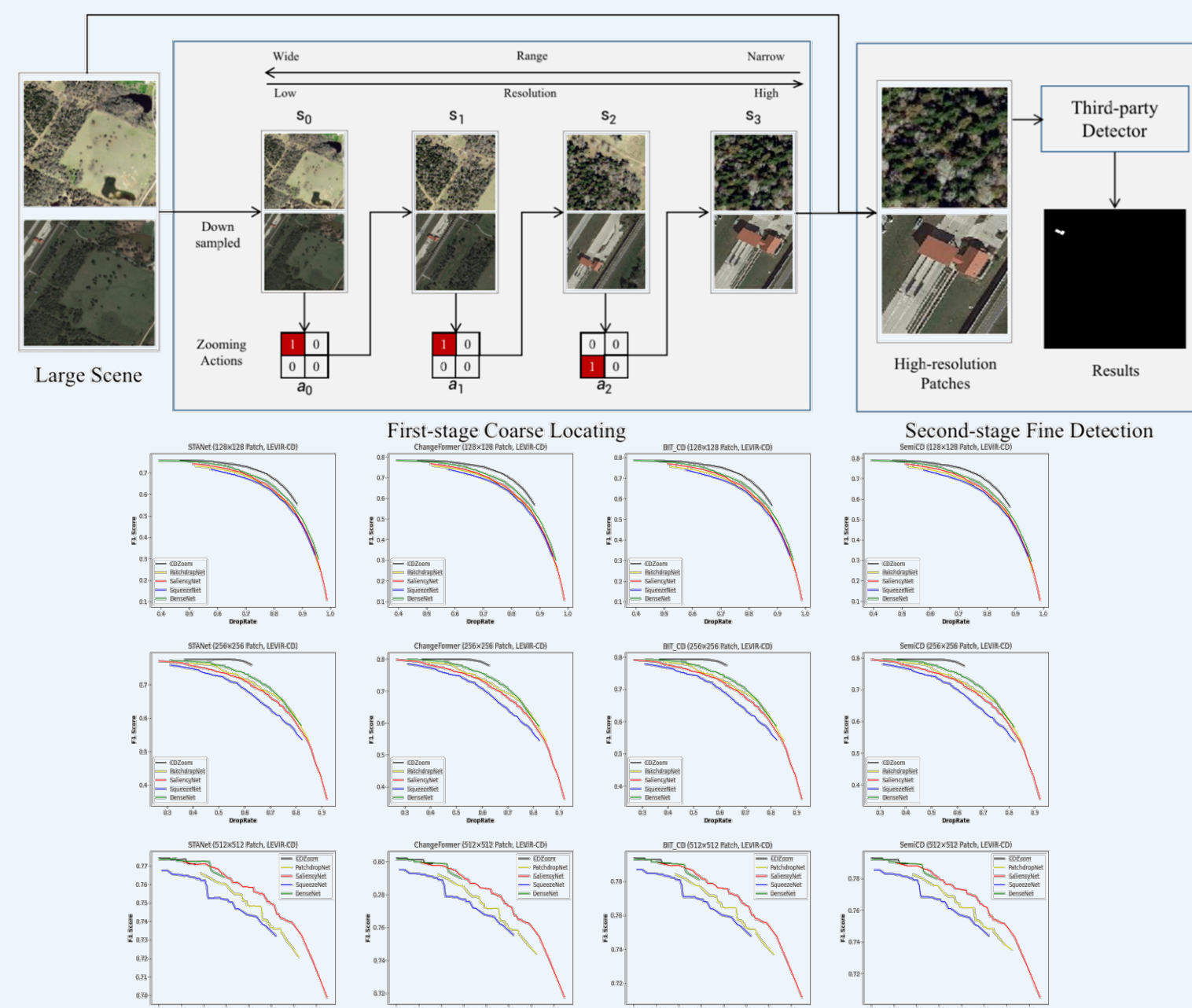
针对机器识别模型的视觉信息偏好差异问题, 构建度量网络, 以将图像投影到目标嵌入空间。随后利用余弦距离衡量图像在面向特定下游模型时的质量, 给出调整策略。

Exposure Method	Target Homography Estimator $\mathcal{H}(s)$					
	HENet	UHENet	SIFT	SURF	ORB	LK
Original Images	9.38	9.58	15.80	17.17	15.89	23.30
Ours(T=3)	8.59	8.94	13.27	14.82	12.62	21.67
Ours(T=5)	8.30	8.54	11.98	13.74	11.00	20.51
Ours(T=10)	7.86	8.11	9.98	12.17	8.39	19.24
Ours(T=20)	<b>7.43</b>	<b>8.07</b>	<b>8.78</b>	<b>11.54</b>	<b>7.04</b>	<b>18.55</b>
HDRNet	12.3	10.65	9.55	16.75	11.75	22.92
RetinexNet	11.48	12.00	15.73	21.71	15.57	23.93
3DLookUp	17.42	19.58	12.86	16.08	13.54	23.86
DeepUPE	16.54	20.34	12.79	16.57	13.63	24.30
DeepLPF	18.04	20.41	12.98	16.40	13.64	24.02

Yijun Lin, Fengge Wu and Junsuo Zhao. Reinforcement Learning-based Image Exposure Reconstruction for Homography Estimation. [J] The International Journal of Applied Intelligence.

### 显著性研究

针对目标视觉信息在大视场下稀疏分布的问题, 通过模仿人类的序列视觉注意行为来收缩状态空间。并提出基于任务分解、优势初始化和知识蒸馏的课程式强化学习框架, 解决高维状态空间带来的稀疏奖励挑战。



Yijun Lin, Fengge Wu and Junsuo Zhao. CDZoom: A Human-like Sequential Zoom Agent for Efficient Change Detection in Large Scenes. [J] The International Journal of Neural Computing and Applications.

## 展望

弥合计算成像与视觉识别任务间的隔阂, 开展全链路一体优化设计, 是提高复杂环境下机器认知能力的关键。其中, 通过仿生人类等智慧生物的主动视觉机理, 建立机器视觉行为决策模型, 是一条值得探索的研究路线。