



# 图像分类任务 (1)

叶山 中国地质大学 (北京)

yes@cugb.edu.cn

# 图像分类任务的应用场景

# 图像分类任务

图像分类任务是计算机视觉研究领域中的**基础性问题**（核心任务），其旨在通过分析图像信息中所蕴含的各种复杂**特征**，实现对不同类别图像的高精度**自动辨识与分类**，进而为图像理解、信息检索、智能决策等高级视觉任务提供支持。

# 图像分类任务

图像分类：在已知的类别标签集合中，为给定的输入图片选定一个类别标签。

类别标签：{狗、猫、牛、轮船、飞机.....}



狗



猫



牛



轮船



飞机

# 图像分类的应用场景



## 电气石 (tourmaline) 宝石称碧玺

### 基本资料

类别	环硅酸盐矿物
化学式	(Ca,K,Na, $\square$ ) (Al,Fe,Li,Mg,Mn) <sub>3</sub> (Al,Cr, Fe,V) <sub>6</sub> (BO <sub>3</sub> ) <sub>3</sub> (Si,Al,B) <sub>6</sub> O <sub>18</sub> (OH,F) <sub>4</sub>
IMA记号	Tur <sup>[1]</sup>
施特龙茨分类	9.CK.05
戴纳矿物分类	61.3.1
晶体空间群	复三方锥体 (3m) (H-M记号相同)

### 性质

颜色	黑色、透明、绿、粉红、红、蓝、黄等，或是2~3种颜色。
晶体惯态	十二面体或假八面体
晶系	三方晶系
解理	不完全解理
断口	不规则/参差断口、贝状断口
韧性/脆性	脆性
莫氏硬度	7-7.5
光泽	玻璃光泽
条痕	白色
比重	3.06 (+.20 -.06)
密度	2.82-3.32g/cm <sup>3</sup>
折射率	n <sub>w</sub> =1.635-1.675, n <sub>e</sub> =1.610-1.650
双折射	-0.018 to -0.040

这是什么矿物？

# 图像分类的应用场景



The Image might have been taken in: Beijing, China. GPS location : 39.904202,116.407394

Join our Discord 

Show Image Location on the Map

<https://picarta.ai/>

# 图像分类任务面临的挑战

# 语义隔阂（语义“鸿沟”）

跨越语义隔阂（semantic gap），建立从像素到语义的映射。



人眼看到的

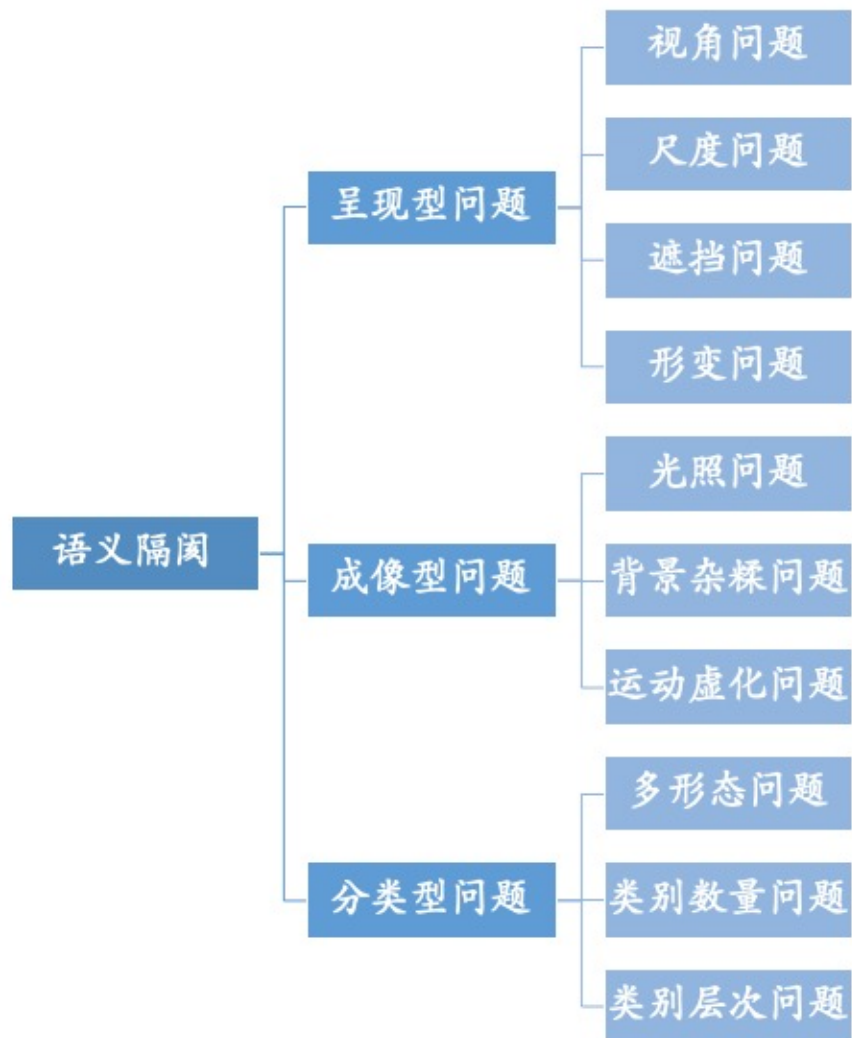
157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	85	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	85	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

机器看到的



# 语义隔阂（语义“鸿沟”）



# 视角问题 Viewpoint variation

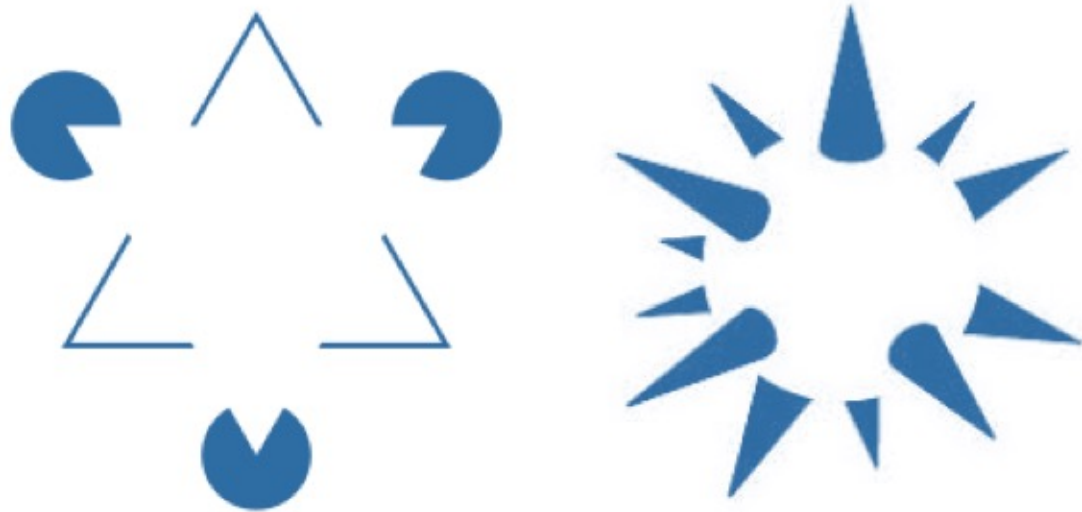


# 尺度问题 Scale



## 遮挡问题 Occlusion

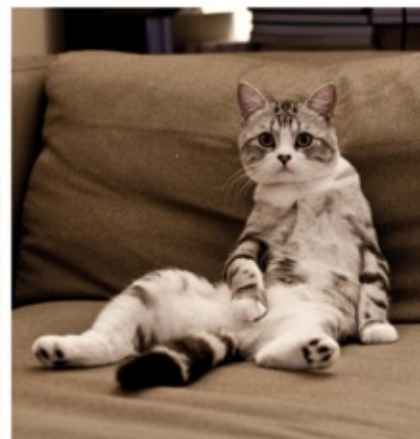
格式塔理论（Gestalt）：该心理学流派强调经验和行为的整体性。在视觉领域，强调人类可以“脑补”出图像中缺失的部分。



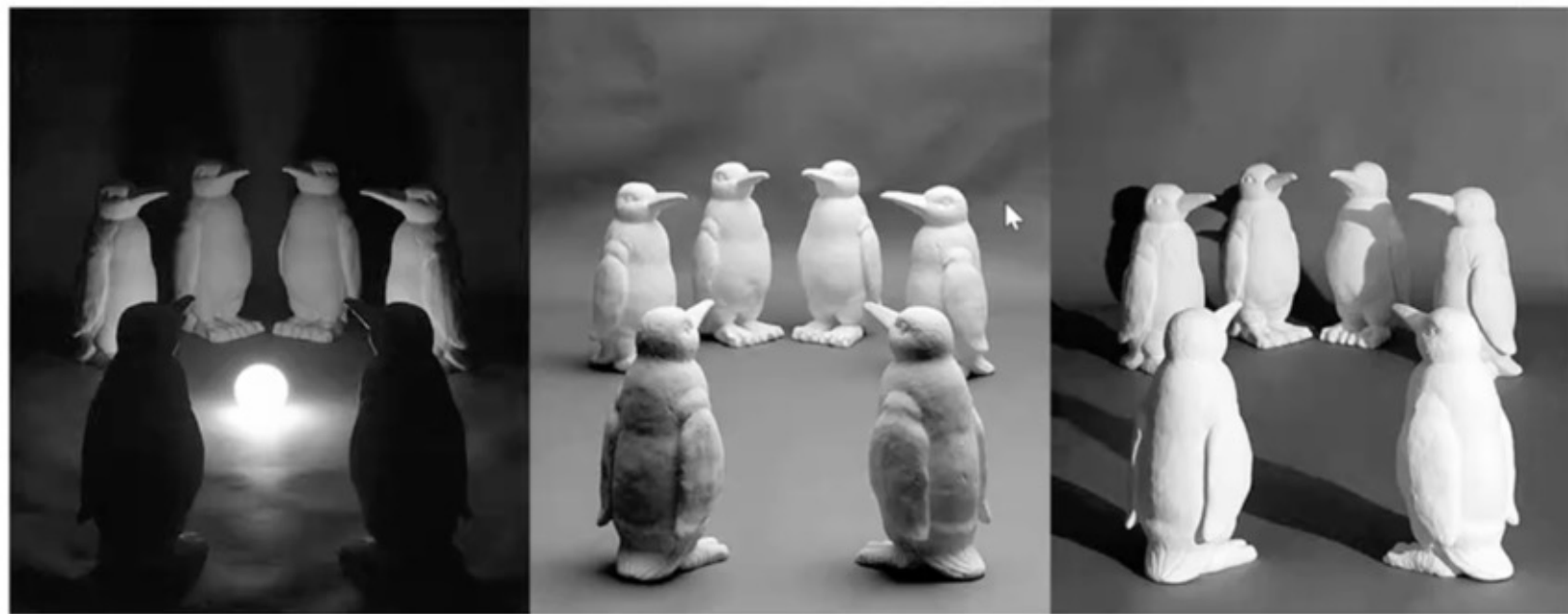
# 遮挡问题 Occlusion



# 变形问题 Deformation



# 光照 Illumination



## 背景杂播 Background clutter





# 组内多形态 Intra-class variation



# 运动模糊Motion blur



# 物品种类 Number of categories



**硬编码是否可行?**

# 基于规则（硬编码）的分类

通过硬编码的方式进行分类

```
def classify_image ( image ):  
    # Do something magical here  
    return class_label
```



对图片进行某些处理

# 基于规则（硬编码）的分类

