



# 图像信息量度量

李学龙<sup>1,2\*</sup>, 何如玢<sup>1,2</sup>

1. 西北工业大学光电与智能研究院, 西安 710072

2. 智能交互与应用工业和信息化部重点实验室 (西北工业大学), 西安 710072

\* 通信作者. E-mail: li@nwpu.edu.cn

收稿日期: 2023-03-22; 修回日期: 2023-07-30; 接受日期: 2023-09-03; 网络出版日期: 2024-06-13

国家重点研发计划 (批准号: 2022YFC2808000) 和国家自然科学基金 (批准号: 61871470) 资助项目

**摘要** 图像不仅承载着丰富的视觉内容, 同时还蕴含潜在的高级语义, 是重要的信息传递媒介。度量图像信息量则是对图像信息进行定性描述和定量计算的抽象过程, 属于对图像处理领域中的科学问题进行理论构建的关键环节。然而, 在现有的研究工作中, 虽然对“图像信息量”的度量有所提及, 但实际上关注的往往是在图像质量或图像特征的层级, 而忽略了图像所具有的高级语义以及人对图像的认知。因此, 为了衡量图像的信息提供价值, 本文基于“图像信息量因人、因任务而异”这一常识, 通过建模“先验知识”相关的图像记忆并提取“给定任务”相关的图像特征, 尝试对图像信息量进行度量。首先, 提出图像信息量度量的理论框架, 具体表现为: 在“能够获得全世界所有的图像”和“能找到一种能够准确表达图像语义的特征”两个假设条件下, 以信息论为基础, 通过构建超球来描述样本点的邻域稠密度, 由此建模语义信息的概率, 进而得到理论假设条件下的图像信息量。其次, 探讨了理论假设无法满足的实际情况, 将“全世界所有的图像”这一假设条件退化为“给定具体的数据集”, 并将“一种能够准确表达图像语义的特征”这一假设退化为“给定任务相关的特征”, 进一步通过实验展示了在实际情况如何有效获得图像信息量的数值结果。最后, 指出了本文度量方法的限定对象和度量结果的边界, 并展望了未来对相关体系进行丰富和完善的可行方向。

**关键词** 图像信息量, 信息量度量, 信容, 信息论, 图像处理

## 1 引言

什么是图像信息? 正如谚语所说, 一图胜千言, 图像承载了丰富的视觉内容和高级的语义信息, 是一种重要的传输媒介。在此基础上, 图像信息量如何理解? 在信容理论中 (信容 = 信息量/数据量, 用以衡量单位数据量的信息提供能力)<sup>[1]</sup>, 图像信息量是其计算元素之一, 通过探索图像信息量的度量方式, 实现对图像信息量这一要素的抽象描述, 可以进一步推动信容理论体系的发展, 以此为图像处理

引用格式: 李学龙, 何如玢. 图像信息量度量. 中国科学: 信息科学, 2024, 54: 1558–1566, doi: 10.1360/SSI-2023-0078

Li X L, He R B. Measuring the information of images (in Chinese). Sci Sin Inform, 2024, 54: 1558–1566, doi: 10.1360/SSI-2023-0078

领域提供具有启发性的理论指引. 进一步地, 在图像信息量度量方面, 曾经有过哪些努力? 根据分析, 以往的相关研究<sup>[2~4]</sup>表明, 一旦给定一幅图像, 其度量结果将是一个确定值, 而度量过程与图像所处的数据集和任务情境均无关. 然而, 由于无法直观地描述图像所提供的信息价值, 这类工作并不属于对图像信息量的度量, 其本质上关注的是图像特征层级的内容. 目前, 如何度量图像信息量仍然是一个尚未深入系统研究的问题.

基于特定“先验知识”和“给定任务”对图像数据进行分析处理后, 可以提取到有价值的图像信息. 对于一幅图像, 其本身可看作图像数据, 而人们在看到图像后, 会根据已有认知并结合当下的情境分析图像, 从而获得图像信息. 也即是说, 图像信息和人的认知理解相关, 这会涉及到图像中的高级语义. 而图像信息量则是通过数学建模, 对图像信息进行定性描述和定量计算的结果, 能反映图像信息给人所提供的价值. 对图像信息进行度量后得到图像信息量, 该过程与人的先验知识和当下情境的任务相关<sup>[5]</sup>. 由于不同的人有不同的“先验知识”, 不同情境对应不同的“给定任务”, 因此图像信息量也因人、因任务而异. 举例来说, 一幅内脏器官的图像, 对医生和普通人的信息量完全不同: 医生对内脏司空见惯, 而对于不常接触这类事物的普通人来说, 见到内脏器官可能会引起不适. 此外, 同一人在不同的任务情境下获得的图像信息量也会有所不同. 医生事业初启之时与累积多年从业经验后, 对内脏图像的信息认知会有显著差异. 再考虑一张白纸, 在日常生活中看到一张白纸, 和打开一个巨大的快递包裹后发现里面只有一张白纸, 这两种情境给人们带来的信息量也将截然不同.

在分析图像的内容方面, 过去有研究工作<sup>[6]</sup>提出了图像信息系统的框架, 该框架涵盖数据编码、图像结构化和空间推理 3 个相互递进的阶段. 在此框架中, 通过对图像进行底层表征并结合先验知识, 可以得到图像更高级的语义信息, 进而再在不同任务情境下根据用户需求提取图像信息. 然而, 这是一种定性工作, 没有对图像给用户带来的信息价值进行度量. 此外, 在度量图像的内容方面, 有部分研究提及了“图像信息量”的度量, 而实际上关注的却是图像质量<sup>[2]</sup>或图像特征<sup>[3,4]</sup>等底层内容. 虽然这类工作在特定的领域具有显著效果, 但其度量的数值结果缺乏直观的意义, 因此在概念上与本文所提的图像信息量有所区别. 例如, 无需考虑特定的数据集和任务情境, 对于给定图像  $a$  和  $b$  本身进行特征分析即可得到两个度量数值, 假设结果分别为 0.1 和 0.9 (此处数值仅用来说明), 这两个不同的数值可以反映  $a$  和  $b$  在特征层级存在的差异, 但却不能直接说明数值为 0.1 的图像  $a$  比数值为 0.9 的图像  $b$  的图像信息量更小. 相较于图像质量和图像特征等属性, 图像信息具有更高级的语义并涉及到人的认知层面上的理解, 因此, 本文需要明确的是图像信息  $\neq$  图像质量/图像特征/...

信息论作为信息度量的理论基础<sup>[7]</sup>, 提供了系统性的框架和数学工具, 其在复杂动态系统的推理<sup>[8]</sup>、辅助研究主观认知<sup>[9]</sup>等领域都发挥着重要的作用, 为相关研究提供了有益的启示. 香农 (Shannon) 信息论的相关理论提出, 假设事件空间  $X$  的概率分布已知, 事件  $x$  的概率为  $P(x)$ , 其带来的信息量为

$$I(x) = -\log P(x). \quad (1)$$

事件  $x$  概率越小, 其发生所提供的信息量  $I(x)$  越大.

为了度量图像信息所提供的价值, 本文基于“图像信息量因人、因任务而异”这一常识, 引入“先验知识” $\mathcal{P}$ 、“给定任务” $\mathcal{T}$  两个要素条件, 构建图像信息量的度量框架. 式 (1) 中用来度量信息量的概率分布  $P(x)$  在图像信息量度量中可用条件概率  $P(x | \mathcal{P}, \mathcal{T})$  进行表征, 从而可以得到  $x$  的图像信息量的表达式  $I(x; \mathcal{P}, \mathcal{T})$ , 即

$$I(x; \mathcal{P}, \mathcal{T}) = -\log P(x | \mathcal{P}, \mathcal{T}). \quad (2)$$

为了更清晰地说明式 (2), 其中“先验知识” $\mathcal{P}$  可以理解为大脑中存储的图像记忆, 在图像处理领

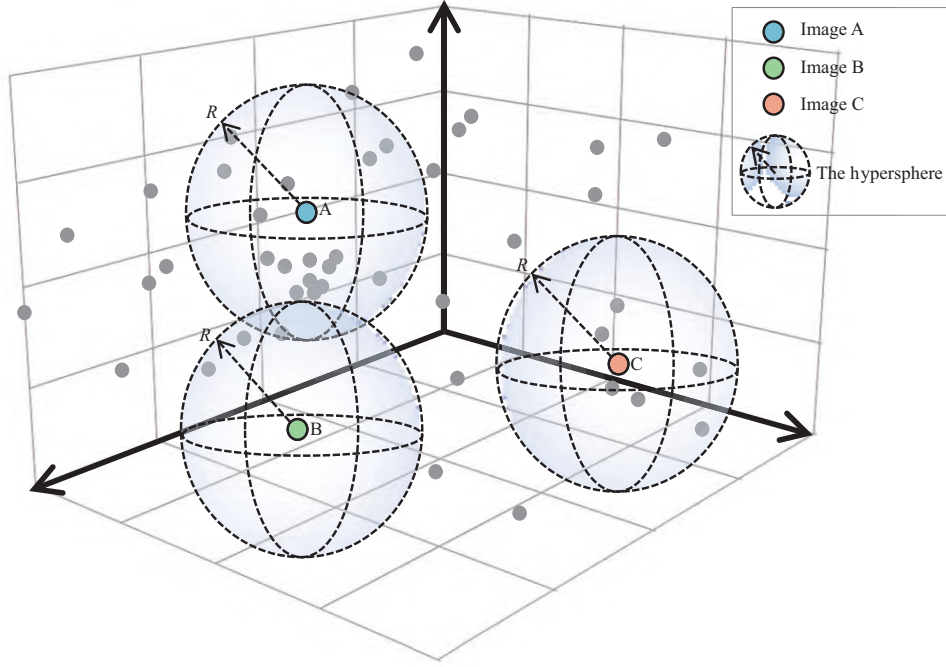


图 1 (网络版彩图) 图像信息量度量示意图. 图示空间为语义特征的映射空间, 图中不同的点表示不同的图像. 在 A, B, C 三幅图像中, 图 A 邻域 (对应圆心为 A 的超球) 的样本点最多, 表示出现的概率相对最大, 其语义的信息量相对最小; 反之, 图 B 邻域 (对应圆心为 B 的超球) 的样本点最少, 出现概率相对最小, 其语义的信息量相对最大

**Figure 1** (Color online) Diagram for measuring the information of images. The space depicted in the diagram serves as a mapping space for the semantic feature, where different sample points represent different images. Among the three images A, B, and C depicted in the figure, image A has the highest number of neighboring sample points (corresponding to the hypersphere with center A), indicating its highest probability of occurrence and thus the least amount of semantic information compared to the other images. Conversely, image B has the lowest number of neighboring sample points (corresponding to the hypersphere with center B), indicating its lowest probability of occurrence and thus the most amount of semantic information.

域可以通过数据集  $\mathcal{D}$  来建模; 而“给定任务” $\mathcal{T}$  则与所关注的图像中的特定内容相关联, 在图像处理领域可以通过图像的特征映射  $f_{\mathcal{T}}$  来建模. 因此式 (2) 等价于

$$I(x; \mathcal{D}, f_{\mathcal{T}}) = -\log P(x | \mathcal{D}, f_{\mathcal{T}}). \quad (3)$$

本文后续章节将分别从理论假设和实际情况两个角度出发, 讨论不同的数据集  $\mathcal{D}$  和与任务相关的图像特征  $f_{\mathcal{T}}$  所对应的图像信息量  $I(x; \mathcal{D}, f_{\mathcal{T}})$ , 由此分别得到  $x$  在理论假设和实际情况下的图像信息量的表达式  $I_{\text{theo}}(x)$  和  $I_{\text{exp}}(x)$ , 以及在本文概率建模方法下的图像信息量  $\hat{I}_{\text{theo}}(x)$  和  $\hat{I}_{\text{exp}}(x)$ .

## 2 理论框架

本节提出图像信息量度量的理论框架, 核心内容如图 1 所示. 框架由 4 个递进的步骤组成, 分别是 2 个核心假设、构建与样本点对应的超球、刻画样本点的概率和计算信息量. 在 2 个核心理论假设条件下, 图像信息量的理论表达式可以由算法 1 得到. 以下是对每个步骤的详细论述.

(1) **假设.** 为了在统一的基准下度量图像信息量, 本节提出以下两个假设:

(a) 假设“能够获得全世界所有的图像”, 构成数据集  $\Omega$ .

**算法 1** 理论假设下度量图像信息量的过程

输入: 全世界所有的图像构成的数据集  $\Omega$ ; 一种能够准确表达图像语义的特征  $f_T^*$ ; 图像  $x$ ; 图像信息量的理论表达式

$$I_{\text{theo}}(x) = -\log P(x | \Omega, f_T^*);$$

- 1: 得到图像  $x$  在语义特征  $f_T^*$  的映射空间中的样本点  $f_T^*(x)$ ;
- 2: 以样本点  $f_T^*(x)$  为中心做半径为  $R$  的超球, 得到超球内样本点的数目  $n_R(x)$ ;
- 3: 利用最大值对  $n_R(x)$  进行归一化处理, 建模概率  $P(x | \Omega, f_T^*) \propto \frac{n_R(x)}{\max_{x \in \Omega} n_R(x)}$ ;
- 4: 使用信息论计算式得到理论条件下基于本节概率建模方式的图像信息量  $\hat{I}_{\text{theo}}(x)$ .

(b) 假设“能找到一种能够准确表达图像语义的特征”, 表征为映射  $f_T^*$ , 在此特征空间中, 数据集可表示为一系列样本点. 其中, 语义相近的样本点彼此会更接近, 例如, 包含“天空”这一语义的图像所对应的样本点将互相靠近, 而包含“汽车”这一语义的图像所对应的样本点互相靠近.

两个假设在理论上对应最大的数据集和最完备的语义特征, 这为刻画样本点的概率提供了统一条件. 换言之, 在本节理论框架中, 对所有图像分析条件概率是在相同条件下进行的. 将假设条件下的数据集  $\Omega$  和语义特征  $f_T^*$  带入到式 (3), 由此得到图像信息量的理论表达式  $I_{\text{theo}}(x)$ , 即

$$I_{\text{theo}}(x) = -\log P(x | \Omega, f_T^*). \quad (4)$$

**(2) 构建与样本点对应的超球.** 为了描述样本点邻域的疏密程度, 构建以样本点为中心的超球. 根据分析, 语义相近的样本点彼此接近, 并且不同的样本点邻域的疏密程度不同. 对图像  $x$  来说, 这意味着它的邻域的样本点与它有相近的语义信息. 在此分析基础上, 以图像  $x$  在特征空间中所对应的样本点  $f_T^*(x)$  为中心, 做半径为  $R$  的超球, 得到超球中包含的样本点数量  $n_R(f_T^*(x))$ , 在不产生歧义的情况下, 下文用  $n_R(x)$  简化表示.

对此进一步分析可以得到如下结论: 在语义特征  $f_T^*$  对应的特征空间中, 图像  $x$  的语义信息越常见, 邻域则越密集, 以样本点  $f_T^*(x)$  为中心的超球中包含的样本点数量  $n_R(x)$  越大, 反之亦然.

举例来说, “天空”这类图像在语义特征的映射空间中的分布状态会以如下形式呈现:

(a) 对于所有包含“天空”语义的图像, 它们对应的样本点会聚集成簇;

(b) “天空”这一语义可进一步划分为“蓝色天空”、“绿色天空”、“红色天空”等语义子集, 这些子集的样本点彼此会聚集成更小的样本簇;

(c) 在上述 3 个语义子集中, “蓝色天空”这一语义信息最常见, 所以“蓝色天空”对应的样本簇最密集, 而“红色天空”、“绿色天空”相对更罕见, 因此对应的样本簇相对更稀疏. 正如生活在非极地地区的人们见到绿色天空会感到新颖, 见到蓝色天空会习以为常: 若图像  $x$  属于“绿色天空”子集,  $n_R(x)$  相对会更小; 若图像  $x$  属于“蓝色天空”子集,  $n_R(x)$  相对会更大.

如图 1 所示, 在语义特征  $f_T^*$  所映射的空间中, 图像 A, B, C 分别与图中示意的超球相对应. 与图像 B 和 C 相比, 图像 A 的超球中的样本点数最多, 则图像 A 的语义信息更常见, 可表示某幅“蓝色天空”的图像; 图像 B 和 C 则可分别表示某幅“绿色天空”和“红色天空”的图像.

**(3) 刻画样本点的概率.** 借助上文中引入的两个核心假设和超球中的样本点数目  $n_R(x)$ , 可定义本文的核心概率

$$P(x | \Omega, f_T^*) \propto \frac{n_R(x)}{\max_{x \in \Omega} n_R(x)}. \quad (5)$$

具体而言, 超球中的样本点总数越大, 式 (5) 右侧值越大, 表示图像信息在该语义空间中越常见.

**(4) 计算信息量.** 在获得图像  $x$  的概率的建模表征后, 使用信息论计算式可得到信息量的数值. 由对数的运算法则, 正比关系的比例系数在计算后会成为常数项, 不影响信息量的相对大小, 由此得到

**算法 2** 实际情况下度量图像信息量的过程

输入: 图像  $x$ ; 具体数据集  $D$ ; 指定任务  $T$ ; 实际情况下的图像信息量表达式  $I_{\text{exp}}(x) = -\log P(x | D, f_T)$ ;

- 1: 选取和任务  $T$  相关的特征  $f_T$ , 对数据集  $D$  进行特征提取;
- 2: 在所确定的特征空间中, 选取距离定义方式和超球半径  $R$ , 得到超球内样本点的数目  $n_R(x)$ ;
- 3: 代入式 (5), 取对数的底数为 2, 得到实际情况下基于本文概率建模方式的图像信息量  $\hat{I}_{\text{exp}}(x) = -\log_2 \frac{n_R(x)}{\max_{x \in D} n_R(x)}$ ;
- 4: 返回  $x$  的图像信息量的数值结果  $\hat{I}_{\text{exp}}(x)$ , 单位为比特.

在理论条件下  $x$  基于本节概率建模方式的图像信息量  $\hat{I}_{\text{theo}}(x)$ , 即

$$\hat{I}_{\text{theo}}(x) = -\log \frac{n_R(x)}{\max_{x \in \Omega} n_R(x)}. \quad (6)$$

### 3 实验分析

目前, 第 3 节理论框架中的两个假设条件 (“全世界所有的图像”  $\Omega$  和 “一种能够准确表达图像语义的特征”  $f_T^*$ ) 均无法满足, 在实际情况中, 假设条件可退化为 “给定具体的数据集”  $D$  和与 “给定任务”  $T$  相关的特征  $f_T$ , 进而将其代入式 (3) 得到实际情况下的图像信息量  $I_{\text{exp}}(x) = -\log P(x | D, f_T)$ . 通过分析, 在实验时使用的数据集  $D$  是从假设条件 “全世界所有的图像” 构成的数据集  $\Omega$  中采样得到的, 即  $D \sim \Omega$ . 针对给定任务  $T$ , 可选择与该任务相关的特定特征  $f_T$  进行处理. 举例来说, 检测任务可以通过提取纹理等特征进行处理, 分类任务可以通过提取类别等特征来实现.

一个特征空间中呈现为某种分布的点集, 通过构建超球建模样本点  $x$  邻域的稠密度进而表征概率, 再计算得到图像信息量的具体数值  $\hat{I}_{\text{exp}}(x)$ , 详细请见算法 2.






对上述过程可进行如下分析: 不同的人对应不同的数据集, 在特征空间中将呈现出不同的点集分布. 正如生活在极地地区和非极地地区的人的大脑中关于 “绿色天空” 这类图像的分布不同, 这导致了样本点邻域的疏密程度不同; 与此同时, 不同的任务情境对应不同的语义空间, 从而导致不同的特征映射结果, 并且也会影响关于邻域的界定范围. 例如某些情境关注 “天空” 这一语义信息, 而其他情境下可能需要进一步细化到 “蓝色天空”、“绿色天空”、“红色天空” 等层次的语义. 进一步地, 对于第 1 节中与常识相关的例子, 可从图像信息量度量的角度做如下解释: 医生和普通人在看到一幅内脏图像时拥有不同的 “先验知识”, 二者大脑中与内脏相关的图像记忆可以抽象表示为两个不同的数据集. 普通人脑中的内脏的图像少, 对应内脏图像少的数据集, 从而在看到内脏时有新鲜的感觉; 经验丰富的医生的大脑中存储的内脏的图像多, 对应内脏图像数目更多的数据集, 因此看到内脏习以为常. 也即是说, 图像信息量的数值结果相对于前者更大. 进一步举例, 医生第一次接触解剖感觉新鲜, 但随着时间和经验的累积, 大脑中的内脏图像不断增多, 导致内脏的图像信息量逐渐降低. 再以任务情境为例, 一张白纸, 在日常情境和在快递包裹里的两种情境下, 对应的 “给定任务” 不同, 因此所提取的特征也不同. 前者对应检测任务: 一张纯白的纸本身没有明显的边缘变化, 因此对人没有信息价值或信息价值很小; 而后者对应分类任务: 快递包裹中的物品类别只有一张白纸, 与一般的快递中的物品类别差异很大, 从而导致即使是一张白纸在这种情境下也有很大的图像信息量.

为了更直接地阐述本文度量图像信息量的核心思路, 我们设计了如下实验: 选取来自数据集 101-ObjectCategories<sup>1)</sup> 的最后 7 个类别 (对应 7 个图像集合, 分别名为 “watch”, “water lily”, “wheelchair”, “wild cat”, “windsor chair”, “wrench”, “yin yang”), 设为  $D_j$  ( $j = 1, \dots, 7$ ), 样本个数为  $N_j$  ( $N_1 = 239, N_2 = 37, N_3 = 59, N_4 = 34, N_5 = 56, N_6 = 39, N_7 = 60$ ), 作为 7 个数据集. 设  $T_1$  是检测任务,  $T_2$  是

1) <https://data.caltech.edu/records/mzrjq-6wc02>.

表 1 5 幅图像对应的信息量 (比特) 度量

Table 1 Five instances of measurement of image information (bit)

										
Dataset (size)	Task $T_1$	Task $T_2$	Task $T_1$	Task $T_2$	Task $T_1$	Task $T_2$	Task $T_1$	Task $T_2$	Task $T_1$	Task $T_2$
$D_1$ ( $N_1 = 239$ )	0.006	0	0.006	7.883	0.006	7.883	0.006	7.883	0.006	7.883
$D_2$ ( $N_2 = 37$ )	0.503	5.210	0.472	0	0.396	5.210	0.514	5.210	0.503	5.210
$D_3$ ( $N_3 = 59$ )	0.530	5.858	0.475	5.858	0.390	0	0.530	5.858	0.519	5.858
$D_4$ ( $N_4 = 34$ )	0.510	5.088	0.516	5.088	0.482	5.088	0.456	0	0.510	5.088
$D_5$ ( $N_5 = 56$ )	0.505	5.808	0.524	5.808	0.519	5.808	0.368	5.808	0.531	0
$D_6$ ( $N_6 = 39$ )	0.390	5.327	0.334	5.327	0.433	5.327	0.262	5.327	0.390	5.327
$D_7$ ( $N_7 = 60$ )	0.531	5.907	0.435	5.907	0.406	5.907	0.496	5.907	0.530	5.907

分类任务,  $f_{T_1}$  是和检测相关的局部二值模式 (local binary pattern, LBP) 这一纹理特征,  $f_{T_2}$  表示和分类任务相关的类别特征, 类别取值范围限定在数据集 101-ObjectCategories 的 101 个标签之中. 分别得到  $x_1, \dots, x_5$  在任务  $T_1$  和  $T_2$  下相对于不同数据集  $D_j$  的图像信息量表达式  $I(x_i | D_j, f_{T_k})$ . 利用算法 2 计算, 进而得到基于本文概率建模方式的图像信息量的数值结果  $\hat{I}(x_i | D_j, f_{T_k})$  ( $i = 1, \dots, 5; j = 1, \dots, 7; k = 1, 2$ ). 为了简单起见, 选用欧氏距离作为特征空间中的距离定义, 实验所得计算结果如表 1 所示. 值得一提的是, 实验旨在展示有效获得图像信息量的数值这一结果, 而高维特征空间中存在的维度诅咒问题不在本文的讨论范畴内.

表 1 第 1 行对应待度量信息量的 5 幅图像 ( $x_1, \dots, x_5$ ); 表中每一个数据表示所选图像在给定数据集、给定任务相关的特征时的图像信息量  $\hat{I}(x_i | D_j, f_{T_k})$ . 例如, 左上角的数值 0.006 表示  $x_1$  在检测任务  $T_1$  和数据集  $D_1$  的情况下, 图像信息量  $\hat{I}(x_1 | D_1, f_{T_1})$  的数值结果为 0.006.

对于分类任务  $T_2$ , 图像  $x_1, \dots, x_5$  分别被划分到 “watch”, “water lily”, “wheelchair”, “wild cat”, “windsor chair” 5 个类别中, 这与数据集  $D_1, \dots, D_5$  的类别分别对应. 因此给定任务  $T_2$ , 图像  $x_i$  相对于数据集  $D_j$  的信息量有如下关系:

$$\hat{I}(x_i | D_j, T_2) = \begin{cases} 0, & \text{当 } i = j, \\ -\log_2 N_j, & \text{当 } i \neq j. \end{cases} \quad (7)$$

当  $i = j$ , 即图像  $x_i$  的类别与数据集  $D_j$  的类别相同时, 此时图像的类别语义为数据集提供的信息量为 0; 而当  $i \neq j$  时, 图像为数据集  $D_j$  带来了全新的语义信息, 根据前文构建的超球模型, 此时  $n_R(x_i) = 1$ ,  $\max_{x \in D_j} n_R(x) = N_j$ . 这种度量表达式有直观的意义, 并且和任务  $T_2$  相关的图像信息量的数值呈现出一种规律的分布形式: 如表 1 所示, 对指定数据集  $D_j$  来说 (表中对应  $D_j$  所在的一整行), 在任务  $T_2$  下 (表中对应  $T_2$  所在列), 图像在与数据集类别相同时信息量为 0, 而其他图像对应的数值都可以由数据集的样本个数  $N_j$  计算得到, 因此数值结果相同. 同时, 通过分析图像  $x_i$  的信息量 (表中对应于图像  $x_i$  下的相邻两列的数值) 可以发现相对于不同的数据集以及在不同的任务下所得到的图像信息量不同, 这从实验角度体现了 “图像信息量因人、因任务而异” 这一常识现象.



## 4 总结与讨论

### 4.1 总结

本文探索了信容 (information capacity) 计算公式 (信容 = 信息量/数据量) 的分子 – 信息量在图像领域的度量方式. 通过强调度量图像信息量需要结合一定的“先验知识”和“给定任务”, 明确了图像信息与图像质量、图像特征等概念之间的区别. 以数值 0.1 和 0.9 为例, 就图像信息量而言, 在人的认知中, 信息量为 0.9 的图像比信息量为 0.1 的图像更罕见, 因此图像信息量更大; 然而, 对于图像质量或图像特征来说, 数值 0.1 和 0.9 用于量化不同图像的差异, 却无法反映图像信息给人所提供的实际价值, 缺乏直观的意义. 同时, 一旦给定图像, 图像质量或特征就会确定, 而图像信息量的度量还需要考虑一定的“先验知识”和“给定任务”, 即“图像信息量因人、因任务而异”. 在此分析基础上, 本文利用数据集和与任务相关的特征映射对“先验知识”和“给定任务”两个要素进行建模, 提出了度量图像信息量的框架. 首先, 在“能够获得全世界所有的图像”和“能找到一种能够准确表达图像语义的特征”两个假设下, 通过构建超球来描述图像的概率分布, 进而推导出图像信息量的理论表达式. 进一步地, 探讨了理论条件不满足的实际情况, 通过将假设条件退化为“给定具体的数据集”和“给定任务相关的特征”, 计算得到了图像信息量的数值结果.

### 4.2 贝叶斯视角

根据分析, 本文关于度量图像信息量的思想还可从贝叶斯 (Bayes) 的视角进行讨论. 以理论框架中“绿色天空”相对于生活在非极地地区的人的图像信息量较大这一情况为例, 基于贝叶斯定理可以得到如下推断关系:

$$P(\text{非极地} | \text{绿色天空}) = \frac{P(\text{绿色天空} | \text{非极地}) \cdot P(\text{非极地})}{P(\text{绿色天空} | \text{非极地}) \cdot P(\text{非极地}) + P(\text{绿色天空} | \text{极地}) \cdot P(\text{极地})}. \quad (8)$$

式 (8) 右侧  $P(\text{绿色天空} | \text{非极地})$  和  $P(\text{绿色天空} | \text{极地})$  对应似然概率, 分别表示生活在非极地地区和极地地区的人的大脑中绿色天空的出现概率. 根据常识, 生活在非极地地区的人看到“绿色天空”感到新颖而极地地区的人们能更多地看到“绿色天空”, 这一现象可通过  $P(\text{绿色天空} | \text{非极地}) < P(\text{绿色天空} | \text{极地})$  表示;  $P(\text{非极地})$  和  $P(\text{极地})$  对应贝叶斯定理中的先验概率, 由于目前地球上生活在非极地地区的人多于生活在极地地区的人数, 因此  $P(\text{非极地}) > P(\text{极地})$ .

式 (8) 左侧  $P(\text{非极地} | \text{绿色天空})$  与后验概率相对应, 表示绿色天空相对于生活在非极地地区的人的概率.  $P(\text{非极地} | \text{绿色天空})$  和  $P(\text{极地} | \text{绿色天空})$  的差异能反映“图像信息量因人、因任务而异”这一常识中的现象, 即“绿色天空”对于生活在非极地地区和极地地区的人的图像信息量不同.

### 4.3 应用探讨

本小节将对图像信息量的可能应用进行初步探讨.

一方面, 本文所提出的图像信息量  $I(x; D, f_T)$  可作为某种任务的衡量指标. 并且, 由于该指标的衡量基准在认知层面, 因此其结果导向与人的直观认知相符合. 例如, 对于异常检测任务  $T$ , 可用  $I(x; D, f_T)$  作为判断异常图像的指标. 当  $I(x; D, f_T) = 0$  时, 表示  $x$  不属于异常图像. 另外, 在某一视频表征为数据集  $D$  的情况下,  $I(x; D, f_T)$  可以作为选取视频关键帧的指标. 越大的数值表示图像帧  $x$  的语义信息变化越显著, 相对于其他帧更能代表视频的内容, 因此可被选取为视频的关键帧. 在此基础上, 可应用信息论中的熵这一概念对视频信息量进行度量. 具体而言, 对视频的每一帧进行图像信

息量度量,再通过加权求和得到信息熵,由此可表示视频的信息量.这样得到的结果能反映视频的冗余性,越小的熵值表示视频各帧之间的差异越小.当计算所得视频的信息量为0时,表示视频每一帧的语义信息都相同,从而表明该视频具有极高的冗余性.

另一方面,图像信息量可用以评估模型生成的图像是否符合人的认知.举例来说,可度量训练集中的图像信息量并基于指标  $I(x; D, f_T)$  训练模型,从而使得生成图像符合用户对图像信息量的预期.

值得一提的是,由于本文旨在提出图像信息量度量这一理论的抽象过程,因此在此不对具体的技术细节详细展开.此外,本节的探讨仅基于本文所提出的图像信息量的度量方法,并未对图像信息量的所有潜在应用领域进行全面考虑.未来随着图像信息量度量体系的丰富和完善,图像信息量这一指标有望为整个图像处理领域的理论和应用提供重要启示.

#### 4.4 未来工作

在未来的工作中,拟运用统计学的相关方法,在不同情境下邀请不同实验者对不同图像的信息量进行评分,通过收集和分析大量的统计数据,对所度量的图像信息量与人的主观一致性进行评估.同时,在此分析基础上,如何优化图像信息量度量的方法等问题还值得深入研究.

值得说明的是,本文度量方法的限定对象是未考虑光电系统因素的图像,若要度量考虑了光电系统(比如相机)成像过程的影像信息量,还需要与成像过程涉及的物理模型进一步结合.图像信息量作为信容计算公式的分子,按本文方法进行度量的结果边界是  $[0, \infty)$ ,而信容取值的边界划定在  $[0, 1]$ ,因此未来亟需探索信容计算公式的分母—数据量的度量方式.此外,还需建立其他模态的信息量和数据量的度量框架,以完善信容这一科学理论体系.本文抛砖引玉,希望能为图像处理、计算机视觉等领域提供一些启发性的思考和探讨方向.

#### 参考文献

- 1 Li X L, Zhao B. Video distillation. *Sci Sin Inform*, 2021, 51: 695–734 [李学龙, 赵斌. 视频萃取. *中国科学: 信息科学*, 2021, 51: 695–734]
- 2 Sheikh H R, Bovik A C. Image information and visual quality. *IEEE Trans Image Process*, 2006, 15: 430–444
- 3 Engelsma J J, Cao K, Jain A K. Learning a fixed-length fingerprint representation. *IEEE Trans Pattern Anal Mach Intell*, 2021, 43: 1981–1997
- 4 Ferraro M, Boccignone G, Caelli T. On the representation of image structures via scale space entropy conditions. *IEEE Trans Pattern Anal Machine Intell*, 1999, 21: 1199–1203
- 5 Li X L. Multi-modal cognitive computing. *Sci Sin Inform*, 2023, 53: 1–32 [李学龙. 多模态认知计算. *中国科学: 信息科学*, 2023, 53: 1–32]
- 6 Chang S K, Hsu A. Image information systems: where do we go from here? *IEEE Trans Knowl Data Eng*, 1992, 4: 431–442
- 7 Shannon C E. A mathematical theory of communication. *Bell Syst Technical J*, 1948, 27: 379–423
- 8 Golan A, Harte J. Information theory: a foundation for complexity science. *Proc Natl Acad Sci USA*, 2022, 119: e2119089119
- 9 Sayood K. Information theory and cognition: a review. *Entropy*, 2018, 20: 706–724



## Measuring the information of images

Xuelong LI<sup>1,2\*</sup> & Rubin HE<sup>1,2</sup>

1. *School of Artificial Intelligence, Optics and ElectroNics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China;*

2. *Key Laboratory of Intelligent Interaction and Applications (Northwestern Polytechnical University), Ministry of Industry and Information Technology, Xi'an 710072, China*

\* Corresponding author. E-mail: li@nwpu.edu.cn

**Abstract** Images possess not only abundant visual content but also inherent high-level semantics, rendering them a pivotal medium for information transmission. Measuring the information of images is an abstract process of qualitative description and quantitative calculation of image information, constituting a fundamental aspect of scientific theories in the field of image processing. While the measurement of “image information” has been mentioned in previous research, it primarily centers on the quality or features of images, neglecting the exploration of the factors linked to high-level semantics and human cognition. Hence, to gauge the informational value that images offer to individuals, this paper seeks to measure image information by modeling the iconic memory associated with individuals’ “prior knowledge” and extracting image features pertinent to “the given tasks”. This approach is grounded in the widely accepted belief that “the information of images differs among individuals and across various tasks”. Firstly, a theoretical framework for measuring the information of images is proposed. It is demonstrated as follows: under the assumptions of “having access to all images in the world” and “possessing a feature that can fully express the semantic content of images”, a hypersphere is constructed to describe the density of the neighborhoods of the sample point. By modeling the probability of semantic information within this framework, the information of images is obtained based on information theory under the theoretical assumptions. Secondly, the practical scenarios when the assumptions specified in the theoretical framework are not feasible are discussed. It is illustrated by reducing the assumption of “having access to all images in the world” to “the given specific dataset” and reducing the assumption of “possessing a feature that can fully express the semantic content of images” to “the given task-related features”. Then the experiments were conducted to show how to effectively obtain numerical results of image information in practical scenarios using the proposed method. Lastly, the limitations of the measurement method and the boundaries of the measurement results are pointed out, and feasible directions for enriching and improving the relevant systems in the future are contemplated.

**Keywords** image information, information measurement, information capacity, information theory, image processing