

自然图像的统计数据

Daniel L Rudermant*

翻译: 林绪虹†

摘要

最近,人们对自然图像的属性重新产生了兴趣。它们的统计数据不仅在图像压缩中很重要,而且在生物学感官处理研究中也很重要,这可以被视为满足某些“设计标准”。这篇评论总结了以前关于图像统计的研究,并提供了我们自己的数据。自然图像最显着的属性可能是尺度不变性。我们提供了数据来支持这一说法,以及自然场景中层次不变性的证据。这些对称性为自然图像提供了强有力的描述,因为它们极大地限制了允许分布的类别。

Keywords: 自然图像统计.

1. 介绍

我们可以轻松区分自然界的图像与人造图像或计算机随机创建的图像。自然图像之所以与众不同,是因为它们包含特定类型的结构。它们远非随机:在计算机上随机构建的图像几乎从不包含自然场景,甚至不包含一棵树。因此,在所有可能的图像中,自然图像非常罕见。我们如何利用这一事实?

图像处理系统(例如压缩算法、模拟存储介质和我们自己的视觉系统)处理这些真实世界的图像。因此,要了解这些系统的典型行为,我们必须首先研究自然场景的结构。然后,我们可以解决诸如“我们应该期望实现多少压缩?”或“播放失真会多久超过临界值?”或“视神经每秒传递多少位信息?”等问题。

在这篇评论中,我们首先总结了已知的自然图像特征,并添加了新发现。这项研究的动机是使用此类图像统计数据的实际例子,主要针对生物视觉。在分析图像时,我们会提出以下问题:

- 自然图像与随机图像有哪些不同?
- 自然图像统计数据是否遵循任何简单的不变性?
- 这些统计数据对于生物视觉系统中的图像处理有何影响?

在第一部分中，我们概述了信号处理中的统计方法，重点介绍了其在视觉方面的应用。我们描述了一个框架，用于根据涉及自然场景统计数据的设计标准来理解视觉性能。

随后，我们概述了几种可以表征这些图像的方法，并介绍了以前的研究成果，这些成果暗示了自然图像结构的一些简单属性。最后，我们对数据进行了详细分析，并尝试量化图像中存在的可预测性或冗余量。我们能够确认其他研究提出的缩放比例，并发现与自然场景的层次结构相关的新不变性。

2. 统计框架

我们每天遇到的图像只是所有可能图像的一个非常稀疏的子集。

大多数图像在自然界中根本不会出现，这可以通过在计算机上创建随机图像轻松证明。想象一下所有 256×256 图像都存在于 65536 维空间中，那么自然图像所占据的空间的“体积”是无穷小的。此外，自然图像不是高斯分布，也不是从图像建模中常用的任何其他基本分布中得出的。用 Field 的话来说 [32]

‘...描述自然场景概率密度的状态空间具有高度可预测性，但没有广泛假设的形状。’

自然图像的分布很复杂。也许它就像啤酒泡沫一样，大部分是空的，但包含一层薄薄的流体网状结构，填充空间，几乎不占据任何体积。流体区域代表那些具有自然特征的图像。我们的直觉应该是，没有“简单”的空间变换可以消除分布的复杂性。空间大部分未被占用这一事实表达了分布中的冗余。

由于自然图像具有高度非随机性，我们可能怀疑以像素为单位来表达它们并不是最方便的选择，因为它适用于完全随机的图像。如果有一种方法可以将最常出现的图像编码为短的位串，将最不可能出现的图像编码为较长的位串（也许使用霍夫曼编码 [41]），那么平均所需的存储空间量可以减少。大多数实用算法不是在整个图像级别上工作，而是考虑子图像，例如水平扫描线（预测编码）或 8×8 像素块（JPEG）。然后将它们编码为独立实体，忽略它们的相互依赖性。事实上，此类程序目前在磁盘空间中节省了一个数量级 [1]，这说明了真实图像中包含的大量冗余和可预测性。最重要的是，这些图像的统计结构决定了最佳压缩算法。但目前我们甚至不知道理论上可以实现多少压缩，因为真实图像的统计数据尚未得到很好的表征。

讨论的核心是图像集合的概念。我们设想每个图像 $I(x)$ 都与一个发生概率相关联¹， $P[I(x)]$ ，它定义了集合。从此分布中随机抽取的图像将完全代表手头的自然环境。实际问题可能与设备或算法在处理从此集合中抽取的图像时的性能有关。

为了给出统计公式的说明性示例，我们询问一组图像如何通过噪声线性编码进行表示。让我们通过考虑一维随机信号 $\phi(x)$ （带限和广义平稳，具有零集合均值）来简化，该信号在每个点处与滤波器 $f(x)$ 卷积。由 $\eta(x)$ 给出的随机噪声被添加到此信号中。最终编码 $y(x)$ 可能表示模拟磁带上记录的信号或一组神经元的响应。具体来说，我们有

$$y(x) = (f * \phi)(x) + \eta(x) \quad (1)$$

¹严格地说，我们应该在图像空间上调用概率密度（具有相关度量），因为它们构成了“事件”的连续体。

我们希望知道 $\phi(x)$ 在多大程度上能被 $y(x)$ 表示。一种方法是从信号 $y(x)$ 重构 $\phi(x)$ 的“最佳”估计值²。

我们将最佳估计 $\phi_{\text{est}}(x)$ 定义为具有最小均方误差的估计，即

$$\langle |\phi_{\text{est}}(x) - \phi(x)|^2 \rangle \quad (2)$$

最小化，其中期望值是所有信号和噪声的。众所周知，这个估计量由后验分布的均值给出 [65]， $P[\phi|y]$ ：

$$\phi_{\text{est}}(x) = \int \mathcal{D}\phi \phi(x) P[\phi|y] \quad (3)$$

其中积分针对的是集合中的所有 $\phi(x)$ 。

集合的统计数据通过贝叶斯定理作为先验分布输入：

$$P[\phi|y] = \frac{P[y|\phi]P[\phi]}{P[y]} \quad (4)$$

并且 $P[y|\phi]$ 取决于滤波器 $f(x)$ 和噪声统计量。所以我们有

$$\phi_{\text{est}}(x) = \frac{\int \mathcal{D}\phi \phi(x) P[y|\phi] P[\phi]}{\int \mathcal{D}\phi P[y|\phi] P[\phi]} \quad (5)$$

这个方程告诉我们如何从测量值 $y(x)$ 得出信号的最佳猜测 $\phi_{\text{est}}(x)$ 。当信号和噪声概率分布均为高斯分布时，估计值是测量值的线性函数。但一般来说情况并非如此，估计量可能相当复杂。在上述公式中，很明显，了解由 $P[\phi]$ 给出的集合对于理解系统运行情况非常重要。

一般而言，必须完全表征 $P[\theta]$ 才能进行上述计算。但是，当面对高维信号时，这是不可能的，因为这需要从分布中收集大量图像³。幸运的是，在信噪比 (SNR) 较低的情况下，只需要分布中的少数相关函数。

可以证明，对于 SNR 的最低阶，最佳估计由以下公式给出：

$$\Phi_{\text{ess}}(k) = \frac{F(k)S(k)}{\mathcal{N}(k)} Y(k) \quad (6)$$

其中 Φ_{est} 、 F 和 Y 分别是 ϕ_{est} 、 f 和 y 的傅里叶变换， $S(k)$ 和 $\mathcal{N}(k)$ 分别是信号和噪声在空间频率 k 处的集合功率谱密度。信号集合的功率谱可以用其二阶相关函数表示为

$$\langle \Phi^*(k)\Phi(k') \rangle = 2\pi \mathcal{S}(k)\delta(k - k') \quad (7)$$

在较高的信噪比下，方程 (6) 将包括图像集合的高阶相关函数。请注意，该方程也是维纳滤波器低信噪比展开式中的一阶项。

²关于线性信号分析和估计的概述可以在 Wiener [88] 的开创性著作中找到。

³想象一下，需要查看多少张 4×4 图像才能填充它们的概率分布。假设每个像素量化为 16 个灰度值。那么可能的图像数量超过 10^{19} ，我们需要许多倍于此的示例才能猜测它们的分布方式。

现在我们可以问这样的问题：“在指定的约束条件下，选择哪种 $f(x)$ 可以最小化预期的重建误差？”一旦指定了问题的统计特征，我们不仅可以量化其典型行为，还可以询问系统相对于给定标准的最佳设计是什么。

近年来，这一范式变得流行起来，并在应用于生物视觉时产生了令人鼓舞的成果，例如预测神经反应 [48, 50, 68]、感受野 [2, 4, 5, 10, 54, 57, 71, 72, 80, 83, 84, 86]、颜色编码 [3, 8, 14, 22, 59, 64]、立体编码 [53]、复眼的设计 [47, 78, 79]，甚至瞳孔反应 [49]。

该方法的最终目标是根据第一原理预测生物系统的测量响应，无论是神经元的活动、复眼中的最佳面间距还是人类的心理物理阈值。在后一类中，Atick 和 Redlich [4] 根据最小化表示冗余的标准预测了自然场景的最佳神经编码。他们的结果与人类在光照水平下数十年的检测阈值相匹配。在构建方法时，他们结合了设计标准、系统约束和图像的测量统计数据。这些最小要素允许进行无参数预测，这意味着效率的基本思想可能在视觉领域得到广泛应用。

Laughlin [48] 询问视觉神经元应如何最好地编码对比度，以便传输尽可能多的信息（以香农度量 [76] 为准）。答案是将对比度转换为响应直方图均匀的方式。此过程消除了英文文本中存在的相同类型的冗余，其中字母表中的字母使用频率并不相同。他的处理方法预测了一条对比度响应曲线，该曲线与苍蝇视觉系统中 LMC 细胞的响应特性非常相似。做出这一预测需要测量自然图像的对比度直方图。我们将在本文后面回到响应直方图这一问题。

所有这些方法的基本思想是，传感系统能够很好地处理自然界中存在的各种信号。许多提出的有效设计标准都是基于统计测量的。Attneave [6] 和 Barlow [7] 认为，减少传感信息的冗余是传感系统的主要目标。

Linsker 列举了一些例子，说明感官编码如何最大限度地提供信息 [55, 56, 57]。其他衡量标准包括重建保真度 [57, 71] 和神经反应直方图的形状 [32, 48]。所有这些标准都涉及神经编码的统计数据。只有首先研究自然场景的属性，我们才能预测如何最好地处理它们。

总之，我们经常想知道图像处理系统在正常条件下的工作情况。人们通常使用统计度量，例如系统提供的信息量的平均重建误差。这将我们带入一个具有统计基础的框架，需要了解自然图像统计知识。

3. 收集自然图像

到目前为止，我们已经为自然图像的研究提供了动机。但仍存在一些重要问题：

- 哪些图像应该包含在整体中？
- 我们应该测量什么？
- 我们应该计算哪些统计数据？

显然，没有“正确”的集合。每个环境都有自己的典型特征，因此也有自己的统计数据。在解释这些统计数据之间的差异时，我们可能会寻找视觉系统与其所居住环境的统计数据有系统差异的生物 [38, 58]。

一项早期研究对电视图像进行了描述 [45]，旨在了解如何对其进行最佳编码。最近的研究则集中于户外自然图像 [31, 73, 82]。

我们的数据包括春天树木繁茂的环境的图像。

一旦选定了环境，我们就必须决定要捕捉哪些图像，也就是将相机指向何处。我们还可以选择使用角分辨率或光谱灵敏度来模拟特定生物的视觉系统⁴。当然，视觉系统是实时运作的，因此分析短片而不是静态图像会增加另一个层次的细节。

最后，我们必须考虑统计分析。最终，人们想知道 $P[I]$ ，即任何图像出现的概率。但是，如前所述，由于数据集所需的大小，这是不可能实现的。另一种可能性是通过实验确定模型 $P[I]$ 的最佳参数，例如马尔可夫随机场 [33, 43]。

也许最直接的方法是将图像分布的相关函数编入目录 [13]。这意味着使用图像数据来评估类似 $\langle I(x_1)I(x_2)\cdots I(x_n) \rangle$ 的表达式。但除了二阶相关函数 ($n = 2$) 之外，这些量很难解释（也很难可视化，因为每个空间索引都会为函数的域添加两个维度）。如上所述，相关函数在低 SNR 下以扰动方式自然进入统计计算。因此，尽管高阶相关函数可能无法提供太多见解，但它们确实具有直接的应用。

表征任意分布可以通过蛮力完成（例如，通过测量相关函数）。但我们也可以在分布中寻找简单的底层结构或“不变性”属性。也就是说，如果图像简单地变换，图像概率也可以简单地变换。平移不变性就是这样一种对称性。对于某些集合，我们期望给定图像出现的概率与其位置偏移无关。这种不变性可以大大降低分布的复杂性。它们通常足以合成量子场论的分布 [39]，并且在图像处理 [51] 中也发挥着重要作用。如果自然场景的分布不包含这样的不变性，那么仅仅收集统计数据将是一项有用但不一定有趣的冒险。然而，我们会发现自然图像确实表现出一些相当令人惊讶的对称性。

由于无法收集足够的数​​据来全面描述图像环境，我们的统计描述将远非完整。我们甚至无法用我们最少的统计知识重现逼真的图像。有趣的是，考虑一下需要多少这类知识才能做一些有用的事情。例如，我们已经看到，在低 SNR 条件下，唯一重要的统计数据是功率谱，它可用于设计最佳的低 SNR 滤波器。在高 SNR 下，统计数据的每个细节都很重要，但有多重要呢？我们能否将我们的知识限制在几个允许近乎最佳性能的统计数据上？

实际上，这就是我们的视觉系统所做的。哺乳动物视觉系统的发育在很大程度上取决于早期的视觉刺激 [11, 61]。具体来说，众所周知，在特殊图像环境中长大的哺乳动物成年后其视觉系统功能会发生改变 [36]。这表明，视觉系统的发育受到其环境统计数据的影响，通过某种尚不为人所知的算法。视觉系统的最终“线路”中包含一组有关生物过去视觉体验的统计数据。了解这些统计数据可能有助于深入了解视觉处理的本质。

4. 图像统计

1952 年，Kretzmer [45] 开创了对现实世界图像的现代分析。

考虑到在电视图像编码中的应用，他列出了一组局部图像统计数据，从中，例如点直方图和二阶相关函数。他通过测量将图像冗余的下限设定为每像素约 3 位。

图像功率谱中幂律缩放首次被提及是在 1957 年 Deriugin [28] 的一篇论文中，他也测量了电视信

⁴希望自然图像的特征对这些类型的细节相当稳健。

号。这一特性在 1978 年被 Cohen 等人 [23] (另见 [20]) 重新发现, 1987 年 Burton 和 Moorhead [16] 以及 Field [31] 再次发现。后来 Tolhurst 等人 [82] 和我们 [73, 74] 研究了这种缩放。发现集合功率谱 (在方向上平均) 的行为大致为

$$\mathcal{S}(k) \propto k^{-2+\eta} \quad (8)$$

其中 k 是空间频率的模数 (例如, 以 deg^{-1} 为单位), 并且 η 被测量为很小。功率谱是空间频率给出的单个角尺度的函数, 并且它随着该尺度的幂而变化。自然图像中没有首选的角尺度, 因为功率谱的形式对于任何基本尺度的选择都是不变的。将空间频率加倍总是会将功率降低 $2^{-2+\eta}$ 倍。例如, 对于形式为 $\mathcal{S}(k) \approx e^{-k/k_0}$ 的功率谱, 情况就不是这样, 其中 k_0 充当“典型”空间频率。

这些研究为自然图像集合中的某种对称性提供了证据: 尺度不变性。尺度不变性仅仅意味着图像统计量不会随角度尺度而变化。无论镜头焦距如何, 这种集合的图片都将具有相同的集合统计量。更一般地, 新的集合可能与原始集合自仿射, 这意味着新图像在重新缩放后也必须乘以合适的常数, 以使统计量与原始图像相同。如果 $Q[\langle(ax)\rangle]$ 是尺度 cz 上 $4(x)$ 的任何集合统计量, 则尺度不变性意味着

$$Q[\phi(x)] = Q[\alpha^\nu \phi(\alpha x)] \quad (9)$$

其中 ν 是通用指数 (即它与 a 和 Q 无关)。因此, 在尺度不变集合中, 我们可以将 $\phi(x) \rightarrow \alpha^\nu \phi(\alpha x)$ 替换为任意期望值中所有 ϕ 实例。

这是一个强有力的说法。它极大地限制了图像分布的形式。这样的属性还让我们对自然场景有了一定的直觉, 而不仅仅是对其统计数据的量化。例如, 它强化了这样一种观念: 自然世界中的物体可以在图像中以任何角度尺度出现 (即它们可以位于任意距离), 这是一种产生尺度不变性的合理机制。

尺度不变性是临界现象 (如铁磁体的居里点) 的一个被广泛研究的性质。物理学家研究具有局部 (马尔可夫) 相互作用的模型, 这些相互作用会产生“长程” (即幂律) 相关性。为了实现尺度不变性, 必须非常精确地选择模型参数 [89]。有趣的是, 在二维中, 局部尺度不变和旋转不变模型也必须是共形不变的 [19]。但由于自然界中的图像甚至不是各向同性的——地平线具有明确的方向——我们不能指望它们具有更普遍的共形不变性。因此, 为了在排除共形不变性的同时对尺度不变统计数据建模, 必须包括远距离相互作用。

如果自然图像形成一个尺度不变的集合, 那么我们应该在功率谱之外的其他统计数据中找到尺度。1992 年, William Bialek 和我收集了自然图像数据并开始进行统计分析 [70, 73, 74]。除了一些值得注意的例外, 下面介绍的工作都是这次合作的成果。我们不仅能够确认尺度结果, 还能找到新不变性的证据。

其他关于自然图像的研究包括局部主成分分析 [34、52、75], 其中寻找最大程度地去相关图像的局部线性滤波器。响应方差最大的滤波器往往与皮质中的定向感受野相似, 这表明可能存在某种去相关过程。去相关通常被认为是减少表示冗余的第一步。

人们还研究了自然场景的光谱反射特性。1947 年, 克里诺夫 [46] 测量了 337 个自然物体 (如草和木头) 在 26 个波长下的光谱反射率。马洛尼发现, 所有这些光谱曲线都可以使用具有 7 个自由参

数的模型进行紧密拟合 [59]，这表示 26 个维度中的 19 个维度存在冗余。丹尼米勒 [26] 发现，随机光子捕获产生的噪声有效地将维度降低到 3，这是人类视网膜中存在的视锥细胞类型的数量。使用色测量（因此特别涉及人类的彩色视觉），伯顿和摩尔黑德 [16] 发现自然图像会引起视锥细胞高度相关的反应，并表明功率谱在三个视锥系统中的比例约为 $1/k^2$ 。通过研究彩色图像的属性，人们可以尝试预测视网膜光感受器的最佳排列 [60]、颜色编码 [3] 和彩色视觉的最佳瞳孔函数 [49]。

5. 测量自然图像

我们选择了春季新泽西州中部的树林作为图像集合。这些树林是昆虫、小型哺乳动物和鸟类的栖息地，因此对于许多不同类型的视觉系统来说，这是一个重要的感官环境。由于它仅构成特定环境，因此我们的结果不一定具有其他环境的特征。另一种方法是收集来自不同环境的图像，并将它们作为一个自然场景的单一集合一起进行分析。

树林由树木、灌木丛、岩石和溪流组成。图 1 中展示了该组合的图像。由于相机指向随机方向且仰角较小，因此该组合具有平移不变性，避免了单独的天空或地面图像。附录中介绍了测量过程的详细信息。

反射场景的亮度与太阳的辐射通量成正比。包括相机在内的视觉系统会适应这个平均背景值，因此它无关紧要。通过考虑背景水平的对数强度波动，我们可以将其从图像中去除。我们将“对数对比度” $\phi(x)$ 定义为

$$\phi(x) = \ln [I(x)/I_0] \quad (10)$$

其中 $I(x)$ 是测量的强度信号， I_0 定义为每幅图像，使得 $\sum_x \phi(x) = 0$ 。这会使每幅图像直方图均值为零。 I_0 的定义是任意的，但我们的许多统计数据都是对数对比度差异，其中常数 I_0 会消失。对数测量很方便，因为它覆盖了整个实轴；另一方面，强度有点难以处理，因为它们是非负的。我们发现使用 ϕ 代替 I 似乎也能改善观察到的不变性。

该数据集包含 45 张焦距为 15 毫米（1 像素对应视角 0.059° ）的图像和 25 张焦距为 80 毫米（ $0.011^\circ/\text{像素}$ ）的图像。数字化图像为 640×480 像素，中央 256×256 子图像即取自这些图像。

在 15 毫米和 80 毫米焦距下，图像分别朝向 15° 和 2.8° 。

6. 统计分析

先前的研究已经暗示了自然图像中的缩放。我们将首先尝试通过评估图像的功率谱来证实这一点。数字化图像是离散采样的连续图像 $\phi(x)$ ，它们来自具有功率谱 $S(k)$ 的平稳分布，如下所示

$$\langle \Phi^*(k)\Phi(k') \rangle = (2\pi)^2 S(k)\delta^{(2)}(k - k') \quad (11)$$

其中 $\Phi(k) = \int d^2x \phi(x)e^{-ik \cdot x}$ 。严格来说，坐标系由方位角和仰角组成。像素化图像是从连续图像中得出的，如下所示



图 1: 树林中的图像: 溪流中的岩石, 背景是树叶。

$$\phi_{m,n} = \phi(ma, na) \quad (12)$$

其中 a 是以度为单位测量的像素间距, m, n 是从 0 到 255 的像素索引。我们希望从离散采样的图像中估计 $\mathcal{S}(k)$ 。

谱估计是一个被充分探索的领域, 关于它的一个有说服力的总结可以在 [65] 中找到。功率谱的估计方法是

$$S_{\text{ext}}(k_{r,s}) = \left(\frac{a}{2\pi M^2} \right)^2 \frac{1}{N} \sum_{i=1}^N \left| \sum_{m,n=0}^{M-1} \phi_{m,n}^i W_{m,n} \exp(-ik_{r,s} \cdot \alpha_{m,n}) \right|^2 \quad (13)$$

其中 $x_{m,n} = (ma, na)$ 、 $k_{r,s} = (2\pi r/Ma, 2\pi s/Ma)$ 、 $M = 256$, 并且 i 遍历数据集中的所有图像。 $w_{m,n}$ 称为“窗口函数”; 其形状决定了估计值与实际频谱的关系。我们在计算中使用二维 Bartlett 窗口, 但结果并不强烈依赖于此选择。

图 2 中显示了自然场景对数对比度功率谱的轮廓图。轮廓以功率的对数以恒定间隔放置。它显示了水平和垂直方向上低空间频率的功率优势, 这显然很特别。绘制方向平均光谱更有启发性, 如 @fig-03 所示。该图由两个叠加的图组成, 每个图来自不同的镜头焦距。尺度在空间频率和功率上都是对数的, 因此绘制的数据几乎是线性的。这意味着功率谱是形式为

$$\mathcal{S}(|k|) = \frac{A}{|k|^{2-\eta}} \quad (14)$$

其中 $\eta = 0.19 \pm 0.01$, 且 $A = (6.47 \pm 0.13) \times 10^{-3} \text{deg}^{(0.19)}$ 。

对于给定焦距, 在高空间频率下测量会变得困难。光学模糊会导致光谱下降, 同时噪声和混叠会导致功率增加。我们通过简单地使用两个焦距来扩展空间频率范围。该图显示了光谱在空间频率上近 2.5 个数量级的缩放。显示缩放的最高频率约为 30 个周期 deg^{-1} , 这相当于人眼敏锐度的一半 [17]。因此, 这些是视觉的相关空间频率。虽然我们的结果来自强度的对数, 但我们发现功率谱中的缩放与其他人在使用强度信号时发现的相同 (我们也发现了这一点)。功率谱中的缩放对于这种变化以及集合选择、光谱灵敏度和图像捕获方法的明显差异具有鲁棒性。但是, 指数因作者和环境而异 (例如, 我们发现海滩图像的指数为 $\eta \approx -0.3$)。

这种功率谱的缩放证实了其他人的发现。但我们还可以说更多。首先, 功率谱本身并不能告诉我们分布是否是高斯分布。

此外, 缩放应该可以通过我们选择的任何统计数据进行测试。这两个问题可以通过“粗粒度”过程同时探索。Kadanoff [44] 最初的缩放思想提出, 粗粒度临界系统应该具有与原始系统相同的统计数据, 除了可能重新缩放场变量之外。

类似地, 我们可以粗粒度化图像以寻找缩放和高斯统计的可能性。4 在尺度 N 上的平均值由高斯统计给出。 ϕ 在尺度 N 上的平均值由以下公式给出

$$\phi_N = \frac{1}{N^2} \sum_{m,n=1}^N \phi_{m,n}. \quad (15)$$

图 4 中显示了 $N = 2$ 的此过程示例。如果 ϕ 是缩放场，则概率 $P_N(\phi_N)$ 应具有与 N 无关的 *shape*。在临界现象理论中，当场具有异常维度（即 $\eta \neq 0$ ）时，必须在长度尺度发生变化时对其进行“重新正则化”。这意味着

$$P_N(\phi) = 1/\phi_N^{\text{RMS}} \mathcal{P}(\phi/\phi_N^{\text{RMS}}) \quad (16)$$

其中 $\phi_N^{\text{RMS}} = \langle \phi_N^2 \rangle^{1/2}$ ， \mathcal{P} 是缩放概率分布。对于具有异常维度 $\eta/2$ 的缩放场， $\phi_N^{\text{RMS}} \propto N^{-\eta/2}$ 。为了比较直方图形状，必须除以这个 RMS 值。

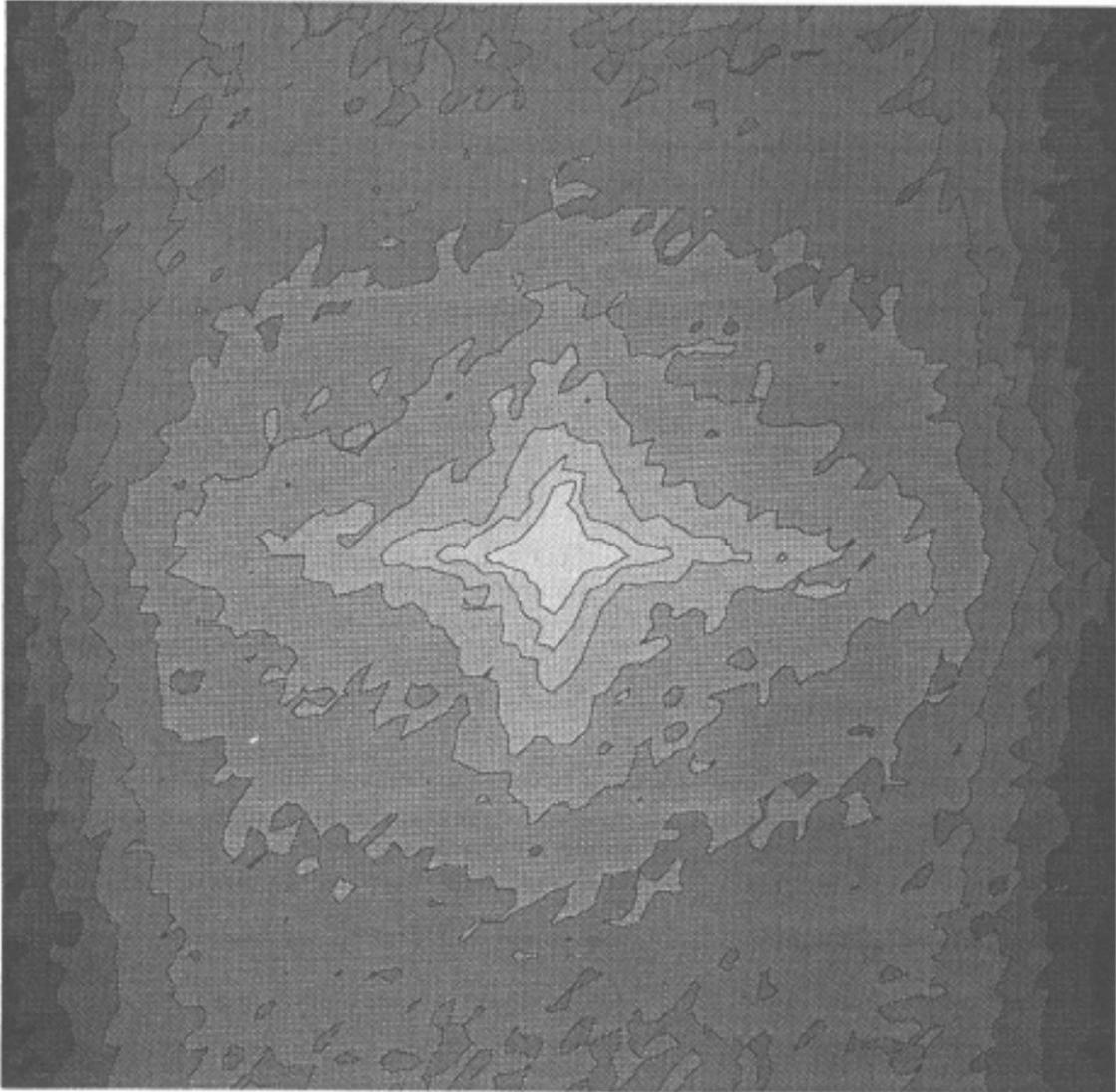


图 2: 焦距为 15 毫米时拍摄的 45 幅图像的集合功率谱轮廓图。图中心为 $k = 0$ 。轮廓以功率对数的等距间隔放置，空间频率绘制在线性轴上。

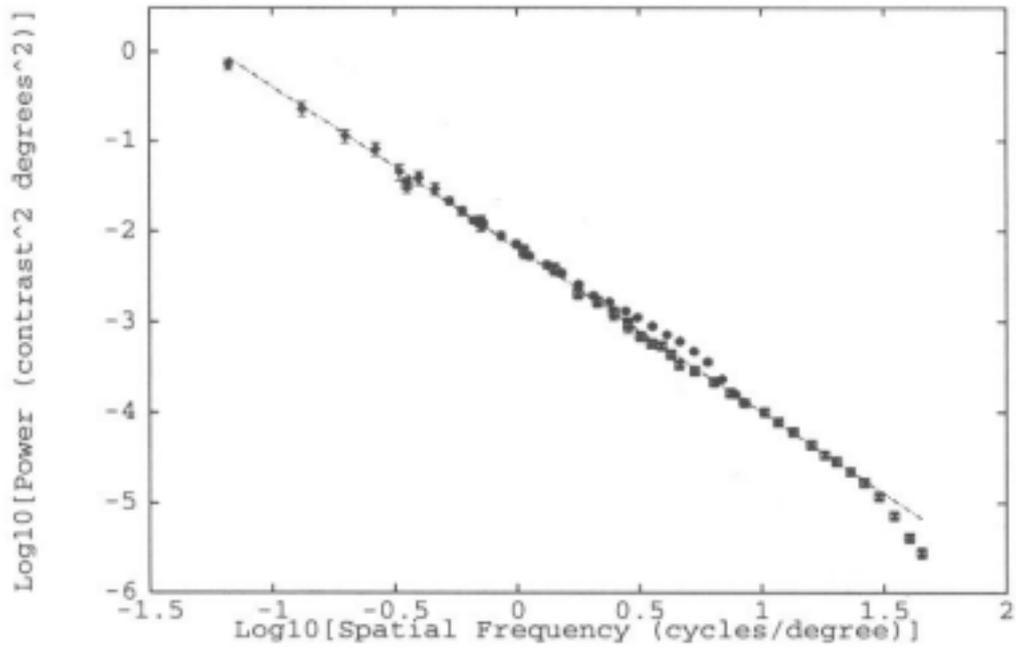


图 3: 15 毫米和 80 毫米焦距数据（重叠）的方向平均功率谱及其标准误差线，以及文中讨论的拟合回归线。

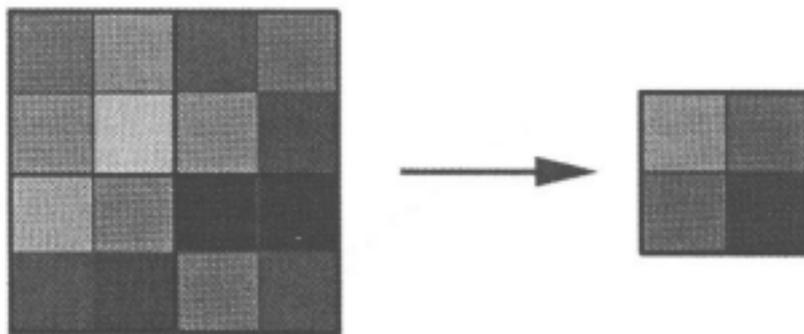


图 4: $N = 2$ 的块平均程序。

如果 $\phi(x)$ 缩放, 则 ϕ_N 的概率分布应始终具有相同的形状, 无论缩放参数 N 的值如何。图 5 演示了对数对比度分布的缩放。该图显示了 $N = 1, 2, 4, 8, 16$ 和 32 的 $P(\phi_N/\phi_N^{\text{RMS}})$ 与 $\phi_N/\phi_N^{\text{RMS}}$ 。这六个图都相互叠放; 它们具有相同的形状。高斯分布将在图中显示抛物线, 而不是我们发现的近乎线性的尾部。直方图缩放比二点函数 (即功率谱) 的缩放更有力, 因为它意味着高阶相关函数也必须缩放。

中心极限定理指出, 当大量独立随机变量 (具有有限方差) 一起平均时, 所得分布将变为高斯分布。即使在将近 1000 个 (32×32) 数据点一起平均后, 我们的直方图也不会变成高斯分布, 这一事实证明了物理学家所说的非高斯缩放不动点。简而言之, 中心极限定理不适用, 因为被平均的变量高度相关。这种扩展相关性是热力学临界点的典型特征, 其中相关性长度是无限的。

作为统计缩放和集成的非高斯特性的另一个例子, 考虑局部梯度的分布。我们首先在

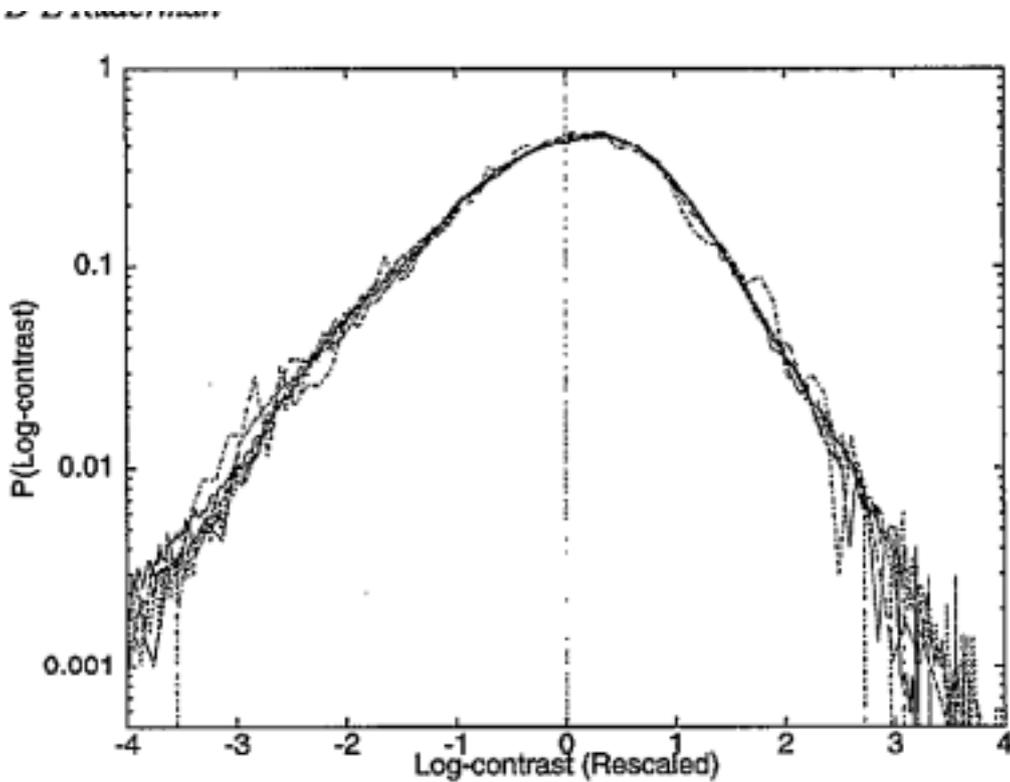


Figure 5. Scaling of log contrast histograms over scales 1, 2, 4, 8, 16, and 32.

图 5: 对数对比度直方图在 1、2、4、8、16 和 32 尺度上的缩放。

$$G_N(m, n) \approx |\nabla \phi_N(m, n)|. \quad (17)$$

如果 ϕ 是缩放场, 则 G 的直方图应具有与 N 无关的形状。图 6 显示了 $P(G_N/\overline{G_N})$ 与 $G_N/\overline{G_N}$ 的关系, 其中 $\overline{G_N}$ 是分布的平均值。如果 ϕ 是高斯分布, 则该分布将采用瑞利形式

$$P(G) = \frac{\pi}{2} \frac{G}{\overline{G}^2} \exp \left[-\frac{\pi}{4} \left(\frac{G}{\overline{G}} \right)^2 \right] \quad (18)$$

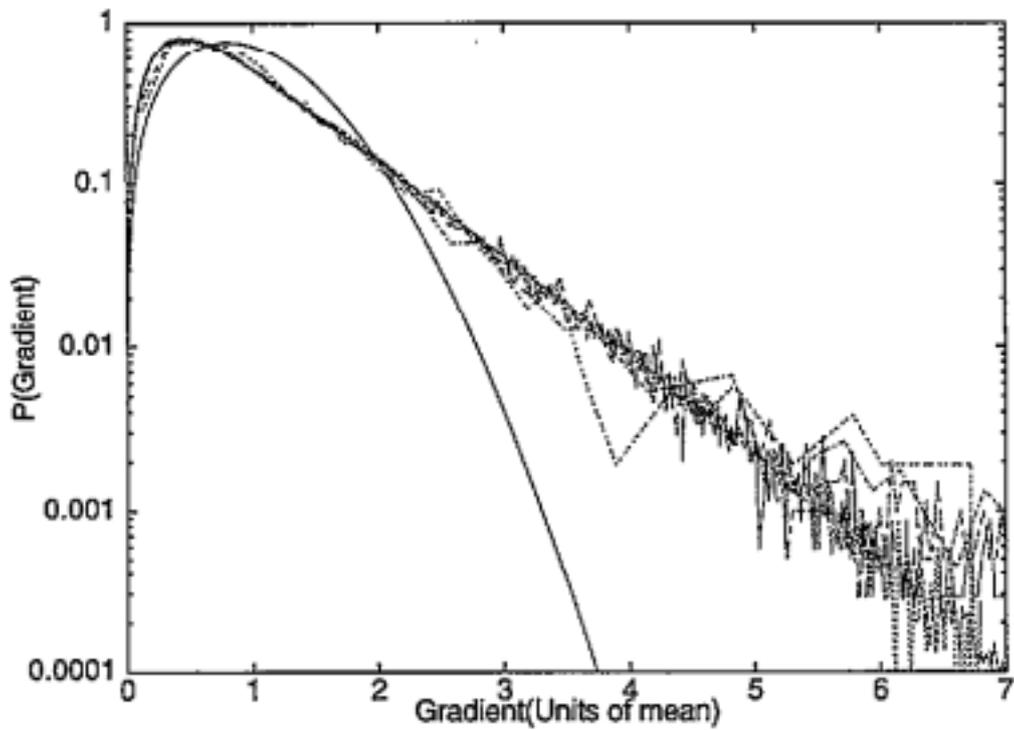


图 6: 梯度直方图的缩放。图中显示 $P(G_N/\overline{G_N})$, 其中 $N = 1, 2, 4, 8, 16, 32$, 并与瑞利分布 (实线) 进行比较。

为了进行比较，绘制在图中。

在近四十年的时间里，长度尺度从 1 到 32 的梯度分布在概率上形状相同。这种分布与瑞利分布形成了鲜明的对比。首先，梯度直方图具有非常长的指数尾部，而瑞利分布的下降幅度要大得多。这意味着图像中的大梯度区域比高斯分布要多得多。此外，小梯度或均匀斑块过多。这些是自然场景的非高斯特征。在热驱动的对流湍流中可以看到这种非常大和非常小的梯度模式，这也会导致具有指数尾部的非高斯概率分布 [21]，并显示出缩放 [67]。

缩放使我们能够用 $\alpha^\nu \phi(\alpha x)$ 替换 $\phi(x)$ ，而不会更改任何统计数据。我们已经看到，重新缩放图像不会改变对数对比度分布的形状，只会改变其宽度 ϕ_N^{RMS} 。根据缩放定律，此宽度应按 $N^{-\nu}$ 缩放。在图 7 中，我们在对数-对数刻度上绘制了 ϕ_N^{RMS} 与 N 的关系。该图呈线性，斜率为 $-\nu \approx -0.2$ 。我们测试的所有局部量都以相同的指数缩放⁵。如果像素彼此独立，则重新缩放因子将下降为 $\{N^{-1}$ ；绘制此线是为了进行比较。当图像在大尺度上平均时，自然图像中的方差仍然比白噪声中的方差大。

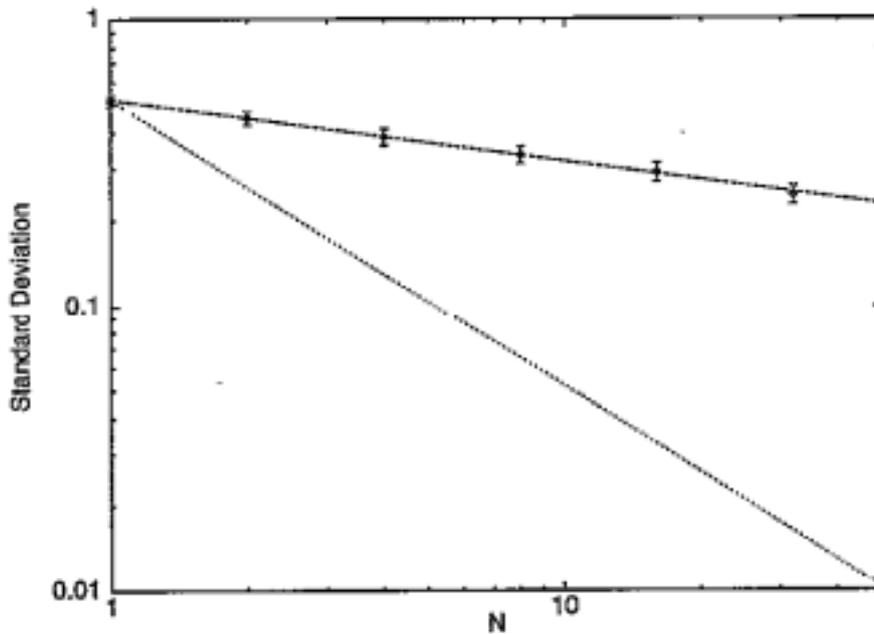


图 7: ϕ_N 分布的标准差是 $N(\log - \log \text{ plot})$ 的函数。如果图像由不相关的噪声组成，则标准差将按 $1/N$ 缩放，如虚线所示。

7. 自然图像的可预测性

图像非随机性的衡量标准之一是其所包含的可预测性的量。

假设我们知道图像的某些部分是什么样子的，并且我们想从中猜测缺失的部分是什么。这能做到

⁵精明的读者可能会想知道为什么 $n \neq 2\nu$ ，因为功率谱涉及场 4 的两个幂。这确实有点神秘，但可以通过考虑缩放可能不是完美和各向同性的事实来解决。计算功率谱时方向的平均方式与在方块上平均时方向混淆的方式不同，这可能解释了指数的差异。

吗？答案在于图像的统计结构，正如人们所预料的那样。如果它们由随机像素组成，那么就没有可预测性。但自然图像具有长程相关性，因此可以预期很大程度的可预测性。

信息论的发明者克劳德·香农 (Claude Shannon) 对统计数据很感兴趣，他想从英文文本中找出 n 个字符的界限，以及当 $n = 100$ 时，每个字母的母语熵位数。这代表单个字母的冗余度约为 75%。

1987 年，Kersten 要求人类受试者对日常图像执行类似的任务。从图像中删除一个像素，并要求受试者替换它。使用香农的方法，他将自然场景中的单个像素冗余度设为 65%。这意味着，在一个像素拥有的所有熵中，65% 是可以根据对图像其余部分的知识预测的。

我们与剑桥的 Horace Barlow 和 Chris Wroe 合作，利用自然图像对可预测性进行了一些评估。人们最终想知道一个图像块能传达多少有关远处另一个图像块的信息。这样的计算将涉及大量图像，以便很好地对分布进行采样。实际上可以完成的最多工作是一次询问几个像素。例如，一个像素能传达多少有关给定距离的另一个像素的信息？

我们计算两个像素值 ϕ_1 和 ϕ_2 之间的香农互信息

$$I = \int d\phi_1 d\phi_2 p_2(\phi_1, \phi_2) \log_2 \frac{p_2(\phi_1, \phi_2)}{p_1(\phi_1)p_1(\phi_2)} \quad (19)$$

其中 p_2 是给定间隔处两个像素的联合分布， p_1 是像素的边际分布。对于给定的位移向量 d ，扫描图像中的所有位置以创建联合概率直方图。根据此分布计算公式 19 的离散近似值。

图 8 的左图显示了两个像素之间的相互信息，它是垂直分离像素之间的距离 d 的函数。该图在对数-对数尺度上几乎呈线性，这意味着信息以分离距离的幂律形式缩放：

$$I(d, \theta) \approx d^{-\alpha(\theta)} \quad (20)$$

其中 $\alpha(\theta)$ 表示斜率对分离轴角度的依赖性。右图显示了系统各向异性与 ϕ 的关系 ($\phi = 0$ 对应于水平分离)。斜率在 $\phi = 90^\circ$ 时最小，这意味着相关性在垂直方向上最强，正如人们在包含树木的图像中所期望的那样。我们较大的图反映了冗余存在于图像的更大区域中的事实。

有趣的是，在英语文本中也发现了类似的信息尺度。Ebeling 和 Poschel 发现，两个字母之间的相互信息作为距离的函数逐渐渗入到测量中 [30]。

请注意，信息随距离缩放是非常偶然的，因为如此复杂的统计数据没有先验理由成为尺度不变的，即使图像集合是尺度不变的。在尺度不变的集合中，任何单幂相关函数都会缩放，

例如 $\langle \phi^2(0)\phi^4(x) \rangle$ 。但对于幂的混合，这并不一定正确，例如 $\langle (\phi(0)\phi(x) + \phi^2(0)\phi^2(x)) \rangle$ ，因为长度尺度的变化会通过不同的指数重新缩放每个项（参见公式 (9)），因此函数会发生变化。上面的信息度量是一个特殊量，即使它不是缩放场 ϕ 的单一幂，它也会缩放。

我们可以尝试另一种相关性度量，即二阶相关系数，其公式如下

$$r = \frac{\langle \phi_1 \phi_2 \rangle}{\sqrt{\langle \phi_1^2 \rangle \langle \phi_2^2 \rangle}} \quad (21)$$

假设 $\langle \phi_1 \rangle = \langle \phi_2 \rangle = 0$ 。事实证明，该统计数据不遵循幂律；相反，它在距离大于约 20 像素时从

The statistics of natural images

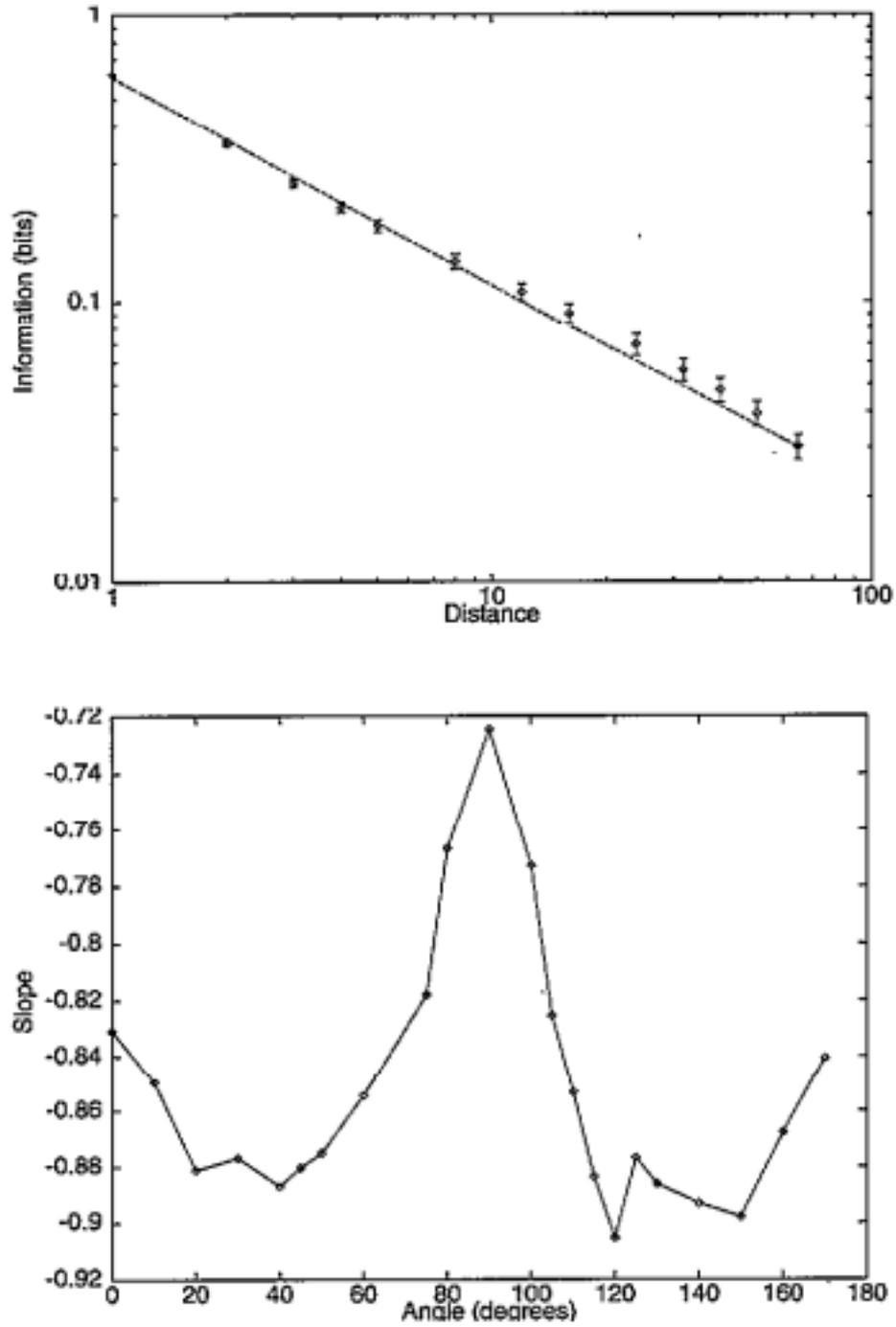


Figure 8. Top: log-log plot of $I(d, \theta)$ versus d for $\theta = 90^\circ$. Bottom: scaling ϵ .

图 8: 顶部: $I(d, \theta)$ 与 d 的对数对数图, 其中 $\theta = 90^\circ$ 。底部: $I(d, \theta)$ 与 θ 的缩放指数。

幂律转变为指数行为。由于缩放是自然图像的基本统计数据，因此香农互信息可能比二阶相关系数更“基本”地衡量相关性。

作为两点图像统计的最后一个例子，我们检查一个像素相对于另一个像素的可重构程度。从一定距离外的像素 ϕ_1 可以多好地估计出 ϕ_2 ? 我们想要一个函数 $\phi_2^{\text{est}}(\phi_1)$ ，它是给定 ϕ_1 时 ϕ_2 的最佳猜测。最小化均方误差的估计值 $\langle |\phi_2 - \phi_2^{\text{est}}(\phi_1)|^2 \rangle$ 是

$$\phi_2^{\text{est}}(\phi_1) = \int d\phi_2 \phi_2 p(\phi_2 | \phi_1) \quad (22)$$

这是给定 ϕ_1 时 ϕ_2 的条件平均值。

图 9 显示了垂直方向上距离为 1、4 和 16 像素的像素的 $\phi^{\text{est}}(\phi)$ 。

图中不同区域的相对重要性由像素概率的叠加图表示。在相关区域中，估计是线性预测，即

$$\phi_2^{\text{est}}(\phi_1) \approx m(d)\phi_1 \quad (23)$$

其中 $m(d)$ 是直线的斜率，是两个像素之间距离的函数。图 10 显示了函数 $m(d)$ 和 $F(d)$ ，重建保真度定义为

$$F = 1 - \frac{\langle |\phi^{\text{est}} - \phi|^2 \rangle}{\langle \phi^2 \rangle}. \quad (24)$$

两者均为幂律形式，指数分别为 -0.31 和 -0.72 。

基于对单个像素的了解进行重建效果并不好。基于了解垂直邻居的 RMS 预测误差几乎是 RMS 像素变化本身的 70%（即如果我们没有这样的知识，我们会得到的误差）。通过使用更多附近的像素作为最佳估计器，可以改善估计值。随着包含的像素越来越多，最佳估计器会以联合像素密度的形式考虑越来越多的图像统计数据。两点像素统计数据并没有捕捉到自然图像的特别之处。

即使没有完整的像素统计数据，我们也可以从像素的最近邻 (NN) 中找到像素的最佳线性估计量

$$\phi^{\text{est}} = \sum_{i \in \text{NN}} a_i \phi_i. \quad (25)$$

它只需要二阶相关性。系数向量 \mathbf{a} 的计算是为了最小化预期均方误差，如下所示

$$\mathbf{a} = C^{-1}\mathbf{y} \quad (26)$$

其中 $C_{ij} = \langle \phi_i \phi_j \rangle$ ， $y_i = \langle \phi \phi_i \rangle$ ，并且索引 i 和 j 覆盖 8 个最近邻居。

系数 a_i 在对角线上略微为负，在水平和垂直方向上的值约为 0.3。此重建提供的 RMS 误差约为 RMS 像素波动的 50%。包括次近邻的重建将误差降至 45%。

8. 新的不变性

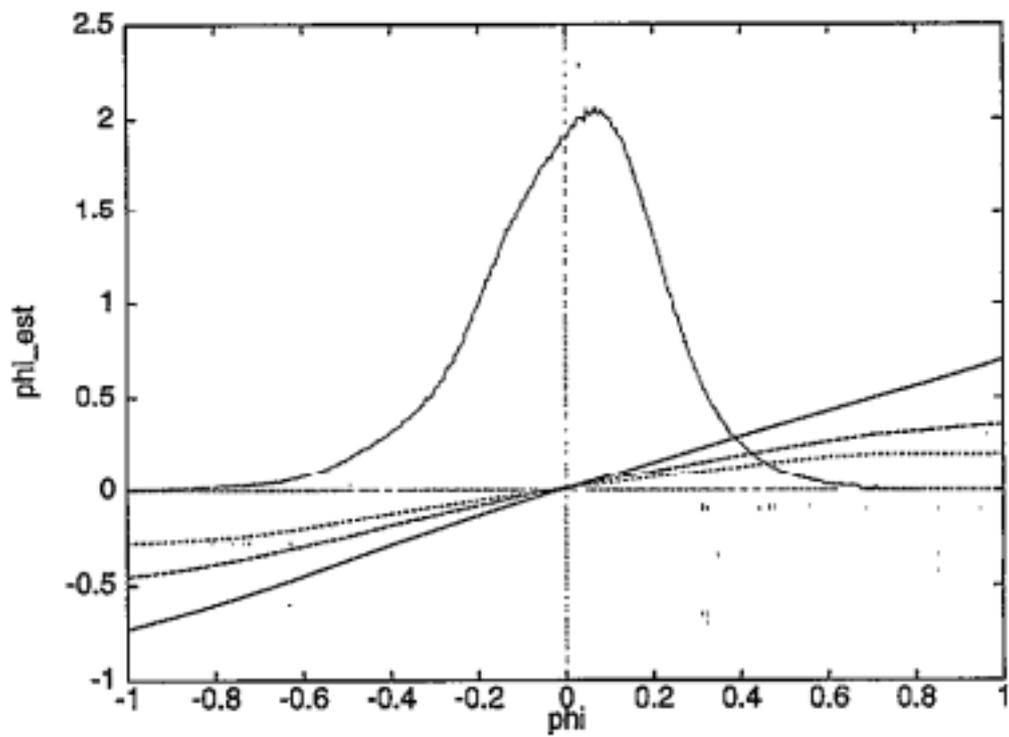


Figure 9 Estimator $\phi_{est}(\phi)$ for pixels separated vertically by 1 (solid), 4 (dashed)

图 9: 垂直方向上相隔 1 个 (实线)、4 个 (虚线) 和 16 个 (点线) 像素的像素估计量 $\phi^{est}(\phi)$ 。绘制了 ϕ 的概率分布以显示感兴趣的区域。

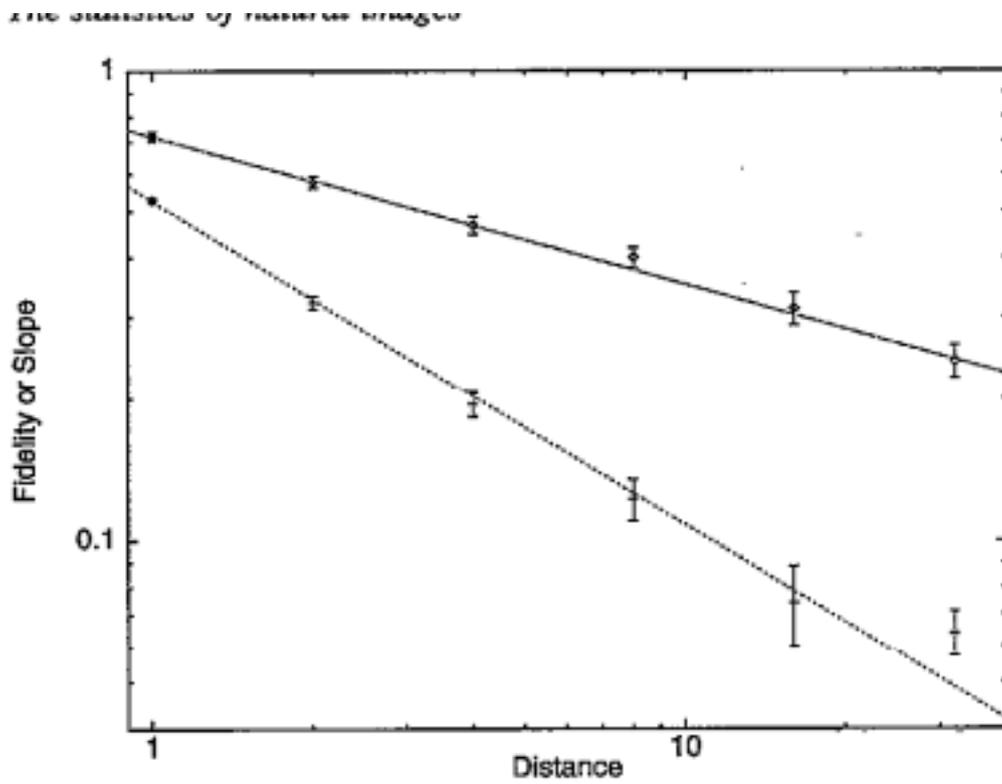


图 10: 重建保真度 F (虚线) 和线性重建斜率 m (虚线) 作为距离 d 的函数。

8.1. 过滤自然图像

我们一直在探索的图像是视觉系统处理的信号的例子。它们的统计数据告诉我们应该如何进行这种处理？视觉的早期阶段，例如视网膜中的阶段，仅限于局部处理图像——没有神经元可以访问整个图像 [29]。传递这些信号的神经元将具有由图像决定的输出统计数据。根据各种效率标准，这些通道的响应应具有某些统计特性。

例如，具有信号方差约束的通道通过发送高斯信号来优化信息传输 [76]。神经元在其函数中具有类似的约束，因为它们的发放率可以在高水平上饱和并且不能变为负值。因此，最佳编码统计数据取决于施加的约束。

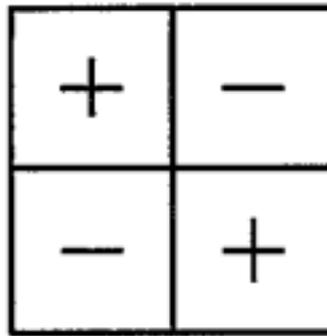


图 11: 局部带通滤波器。

首先，我们应该确定对数对比度进行简单线性过滤后得出的统计数据类型（过滤强度可得到类似的结果）。考虑图 11 中所示的过滤器。此 2×2 过滤器不会通过零空间频率的信号，因此 I_0 会消失。图 12 中以半对数刻度显示了此过滤器输出的直方图。它在四十年内具有近乎完美的指数尾部。事实上，我们尝试的任何局部线性变换（包括 Gabor 和中心环绕过滤器）似乎都会产生指数尾部直方图，尽管它们的形状可能略有不同。

Daugman [27]、Barlow 和 Tolhurst [9] 以及 Field [32] 也发现了自然场景中的长尾直方图。Field 指出，这种“稀疏编码”是自然图像的傅里叶相位的特定排列的结果。他认为，长尾只会激活编码图像的神经元中非常稀疏的子集。Burr 和 Morrone [15] 认为，图像的这种特性与自然场景中边缘的存在有关，并且“发出视觉感兴趣的信号”。从高斯分布中绘制的图像具有完全随机的相位，并且与自然场景没有任何结构相似性 [32, 70]。

单侧指数分布可最大化给定平均激活水平的信息传输，就像高斯分布对于固定方差是最优的一样。如果编码视觉刺激的神经元具有平均发放率约束，那么这些指数直方图就是理想的；我们只需对它们进行校正，使它们成为单侧的。

有没有办法消除指数分布？没有局部线性变换可以解决这个问题。但是局部非线性变换呢？一种方法就是简单地找到图像的逐点变换，产生我们想要的分布，就像 Laughlin 预测对比度编码时所做的那样 [48]。但他的方法并没有找到直方图的原因，只是重塑了它。我们更愿意寻找产生长尾的机制，并系统地“消除”它。

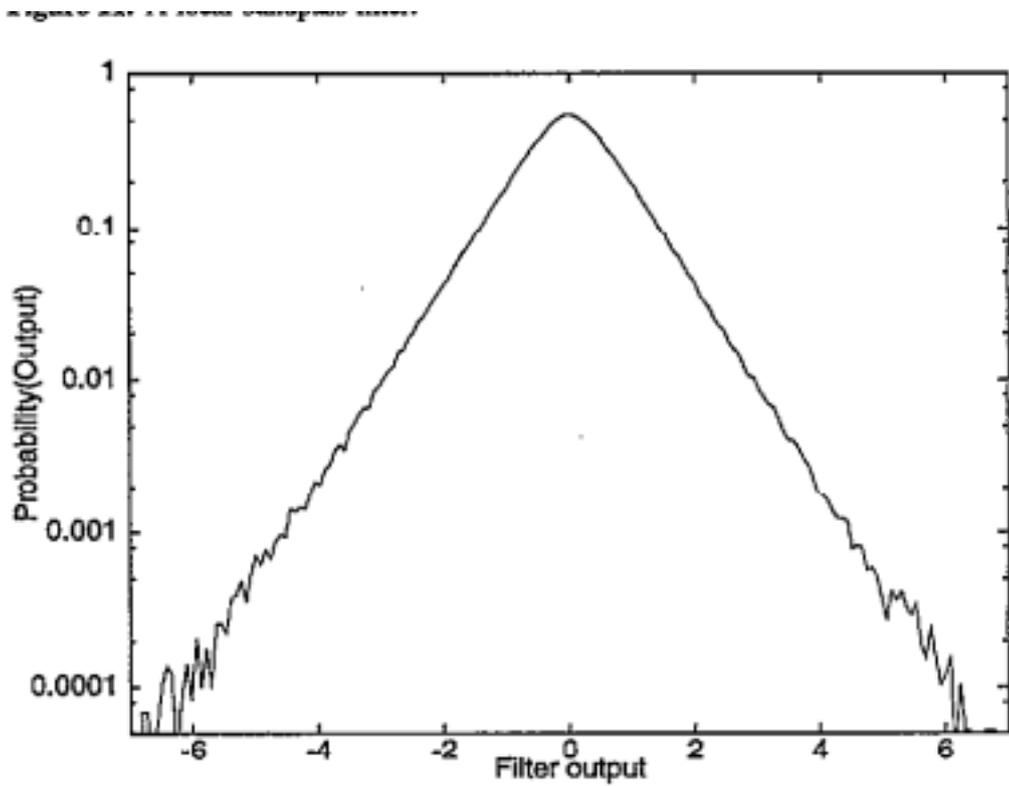


Figure 12 Histogram of the output of a local filter

图 12: 局部滤波器输出的直方图。

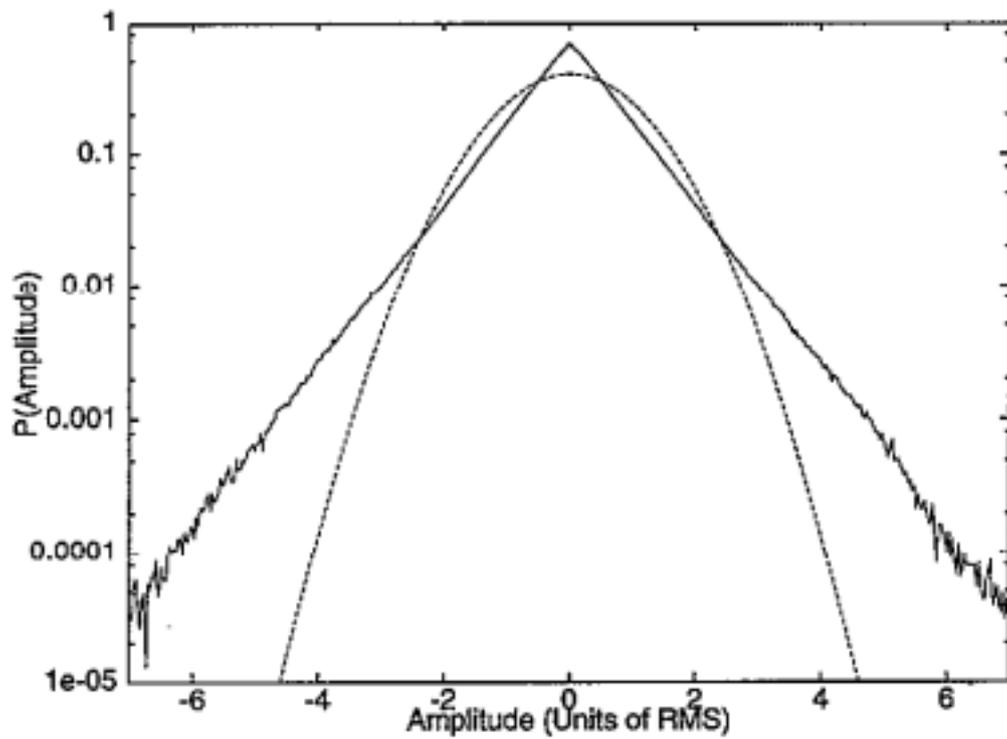


Figure 13: Histogram of amplitude of 5 min sequence from Strauss' 'The Blue Danube'.

图 13: 施特劳斯《蓝色多瑙河》中 5 分钟序列的振幅直方图, 该序列来自一张光盘录音, 采样率为 44 kHz, 使用 16 位分辨率的线性模拟数字转换器 (ADC)。高斯分布突出显示了直方图峰值和尾部的过度部分。

8.2. 添加非线性

为了寻找可能的候选者，我们应该考虑直方图尾部过多的可能原因。考虑与音乐的类比，音乐是具有与图像相似属性的合奏。

音乐声压的振幅也具有指数尾部（见图 13）。长尾部的来源是乐谱的动态；在一段时间内，有些部分声音很大，有些部分很安静。如果安静的段落被放大，而响亮的段落被减弱，那么分布的尾部和峰值处的过量将移至更“典型”的值，从而减少峰值和尾部。

这将使分布具有更“圆润”或高斯特性。也许在自然场景中也会出现类似的动态，其中局部相关区域要么纹理平坦（即安静），要么非常动态（响亮）。这表明了长指数尾部的起源：直方图是许多不同方差分布的叠加。

将声音波形除以其最近的响度是一种局部非线性运算。我们可以在图像上尝试类似的程序，通过相对于其局部标准差对对数对比度波动进行归一化。这创造了一个新的领域

$$\psi(x) = \frac{\phi(x) - \bar{\phi}(x)}{\sigma(x)} \quad (27)$$

其中 $\bar{\phi}(x)$ 是位置 a 周围 $N \times N$ 块内的局部均值， $\sigma(x)$ 是块内 $4>$ 的标准偏差。此过程的作用是从零对数对比度中去除平均位移，并对对数对比度的局部方差进行归一化。局部对比度较小的区域将扩大，而高对比度区域将变淡。分子就像一个中心环绕机制，环绕是中心（1 个像素）的 N 倍。在图 1 中的图像上运行此过程（使用 $N = 5$ ）会得到一个方差修改图像 ψ （显示在图 14 中）和一个标准偏差图像 σ （显示在图 15 中）。方差归一化图像比原始图像更加均匀。例如，岩石上的小波动已经扩展到更高的对比度。除了一些残留物体边界外，该图像几乎看起来像噪声图案。另一方面，标准偏差图像 σ 似乎几乎突出了物体的边界并减弱了物体的纹理。这很粗略地就像方差归一化将物体与其纹理分开一样。

我们发现，当 $N = 5$ 时， $\psi(x)$ 的直方图最接近高斯分布，因为其峰度最接近高斯分布 ($3\sigma^4$)。当 N 较小时，峰度较大，当 N 较大时，峰度较小。图 16 显示了 ψ 的统计数据。方差修正图像并非完全符合高斯分布，但它们显示出高斯特征： ψ 的直方图尾部迅速下降，其梯度呈瑞利分布。因此，这种新信号适合通过动态受限通道传输。

那么 $\sigma(x)$ 呢？根据方差 $\sigma^2(x)$ 生成图像，取平均值并以 5×5 块为单位进行采样，可以得到一组缩小尺寸的非负图像。我们可以通过查看对数对比度，以与原始图像数据相同的方式处理这些图像。定义

$$\xi(\mathbf{x}) = \ln \left[\frac{\sigma^2(\mathbf{x})}{\sigma_0^2} \right] \quad (28)$$

其中 σ_0^2 类似于 I_0 。 ξ 的统计数据与 ϕ 的统计数据非常相似，如图 17 所示。原始图像和方差图像的对数对比度和对数对比度梯度分布都非常相似。在子采样之前，对 ϕ 和方差图像之间的功率谱进行了比较，因此它们处于相同的空间频率尺度上。频谱的斜率在低频下完全一致（它们已被移动以在垂直方向上匹配）。在高频下，方差图像频谱下降，正如统计数据的低通性质所预期的那样。

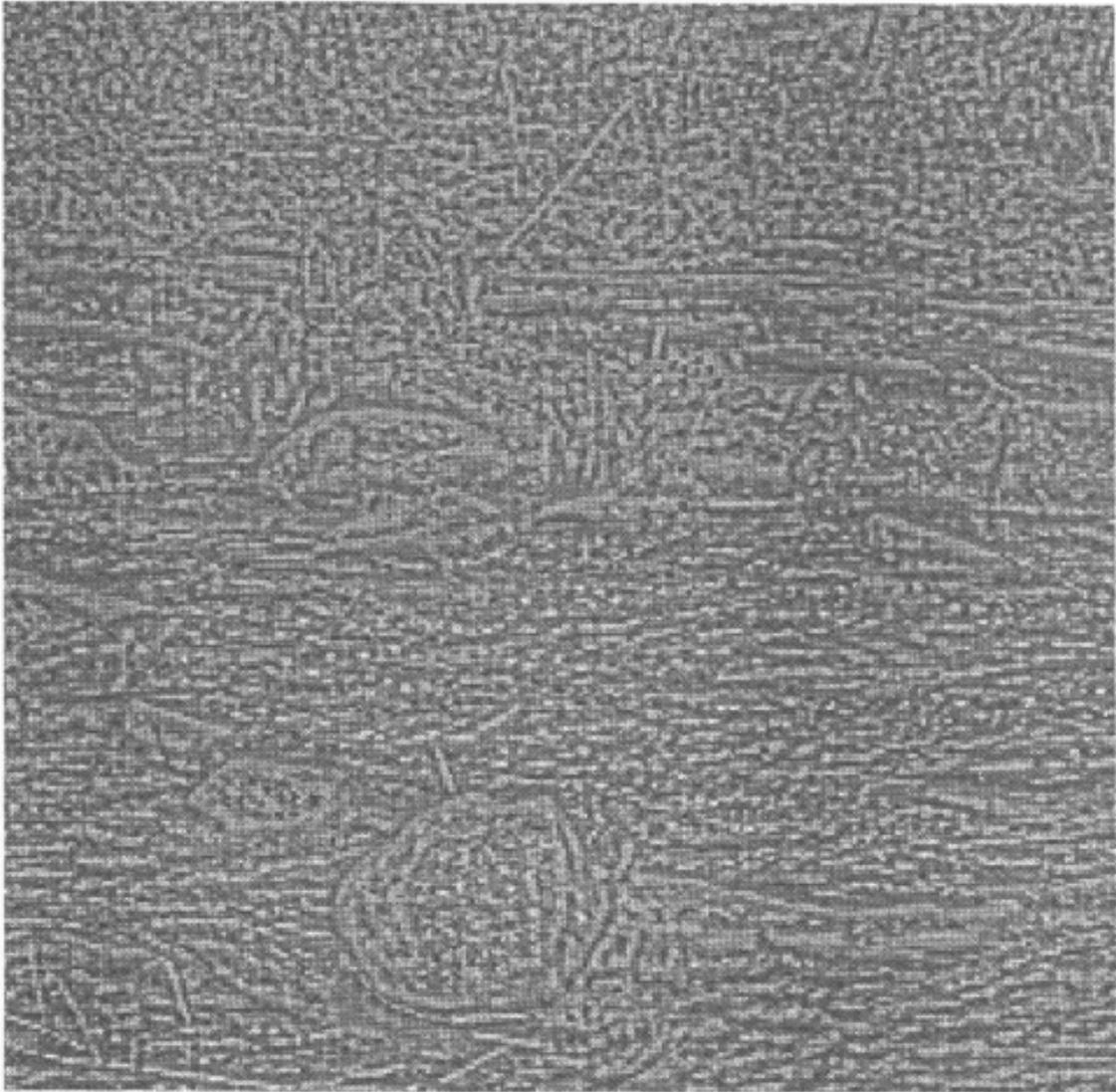


图 14: 方差修正图像, $\psi(x)$.

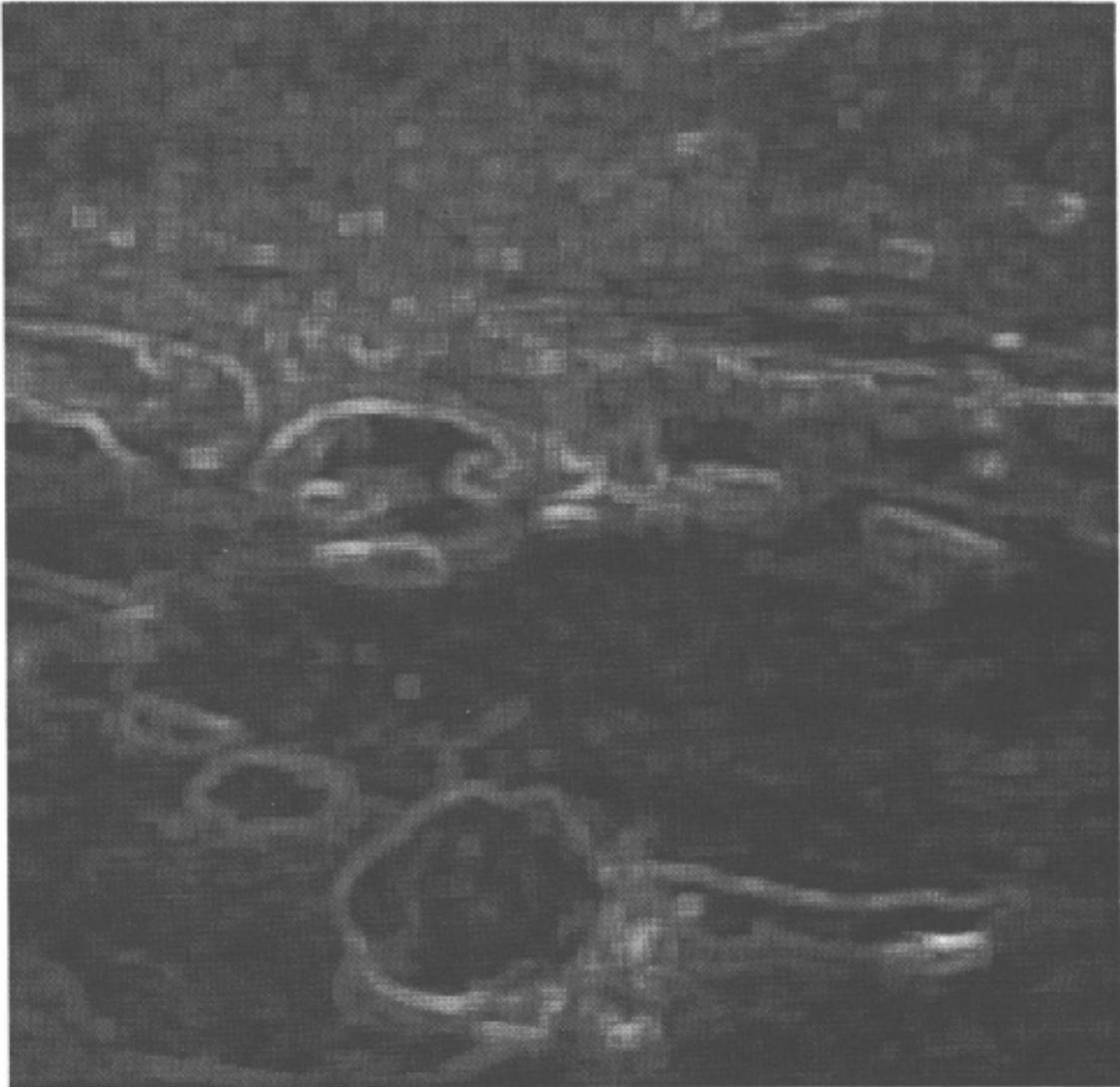


图 15: 标准差图像, $\sigma(x)$.

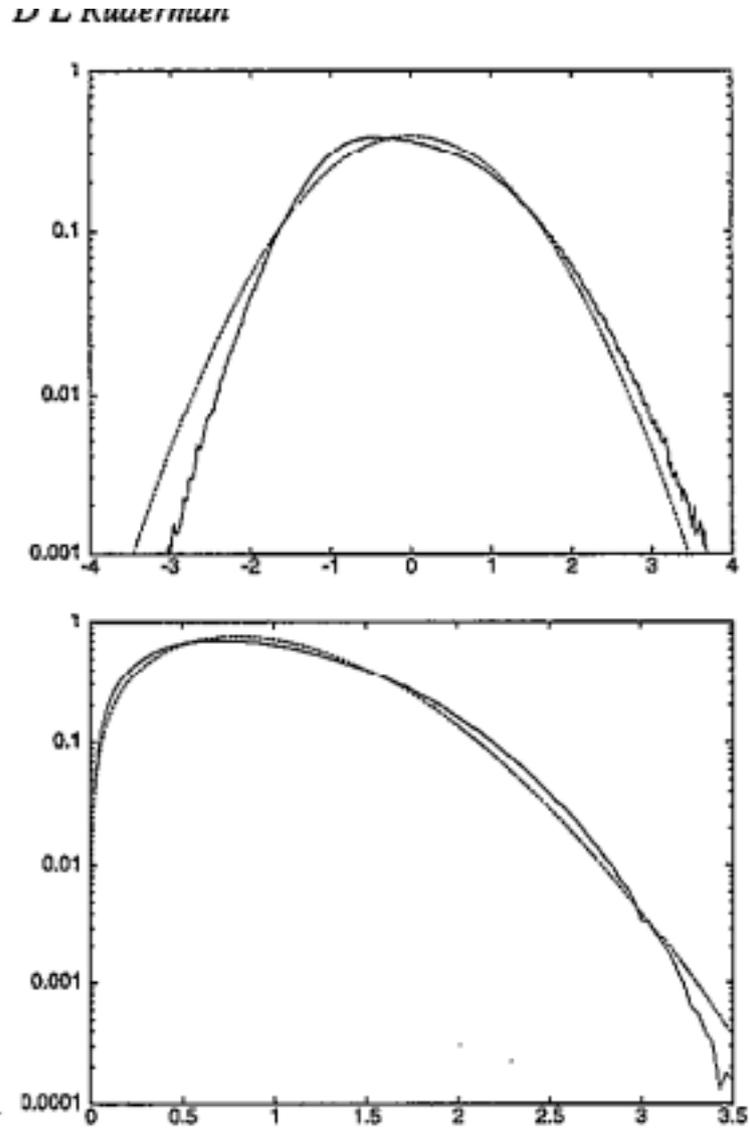


Figure 16. Semi-logarithmic plots of the statistics of variance

图 16: 方差修正图像统计的半对数图, ψ 。顶部: ψ 的直方图 (重新缩放为单位方差) 与高斯分布进行比较, 底部: ψ 的梯度直方图与瑞利分布进行比较。

8.3. 重复该过程

对于我们测量的每个统计数据，方差图像与原始图像相同。这意味着自然图像中的局部方差模式在统计上与强度模式非常相似。这是否意味着整个过程可以重复？答案是肯定的，但结果并不像以前那么清晰。从全分辨率方差图像开始，生成两个数据集 $f(x)$ 和 $E(x)$ ，它们分别是 f 的方差正则化图像和标准偏差图像（参见图 18）。通过在 11×11 块中取平均值和采样可以找到最符合高斯分布的 f 统计数据（参见图 19）。这是 5 的峰度与零相交的长度尺度。类似地，当对 5 进行采样并在 11×11 块上取平均值时，梯度分布最接近瑞利分布；如图 19 所示。该过程经过第二次迭代，产生了另一组近似高斯的数据以及一组低分辨率的方差。这些新的差异与原来的差异相比如何

$$z(x) = \ln \left[\frac{\Sigma^2(x)}{\Sigma_0^2} \right]. \quad (29)$$

其统计数据（对 11×11 个块取平均值）显示在 @fig-20 中，以及原始图像对数对比度的统计数据。同样，匹配几乎完美。请注意，这不是该过程的真正重复，因为那意味着对 δ 的缩小图像中的 5×5 个块取平均值。但后一种方法不能很好地再现统计数据。

由于缺乏数据，该过程无法合理地进行第三次。如果可以无限期地继续下去，那么我们就发现了图像集合的非线性不变性。局部方差本身具有具有相同统计数据的局部方差。因此，不仅图像强度之间存在相关性，而且局部方差之间也存在相关性。方差归一化过程类似于光谱白化，因为它消除了相关性，尽管它是分层完成的。不变性对编码有影响：可以迭代该过程以在每个阶段产生一组可以有效通信的高斯信号。最终，图像可以完全变成高斯图像。最重要的是，这一切都是通过局部处理实现的。这些结果也可能意味着可以通过反转此过程从高斯噪声中生成自然场景。实现这一点意味着已经完全理解了 $P[\phi]$ 的结构。

简而言之，我们在自然场景中发现了一种新型的迭代不变性。这种方差归一化过程让人想起了视觉皮层中的“对比度增益控制”机制 [35, 63]。它可能只是通过减少神经反应分布的尾部来限制神经反应的方差。迭代该过程会一次又一次地重现相同的统计数据，这意味着通用算法可能非常适合处理这些图像组件。

请注意，上述过程并不特别先进或复杂。我们尝试的简单方形区域可能不是产生这种效果的最佳选择——也许圆形区域或加权平均值效果会更好。此外，该技术可用于对对数对比度图像和原始强度图像进行高斯化（数据未显示）。无需对该过程进行微调，这表明方差归一化的效果既基本又稳健。

9. 视网膜中的信息

我们已经看到，自然场景在统计上与白噪声截然不同。它们结构严谨，在大范围内相互关联。在所有可能的图像空间中，它们占据的体积极小，这一概念可以用熵的概念来量化。自然场景的集合非常有限，因此其熵比白噪声集合小得多。

自然场景的熵是一个重要量，它与表示这些场景所需的“空间”大小有关。如果每幅图像都由均匀的灰色组成，那么只有一个数字可以指定整幅图像：其灰度。另一方面，如果整体真的是白噪声，那么每个像素都需要保留，因为它们完全不相关。自然图像介于两者之间。

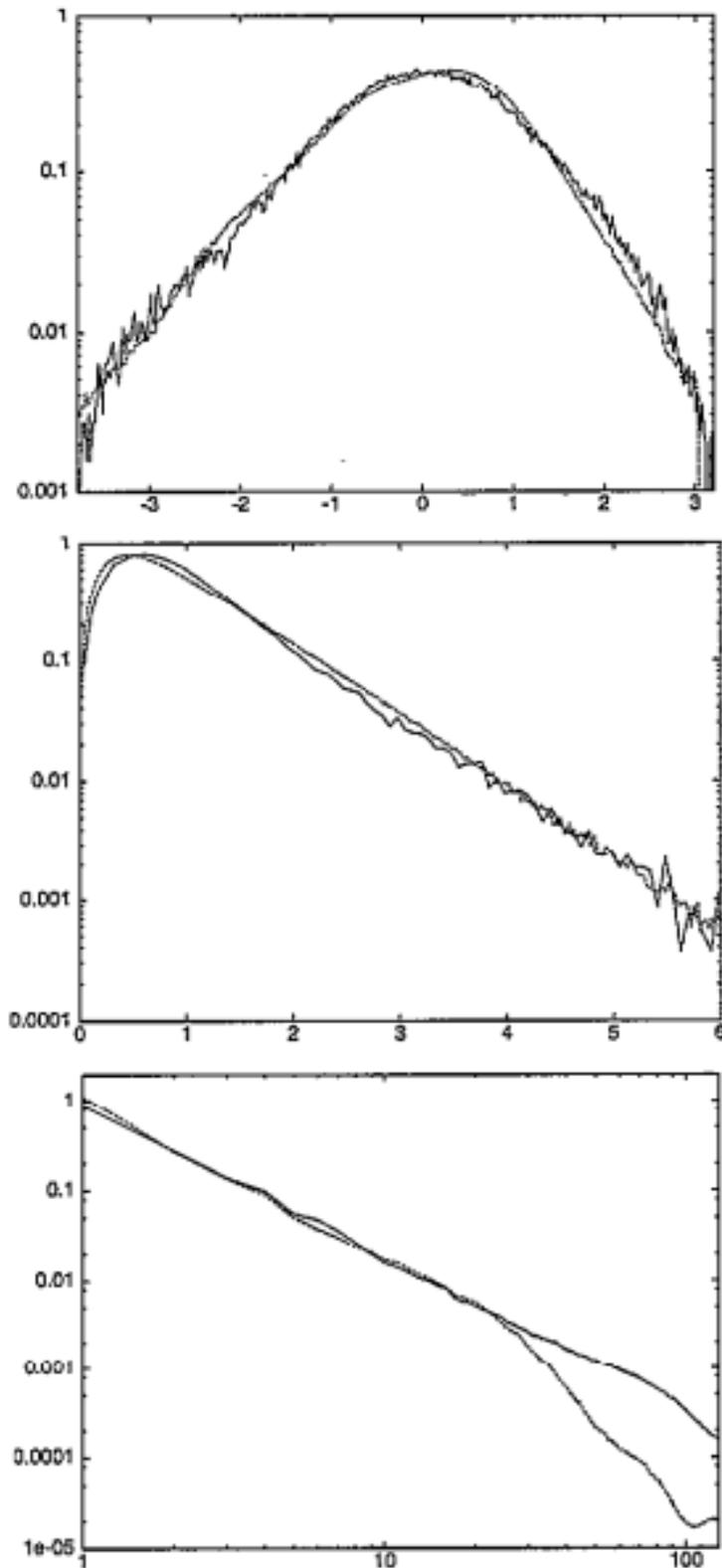
The statistics of natural images

Figure 17. Statistics of ξ images compared to the original images, ϕ .

图 17: ξ 图像与原始图像 ϕ 的统计数据对比。顶部: 对数对比直方图 (缩放至单位方差, 半对数图)。中间: 梯度直方图 (缩放至单位均值, 半对数图)。底部: 图像的功率谱 (在高频处下降) 和原始图像 (任意空间频率单位, 对数对数图)。

D. L. SILKMAN

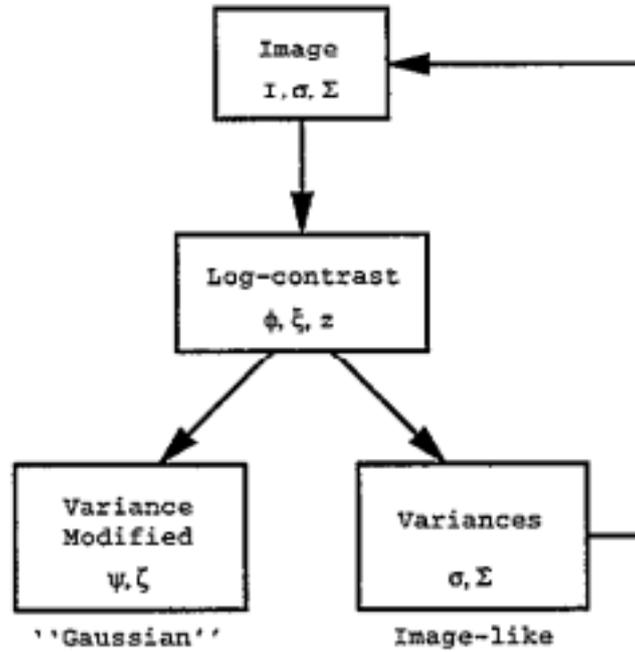


图 18: 迭代方差归一化程序

为了给熵赋值，我们必须调用包含噪声的实际图像表示；否则连续分布的熵是无限的。对于视觉，主要表示是在光感受器阵列的级别，场景首次被捕获。图像熵的问题可以表述为：“给定光感受器对图像的编码，需要什么通道容量才能将它们传输到大脑？”这个问题与视觉系统有关，因为传递自然图像是视神经的作用。每根神经纤维需要能够发送多少信息？

假设光感受器将图像表示为一组响应 $\{y_n\}$ 。

编码中的信息是

$$\begin{aligned}
 I[\{y_n\}, \phi] &= \int D\phi P[\phi] \int D\{y_n\} P[\{y_n\}|\phi] \log \left[\frac{P[\{y_n\}|\phi]}{P[\{y_n\}]} \right] \\
 &= H[\{y_n\}] - H[\phi]
 \end{aligned}
 \tag{30}$$

其中 H 是随机变量的熵。该方程的第二部分在假设光感受器噪声是附加的且与信号无关的情况下有效。我们使用线性模型进行编码⁶。首先，根据衍射限制的非相干点扩展函数，光学器件对图像进行低通滤波，其空间传递函数（大约）为 [12]

$$M(k) = 1 - |k|/k_e
 \tag{31}$$

⁶响应在图像强度中更实际地呈线性，而不是对数对比度。但是我们发现它们都具有几乎相同的功率谱，因此结果将是相同的。因此，出于此计算的目的，上面的 k 可以自由地与 I 互换。

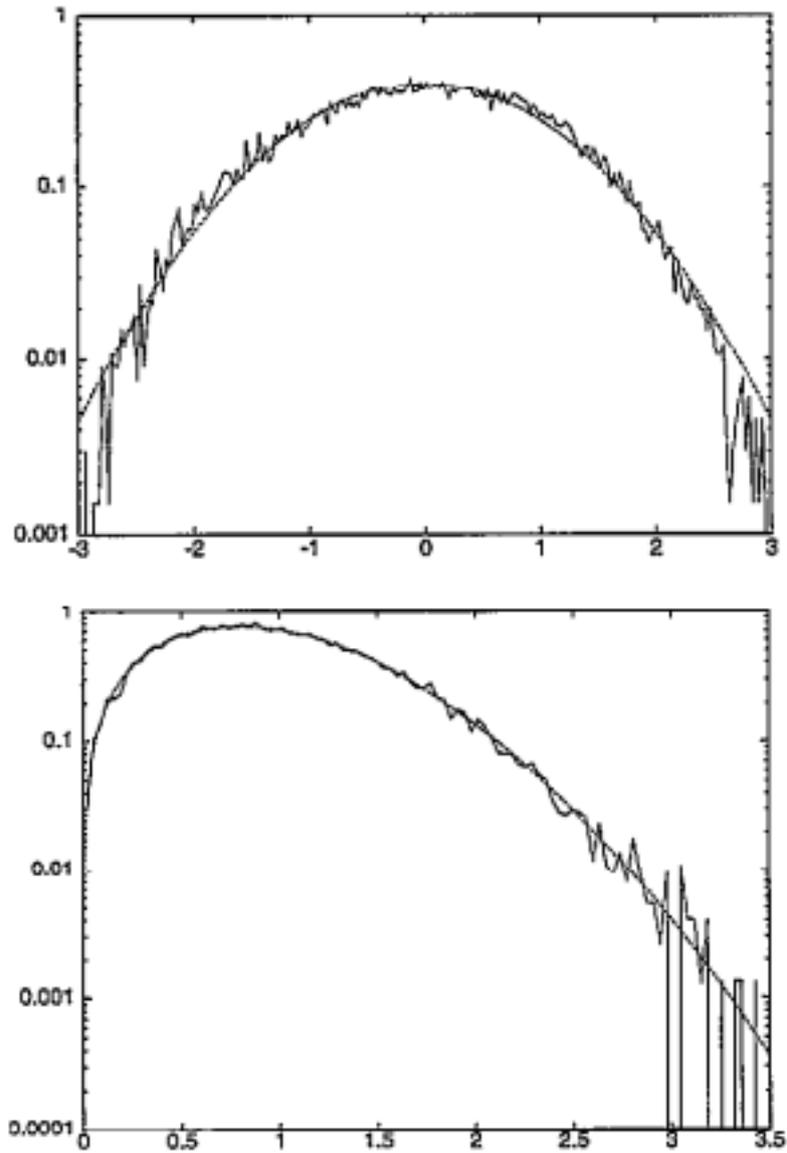


Figure 19. Statistics of ξ images compared with those of a Gaussian f

图 19: ξ 图像的统计数据与高斯场的统计数据进行比较。顶部: 直方图 (重新缩放为单位方差, 半对数图)。底部: ξ 梯度的直方图 (重新缩放为单位均值, 半对数图)。

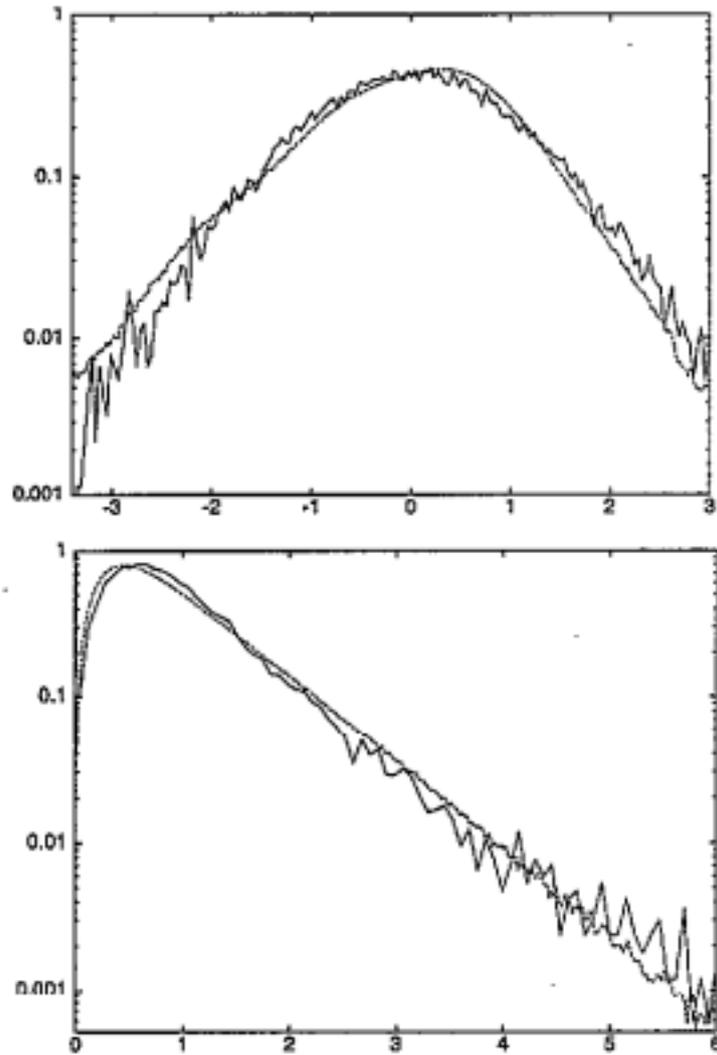


Figure 20 Statistics of z images compared with the original im

图 20: z 图像与原始图像 ϕ 的统计数据对比。顶部: 对数对比的直方图 (重新缩放为单位方差, 半对数图)。底部: z 梯度的直方图 (重新缩放为单位均值, 半对数图)。

对于 $|k| < k_e$, 否则为零。然后光感受器对该信号进行采样, 并将方差为 σ^2 的高斯白噪声添加到它们的响应中。因此

$$y_n = \int d^2x M(\omega - \omega_n)\phi(\omega) + \eta_n \quad (32)$$

其中 x_n 是第 n 个受体的位置, $\langle \eta_m \eta_n \rangle = \sigma^2 \delta_{m,n}$ 。这定义了图像编码。

每个受体的信息由下式给出:

$$\mathcal{J} = \lim_{N \rightarrow \infty} \frac{1}{N} I_N[\{y_n\}, \phi]. \quad (33)$$

可以通过假设图像(因此响应)具有比实际更大的熵来找到该数量的上限。这是通过分配具有与自然场景测量的相同功率谱的图像高斯分布来实现的。该分布具有与该功率谱一致的最大熵, 因此信息被高估了。

受体呈六边形排列(即三角形格子), 就像视网膜中央凹一样 [25]; 它们之间的距离尽可能远, 这样就不会出现混叠, 而视网膜中央凹的情况几乎就是如此。[这种结构在人类视网膜中央 1° 左右以外并不明显, 那里的情况因视杆细胞的存在而变得复杂。视锥细胞在外围变得更加稀疏和随机排列。我们的模型受体格子具有非常大的视网膜中央凹的特性。] [18]。因此, 间距为

$$a = \frac{2}{\sqrt{3}} \frac{\pi}{k_e}. \quad (34)$$

如果受体位于方形晶格上, 则无混叠条件将需要受体密度增加 13%, 因为间距恰好为 π/k_e 。六边形排列的光受体晶格中单元格的面积为 $A_c = (\sqrt{3}/2)a^2$ 。

在无限格子的极限中, 平稳高斯信号的傅里叶分量是独立的, 总信息是每个分量中信息的总和:

$$\mathcal{J} = \frac{A_c}{4\pi} \int_0^{k_e} dk \, k \log \left[1 + \frac{1}{A_c \sigma^2} |M(k)|^2 S(k) \right].$$

这里 \mathcal{J} 是每个受体的信息, A_c 是晶格中单元格的面积, σ^2 是噪声的方差。

我们使用 $S(k) \propto 1/k^{2-\eta}$, 其中 η 取其测量值, 我们用受体中的 SNR 来表示噪声水平。有一个信息度量来比较很重要。忽略受体之间的相关性会在每个受体中产生独立的高斯信号, 从而最大化给定 SNR 下的信息速率。空间冗余度以差值 $Z_{\text{ind}} - Z$ 来衡量, 它告诉我们由于空间相关性而浪费了多少信息容量。对于给定 SNR (信号方差与噪声方差之比) 的高斯信道,

$$\mathcal{J}_{\text{ind}} = \frac{1}{2} \log [1 + \text{SNR}]. \quad (35)$$

图 21 中比较了信噪比从 1 到 1000 的量。空间冗余度始终大于 2 的倍数。也许更有趣的是, 每个受体每幅图像仅传达几比特信息, 这似乎非常小。在中央凹, 大约有一条神经节细胞纤维离开眼睛到达每个受体 [24]。如果每秒有 20 张新图像呈现在眼睛上(上限), 那么每个神经节细胞需要大约 50 波特的信息容量, 这完全在某些神经元中估计的 300 波特左右的容量范围内 [69]。如果一百万左右的视神经纤维必须各自传达这么多的信息, 那么整个视神经应该以大约 50 Mbaud 的速度运行, 这是以

太网电缆容量的五倍。当然，新的随机图像不会每 50 毫秒出现一次；相反，它们随着时间的推移高度相关。因此，真实数字肯定比这个小得多，可能要小很多数量级。

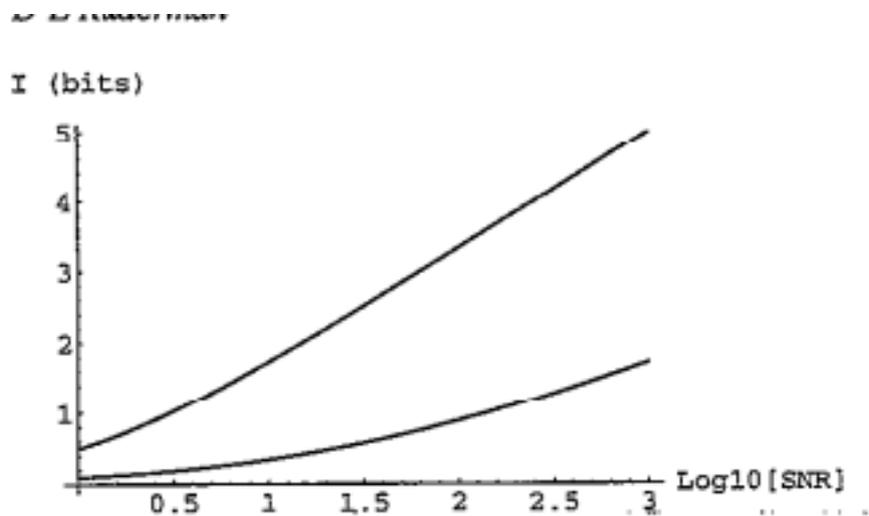


Figure 21: Lines bound to the information per image per receptor as a

图 21: 对于自然场景, $\eta = 0.19$ 的 SNR 函数表示每个受体每个图像的信息上限 (下曲线), 以及每个受体的信息容量 (上曲线)。

10. 结论

先前的研究表明, 自然场景具有尺度不变性。我们还通过功率谱、局部直方图和像素信息的缩放证明了这一点。我们的数据还展示了一种新的对称性的近似不变性: 方差归一化。局部像素方差的图像似乎具有与图像强度本身相同的统计数据。此外, 正如我们的数据所暗示的那样, 这个过程可能以递归方式进行。

理解尺度不变性的表象很简单。图像中的物体可以出现在任何距离, 这似乎是合理的, 例如, 有些物体甚至大小不一, 例如树木和岩石。当它们被捕捉到图像中时, 这意味着它们跨越许多角度尺度, 这可以产生尺度不变的集合。另一种可能性是, 物体和场景本身是自相似的, 正如分形图像生成 [66] 和分形图像压缩方法 [40] 的成功所表明的那样。

然而, 方差正则化不变性的原因并不直观。我们寻求一种去除非高斯图像统计数据的方法。同时出现了不变性。要么是我们很幸运, 要么这种类型的程序有一些根本性的东西。自然图像的一个非常显著的特性可能是它们可以通过这种迭代方案“高斯化”。随着视觉系统中存在的非线性开始被系统地研究, 了解它们对图像编码的影响非常重要。方差正则化的结果表明, 这些非线性可能是一个非常丰富的研究领域。

我们测量的统计数据是高维概率分布的低维投影。再次考虑啤酒泡沫的类比。假设中间的气泡比其他地方的气泡小, 因此那里的流体密度更大。泡沫的二维投影 (即通过对一个变量取平均值而产生的边际分布) 将显示密度的增加, 但不会显示它是由气泡组成的事实。同样, 为了直观地看到自然图像分布的真实复杂性, 可能必须在非常高的维度上进行搜索。找到分布中的不变性有助于我们降低其

固有维数。

图像统计数据本身就很有趣，而且也有实际用途。图像压缩算法通过结合图像统计数据和心理物理阈值来工作 [42]。图像恢复过程最终依赖于图像集合的先验或“贝叶斯”知识 [81]。但场景统计数据最令人兴奋的应用或许是理解视觉中的感觉处理。为了回答“生物的视觉系统设计得有多好？”这个问题，我们需要三样东西：一个优点标准、一组设计约束和自然图像统计数据。测量自然场景对于真正从生态角度理解视觉至关重要。

致谢

我要感谢我的合作者 William Bialek、Horace Barlow 和 Chris Wroe 的启发性互动。我感谢 Joseph Atick、Roland Baddeley、David Field、Simon Laughlin、Albert Libchaber、Adar Pelah 和 Yoav Tadmor 的启发性讨论。这项工作得到了 Fannie 和 John Hertz 基金会、NEC 研究所、NSF 拨款 INT-9301746 和剑桥生理实验室的部分支持。

附录

所有图片均在公园内随机位置以与眼睛齐平的高度（约 1.7 米）拍摄。由于存在携带莱姆病的鹿蜱，大多数图片均在小径上拍摄；这可能会导致图像内容出现系统性偏差。镜头设置为无限远对焦，所有照片的光圈均为 $f/5.6$ 。由于光学元件的景深有限，附近的物体会显得失焦。如果场景中存在任何物体，则相机的方位角会改变到足以将令人讨厌的物体从视野中移除。

在任何图像中，相机的仰角均不超过约 $+10^\circ$ 。

使用配备 9.5-123.5 毫米变焦镜头的 Sony Mavica MVC-5500 静态视频 CCD 相机收集图像。该设备将模拟视频帧写入小磁盘，然后从播放单元（Sony MVR-6500）读取。视频信号为 NTSC 格式 RGB。这三个信号（红色、绿色和蓝色）被数字化为 8 位（0-255）使用 Silicon Graphics VideoLab 板。为了减少模拟播放噪音的影响，我们对 32 帧捕获的结果取平均值，以产生 0-255 范围内的浮点数。请注意，由于模拟噪音在数字化之前就存在于系统中，因此重复量化后再取平均值可以让我们避开量化噪音。

我们使用 CIE 亮度作为信号。它是使用公式 [62] 从数据中得出的。

$$Y = 0.59G + 0.3R + 0.11B \quad (\text{A1})$$

该量是根据已知反射率的灰卡校准的，以产生强度信号 $I(Y)$ ，该信号与通过镜头的光源通量成比例。由于相机的动态范围有限，因此可能会出现饱和。我们丢弃了饱和像素超过 1% 的图像。饱和的像素要么设置为最低校准亮度值，要么设置为最高亮度值。

参考文献

- [1] Arps R B and Truong T K 1994 Comparison of international standards for lossless still image compression Proc. IEEE 82 889-99
- [2] Atick J J 1992 Could information theory provide an ecological theory of sensory processing? Network 3 213-51
- [6] Attneave F 1954 Some informational aspects of visual perception Psychol. Rev. 61 183-93
- [7] Barlow H B 1961 Possible principles underlying the transformation of sensory messages Sensory Communication ed W A Rosenblith (Cambridge, MA: MIT Press) p 217
- [8] Barlow H B 1982 What causes trichromacy? A theoretical analysis using comb-filtered spectra Vision Res. 22 635-43
- [9] Barlow H B and Tolhurst D J, Private communication
- [10] Bialek W, Ruderman D L and Zee A 1991 Optimal sampling of natural images: A design principle for the visual system? Advances in Neural Information Processing Systems 3 ed D Touretzky and J Moody (San Mateo, CA: Morgan Kaufmann) pp 363-9
- [11] Blakemore C B and Cooper G 1970 Development of the brain depends on the visual environment Nature 28 477-8
- [12] Born M and Wolf E 1989 Principles of Optics (Oxford: Pergamon) 6th edn
- [13] Brillinger D R 1981 Time Series: Data Analysis and Theory (San Francisco: Holden-Day)
- [14] Buchsbaum G and Gottschalk A 1983 Trichromacy, opponent colour coding and optimum colour information transmission in the retina Proc. R. Soc. B 220 89-113
- [15] Burr D C and Morrone M C 1990 Feature detection in biological and artificial visual systems Vision: Coding and Efficiency ed C Blakemore (Cambridge: Cambridge University Press)
- [16] Burton G J and Moorhead I R 1987 Color and spatial structure in natural scenes Appl. Opt. 26 157-70 [17] Campbell F W and Green D G 1965 Optical and retinal factors affecting visual resolution J. Physiol. 181 576-93
- [18] Campbell F W and Gubisch R W 1966 Optical quality of the human eye J. Physiol. 186 558-78
- [19] Cardy J L 1986 conformal invariance and critical behavior Statistical Physics ed H E Stanley (Amsterdam: North-Holland) pp 219-241.
- [20] Carlson C R 1978 Thresholds for perceived image sharpness Photographic Sci. Eng. 22 69-71
- [21] Castaing B, Gagne Y and Hopfinger E J 1990 Velocity probability density functions of high Reynolds number turbulence Physica 46D 177-200
- [22] Chittka L and Menzel R 1992 The evolutionary adaptation of flower colours and the insect pollinators' colour vision J. Comp. Physiol. A 171 171-81

- [23] Cohen R W, Gorog I and Carlson C R 1975 Image descriptors for displays Technical Report Cont. No NOOO14-74-C-0184, Office of Naval Research
- [24] Curcio C A and Allen K A 1990 Topography of ganglion cells in human retina *J. Compar. Neurol.* 300 5-25
- [25] Curcio C A and Sloan K R 1992 Packing geometry of human cone photoreceptors: Variation with eccentricity and evidence for local anisotropy *Visual Neurosci.* 9 169-80
- [26] Dannemiller J L 1992 Spectral reflectance of natural objects—how many basis functions are necessary? *J,Opt. Soc. Am. A* 9 507-15
- [27] Daugman J G 1988 Complete discrete 2D Gabor transforms by neural networks for image analysis and compression *IEEE Trans. Acoust., Speech, Signal Process.* 36 1169-79
- [28] Deriugin N G 1956 The power spectrum and the correlation function of the television signal *Telecommunications* 1 (7) 1-12
- [29] Dowling J E 1987 *The Retina: An Approachable Part of the Brain* (Cambridge, MA: Belknap Press of Harvard University)
- [30] Ebeling W and Pöschel T 1994 Entropy and long-range correlations in literary English *Europhys. Lett.* 26 241-46
- [31] Field D J 1987 Relations between the statistics of natural images and the response properties of cortical cells *J. Opt. Soc. Am.* 4 2379
- [32] Field D J 1994 What is the goal of sensory coding? *Neural Comput.* 6 559-601 [33] Geman S and Geman D 1984 Stochastic relaxation. Gibbs distribution, and the Bayesian restoration of images *IEEE Trans. Patt. Anal. Mach. Intel. PAMI-6* 721-41
- [34] Hancock P J B, Baddeley R J and Smith L S 1992 The principal components of natural images *Network* 3 61-70
- [35] Heeger D J 1992 Normalization of cell responses in cat striate cortex *Vis. Neurosci.* 9 181-97
- [36] Hirsch H V B and Spinelli D N 1971 Modification of the distribution of receptive field orientation in cats by selective visual exposure during development *Exp. Brain Res.* 13 509-27
- [37] Hsui K J and Hsui A J 1990 Fractal geometry of music *Proc. Natl Acad. Sci. USA* 87 938-41 [38] Hughes A 1977 The topography of vision in mammals of contrasting life style: comparative optics and retinal organisation *Handbook of Sensory Physiology* vol VII/5, ed F Crescitelli (Berlin: Springer) pp 613-756
- [39] Itzykson C and Zuber J-B 1980 *Quantum Field Theory* (New York: McGraw-Hill) [40] Jacquin A E 1993 Fractal image coding: a review *Proc. IEEE* 81 1451-465
- [41] Jain A K 1989 *Fundamentals of Digital Image Processing* (New York: Prentice Hall) pp 480-83 [42] Jayant N, Johnston J and Safranek R 1993 Signal compression based on models of human perception *Proc.IEEE* 81 1385-422

- [43] Jeng F-C and Woods J W 1991 Compound Gauss-Markov random fields for image estimation *IEEE Trans. Signal Process.* 38 683-97
- [44] Kadanoff L P 1966 Scaling laws for ising models near T_c *Physics* 2 263-72.
- [45] Kretzmer E R 1952 Statistics of television signals *Bell Syst. Tech. J.* 31 751-631.
- [46] Krinov E L 1947 Spectral reflectance properties of natural formations Technical Report Technical Translation T-439. National Research Council of Canada
- [47] Land M F 1990 The design of compound eyes *Vision: Coding and Efficiency* ed C Blakemore (Cambridge: Cambridge University Press) pp 55-64
- [48] Laughlin S B 1981 A simple coding procedure enhances a neuron's information capacity *Z. Naturforsch.* 36e 910-2
- [49] Laughlin S B 1992 Retinal information capacity and the function of the pupil *Ophthalm. Physiol. Opt.* 12 161-4
- [50] Laughlin S B 1994 Matching coding, circuits, cells, and molecules to signals-general principles of retinal design in the fly's eye *Prog. Retinal Eye Res.* 13 165-96.
- [51] Lenz R 1990 Group invariant pattern recognition *Patt. Recogn.* 23 199-217 [52] Lenz R 1994 Statistical properties of TV-signals and approximations of the Karhunen-Loève-transform Technical Report LiTH-ISY-R-1617, Department of Electrical Engineering, Linköping University, Sweden
- [53] Li Z and Atick J J 1994 Efficient stereo coding in the multiscale representation *Network* 5 157-74
- [54] Li Z and Atick J J 1994 Toward a theory of the striate cortex *Neural Comput.* 6 127-46
- [55] Linsker R 1989 How to generate ordered maps by maximizing the mutual information between input and output signals *Neural Comput.* 1 402-11
- [56] Linsker R 1990 Perceptual neural organization: Some approaches based on network models and information theory *Ann. Rev. Neurosci.* 13 257-81
- [57] Linsker R 1994 Sensory processing and information theory *From Statistical Physics to Statistical Inference and Back* ed P Grassberger and J-P Nadal (Dordrecht: Kluwer) pp 237-247
- [58] Lythgoe J N 1979 *The Ecology of Vision* (Oxford: Oxford University Press)
- [59] Maloney L T 1986 Evaluation of linear models of surface spectral reflectance with small numbers of parameters *J. Opt. Soc. Am. A* 3 1673-83
- [60] Mollon J D and Bowmaker J K 1992 The spatial arrangement of cones in the primate fovea *Nature* 360 677-9
- [61] Movshon J A and Van Sluyters R C 1981 Visual neural development *Ann. Rev. Psychol.* 32 477-522.
- [62] Netravali A N 1988 *Digital Pictures: Representation and Compression* (New York:

Plenum)

[63] Ohzawa I, Sciar G and Freeman R D 1985 Contrast gain control in the cat's visual system *J. Neurophysiol.* 54 651-67

[64] Osorio D and Bossomaier T R 1992 Human cone-pigment spectral sensitivities and the reflectances of natural surfaces *Biol. Cybern.* 67 217-22.

[65] Papoulis A 1991 *Probability, Random Variables, and Stochastic Processes* (New York: McGraw-Hill) 3rd edn

[66] Peitgen H and Saupe D (ed) 1988 *The Science of Fractal Images* (Berlin: Springer)

[67] Procaccia I 1984 Fractal structures in turbulence *J. Stat. Phys.* 36 649-64.

[68] Rieke F, Owen W G and Bialek W 1991 Optimal filtering in the salamander retina *Advances in Neural Information Processing Systems* 3 ed D Touretzky and J Moody (San Mateo, CA: Morgan Kaufmann) p 377-83

[69] Rieke F M, Warland D and Bialek W 1993 Coding efficiency and information rates in sensory neurons *Europhys. Lett.* 22 151-56

[70] Ruderman D L 1993 *Natural ensembles and sensory signal processing* PhD Thesis University of California, Berkeley

[71] Ruderman D L 1994 Designing receptive fields for highest fidelity *Network* 5 147-55

[72] Ruderman D L and Bialek W 1992 Seeing beyond the Nyquist limit *Neural Comput.* 4 682-90

[73] Ruderman D L and Bialek W 1994 Statistics of natural images: Scaling in the woods *Advances in Neural Information Processing Systems* 6 ed J D Cowan, G Tesauro and J Alsppector (San Mateo, CA: Morgan Kaufmann)

[74] Ruderman D L and Bialek W 1994 Statistics of natural images: scaling in the woods *Phys. Rev. Lett.* 73

[75] Sanger T D 1989 Optimal unsupervised learning in a single-layer linear feedforward neural network *Neur.*

Networks 2 459-473

[76] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* 27 379
s

[77] Shannon C E 1951 Prediction and entropy of printed English *Bell Syst. Tech. J.* 30 50-64 [78] Snyder A W 1977 Acuity of compound eyes: Physical limitations and design *J. Comp. Physiol.* 116 161-824.

[79] Snyder A W, Stavenga D G and Laughlin S B 1977 Spectial information capacity of compound eyes *J. Comp. Physiol.* 116 183-207

[80] Srinivasan M V, Laughlin S B and Dubs A 1982 Predictive coding: A fresh view of

inhibition in the retina Proc. R. Soc. B 216 427-59

[81] Stark H (ed) 1987 Image Recovery: Theory and Application (New York: Academic)

[82] Tolhurst D J, Tadmor Y and Chao T 1992 Amplitude spectra of natural images Ophthal. Physiol. Opt. 12 29229

[83] van Hateren J H 1992 Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation J. Comp. Physiol. A 171 157-70

[84] van Hateren J H 1992 A theory of maximizing sensory information Biol. Cybern. 68 23-9 [85] Voss R F and Clarke J 1978 1/f noise in music; Music from 1/f noise J. Acoust. Soc. Am. 63 258-63 [86] Webber C J S 1991 Competitive learning, natural images and cortical cells

Network 2 169-87

[87] Webster R J 1993 Ambient noise statistics IEEE Trans. Sig. Proc. 41 2249-53

[88] Wiener N 1949 Time Series (Cambridge, MA: MIT Press)

[89] Wilson K G 1983 The renormalization group and critical phenomena Rev. Mod. Phys. 55 583

Affiliation:

Daniel L Rudermant⁷

E-mail: dlr1002@cus.cam.ac.uk

翻译: 林绪虹⁸

E-mail: linxuhong@yahoo.com

Silkman Statistical Journal

published by the Funny Project of Silkman Press

May 2024, Volume 99, Issue 2

[doi:10.18637/ssj.2024.test](https://doi.org/10.18637/ssj.2024.test)

<http://cookwhy.com/>

<http://cookwhy.com>

Submitted: 2024-04-12

Accepted: 2024-05-12

⁷Rudermant Physiological Laboratory, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

⁸软件工程师, 数学史爱好者, 本文基于原作者 2003 年发表于 NETWORK: COMPUTATION IN NEURAL SYSTEMS 的原文翻译, 翻译时间 2024 年 4 月)