

Scoot: 一种用于人脸素描的感知指标

范登平^{1,2} 张声传⁴ 吴宇寰¹ 刘云¹ 程明明¹ 任博¹ Paul L. Rosin³ 纪荣嵘^{4,5}

¹ 南开大学 ² 起源人工智能研究院 (IIAI) ³ 卡迪夫大学 ⁴ 厦门大学 ⁵ 鹏城实验室

<http://mmcheng.net/scoot/>

摘要

人类的视觉系统具有快速评估两幅人脸素描图像之间感知相似度的能力。然而，目前广泛使用的面部素描指标，比如，SSIM 和 FSIM 的设计初衷是评估图像的局部畸变，而在处理两张人脸图像之间的相似度时常常失效。在本文中，我们设计了一个名为结构共生纹理 (Scoot¹) 感知指标，这种度量指标同时考虑了块级别的空间结构和共生纹理统计。为了测试所提指标的准确度，我们基于多个可靠属性提出了 3 种新的元度量方式。大量的实验证实，我们的 Scoot 指标性能上超过了先前的指标。此外，我们构建了第一个最大规模 (152K 次判断) 的基于人类感知的素描数据库。这个数据库可以用来评估一个度量指标与人类感知一致的程度。我们的测试结果表明，“空间结构”和“共生纹理”是人脸素描合成中两个普遍适用的感知特征。

1. 引言

度量指标能够根据参考数据正确地评判不同算法的输出 [81] 质量，这对于计算机和图像处理领域 [5, 8, 90] 是至关重要的。对于各种终端用户应用，比如，人脸素描 [50]，图像风格转换 [28]，图像质量评价 [67]，显著性目标检测 [11–13, 91]，分割 [42–44]，疾病分类 [74]，图像去噪 [72] 等等，这类比较本质上上是计算“感知距离”，也就是用一种和人类感知高度相关的方法评价两幅图像之间的相似度。

在本文中，我们研究了人脸素描并且说明了人类的判断经常不同于目前的评价指标。我们根据人类的

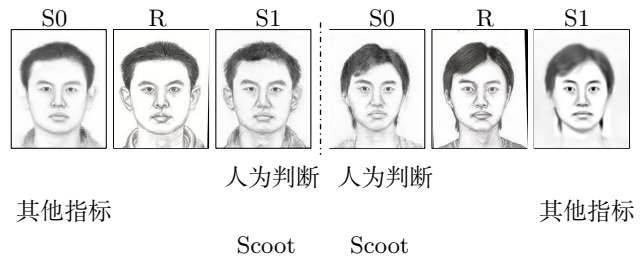


图 1: 哪张合成的素描图像和中间的素描更近似? 对于右边的例子，对照参考素描 (R)，素描 0 (S0) 在结构和纹理方面比素描 1 (S1) 更相似。素描 1 几乎完全破坏了头发的纹理。现有广泛使用 (SSIM [67], FSIM [77]), 经典的 (IFC [41], VIF [40]) 和以及最近的 (GMSD [73]) 指标都于认为判断不一致。只有本文的 Scoot 指标和人为判断一致。

选择原则为素描提供了一种全新的感知距离，这是该领域的首次尝试。正如在 [81] 中提到，人类对于相似性的判断取决于高阶的图像结构。人脸素描是由许多纹理组成，并且已经有很多素描合成算法，这就为我们研究这个问题提供了很好的基础。然而，在评价人脸素描图像时，设计一个好的感知指标应该考虑人类感知因素，例如：

- 和人的感知相匹配，以便使高质量的素描可以直接用于各种主观应用，如，执法和娱乐。
- 对轻微的不匹配 (例如，缩放、旋转) 不敏感，因为艺术家绘画的现实世界的素描并不能精确地将每个像素与原始照片匹配。
- 具有能够捕捉整体内容的能力，即青睐完整的素描，而不是只有笔画的素描 (失去面部一些要素)。

据我们所知，之前的度量指标还不能同时满足上述所有性质。

¹本文为 ICCV2019 [14] 的中文翻译版。程明明为本文的通讯作者 (cmm@nankai.edu.cn)

序号.	模型	发表年份	Sj.	Rr.	Ob.	No.	模型	发表年份	Sj.	Rr.	Ob.
1	ST	[50]	03'ICCV	VRR		2	STM	[51]	04'TCSVT	VRR	
3	LLE	[32]	05'CVPR	VRR		4	BTI	[33]	07'IJCAI		RMSE
5	E-HMMI	[23]	08'NC	VRR	UIQI	6	EHMM	[22]	08'TCSVT	VRR	
7	MRF	[65]	08'PAMI	VRR		8	SL	[71]	10'NC	VRR	UIQI
9	RMRF	[87]	10'ECCV	VRR		10	SNS-SRE	[21]	12'TCSVT	VRR	
11	MWF	[92]	12'CVPR	VRR		12	SCDL	[61]	12'CVPR		PSNR
13	Trans	[58]	13'TNNLS	VRR		14	SFS-SVR	[57]	13'PRL	VRR	VIF
15	Survey	[59]	14'IJCV		RMSE, UIQI, SSIM	16	SSD	[46]	14'ECCV	SV	
17	SFS	[84]	15'TIP	VRR	FSIM, SSIM	18	FCN	[76]	15'ICMR	ES	
19	RFSSS	[82]	16'TIP	VRR	FSIM, SSIM	20	KD-Tree	[89]	16'ECCV		VIF, SSIM
21	MrFSPS	[37]	16'TNNLS	VRR	FSIM, VIF, SSIM	22	2DDCM	[52]	16'TIP	VRR	FSIM, SSIM
23	RR	[60]	17'NC	VRR	VIF, SSIM	24	Bayesian	[56]	17'TIP	VRR	VIF, SSIM
25	RFSSS	[83]	17'TCSVT	VRR	FSIM, SSIM	26	S-FSPS	[36]	17'TCSVT	VRR	FSIM, VIF, SSIM
27	ArFSPS	[30]	17'NC	VRR	FSIM	28	BFCN	[75]	17'TIP	SV	
29	DGFL	[93]	17'IJCAI	VRR	SSIM	30	FreeH	[31]	17'IJCV	SV	
31	Pix2pix	[28]	17'CVPR			32	CA-GAN	[19]	17'CVPR		SSIM
33	ESSEFA	[15]	17'TOG			34	PS ² MAN	[53]	18'FG		FSIM, SSIM
35	NST	[39]	17'NPAR			36	CMSG	[78]	18'TC	SV	
37	RSLCR	[54]	18'PR	VRR	SSIM	38	MRNF	[79]	18'IJCAI		VIF, SSIM
39	ρ GAN	[85]	18'IJCAI		FSIM	40	FSSN	[29]	18'PR		PSNR, SSIM
41	MAL	[86]	18'TNNLS		FSIM, SSIM	42	MRNF	[80]	18'AAAI	VRR	VIF, SSIM

表 1: 42 种典型的人脸素描合成算法概要。Sj.: 主观指标. Rr.: 识别率. Ob.: 客观指标. SV = 主观投票. ES = 经验研究. VRR = 多种识别率方法, 如, null-space LDA [4], 随机采样 LDA [63,64], 对偶空间 LDA [62], LPP [27], 稀疏表示和分类 [69]. 注意 UIQI [66] 是 SSIM [67] 的特殊情况.

比如, 在素描合成中, 其目标是让人类受试者无法区分真实素描与合成素描, 虽然两幅素描图像之间的像素值可能会失配。请看一下 Fig. 1, 在这些例子当中, 哪个更接近中间的参考图? 虽然这些对比任务对于人类来说微不足道, 但目前广泛应用的指标和人类判断并不一致。正如 [81] 提到, 这种现象主要归结为两个原因: 1) 视觉模式的维度非常高; 2) 视觉相似性的概念往往是主观的。

本文对于人脸素描领域的贡献主要总结为以下三点: 首先, 正如 Sec. 3 中所描述, 我们为素描合成提出一种结构共生纹理感知指标, 它提供了一种综合考虑结构和纹理的统一评价。

其次, 正如 Sec. 4.2 中所述, 我们在上述三种合理假设的基础上设计了三种元度量。大量的基于三种元度量实验也证明了我们所提出的 Scoot 指标的性能超越了先前的工作。同时, 我们的实验也证明了“空间结构”和“共生”纹理是人脸素描合成中两个普遍适用的感知特征。

最后, 我们(在指标中)探索了利用不同的纹理统计方法(如, Gabor, Sobel, 和 Canny 等等)。我们发现简单纹理特征 [17,18] 的性能远远好于这些文献中常用的度量指标。基于我们的发现, 我们构建了第一个大规模的基于人类感知的素描数据库, 它可以评估一个指标与人类感知的一致性程度。

上述三个贡献提供了一个完整的度量评测工具, 为从人类感知的角度分析数据相似性提供了新的视角和实用的工具(如度量、元度量和数据库)。

2. 相关工作

从表格 1 中, 我们可以看到一些工作利用识别率 (Rr.) 去评价合成的素描图像的好坏。然而, 识别率并不能完全的反映合成的素描 [55] 的视觉质量。在人脸素描合成 (FSS) 领域中, 广泛应使用的感知指标, 如 SSIM [67], FSIM [77] 和 VIF [40], 最初都是为图像质量评估 (IQA) 而设计的, 其目的是评估图像由如高斯模糊、jpeg 和 jpeg2000 压缩而引起的失真。鉴于任务性质的不同, 直接将 IQA 度量指标引入到人脸素描合成领域是不合适的 (见 Fig. 1)。

心理物理学 [95] 和先前的工作, 如线描画 [16,24] 证明了人类对素描相似度的感知主要取决于两个关键因素, 即图像结构 [67] 和图像纹理 [55]。然而, 这些所谓的感知特征到底有多感性? 这些“感知特征”与人类的视觉感知的对应程度又如何? 正如 Wang 等人 [55] 在文中提到, 目前在人脸素描合成领域没有可靠的感知指标。我们以有限的篇幅回顾与人脸素描最相关的话题。

基于启发式的度量指标。在人脸素描合成中, 最广泛应用的度量指标是 Wang [67] 提出的 SSIM。SSIM 采用一个滑动窗口的方式在图像的局部块上计算结构相似度, 亮度和对比度。Sheikh and Bovik [40] 提出 VIF 度量指标, 它通过量化两种信息来评价图像质量。一种是通过人类视觉系统通道获得的, 并包含输入真实图像和输出参考图像信息; 另一种是通过名为失真信息的失真通道实现的。这两种信息的比率即为指标评价结果。对人类视觉系统 (HVS) 的研究发现, 在不

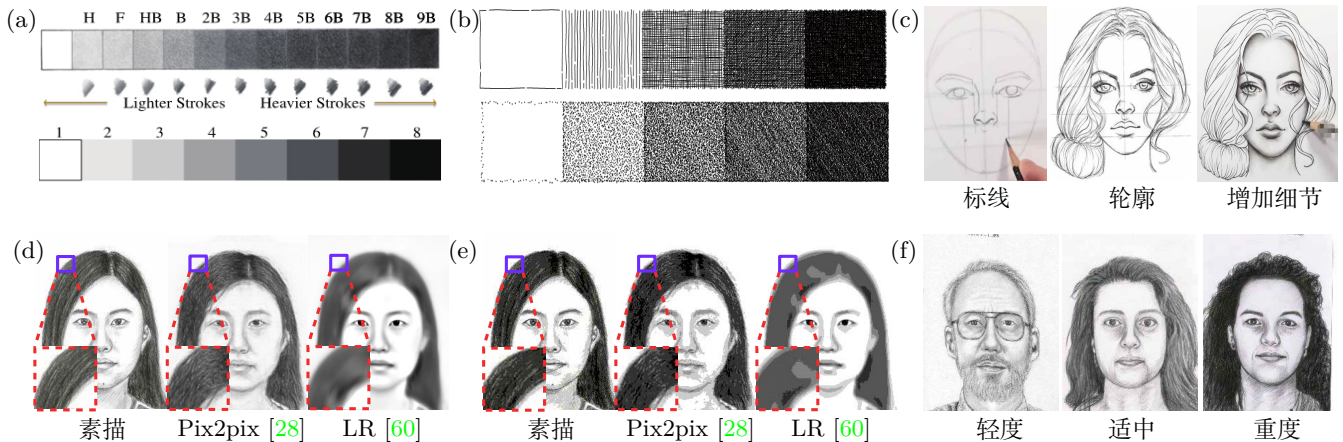


图 2: 本文 Scoot 指标的动机。(a) 铅笔等级和它们的笔画。(b) 使用笔画笔调来表示纹理。笔画纹理的笔调使用, 从上到下, 分别为: “交叉影线”, “点刻”。笔画的属性, 从左到右, 分别是: 稀疏到稠密。图像来源于文献 [68]。(c) 艺术家从标线到细节来绘制素描图像。(d) 原始素描图像。(e) 量化的素描图像。(f) 在笔尖上采用不同的压力 (如, 轻度到重度) 得到的不同笔画笔调。

同的频率下, 人类视觉感知的特征与傅立叶级数的相位是一致的。因此, Zhang 等人. [77] 选择相位一致性作为主要特征。然后, 他们提出了一种名为 FSIM 的底层特征相似性度量。最近, Xue 等人. [73] 设计了一个名为梯度幅度相似性偏差 (GMSD) 的简单指标, 它通过像素级的梯度幅度相似性去捕捉图像局部质量。计算总体梯度幅度相似性图的标准差作为最终图像质量指标。与其他度量指标相比, GMSD 达到了业内最好评价效果 (SOTA)。

基于学习的度量指标。除了基于启发式的度量指标, 还有许多基于学习的度量指标 [7, 20, 49] 用于以基于感知的方式比较图像, 并且这些度量指标已经被用来评价图像压缩和许多其他成像任务。我们建议读者参考最近的一篇综述 [81], 对感知度量所采用的多种深度学习特征进行全面综合的了解。本文的重点在于说明为什么人脸素描图像需要一个特定的感知距离度量, 该度量不同于或改进了以前基于启发式的度量方法。

2.1. 研究动机

我们观察了素描图像形成的基本元素, 发现“石墨铅笔等级”和“铅笔笔画”是素描的两个基本元素。

2.2. 石墨铅笔等级。

在欧洲体系中, “H” & “B” 分别代表“硬” & “软”铅笔。Fig. 2(a) 展示了石墨铅笔的等级。素描是通过有限的 (无颜色) 介质 (石墨铅笔) 进行表达。正如 Illustrator Sylwia Bomba [48] 所说: “(绘画时) 如果你手靠近铅笔尖, 你就能得到更黑的线条。当笔握的更

紧, 印记就越明显”。此外, 经过长时间的练习, 艺术家就会形成他们自己固定的压力 (例如 Fig. 2(c) 中从标线到细节) 风格。换句话说, 笔划印记可以通过改变笔尖的压力来改变 (例如, Fig. 2(f) 中从轻度到重度)。需要注意的是, 不同的笔尖压力会导致不同的印记类型, 这种名为灰度色调的印记也是一种可量化的因子。

灰度色调。灰度色调的量化可减少轻微噪音和对素描图像中细微的灰色色调梯度变化过度敏感的影响。因此, 我们在评价灰度色调相似性时引入了强度量化。受到以往工作的启发 [6], 我们量化输入素描图像 I 到 N_I 个不同的等级, 以减少需要考虑的强度级数: $I' = \Omega(I)$ 。Fig. 2(d, e) 给出了这种量化方式的典型例子。在量化素描图像前 (Fig. 2(d)) 后 (Fig. 2(e)), 人类在评价感知相似性时始终保持一致, 即 Pix2pix 方法得分高于 LR 方法。虽然量化可能会带来伪影, 但是我们的实验 (Sec. 6) 也同时证明了量化可以降低对微小强度变化的敏感度, 并在性能和计算复杂度之间达到平衡。

2.3. 铅笔笔画。

众所周知, 所有的素描图像是随着笔尖在纸张上的移动生成的, 并且笔尖的不同移动路径会产生不同的笔画形状。Fig. 2(b) 给出了一个例子, 在图中我们可以看到笔画的不同空间分布可以产生不同的纹理, (比如稀疏的或者稠密的纹理)。因此, 笔画是另外一个可量化的因子。

笔画笔调。笔划笔调和灰度色调并不是完全独立的概念。在素描图像中, 灰度色调是在不同的灰度级的

笔画上延伸而来，因此笔划笔调也可以定义为灰度色调的空间分布。

Fig. 2(d) 给出了另一个例子。因为 Pix2pix [28] 保留头发的纹理（笔画笔调）以及面部细节，所以 Pix2pix [28] 直观上来讲要明显好于 LR [60]。LR 呈现出过于平滑的结果，失去了素描图像该有的风格。

3. 提出的算法

本章节解释提出的指标，这个指标在块级空间结构中获取共生纹理统计特性。

3.1. 共生纹理

有了上述两个量化因子，我们可以开始描述其细节内容。为了同时提取关于“笔画笔调”及其与周围“灰度色调”关系的统计数据，我们需要表征它们的空间相互关系。前人在纹理方面的工作 [26] 证明了共生矩阵中各种强大的统计特征能够有效地获取纹理特征。由于素描和纹理在一定程度上是类似的，我们自然可以使用共生矩阵来作为灰度色调和笔画笔调的提取器。具体来说，该矩阵 \mathcal{M} 可以被定义为：

$$\mathcal{M}_{(i,j)|d} = \sum_{y=1}^H \sum_{x=1}^W \begin{cases} 1, & \text{if } I'_{x,y} = i \text{ and } I'_{(x,y)+d} = j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

其中， i 和 j 都表示灰度值； $d = (\Delta x, \Delta y)$ 是到 (x, y) 的相对距离； x 和 y 为所给量化后的素描图像 I' 的空间位置， $I'_{x,y}$ 表示素描图像 I' 中 (x, y) 坐标处的灰度值； W 和 H 则分别表示素描图像 I' 的宽和长。为了提取素描图像中的感知特征，我们测试了三种最常用 [25] 的统计特征：同质性 (\mathcal{H})，对比度 (\mathcal{C}) 和能量 (\mathcal{E})。

同质性反映了在图像局部区域纹理变化的大小，每个像素对的灰度色调越相似，则其值越高。同质性可以定义为：

$$\mathcal{H} = \sum_{j=1}^{N_i} \sum_{i=1}^{N_i} \frac{\mathcal{M}_{(i,j)|d}}{1 + |i - j|}, \quad (2)$$

对比度则表示整个素描图像中每个像素点（图像 I' 中）和其邻域内的像素点之间的差异之和。这也反映了低对比度素描的特点不是灰度值低，而是低的空间频率。也就是对比度与空间频率高度相关。对于常数色调的素描，对比度为 0。

$$\mathcal{C} = \sum_{j=1}^{N_i} \sum_{i=1}^{N_i} |i - j|^2 \mathcal{M}_{(i,j)|d} \quad (3)$$

能量可以测量纹理的均匀性。当素描图像块 (I') 中只包含了相似色调的像素时， \mathcal{M} 中的一些元素将趋

算法 1: 结构共生纹理度量

输入: 合成的素描 X , 真实素描 Y

步骤 1: 将 X 和 Y 量化为 N_i 个灰度级

步骤 2: 根据公式 1 计算共生矩阵 $\mathcal{M}(X)$ 和 $\mathcal{M}(Y)$

步骤 3: 将整幅素描图像划分为 $k * k$ 格共 k^2 个块

步骤 4: 根据公式 3 和 4, 从每个分块中提取 \mathcal{CE} 特征并将其合并

步骤 5: 根据公式 5 计算四个方向的平均特征

步骤 6: 根据公式 6 评价 X 和 Y 之间的相似度

输出: Scoot 分数;

近于 1，而剩下的元素则趋近于 0。如果在素描块 (I') 中只有一个灰度色调，则能量将达到最大值。因此，高能量对应的是素描图像的灰度色调分布具有周期性或者恒定性。

$$\mathcal{E} = \sum_{j=1}^{N_i} \sum_{i=1}^{N_i} (\mathcal{M}_{(i,j)|d})^2 \quad (4)$$

3.2. 空间结构

为了全面的描述空间结构，我们根据空间包络策略 [9, 35] 从素描图像的“块级”结构上提取统计特征。首先，我们将整个素描图像分为 $k \times k$ 个网格，共 k^2 个块。同时，我们的实验也证明了该过程有助于提取图像的内容信息。其次，我们计算所有块的共生矩阵 \mathcal{M} ，并对每个矩阵进行归一化，使得其分量之和为 1。最后，我们将 $k * k$ 个块的统计特性（比如， $\mathcal{H}, \mathcal{C}, \mathcal{E}$ ）合并为一个向量 $\vec{\Phi}(I'_s|d) \in \mathbb{R}^{p \times k \times k}$ 。

需要注意的是，以上所述的统计特性都是基于一个方向（比如， 90° ，即 $d = (0, 1)$ ）。由于空间分布的方向对于类似“头发方向”和“阴影笔画方向”的风格捕捉也非常重要。因此，为了利用上述观察结论从而有效的提取笔画的方向样式，我们计算了 T 个方向向量的平均特征 $\vec{\Psi}(I'_s)$ 来获取更多的方向信息：

$$\vec{\Psi}(I'_s) = \frac{1}{T} \sum_{i=1}^T \vec{\Phi}(I'_s|d_i), \quad (5)$$

其中 d_i 表示第 i 个方向，且 $\vec{\Psi}(I'_s) \in \mathbb{R}^{p \times k \times k}$ 。

3.3. Scoot 指标

在获取了参考素描图像 Y 和合成的素描图像 X 的感知特征向量之后，需要利用公式计算它们之间的相似度。我们测试了多种形式的计算公式，例如欧氏距离或者指数函数等等，最后我们发现欧氏距离是一个简单且有效的形式，在我们的实验中也是效果最好的。因此，本文的感知相似度度量指标 Scoot 定义为：

$$E_s = \frac{1}{1 + \|\vec{\Psi}(X'_s) - \vec{\Psi}(Y'_s)\|_2}. \quad (6)$$

其中 $\|\cdot\|_2$ 表示 l_2 -范数。 X'_s, Y'_s 分别表示量化后的 X_s, Y_s 。 $E_s = 1$ 则表示风格一致。

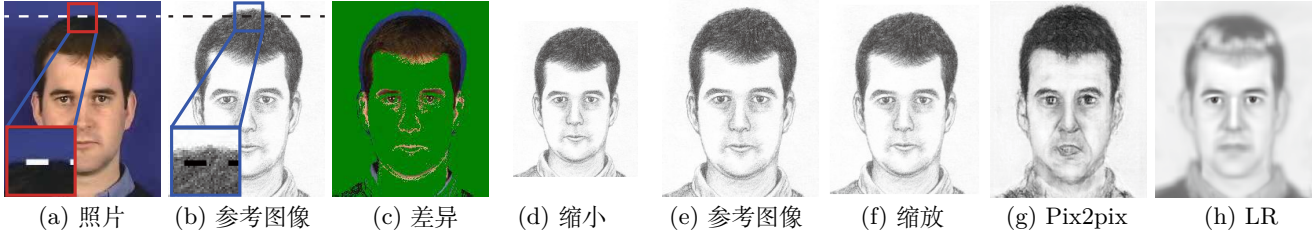
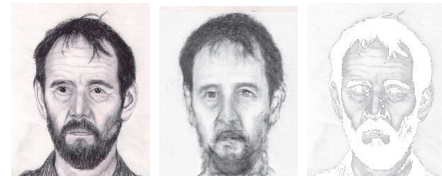


图 3: 元度量 1: 轻微缩放的稳定性。



(a) 参考图像 (b) R-参考图像 (c) Pix2pix (d) MWF

图 4: 元度量 2: 旋转敏感性。



(a) 参考图像 (b) 合成图像 (c) 亮笔素描

图 5: 元度量 3: 内容获取能力。

4. 实验

4.1. 实现细节

为了取得最好的结果，我们把 Sec. 3.2 中的空间结构 k 设置为 4。Eq. (6) 中的量化参数 N_l 设置为 6 个等级。另外，我们也证明了公式中 $p = 2$ (比如，公式 3 中的 C 和公式 4 中的 \mathcal{E} 合并) 可以达到最优性能 (见 Sec. 6)。由于共生矩阵 $\mathcal{M}(i, j)$ 的对称性，四个方向的统计特征实际上等于 1 个像素距离的 8 个相邻方向。根据经验，我们设置 $T = 4$ 个方向 $d_i \in \{(0, 1), (-1, 1), (-1, 0), (-1, -1)\}$ 以达到鲁棒的性能。

4.2. 元度量

如文献 [34] 所述，设计一个指标最具挑战的任务是证明其性能。和文献 [38] 一样，我们使用元度量方法，它通常被用来评估一个指标。受文献 [9, 10, 34] 启发，我们在 Sec. 1 中描述的 3 个特性的基础上提出了对应的三个元度量。

元度量 1: 尺寸微调的稳定性。 第一个元度量 (MM1) 描述为合成素描图像的排名不应随参考素描图像的细微变化而改变太大。因此，我们使用最近邻内插对参考素描进行了 5 个像素的小幅缩小。Fig. 3 给出了一个示例。艺术家绘制的参考素描 (b) 的头发和原始图像 (a) 相比有一定的偏差。我们注意到在图像的边缘有 5 个像素的偏差 (Fig. 3(c))，这是很正常的。虽然两张素描图像 (e) & (f) 几乎相同，但常用的度量指标如 SSIM [67], VIF [40], 以及 GMSD [73], 当这些指标以 (e) 或者 (f) 评测时，则会调换两幅合成的素描图像 (g

和 h) 的排名。然而，本文所提的 Scoot 指标能保持一致的排名。

特别需要说明，关于这个元度量，我们采用 $\theta = 1 - \rho$ [2] 测度去衡量指标在参考素描被缩小前后的排名稳定性。 θ 的值则在区间 $[0, 2]$ 内。

Tab. 2 展示了最终的结果：值越低，则指标对于细微缩小的稳定性越高。我们可以看到在 CUFS 和 CUFSF 数据集上，scoot 指标较 SSIM, FSIM, GMSD 等有显著的提升 ($\approx 77\%$ 以及 83%)。这种性能的提升主要归功于本文所提指标考虑的是“块级”统计特性而不是“像素”级统计特性。

元度量 2: 旋转敏感性。 在现实世界，艺术家绘制的素描图像相较于原始的图像可能存在细微的旋转。因此，本文所提的第二个元度量 (MM2) 描述为评价指标的对于参考图像的旋转敏感性。我们对参考图像做了一个细微的 (5°) 的逆时针旋转。Fig. 4 给出一个示例。当将有细微旋转的参考素描 (b) 替换为没有旋转的参考素描 (a) 时，排名不应该有太大的变化。在 MM2 中，我们分别根据无细微旋转变换的参考素描 (a) 和有细微旋转变换的参考素描 (b) 获得每个指标的排名结果。我们采用与 MM1 相同的测度 (θ) 去评价旋转敏感性。

Tab. 2 展示了旋转敏感性的结果。值得注意的是，MM2 和 MM1 正是 Sec. 4.2 中描述的期望性能的两个方面。在 CUFS 和 CUFSF 数据库上，我们的指标再次显著优于当前其他指标。

元度量 3: 内容获取功能。 元度量 (MM3) 描述为一个好的指标应该给 SOTA 模型生成的完整素描图像

指标	CUFS [65]			RCUFS Jud↑ 评价	CUFSF [88]			RCUFSF Jud↑ 评价
	MM1↓ 缩放	MM2↓ 旋转	MM3↑ 内容		MM1↓ 缩放	MM2↓ 旋转	MM3↑ 内容	
经典 & 广泛应用								
IFC [41]	0.256	0.189	1.20%	26.9%	0.089	0.112	3.07%	25.4%
SSIM [67]	0.162	0.086	81.4%	37.3%	0.073	0.074	97.4%	36.8%
FSIM [77]	0.268	0.123	14.2%	50.0%	0.151	0.058	32.4%	37.5%
VIF [40]	0.322	0.236	43.5%	44.1%	0.111	0.150	22.2%	52.8%
GMSD [73]	0.417	0.210	21.9%	42.6%	0.259	0.132	63.6%	58.6%
Scoot (本文)	0.037	0.025	95.9%	76.3%	0.012	0.008	97.5%	78.8%
基于纹理 & 基于边缘								
Canny [3]	0.086	0.078	33.7%	27.8%	0.138	0.146	0.00%	0.10%
Sobel [45]	0.040	0.037	0.00%	32.8%	0.048	0.044	0.00%	52.6%
GLRLM [18]	0.111	0.111	18.6%	73.7%	0.125	0.079	64.6%	68.0%
Gabor [17]	0.062	0.055	0.00%	72.2%	0.089	0.043	19.3%	80.9%
Scoot (本文)	0.037	0.025	95.9%	76.3%	0.012	0.008	97.5%	78.8%
特征合并								
$\mathcal{H}\mathcal{E}\mathcal{C}$	0.034	0.024	95.9%	76.3%	0.011	0.008	97.4%	78.7%
\mathcal{H}	0.007	0.005	61.5%	77.5%	0.003	0.003	79.1%	77.8%
\mathcal{E}	0.200	0.104	98.5%	73.1%	0.044	0.026	99.2%	77.4%
\mathcal{C}	0.010	0.007	54.4%	74.6%	0.009	0.006	64.7%	73.4%
$\mathcal{H}\mathcal{C}$	0.011	0.007	60.1%	74.6%	0.007	0.005	78.1%	73.7%
$\mathcal{H}\mathcal{E}$	0.156	0.088	97.9%	75.7%	0.030	0.017	98.8%	80.3%
$\mathcal{C}\mathcal{E}$ (Scoot)	0.037	0.025	95.9%	76.3%	0.012	0.008	97.5%	78.8%

表 2: 基于经典和替代纹理/边缘度量的评测结果。最好的结果用黑体高亮, 并且这些差异均在 $\alpha < 0.05$ 水平上具有统计学意义。其中 \uparrow 表示分数越高, 指标性能越好, 反之亦然 (\downarrow)。

一个比任何只保留不完整的笔画的素描图像更高的分数。如 Fig. 5 所示。我们希望一个指标应该更倾向 SOTA 模型合成结果 (b), 而不是亮笔画素描² 结果 (c)。对于 MM3, 我们计算了 10 个 SOTA 素描合成模型 [28, 32, 46, 54, 56, 60, 65, 76, 92, 93] 的平均得分。对于某个模型生成了一个较差的结果时, 使用平均分会更加鲁棒。我们记录了 SOTA 合成算法的平均得分高于亮笔画素描得分的次数。

Fig. 5 展示的案例中, 当前广泛应用的指标 (SSIM, FSIM, VIF) 都比较倾向于亮笔素描。只有本文所提 Scoot 指标给出了正确的排序。在像素级匹配方面, 很明显的看到去除黑色笔画的区域和 (a) 中对应区域有很大的差别。但是在其他区域, 图中的像素与参考素描中对应的像素是完全一样的。之前的指标只考虑像素级别的匹配, 因此亮笔画素描会有更高的排名。然而, 合成的素描图像 (b) 无论是在风格还是内容方面都要优于亮笔画素描 (c)。观察 Tab. 2 中 CUFS 数据集上的表现, 可以发现我们的指标相对于其他指标性能上有一个显著的提升 ($\geq 14\%$)。而在 CUFSF 数据集上也同样获得了较小的提升。

²为了测试第三个元度量, 我们使用了一个简单的灰度阈值 (如 170) 将素描图像 (Fig. 5 参考图像) 分割为暗笔画 & 亮笔画。亮笔画图像缺失了面部主要的纹理特征 (如, 头发, 眼睛, 胡须), 导致了素描的不完整。

5. 提出的感知相似度数据集

参考 Zhang 等人. [81], 我们利用强制 2 选 1 法则 (2AFC) 构建了一个人脸素描感知相似度数据集。这些被用来选择的合成素描主要源自一个广阔的失真空间和实际的算法合成。我们用这些结果来进行感知判断实验是因为合成的素描图像评估指标的测试是在实际的问题和实际的算法上进行的。

5.1. 收集多种人脸素描

原始图像。实际算法的数据是非常有限的, 因为每个合成模型都有其自己不同的属性。为了获取更多的失真数据, 我们收集了 338 对 (CUFS) 和 944 对 (CUFSF) 测试数据集图像作为原始图像, 这些数据按照文献 [54] 中的划分方案。

失真类型。我们采用各种各样的合成方法 [28, 32, 46, 54, 56, 60, 65, 76, 92, 93] 来提供更多来自不同合成算法的多样化失真类型。如 Fig. 6 所示, 我们引入了 10 种失真类型, 比如亮度变化, 前景噪音, 偏移, 线性扭曲, 结构损伤, 对比度变化, 模糊, 内容丢失, 重影, 以及棋盘效应。

5.2. 相似度评估

数据筛选。我们邀请了接受过 50 对图像排名训练的 21 位观察员根据两个标准对合成素描进行排名。为了尽可能的降低人为排序的不确定性, 我们遵循投票策略 [55] 通过以下阶段进行实验 ($\sim 152K$ 次判断):

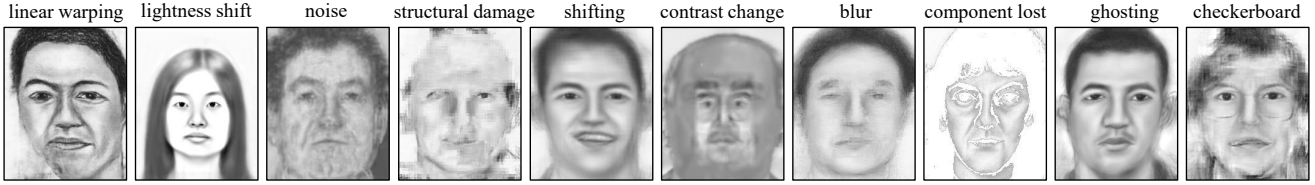


图 6: 畸变。这些畸变发生在多种实际合成算法 [28, 32, 46, 54, 56, 60, 65, 76, 92, 93] 中。

- 我们首先让第一组观察员 (7 个受试者) 根据每张原始图像从 10 张合成素描图像中选择出 4 张。这 4 张素描应该包含两张好的以及两张差的。这样, 我们分别为 CUFS 和 CUFSF 留下了 1352(4×338) 和 3776(4×944) 张素描图像。
- 对于被选出的四张素描图像, 第二组观察员 (7 人) 被要求进一步从中选择三张观察员可以轻松排序的素描图像。基于投票的结果, 我们挑出 3 张最容易被选择的素描图像。
- 过于相似的素描图像会使观察员很难去判断哪个素描图像更好。观察员可能会随机给出他们的选择。为了避免这种随机选择, 我们要求最后一组观察员 (7 人) 挑出最容易排名的一对素描图像。

2AFC 相似度判断 [81]。2AFC 相似度判断。对于每张图像, 我们都有一张艺术家绘制的参考素描图像 r 和两个失真图像 s_0, s_1 。我们询问观察员哪张素描更加接近参考素描 s , 并将其记录为 $q \in \{0, 1\}$ 。平均来看, 观察者每次判断需要 2 秒时间。请注意, 我们有 5 名志愿者参与整个过程来交叉核对排名。比如, 如果 ≥ 4 个观察员偏好 s_0 , ≤ 1 个观察员偏好 s_1 , 则最终的排名将会是 $s_0 > s_1$ 且 $q = 1$ 。那些明显被多数人选择的三元组将会保留下来, 而其他三元组则会被遗弃。最终, 我们建立了两个新的人为定序的³数据集: RCUFS 以及 RCUFSF。完整数据集见我们的网站 [website](#)。

5.3. 人类判断

在此, 我们评估 Scoot 和其他对比的指标的性能。RCUFS 和 RCUFSF 分别包含了 338 个和 944 个人为判断过的三元组。为了降低固有噪声的影响, 我们计算一个度量指标与每个三元组的一致性, 并采用数据集中的平均统计量作为最终性能。

³这两个数据集分别包含了 1014 (3×338 三元组), 和 2832 (3×944 三元组) 数据集。最近的工作 [47, 70] 表明了一个数据集的规模是非常重要的。据我们所知, 这是 FSS 领域第一个大型公开人为定序数据集。

经典指标和 Scoot 的表现如何? 表格 2 展示了多种经典指标 (如, IFC, SSIM, FSIM, VIF 和 GMSD) 的性能。有趣的是, 这些指标性能都非常差 (如, $\leq 59\%$)。尽管这些评价指标在素描合成中经常使用, 但是这些指标不是为度量素描合成领域中像素级失配而设计的。和之前最好的指标相比, 本文提出的 Scoot 指标在 RCUFSF 数据集上得到了一个显著 ($\sim 26.3\%$) 的性能提升。这一改进归功于我们对结构和纹理相似性的考虑, 这是人类在评价素描时可以被人类感知到的两个基本因素。

6. 讨论

成功的关键因素是哪个? 在 Sec. 3.1 中, 我们考虑了 3 个广泛应用的统计特征: 同质性 (\mathcal{H}), 对比度 Contrast (\mathcal{C}) 和能量 (\mathcal{E})。为了达到最好的性能, 我们需要探索这些统计特征的最佳组合。我们使用了三个元度量以及人为判断来测试 Scoot 指标的性能, 并使用单一特征、特征对以及三个特征的组合进行试验。

结果如 Tab. 2 所示。所有可能性 ($\mathcal{H}, \mathcal{E}, \mathcal{C}, \mathcal{CE}, \mathcal{HE}, \mathcal{CH}, \mathcal{HEC}$) 在 Jud (人类的判断) 上表现出色。 \mathcal{H} 和 \mathcal{C} 对于尺寸变化 (MM1) 和旋转变换 (MM2) 是不敏感的, 然而它们在内容获取上的性能并不好。 \mathcal{E} 与 \mathcal{H} 和 \mathcal{C} 恰好相反。因此, 只采用一个单一的特征效果不佳。联合两种特征的结果表明, 如果 \mathcal{H} 和 \mathcal{E} 合并, 在克服 \mathcal{E} 的弱点的同时, 对调整大小和旋转的敏感度仍然很高。相较于 “ \mathcal{CE} ”, $\mathcal{H} + \mathcal{E} + \mathcal{C}$ 三者合并没有任何提升。文献 [1] 中的工作也发现能量 (\mathcal{E}) 和对比度 (\mathcal{C}) 是区分纹理图像最有效的特征。因此, 我们选择 “ \mathcal{CE} ” 特征作为最终的组合来提取感知特征。

这些 “感知特征” 实际上与人类视觉的对应程度如何? 正如章节 3.1 所描述, 素描图像非常接近纹理图像。目前有许多其他纹理 & 边缘特征 (如 GLRLM [18], Gabor [17], Canny [3], Sobel [45])。这里, 我们选择最广泛应用的特征作为我们的候选项去取代我们的 “ \mathcal{CE} ” 特征。对于 GLRLM, 我们选择最原始的版本中提到的

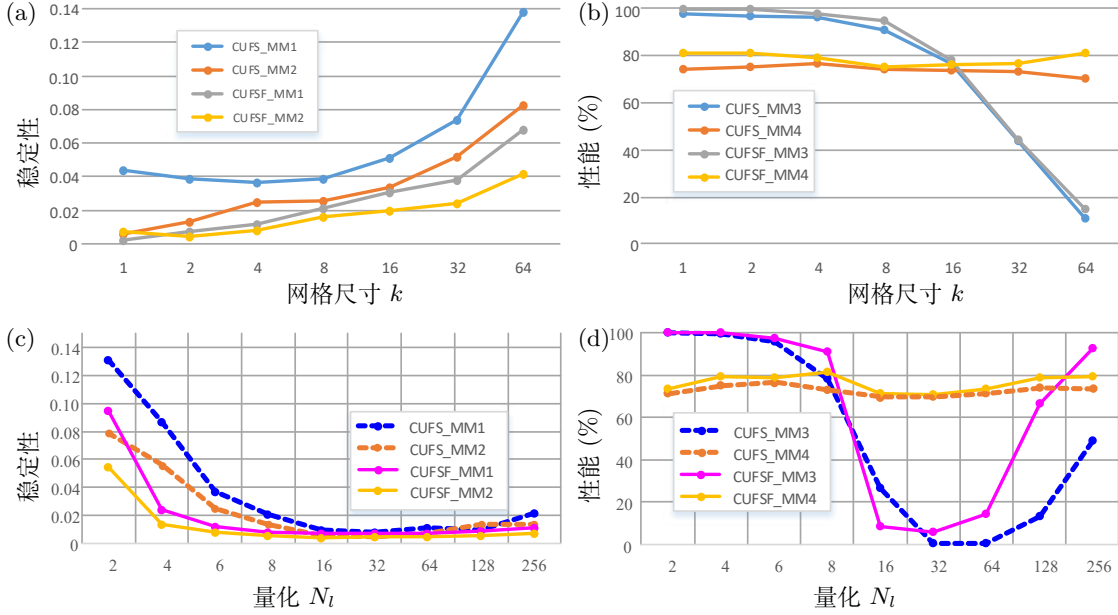


图 7: 空间结构 (上) 和量化 (下) 的敏感度实验. 对于 MM1 & MM2, 值越低越好. 对于 MM3 & MM4, 值越高越好.

五种统计特性。如 Tab. 2所示。Gabor 和 GLRLM 都是纹理特征，其他两个则是边缘特征。所有的纹理特征 (GLRLM, Gabor) 和本文所提的 Scoot 指标给出了一个与人为排名一样好 (比如, $\geq 68\%$) 的结果 (Jud)。在这些纹理特征中, Scoot 得到了一个与人为排名一致且保持较高的平均性能 (Jud)。GLRLM 在 MM1 & 2 & 3 上则可以表现较好。Gabor 在 MM1 & 2 上表现较为合理, 但是在 MM3 上效果较差。对于基于边缘的特征, Canny 在所有的元度量上都失败了。对于轻微缩小 (MM1) 和旋转 (MM2), Sobel 非常稳定, 但不能捕获内容 (MM3), 与人的判断 (JUD) 不一致。有趣的是, Canny、Sobel 和 Gabor 给不完整的笔画打了比 SOTA 算法生成的素描图像更高的分数。换句话说, 指标完全颠倒了所有测试用例的排名结果。就总体结果而言, 我们得出的结论是, “ \mathcal{CE} ” 特征相较于其他特征具有更好的鲁棒性。

空间结构的敏感性如何? 为了分析空间结构的影响, 我们引出了 7 组实验, 可将素描图像分为不同大小的网格, 比如 k 被设置为 1, 2, 4, \dots , 64。Fig. 7(b) 中, MM3 & 4 的结果表明, $k=1$ 具有最优表现。然而, $k=1$ 的不足在于它只能获取图像级别的统计特性, 并且忽视了素描的纹理。也就是说, 一张经过随机组合的素描图像也可以得到一个较高的分数。Fig. 7(a) 中, 元度量 1 的实验清楚地表明, 在 CUFS 数据集上, $k=4$ 时可以

达到最佳性能。根据这两个实验, $k=4$ 时可以取得最鲁棒的结果。

量化的敏感性如何? 为了确定量化参数 N_l (基线: $N_l = \{2, 4, 6, 8, 16, 32, 64, 128\}$) 取哪个值可以获得最优性能, 我们进一步做了一个敏感性测试。从 Fig. 7(c)&(d) 中, 我们可以观察到将输入素描图像量化到 32 个灰度级可以取得最佳结果。然而, 对于 MM3 & MM4 实验却得到了一个很差的表现。综合考虑这些实验, $N_l=6$ 可以取得最鲁棒的结果。

7. 总结

在本文中, 我们探索了人类感知问题, 比如人类选择和指标之间存在哪些不同。一个用来分析上述问题的途径是人脸素描。我们提出了一个特定的指标, 名为 Scoot (结构共生纹理), 该指标可以获取人类感知, 并在提出的三个元度量上进行了分析。最后, 我们构建了第一个人类感知素描数据集, 它可以用来评估一个指标和人类感知之间的契合度。我们在 Scoot 架构上系统地评估了基于纹理/基于边缘的特征, 并将它们与经典指标进行了比较。研究结果表明, “空间结构” 和 “共生” 纹理是人脸素描中两种普遍适用的感知特征。在未来, 我们将继续开发和应用 Scoot, 以进一步推动前沿研究, 如, 背景移除评估 [94]。

参考文献

- [1] Andrea Baraldi and Flavio Parmiggiani. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE T Geosci. Remote.*, 33(2):293–304, 1995.
- [2] DJ Best and DE Roberts. Algorithm AS 89: the upper tail probabilities of Spearman’s rho. *J R STAT SOC C-APPL*, 24(3):377–379, 1975.
- [3] John Canny. A computational approach to edge detection. *IEEE TPAMI*, 8:679–698, 1986.
- [4] Li-Fen Chen, Hong-Yuan Mark Liao, Ming-Tat Ko, Ja-Chen Lin, and Gwo-Jong Yu. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33(10):1713–1726, 2000.
- [5] Ming-Ming Cheng, Yun Liu, Wen-Yan Lin, Ziming Zhang, Paul L Rosin, and Philip HS Torr. BING: Binarized normed gradients for objectness estimation at 300fps. *Computational Visual Media*, 5(1):3–20, 2019.
- [6] David A Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Can. J Remote. Sens.*, 28(1):45–62, 2002.
- [7] Alexey Dosovitskiy and Thomas Brox. Generating images with perceptual similarity metrics based on deep networks. In *NIPS*, pages 658–666, 2016.
- [8] Mohamed Elhoseiny, Yizhe Zhu, Han Zhang, and Ahmed Elgammal. Link the head to the “beak” : Zero shot learning from noisy text description at part precision. In *IEEE CVPR*, 2017.
- [9] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A New Way to Evaluate Foreground Maps. In *IEEE ICCV*, pages 4548–4557, 2017.
- [10] Deng-Ping Fan, Cheng Gong, Yang Cao, Bo Ren, Ming-Ming Cheng, and Ali Borji. Enhanced-alignment Measure for Binary Foreground Map Evaluation. In *IJCAI*, pages 698–704, 2018.
- [11] Deng-Ping Fan, Zheng Lin, Jia-Xing Zhao, Yun Liu, Zhao Zhang, Qibin Hou, Menglong Zhu, and Ming-Ming Cheng. Rethinking RGB-D salient object detection: Models, datasets, and large-scale benchmarks. *arXiv preprint arXiv:1907.06781*, 2019.
- [12] Deng-Ping Fan, Jiang-Jiang Liu, Shang-Hua Gao, Qibin Hou, Ali Borji, and Ming-Ming Cheng. Salient objects in clutter: Bringing salient object detection to the foreground. In *ECCV*, pages 1597–1604. Springer, 2018.
- [13] Deng-Ping Fan, Wenguan Wang, Ming-Ming Cheng, and Jianbing Shen. Shifting more attention to video salient object detection. In *IEEE CVPR*, pages 8554–8564, 2019.
- [14] Deng-Ping Fan, ShengChuan Zhang, Yu-Huan Wu, Yun Liu, Ming-Ming Cheng, Bo Ren, Paul L Rosin, and Rongrong Ji. Scoot: A perceptual metric for facial sketches. In *IEEE ICCV*, pages 5612–5622, 2019.
- [15] Jakub Fišer, Ondřej Jamriška, David Simons, Eli Shechtman, Jingwan Lu, Paul Asente, Michal Lukáč, and Daniel Šykora. Example-based synthesis of stylized facial animations. *ACM TOG*, 36(4):155, 2017.
- [16] William T Freeman, Joshua B Tenenbaum, and Egon C Pasztor. Learning style translation for the lines of a drawing. *ACM TOG*, 22(1):33–46, 2003.
- [17] Dennis Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.
- [18] Mary M Galloway. Texture analysis using grey level run lengths. *NASA STI/Recon Technical Report N*, 75, 1974.
- [19] Fei Gao, Shengjie Shi, Jun Yu, and Qingming Huang. Composition-aided sketch-realistic portrait generation. *arXiv preprint arXiv:1712.00899*, 2017.
- [20] Fei Gao, Yi Wang, Panpeng Li, Min Tan, Jun Yu, and Yani Zhu. Deepsim: Deep similarity for image quality assessment. *Neurocomputing*, 257:104–114, 2017.
- [21] Xinbo Gao, Nannan Wang, Dacheng Tao, and Xuelong Li. Face sketch-photo synthesis and retrieval using sparse representation. *IEEE TCSVT*, 22(8):1213–1226, 2012.
- [22] Xinbo Gao, Juanjuan Zhong, Jie Li, and Chunna Tian. Face sketch synthesis algorithm based on E-HMM and selective ensemble. *IEEE TCSVT*, 18(4):487–496, 2008.
- [23] Xinbo Gao, Juanjuan Zhong, Dacheng Tao, and Xuelong Li. Local face sketch synthesis learning. *Neurocomputing*, 71(10-12):1921–1930, 2008.
- [24] Stéphane Grabli, Emmanuel Turquin, Frédo Durand, and François X Sillion. Programmable style for NPR

- line drawing. *Rendering Techniques (Eurographics Symposium on Rendering)*, 2004.
- [25] Robert M Haralick et al. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.
- [26] Robert M Haralick, Karthikeyan Shanmugam, et al. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, pages 610–621, 1973.
- [27] Xiaoferi He and Partha Niyogi. Locality preserving projections. In *NIPS*, pages 153–160, 2004.
- [28] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *IEEE CVPR*, pages 1125–1134, 2017.
- [29] Licheng Jiao, Sibozhang, Lingling Li, Fang Liu, and Wenping Ma. A modified convolutional neural network for face sketch synthesis. *PR*, 76:125–136, 2018.
- [30] Jie Li, Xinye Yu, Chunlei Peng, and Nannan Wang. Adaptive representation-based face sketch-photo synthesis. *Neurocomputing*, 269:152–159, 2017.
- [31] Yi Li, Yi-Zhe Song, Timothy M Hospedales, and Shaogang Gong. Free-hand sketch synthesis with deformable stroke models. *IJCV*, 122(1):169–190, 2017.
- [32] Qingshan Liu, Xiaoou Tang, Hongliang Jin, Hanqing Lu, and Songde Ma. A nonlinear approach for face sketch synthesis and recognition. In *IEEE CVPR*, volume 1, pages 1005–1010, 2005.
- [33] Wei Liu, Xiaoou Tang, and Jianzhuang Liu. Bayesian Tensor Inference for Sketch-Based Facial Photo Hallucination. In *IJCAI*, pages 2141–2146, 2007.
- [34] Ran Margolin, Lih Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps? In *IEEE CVPR*, pages 248–255, 2014.
- [35] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *IJCV*, 42(3):145–175, 2001.
- [36] Chunlei Peng, Xinbo Gao, Nannan Wang, and Jie Li. Superpixel-based face sketch-photo synthesis. *IEEE TCSVT*, 27(2):288–299, 2017.
- [37] Chunlei Peng, Xinbo Gao, Nannan Wang, Dacheng Tao, Xuelong Li, and Jie Li. Multiple representations-based face sketch-photo synthesis. *IEEE TNNLS*, 27(11):2201–2215, 2016.
- [38] Jordi Pont-Tuset and Ferran Marques. Measures and meta-measures for the supervised evaluation of image segmentation. In *IEEE CVPR*, pages 2131–2138, 2013.
- [39] Amir Semmo, Tobias Isenberg, and Jürgen Döllner. Neural style transfer: a paradigm shift for image-based artistic rendering? In *ACM NPAR*, page 5, 2017.
- [40] Hamid R Sheikh and Alan C Bovik. Image information and visual quality. *IEEE TIP*, 15(2):430–444, 2006.
- [41] Hamid R Sheikh, Alan C Bovik, and Gustavo De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE TIP*, 14(12):2117–2128, 2005.
- [42] Jianbing Shen, Yunfan Du, Wenguan Wang, and Xuelong Li. Lazy random walks for superpixel segmentation. *IEEE TIP*, 23(4):1451–1462, 2014.
- [43] Jianbing Shen, Xiaopeng Hao, Zhiyuan Liang, Yu Liu, Wenguan Wang, and Ling Shao. Real-time superpixel segmentation by DBSCAN clustering algorithm. *IEEE TIP*, 25(12):5933–5942, 2016.
- [44] Jianbing Shen, Jianteng Peng, and Ling Shao. Sub-modular trajectories for better motion segmentation in videos. *IEEE TIP*, 27(6):2688–2700, 2018.
- [45] Irvin Sobel. An isotropic 3×3 image gradient operator. *Machine Vision for Three-dimensional Scenes*, pages 376–379, 1990.
- [46] Yibing Song, Linchao Bao, Qingxiong Yang, and Ming-Hsuan Yang. Real-time exemplar-based face sketch synthesis. In *ECCV*, pages 800–813. Springer, 2014.
- [47] Xiaoxiao Sun, Jufeng Yang, Ming Sun, and Kai Wang. A benchmark for automatic visual classification of clinical skin disease images. In *ECCV*, pages 206–222. Springer, 2016.
- [48] Bomba Sylwia, Cai Rovina, Croes Brun, Gerard Justin, and Lewis Marisa. *Beginner’s Guide to Sketching*. 3dtotal Publishing, 2015.
- [49] Hossein Talebi and Peyman Milanfar. NIMA: Neural image assessment. *IEEE TIP*, 27(8):3998–4011, 2018.
- [50] Xiaoou Tang and Xiaogang Wang. Face sketch synthesis and recognition. In *IEEE CVPR*, pages 687–694, 2003.
- [51] Xiaoou Tang and Xiaogang Wang. Face sketch recognition. *IEEE TCSVT*, 14(1):50–57, 2004.
- [52] Ching-Ting Tu, Yu-Hsien Chan, and Yi-Chung Chen. Facial Sketch Synthesis Using 2D Direct Combined

- Model-Based Face-Specific Markov Network. *IEEE TIP*, 25(8):3546–3561, 2016.
- [53] Lidan Wang, Vishwanath Sindagi, and Vishal Patel. High-quality facial photo-sketch synthesis using multi-adversarial networks. In *IEEE FG*, pages 83–90, 2018.
- [54] Nannan Wang, Xinbo Gao, and Jie Li. Random sampling for fast face sketch synthesis. *Pattern Recognition*, 76:215–227, 2018.
- [55] Nannan Wang, Xinbo Gao, Jie Li, Bin Song, and Zan Li. Evaluation on synthesized face sketches. *Neurocomputing*, 214:991–1000, 2016.
- [56] Nannan Wang, Xinbo Gao, Leiyu Sun, and Jie Li. Bayesian face sketch synthesis. *IEEE TIP*, 26(3):1264–1274, 2017.
- [57] Nannan Wang, Jie Li, Dacheng Tao, Xuelong Li, and Xinbo Gao. Heterogeneous image transformation. *PRL*, 34(1):77–84, 2013.
- [58] Nannan Wang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Jie Li. Transductive face sketch-photo synthesis. *IEEE TNNLS*, 24(9):1364–1376, 2013.
- [59] Nannan Wang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Jie Li. A comprehensive survey to face hallucination. *IJCV*, 106(1):9–30, 2014.
- [60] Nannan Wang, Mingrui Zhu, Jie Li, Bin Song, and Zan Li. Data-driven vs. model-driven: Fast face sketch synthesis. *Neurocomputing*, 2017.
- [61] Shenlong Wang, Lei Zhang, Yan Liang, and Quan Pan. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *IEEE CVPR*, pages 2216–2223, 2012.
- [62] Xiaogang Wang and Xiaoou Tang. Dual-space linear discriminant analysis for face recognition. In *IEEE CVPR*, volume 2, pages II–II, 2004.
- [63] Xiaogang Wang and Xiaoou Tang. Random sampling lda for face recognition. In *IEEE CVPR*, pages 259–265, 2004.
- [64] Xiaogang Wang and Xiaoou Tang. Random sampling for subspace face recognition. *IJCV*, 70(1):91–104, 2006.
- [65] Xiaogang Wang and Xiaoou Tang. Face photo-sketch synthesis and recognition. *IEEE TPAMI*, 31(11):1955–1967, 2009.
- [66] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE SPL*, 9(3):81–84, 2002.
- [67] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.
- [68] Georges Winkenbach and David H Salesin. Computer-generated pen-and-ink illustration. In *ACM SIGGRAPH*, pages 91–100, 1994.
- [69] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE TPAMI*, 31(2):210–227, 2009.
- [70] Xiaoping Wu, Chi Zhan, Yu-Kun Lai, Ming-Ming Cheng, and Jufeng Yang. Ip102: A large-scale benchmark dataset for insect pest recognition. In *IEEE CVPR*, pages 8787–8796, 2019.
- [71] Bing Xiao, Xinbo Gao, Dacheng Tao, Yuan Yuan, and Jie Li. Photo-sketch synthesis and recognition based on subspace learning. *Neurocomputing*, 73(4-6):840–852, 2010.
- [72] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. In *IEEE ICCV*, 2017.
- [73] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE TIP*, 23(2):684–695, 2014.
- [74] Jufeng Yang, Xiaoxiao Sun, Jie Liang, and Paul L Rosin. Clinical skin lesion diagnosis using representations inspired by dermatologist criteria. In *IEEE CVPR*, pages 1258–1266, 2018.
- [75] Dongyu Zhang, Liang Lin, Tianshui Chen, Xian Wu, Wenwei Tan, and Ebroul Izquierdo. Content-adaptive sketch portrait generation by decompositional representation learning. *IEEE TIP*, 26(1):328–339, 2017.
- [76] Liliang Zhang, Liang Lin, Xian Wu, Shengyong Ding, and Lei Zhang. End-to-end photo-sketch generation via fully convolutional representation learning. In *ACM ICMR*, pages 627–634, 2015.
- [77] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE TIP*, 20(8):2378–2386, 2011.
- [78] Mingjin Zhang, Jie Li, Nannan Wang, and Xinbo Gao. Compositional model-based sketch generator in facial entertainment. *IEEE TOC*, 48(3):904–915, 2018.

- [79] Mingjin Zhang, Nannan Wang, Xinbo Gao, and Yunsong Li. Markov random neural fields for face sketch synthesis. In *IJCAI*, pages 1142–1148, 2018.
- [80] Mingjin Zhang, Nannan Wang, Yunsong Li, Ruxin Wang, and Xinbo Gao. Face sketch synthesis from coarse to fine. In *AAAI*, 2018.
- [81] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE CVPR*, pages 586–595, 2018.
- [82] Shengchuan Zhang, Xinbo Gao, Nannan Wang, and Jie Li. Robust face sketch style synthesis. *IEEE TIP*, 25(1):220–232, 2016.
- [83] Shengchuan Zhang, Xinbo Gao, Nannan Wang, and Jie Li. Face sketch synthesis from a single photo-sketch pair. *IEEE TCSVT*, 27(2):275–287, 2017.
- [84] Shengchuan Zhang, Xinbo Gao, Nannan Wang, Jie Li, and Mingjin Zhang. Face sketch synthesis via sparse representation-based greedy search. *IEEE TIP*, 24(8):2466–2477, 2015.
- [85] Shengchuan Zhang, Rongrong Ji, Jie Hu, Yue Gao, and Chia-Wen Lin. Robust face sketch synthesis via generative adversarial fusion of priors and parametric sigmoid. In *IJCAI*, pages 1163–1169, 2018.
- [86] Shengchuan Zhang, Rongrong Ji, Jie Hu, Xiaoqiang Lu, and Xuelong Li. Face sketch synthesis by multidomain adversarial learning. *IEEE TNNLS*, 2018.
- [87] Wei Zhang, Xiaogang Wang, and Xiaoou Tang. Lighting and pose robust face sketch synthesis. In *ECCV*, pages 420–433. Springer, 2010.
- [88] Wei Zhang, Xiaogang Wang, and Xiaoou Tang. Coupled information-theoretic encoding for face photo-sketch recognition. In *IEEE CVPR*, pages 513–520, 2011.
- [89] Yuqian Zhang, Nannan Wang, Shengchuan Zhang, Jie Li, and Xinbo Gao. Fast face sketch synthesis via kd-tree search. In *ECCV*, pages 64–77. Springer, 2016.
- [90] Jiaying Zhao, Ren Bo, Qibin Hou, Ming-Ming Cheng, and Paul L. Rosin. FLIC: fast linear iterative clustering with active search. *Computational Visual Media*, 4(4):333–348, 2018.
- [91] Jia-Xing Zhao, Yang Cao, Deng-Ping Fan, Ming-Ming Cheng, Xuan-Yi Li, and Le Zhang. Contrast prior and fluid pyramid integration for RGBD salient object detection. In *IEEE CVPR*, 2019.
- [92] Hao Zhou, Zhanghui Kuang, and Kwan-Yee K Wong. Markov weight fields for face sketch synthesis. In *IEEE CVPR*, pages 1091–1097, 2012.
- [93] Mingrui Zhu, Nannan Wang, Xinbo Gao, and Jie Li. Deep graphical feature learning for face sketch synthesis. In *IJCAI*, pages 3574–3580, 2017.
- [94] Yizhe Zhu and Ahmed Elgammal. A multilayer-based framework for online background subtraction with freely moving cameras. In *IEEE ICCV*, 2017.
- [95] Steven W Zucker, Allan Dobbins, and Lee Iverson. Two stages of curve detection suggest two styles of visual computation. *Neural computation*, 1(1):68–81, 1989.