arXiv:2303.02547v1 [cs.HC] 5 Mar 2023

结合用户反馈在概念空间中构建情绪板

Shin Sano [1, 3*] 和 Seiji Yamada [1,2]

[1]综合研究大学院大学（SOKENDAI）跨学科科学学院信息学系，日本叶山[2]国立信息学研究所数字内容与媒体科学研究部门，日本东京[3]创意整合研究所，美国加利福尼亚州奥克兰

摘要

我们提出了情绪板构建器（MBC），它支持概念设计师在二维概念空间中检索和编排图像，以传达设计概念。MBC 允许用户直观地迭代进行自适应图像检索。我们对情绪板工具的创新之处在于，根据用户在二维空间中重新排列图像的方式，为下一轮迭代调整查询向量。该算法通过计算情绪板上图像的平均向量，并乘以网格各单元所分配的权重，突出横轴和纵轴标签的含义。随后，系统从所得平均向量中获取最相似的词语，并将其用作新查询，以执行下一次图像搜索。除了上述算法，我们还开展了参与者实验，将其与另外两种交互算法进行比较。第一种算法允许用户删除不需要的图像，然后继续进行下一轮搜索。第二种算法利用每张图像上的语义标签，用户可据此提供负反馈，以修改后续搜索的查询。尽管我们未观察到所提三种算法之间存在显著差异，但对420个案例的情绪板创作进行的实验，通过创造力支持指数（CSI）得分证实了自适应迭代的有效性。

关键词 情绪板工具 · 智能交互系统 · 工业设计 · 概念设计 · 创造力支持工具 · 产品语义 · 设计美学 · 人机交互 · 相关性反馈 · 词汇语义学

1 引言

情绪板是常用的视觉制品，可作为设计开发工具，在利益相关者之间传达和分享设计创意[1]。设计师会探索各种各样的视觉图像，并挑选出最能代表原始设计简报或其后续版本所唤起的情绪、感受或“氛围”的图像集。情绪板通常由搜集或创作的图像组成，用于展示。与可能引发强烈字面联想的具象图像相比，抽象图像往往更适合用于此目的[2]。以往许多研究探讨了如何通过流程和方法，帮助设计师创作能够传达预期设计概念的视觉情绪板。我们将首先介绍有关设计教育和行业实践中情绪板的研究。

在图形软件普及之前，人们通常采用模拟与实体方法制作情绪板。传统上，人们会将不同类型的媒介粘贴在一起制作情绪板，包括杂志图片、照片、织物、灵感物件等。 [3]

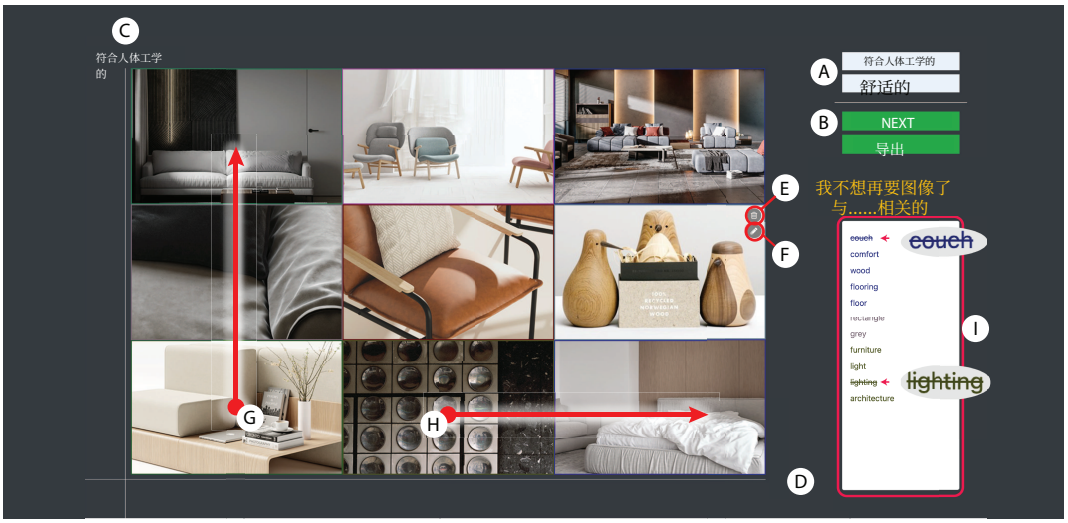

图1：采用“参考2”算法的 MBC 系统用户界面，该算法涉及最详尽的用户反馈。拟议系统不具备标签反馈功能(I)。拟议系统允许用户在 3×3 矩阵内移动任意图像。向上移动图像(G)会增强其语义，使其更接近词语1(C)“符合人体工学的”；向右移动图像(H)则会增强其语义，使其在进行下一次搜索时更接近词语2(D)“舒适的”。

正如 Cassidy [4] 指出的，情绪板常用于时尚产业以及与时尚相关或面向消费者的行业，以视觉方式传达信息。这是因为，相较于设计的功能性和实用性要素，情绪板在传达设计的造型和审美要素方面发挥着至关重要的作用。

情绪板在设计实践和教育中发挥着多种作用，例如思维外化、意义获取和概念推理 [5]。它们也被用作定性设计研究工具，以促进创造性思维、展示和传达产品 [4]，，并表达和传达设计师的想象力及其追求的理念 [3]。设计师将其用作与非设计师沟通的工具 [6]。对设计师而言，创作情绪板的过程比最终结果更重要，因为情绪板是中间媒介，而非设计的最终产品。Bouchard等人 [7] 探讨了中间表征（IR）在设计中的作用：设计师需要在不同抽象层级上创建多种表征。

2 相关工作

与计算方法一样，一些研究人员尝试以交互方式支持设计师构建具有不同侧重点的视觉情绪板。情绪板构建任务涉及需要解决的各种算法问题，例如图像检索、搜索策略、计算机视觉、语义特征工程、自然语言处理，以及基于用户反馈的查询扩展与修改。

Setchi等人 [8] 关注语义鸿沟问题，即低层图像特征有限的描述能力与用户希望操作的丰富语义之间的差异 [9]。他们提出了一种基于语义的图像检索方法，依靠目标图像周围的文本信息，避免使用给定图像的低层次和字面标签。该方法利用TF-IDF提取文档中最相关的词语，并使用通用本体扩展查询，以查找更多相关图像。

Koch 等人 [10] 关注初始设计构思中的问题：探索与利用之间的权衡，即在坚持过去带来最高收益的选项，与探索未来可能带来更高回报的新选项之间取得平衡[11]。他们创建了一款交互式数字工具，以支持设计师

创建情绪板，并采用经合作式上下文赌博机强化学习算法优化的探索-利用策略。用户可以通过为推荐的图像添加标签来提供反馈，系统则会根据每轮更新后的相关推荐代理概率分布，采取不同的查询策略。

Koch 等人 [12] 进一步改进了数字情绪板工具，并创建了语义拼贴。该工具利用 Google Vision API 为其推荐的每张图像分配语义标签，帮助设计师将模糊的视觉构想转化为搜索词，并理解和传达自己的设计。设计师可以通过文字和视觉查询搜索图像，并查看每张图像的语义标签。此外，用户还可以通过调整这些标签的权重向系统提供反馈，从而影响下一次搜索。语义拼贴的一个独特之处在于，系统不仅会为画布上的单张图像显示标签，还会呈现画布上所有图像的整体语义，使设计师能够获得向受众解释情绪板的提示，或通过情绪板发现新的含义。

上述两项研究的显著之处在于，它们通过密切观察用户执行情绪板创作任务时的行为，以务实的方式融入用户反馈。然而，先前的研究都未假设一种语义空间模型，让设计师能够相对于目标设计概念的文字表征来定位自己的构思。

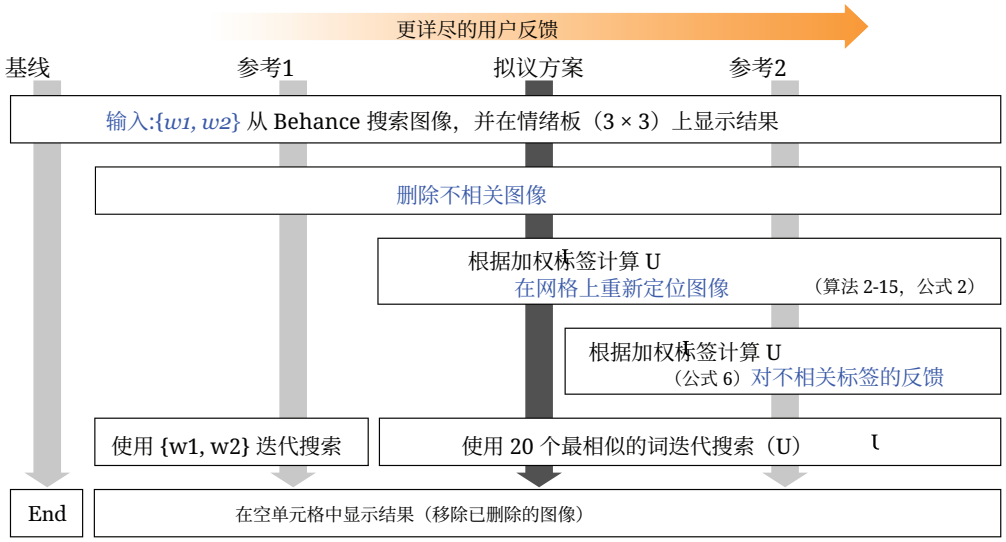

图 2：情绪板构建器（MBC）算法的变化。蓝色文本显示用户的操作

3 方法

MBC 是一种人工智能辅助交互式网络应用，可帮助设计师根据设计概念短语（DCP）[13]构建情绪板。MBC 构建的情绪板由 ×b 个网格单元组成。近期的一些研究表明，研究中的参与者通常会为每个情绪板处理 5 至 12 张图像[12, 10,14]。Aliakseyeu 等人 [15] 通过采用不同大小的数字图像堆进行实验组，比较人类在导航、重新定位和重组任务中的表现，发现两种不同图像堆大小（15 和 45）之间的任务表现存在显著差异。我们选择了九张图像，并采用 3 × 3 网格，以便用户快速浏览和进行迭代，同时保留通过这些图像表达原始设计概念的能力。我们还从实验安排的角度考虑了情绪板的大小，因为我们计划开展大规模参与者实验。

MBC 旨在供希望通过视觉呈现探索并传达其设计概念的概念设计师使用。它还希望以角色空间（CS）和设计概念

字符空间构建

短语（DCP） [13], 为基础，让用户在词汇语义空间中探索设计概念。MBC 接受一个 DCP，并从 Adobe Behance[16],（一个面向专业及商业艺术家的网络作品集服务）中搜索图像。用户界面呈现角色空间（CS）的右上象限，该象限由语义轴上的词 1 和词 2 作为属性构成。因此，情绪板中的图像代表了 CS 右上象限中表达的设计概念（图 1-C、D）。

拟议的 MBC 系统旨在鼓励用户反复探索图像，直到他们对情绪板的整体构成感到满意。多项研究发现，设计过程包括内部思考与外部可视化[17] ，以及不断改进的迭代式心理与手动过程[18] [19]。各种成本因素可能会阻碍这些迭代过程，例如收集素材所需的时间和精力、尝试不同搜索查询和资源，以及摸索构图方式。Edwards 等人 [3] 对实体与数字情绪板的制作开展了一项比较研究。他们指出，即使是在制作数字情绪板时，数字资源提供的图像数量庞大，也可能抑制迭代过程。他们进一步认为，图像一经选定，用户便会建立起信心，觉得继续搜索新素材也无济于事。这是由一种确认偏误造成的：在一系列信息中，越早出现的信息会受到越多重视[20]。为验证迭代过程的积极作用并克服迭代成本，我们以以下方式设计了实验组。

我们首先设置了一项比较实验，将不涉及用户迭代创作情绪板的基准系统与拟议系统进行比较；拟议系统允许用户按需进行迭代。我们的目标是实现低成本、高参与度交互，让用户能够轻松尝试合适的迭代次数，从而在情绪板创作任务中获得最佳体验。此外，我们还开发了多个版本的 MBC，作为支持不同交互式反馈功能的可比参考，以便找出有前景的因素。下一节将讨论这些功能和算法的细节。

3.1 系统概述

情绪板创作工具由一个使用 JavaScript、HTML 和 CSS 构建的前端 Web 应用，以及一个使用 Python 构建、托管在 Amazon Elastic Compute Cloud（EC2）上的后端 Web 服务器组成[21]。除了不支持迭代的基准系统外，我们还开发了所提算法，以及分别在独立系统中实现的两种参考算法。因此，工具因素共有四种变体。图 1 展示了各变体所采用流程的整体差异。

3.2 基线搜索流程

算法 1 和图 2（基线）展示了基线搜索流程。系统首先接收用户在 UI 右上方两个搜索框中输入的查询（Q），即两个形容词（w_1,w_2）。按下“START”按钮后，系统会在 Behance 的三个“领域”中搜索图像，分别为“工业设计”“建筑”和“时尚”。之所以选择这三个领域，是因为形式和 CMF（色彩、材料与表面处理）等风格元素对这些领域的设计十分重要，而且其中许多图像都以视觉方式呈现了这些元素。此外，网页设计、平面设计等其他领域通常包含标志和文案等文本信息。包含会让人联想到特定品牌或图像的文本信息的图像，不适合作为情绪板素材。候选图像（图像）会按相关性排名，并按领域排序。随后，排名前九的图像会随机分配到情绪板的 3×3 图像集（D）中的空网格（网格）内。本次操作到此结束，用户可以将情绪板导出为 PNG 文件并保存到本地客户端。

3.3 拟议方案算法 - 平均向量计算的查询更新

所提算法（算法 2）会根据用户反馈修改查询。对于当前情绪板上的每张图像，系统都会从 Google Vision API[22]获取语义标签。Vision API 使用预

算法 1：基线搜索情绪板生成

```
1：函数 基线_搜索
2:      Q :=[], D :=[], 图像 :=[]
3:      网格:=
4:  [(3, 3), (2, 3), (3, 2), (2,    2)、(1, 3)、(3, 1)、(1, 2)、(2, 1)、(1, 1)]
5:
6:      Q：= 输入 (w1 +w2)
7:      从 Behance 获取图像(Q)
8:      按相关性排序前_9个图像
9:      对于 D 中的每个网格
10:         网格.append(随机(图像))
11:     结束循环
12:     导出(D)
13: 结束函数
```

先训练好的机器学习模型为图像分配标签，并将其分类到数百万个预定义类别中。所提系统会获取情绪板上每张图像置信度分数最高的五个标签。设 D (d 1, d 2, ..., d $_m$) 为当前情绪板上的图像集，其中 d $_i$ 是情绪板上的第 i 张图像，L^i (l $^i_1$, l $_i$2, ...l $_{ik}$) 为每张图像的标签，其中 l $_{ij}$ 是图像 d $_i$的第 j 个标签；S $_i$ (s $_i$1, s $_i$2, ..., s $_{ik}$) 为 Vision API 为每个标签分配的置信度分数，其中 s^i_j 是图像 d $_i$的第 j 个标签的分数。对于情绪板 D 上每张图像 d $_i$ 所包含的图像标签集 L^i 中的每个图像标签 l^i_j，系统都会使用 Concept Net Numberbatch 词嵌入为标签分配向量。设 V $_i$ (v $_i$1, v $_i$2, ..., v^i_k) 为标签 L^i (l $_i$1, l $_i$2, ...l^i_k) 的向量，其中 v $_{ij}$ 是图像 d $_i$的第 j 个标签的向量。图像 d $_i$ 的平均向量 v $_i$ 可按如下方式计算：

$$\begin{aligned}\bar{\mathbf{v}}_i &= \frac{(s^i_1 \boldsymbol{v}^i_1 + s^i_2 \boldsymbol{v}^i_2 +, ..., s^i_k \boldsymbol{v}^i_k)}{k} \\ &= \frac{1}{k}\sum_{j=1}^{k}\{s^i_j \boldsymbol{v}^i_j\}\end{aligned} \tag{1}$$

其中，k 是图像 d $_i$的标签总数。

算法 2 拟议方案（更新查询）

```
1: 函数 NewQuery
2:      查询new := []
3:      L、S、V：= []
4:      权重函数：= []
5:      v̄i := [], Weighted v̄i := []
6:      U：= []
7:
8:      对每个 di ∈D 执行
9:          L.append(VisionAPI(di))
10:         S .append(VisionAPI(di))
11:         Wt.append(OnDropWeight(x, y))
12:         对每个 li ∈ L 执行
13:             V.append(ConceptNet向量(li, si))
14:             如果余弦相似度(l i, w1) > 余弦相似度(l i, w2)，则
15:                 加权vi = v̄i × 权重函数(β)
16:             else
17:                 加权vi = v̄i × 权重函数(α)
18:             结束如果
19:         结束循环
20:         U := 均值(Weighted v̄ i)
21:     结束循环
22:     Qnew.append(MostSimilarWords(U))
23：结束函数
```

拟议的系统允许用户在情绪板的 3×3 矩阵上重新定位图像。此操作将图像标签分为两类：w_1_标签和 w_2_标签，从而确定哪些图像标签应朝词语 1 或词语 2 的语义方向增强。随后，只将分配给每个网格的其中一组位置权重 $Wt(\alpha,\beta)$（图 3）乘到被归类为相应类别的标签上。分类时，会比较每个标签与 w_1 和 w_2 向量的余弦相似度（余弦相似度）（算法 2-14）。例如，假设某个标签向量与 w_1的含义更相似。在这种情况下，该标签会被归类为 w_1 标签，并且该标签向量只乘以位置权重对 $Wt(\alpha,\beta)$ 中的 β 值（y 轴一侧的 w_1 ）。这样，用户将图像沿矩阵中的特定方向重新定位，就能向系统提供反馈（图 1）。实际上，系统会在后续搜索中检测用户增强特定语义的意图，而无需显式修改查询。重新定位后的图像 d_i的加权 位置加权平均向量 v^i 将按算法 2（14-17）所述进行更新。至于 3×3 网格中每个位置对应、将与标签向量相乘的成对权重，我们通过若干初始查询测试了两种方案。图 4 展示了我们采用的权重数组。它既能使图像与用户意图保持较为接近，又能拓展语义空间以供探索。

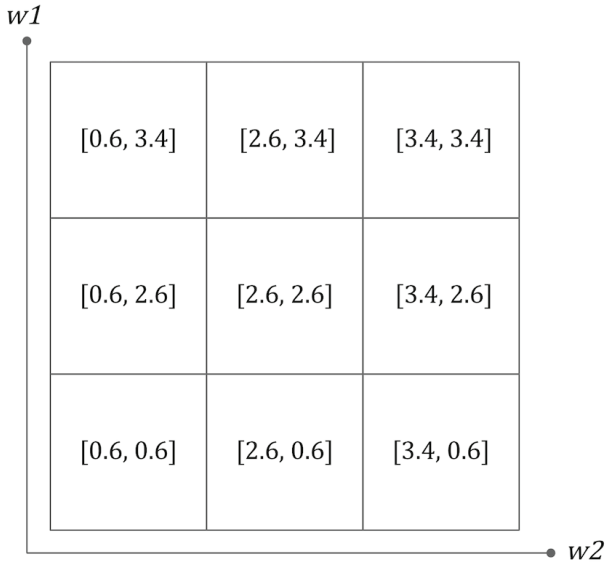

图 3：情绪板矩阵上位置权重 $Wt(\alpha,\beta)$ 的配对。这些权重在将图像放置到 (x, y) 坐标时赋予。

更新新查询之前的最后一步，是获取板上所有当前图像的平均向量，可按如下方式计算。设 U 为板上图像 $\{Weighted\ \underline{v}_1, 的加权\ \underline{v}_2, ...、\ 加权 \underline{v}_m\}$ 向量的平均值。

$$\mathbf{U} = \frac{1}{m}\sum_{i=1}^{m} Weighted\ \overline{\mathbf{v}}_i \tag{2}$$

其中，m 是当前情绪板上的图像数量。

3.3.1 计算最相似的词

为了更新查询以进行下一次搜索，系统会根据输入获取最相似的前 20 个词；在本例中，输入为 U，即情绪板 D 上所有图像的加权向量的平均向量。系统会计算归一化输入向量的余弦相似度（余弦相似度），并输出余弦相似度最高的前 N 个词。此函数

通过 Python 包 gensim.models[23] 中的一个方法实现。该方法按以下步骤，计算给定词的投影权重向量的简单均值与模型中每个词对应向量之间的余弦相似度。

设预训练模型中的词数为 N，预训练模型中归一化向量的维度为 M；那么，已训练模型中各词对应的 $N \times M$ 矩阵为 $X = (x_1, x_2, ..., x_N)$，输入词可表示为 M 维向量 y。X_1 与 y 之间的余弦相似度可计算如下。

$$CosSim(x_i, y) = \frac{x_i \cdot y}{\| x_i \| \cdot \| y \|} \tag{3}$$

其中，X_1 是归一化向量，因此 $\| x_i \| = 1$。所以，余弦相似度的计算可改写为：

$$CosSim(x_i, y) = \frac{x_i \cdot y}{\| y \|} \tag{4}$$

此外，与逐元素计算相比，使用矩阵向量乘积计算余弦相似度更加高效。因此，最终的 M 维余弦相似度向量 S 使用矩阵向量乘积计算如下。

$$\boldsymbol{S} = \frac{X \cdot y}{\| y \|} \tag{5}$$

3.4 参考1 - 无查询修改

参考1（图2）算法涉及最简单的迭代过程。用户可以点击垃圾桶图标（图1(E)）删除不相关图像，并按下“下一步”按钮（图1(B)）尝试下一次搜索。系统会用新图像填充空白单元格。该过程可以重复，直到用户对情绪板感到满意为止。NewQuery 函数尚未实现，并会继续使用相同的初始查询$Q(w_1 + w_2)$。

3.5 参考2 - 涉及标签反馈的查询修改

除了拟议算法之外，参考2版本还接受用户对语义标签的反馈（图2）。用户点击每张图像上的铅笔图标（图1(F)）后，与每张图像关联的语义标签会显示在右侧栏中（图1(I)）。用户可以划掉任何不相关的标签，将其排除。该方法使正向词对相似度计算产生正向影响，使负向词产生负向影响。令 n 为标签中被划掉并作为负向词的词数，例如图1(I)中的“沙发”和“照明”。由于 m 是当前情绪板上具有位置加权平均向量的图像数量，因此新的 U 是情绪板上 m 张图像与 n 个负向词的平均向量。根据模型中相似度最高的前N个词，新的 U 可通过将 1 乘以 m 个向量、将 -1 乘以 n 个向量，再除以向量总数 $m + n$ 来计算：

$$\begin{aligned} \mathbf{U} &= \frac{1}{m+n} m \overline{\mathbf{v}}_i * 1 + n \, v^{negative} * (-1) \\ &= \frac{m \overline{\mathbf{v}}_i - n \, \boldsymbol{v}^{negative}}{m+n} \end{aligned} \tag{6}$$

b 将平均向量 U 缩放至单位长度 U 后，系统会计算 syn0norm（ConceptNet Number Batch 词嵌入中所有词的归一化向量集合）的点积，得出与平均向量最相似的前N个词。

4 实验设计

为探究情绪板创作系统对情绪板创作任务的支持效果，我们通过在线招募平台和问卷平台开展了一项大规模参与者实验。以下是实验组的具体设置。

4.1 参与者与自变量

通过 Prolific 招募了251名工作职能为“艺术、设计、娱乐和休闲”且英语流利的参与者。其中32人（12.75%）因系统故障或未知原因未完成研究。最终共有219名参与者完成了研究（112名男性、96名女性、8名非二元性别者），平均年龄为33.16岁（σ = 11.07）。所有完成研究的参与者均使用自己的笔记本电脑或台式电脑。完成研究的参与者获得了12美元的报酬。所有参与者都被要求使用同类型的 MBC 系统完成两次情绪板创建任务；因此，我们收集了438个案例。参与者间因素是所使用 MBC 系统的差异（图2），参与者内因素则是分配给他们、用作初始查询 Q 的两个不同设计概念短语（DCP）。这两个 DCP 中纳入的因素是 DCP 中单词1与单词2之间的余弦相似度。近端 DCP 为“符合人体工学且舒适”，远端 DCP 为“放松且熟练”。这两个 DCP 的余弦相似度分别为0.4528和0.0053。他们在两项任务中使用的 DCP 顺序以随机、平衡的顺序分配。

4.2 因变量

我们使用创造力支持指数（CSI）[24] 作为任务后心理测量工具，对比四种条件（包括一个基线情绪板创作系统和三个实验组情绪板创作系统）在情绪板创作任务中对创造力的支持情况。CSI 可通过针对六个因素的心理测量量表，对用户的认知过程进行量化。这六个因素分别是：探索、表现力、沉浸感、愉悦感、成果与付出的努力相称，以及协作。CSI 评估的是创造力支持工具本身，关注用户使用工具进行创作的体验，而非试图直接评估创意产品的某项属性。相反，CSI 会结合用户付出的努力来评估创作成果，例如“我对从这个系统或工具中获得的成果感到满意。”这适用于面向有经验用户设计的工具，因为他们了解创意成果应是什么样，也知道理想的创作体验是什么；而面向新手用户设计的工具则不同，这类用户的创造力尚显不足，因此需要支持。CSI 还具有灵活性，可长期应用于各种工具和场景，并能提供标准化测量。CSI 的设计也十分稳健：六个因素各有两条不同的陈述，有助于提高 CSI 所采用调查问卷的统计功效。结合针对六个因素逐一开展的配对因素比较，CSI 能够揭示创造力支持的哪些方面可能需要关注。

CSI 具有一套严谨的协议，研究人员应遵循该协议，以确保测量和分析具有普适性与可靠性。例如，我们的 CSC 工具并非为协作任务而设计；然而，CSI 协议不建议研究人员跳过协作因素中的陈述，而是允许参与者对这些协作陈述进行评分。相反，CSI 协议允许在属于协作因素的陈述中添加“N/A”选项，我们已将其纳入调查问卷。

此外，我们在每项任务开始前和结束后立即使用单项测量——油箱问卷（GTQ）[25], ——来评估剩余心理资源。GTQ旨在测量参与者执行任务时的认知负荷，同时避免通过提出多个问题给他们增加负担。Hart 和 Staveland [26] 证明，广泛使用的工作负荷指标 NASA-TLX 会增加参与者完成问卷时的工作负荷，因此他们开发了 GTQ 作为一种低负担替代方案，尤其是在多次测量时，不会对测量本身造成负面影响。GTQ 提出一个问题：“想象你的大脑是一台发动机。请滑动下方的油量指示器，显示你现在还剩多少油。”在我们的实验中，GTQ 问题在每项任务开始前立即提出，并

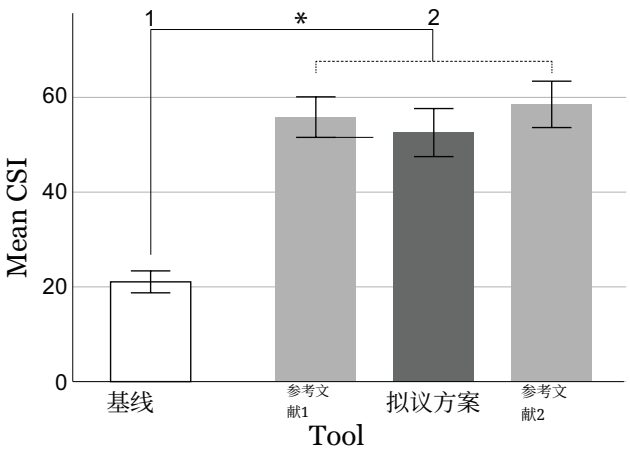

图 4：各工具的平均创造力支持指数。误差线表示标准误。

在每项情绪板创作任务结束后立即再次提出，请参与者滑动量表。量表范围为 0 到 100。我们将任务前后的油箱数值之差作为衡量执行任务所消耗心理资源的数值。

4.3 刺激材料与任务

所有组别的参与者均通过 Prolific 招募，并被引导至 SurveyMonkey 问卷平台。他们被随机分配到八个组别中的任意一组（针对每个基线、拟议方案、参考1和参考2类型的工具，分别设置两个采用不同距离 DCP 的平衡组），并接受有关实验的说明。在获得每位参与者的同意后，研究人员提供了相关说明。这些说明既以嵌入 YouTube 的视频形式提供，也以 Google 云端硬盘上的文本文件形式提供。参与者可以选择使用两种说明或其中任一种，以确保他们理解任务的性质以及如何使用情绪板创作系统。

研究人员通过网页链接向参与者提供情绪板创作工具及 DCP。参与者需要将其创建的情绪板下载到本地计算机，并将其上传至 SurveyMonkey 上的问卷。随后，他们完成了所有 CSI 问卷，接着完成第二份任务前 GTQ、使用同一工具和另一个 DCP 的第二项任务，以及第二份任务后 GTQ。最后，他们完成了成对因素比较，以便为两项任务中的各类 CSI 评价赋予权重。

4.4 参与者概况

我们在设计 MBC 时以专业设计师为对象。我们关注考察 CSI 评分与从业年限之间的关系，以及参与者执行 MBC 所支持的情绪板创作任务的频率。对于前者，我们询问参与者在“艺术、设计、娱乐和休闲”类工作中的从业年限。对于后者，我们将参与者执行情绪板创作任务的频率作为分类变量进行询问，具体如下：每季度少于一次或从不、每季度一次、每月一次、每周一次以及每天一次。

5 结果

请注意，在全部438个案例中，有13个案例被取消资格，因为他们在整个调查问卷中对CSI问卷的回答分数完全相同（全为0或全为10）；另有两个案例因未上传有效的情绪板而被排除。还有三个案例作为离群值被剔除，因为其CSI分数与均值的偏差超过两个标准差（无论高于还是低于均值），最终用于分析的案例数为420个。

表1：各组的 Tukey HSD 事后比较。

Tool	均值差（p）		
	拟议方案	参考文献1	参考文献2
基线	−30.66 (< 0.01**)	−33.13 (< 0.01**)	−36.05 (< 0.01**)
拟议方案		−2.46 (0.861)	−5.38 (0.324)
参考文献1			−2.92 (0.778)

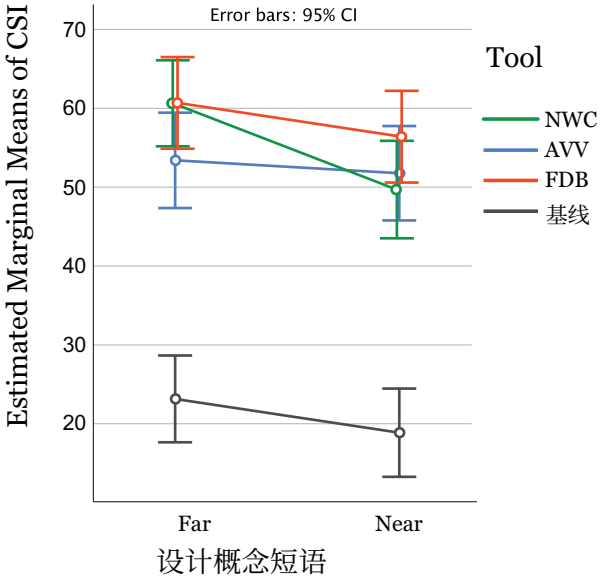

图5：两个不同 DCP 的平均创造力支持指数方差。远端 DCP 为“放松且熟练”（余弦相似度 = 0.0029），近端 DCP 为“符合人体工学且舒适”（余弦相似度 = 0.4528）

5.1 创造力支持指数

图4显示了各工具对应的CSI分数均值。单因素方差分析表明，所有实验组（拟议方案、参考1和参考2）构成同质子集，其平均CSI分数分别为55.83（σ =21.92）、52.57（σ = 24.98）和58.54（σ = 24.95），显著高于基线组（21.05，σ = 12.36）（ α = 0.05的子集）。表1展示了通过Tukey HSD进行的事后多重比较。均值差表示行中工具对应的CSI分数减去列中工具对应的CSI分数。

拟议方案与参考工具之间的 CSI分数没有显著差异 .

关于参与者使用工具完成任务时所进行的迭代次数差异，在最不完善的工具——即参考1——下，参与者进行迭代的次数显著更多（M = 7.08，σ = 5.08），高于使用拟议方案子集（M = 4.35，σ = 4.04）和参考2（M = 4.56，σ = 3.90）时的次数。至于参与者完成任务所消耗的心理资源，以任务后 GTQ 与任务前 GTQ 之差进行测量，基线（−0.86，σ = 9.9）与所提工具（2.85，σ = 8.74）之间存在显著差异（p = 0.04）。基线（−0.86，σ = 9.9）与参考2（2.57，σ = 13.99）之间也观察到差异，并呈现显著趋势（p = 0.065）。

5.2 参与者内因素

图5展示了两种不同 DCP 下工具的 CSI分数估计边际均值的方差。无论使用基线还是实验组中的工具，远端 DCP 下的总体 CSI分数（49.02，$\sigma = 25.04$）均显著高于近端 DCP 下的分数（43.41，$\sigma = 27.62$）（$p = 0.04$）。

6 讨论

所提算法允许参与者以交互方式反复进行图像搜索，CSI分数表明，该算法有助于情绪板创作任务中的创造力。然而，我们希望将所提算法与更复杂和更简单的反馈算法进行比较，以期找出最优化的方案，尽管各算法之间并无显著差异。三种实验工具的任务前和任务后 GTQ数值表明，使用所提算法和参考2算法时，用户可能会因必须在 MBC 上执行的操作过于复杂而感到疲惫或不知所措。不过，根据工具计算的加权结果价值与付出因素评分，参考2（$M = 41.88, \sigma = 22.48$）和参考1（$M = 40.47, \sigma = 21.58$）的评分显著高于单独使用所提工具（$M = 32.79, \sigma = 22.43$）的评分（$\alpha = 0.05$子集）。这表明，用户可能认为参考1（删除不相关图像）和参考2（标签反馈）的流程价值更为明确。另一方面，所提系统可能没有让用户清楚了解如何在网格上重新定位图像。参考2系统会在情绪板的每张图像上附加标签，因此用户实际进行的操作更加直观：划掉标签，使其不再看到与这些标签关联的图像。

透明度以及提供更清晰的指引来帮助用户理解系统的工作原理，这一点至关重要。实际上，在实验开始前，我们如何向参与者解释系统，会显著影响他们使用数字工具进行创作的体验。对于参考2算法，我们以文字和视频的形式向参与者提供了以下说明：“如果您不喜欢某张图像，可以将与该图像相关的标签从搜索池中彻底移除。点击图像上的铅笔图标，并划掉您不想看到的标签。系统在之后的搜索中会避免返回包含这些标签的图像。”然而，结合我们的方法来看，这种解释并不准确。我们采用了公式 6 中的词嵌入向量计算方法：将收到负反馈的 n 个向量乘以 -1，再将当前情绪板中每张图像的 m 个平均向量乘以 1，然后除以 m+n。这相当违反直觉，因为标签级别的向量会从图像向量中减去。该方法运用了向量空间词表示的原理，即可以对词向量进行代数运算，从而使“King”向量 - (“Man”向量) + “Woman”向量得到的结果最接近“Queen”一词的向量表示[27, 28]。然而，准确地向参与者解释这一原理的运作方式，可能会令他们望而生畏，甚至打消他们使用该功能的念头。即使是“反馈词会对下一次搜索产生负面影响”这样含糊的解释，也说不通。

如果创造力支持效应足够显著，即使存在一些模糊性和不透明性，参与者也可能会留下深刻印象。然而，在本实验中，基线与所提出的工具之间 GTQ 值的显著差异表明，创造力支持效应还不足以抵消任务负荷和模糊性的增加。

就被试内因素而言，无论使用何种工具，我们都观察到远端 DCP 显著优于近端 DCP。这意味着，在用户进行视觉任务时，远端 DCP 默认可能产生更具创造性的成果，且更容易被识别。事实上，CSI 的部分因素得分存在显著差异；成果值得付出努力（$p = 0.037^{*}$）：远端 DCP（$M = 36.98, \sigma = 21.26$），近端 DCP（$M = 32.42, \sigma = 23.12$）；愉悦感（$p = 0.042^{*}$）：远端 DCP（$M = 26.48, \sigma = 22.72$），近端 DCP（$M = 22.00, \sigma = 22.29$）。请注意，表现力也呈现出差异趋势（$p = 0.055$）。

此外，我们检索并分析了参与者实验的日志数据，以了解在情绪板创作任务中表现优异者和低绩效者的情况。对于采用所提出的参考

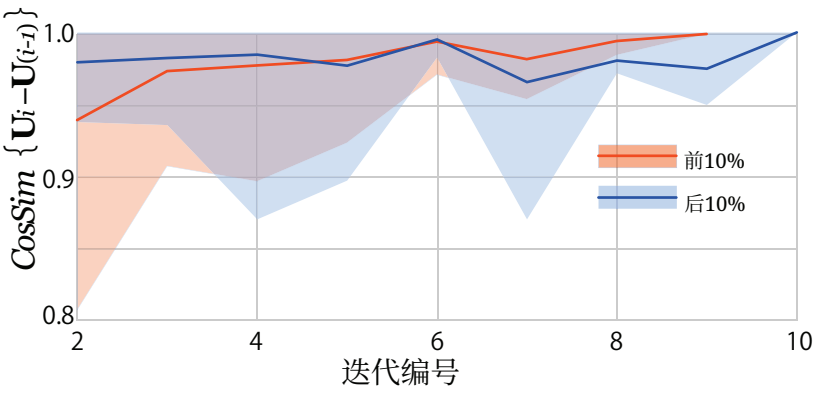

图6：按 CSI 分数划分的各次迭代之间余弦相似度（CosSims）的范围和均值。CSI 前10%的案例在初期倾向于探索更广阔的语义空间，并随着迭代推进不断收敛。CSI 后10%的案例表明，各次迭代之间的平均相似度一直较高，探索较少，可能因此导致结果不尽如人意。

1和参考2算法的情绪板创作工具，我们实现了系统日志，用于记录每次迭代中的以下数据：w_1 和w_2, 迭代编号（初始搜索（w_1, w_2）除外）、图像标签：L^i、置信度分数：S_i、余弦相似度$\{w_1, v_i(a, b)\}$、余弦相似度$\{w_2, v_i(\bar{a}, b)\}$、余弦相似度$(\bar{w_1}, U)$、余弦相似度(w_2, U)、时间戳、U 的前 –N– 个词，以及负反馈词。

图6展示了情绪板平均向量 $U_i - U_{(i-1)}$ 之间余弦相似度（CosSim）的变化，其中 U_h 是第 i 次迭代中的情绪板平均向量。红线表示 CSI 分数前10%的案例在多次迭代中的余弦相似度变化，表明情绪板语义在迭代初期变化较大，随后逐渐减小。这说明这些案例中的参与者在初期更换了较多图像，临近迭代结束时更换的图像较少，以进行细微调整并完成情绪板。另一方面，蓝线表示 CSI 分数后10%的案例在多次迭代中的余弦相似度变化，说明情绪板的语义从初期到结束都没有太大变化。这意味着这些案例中的参与者一直只更换情绪板中的一小部分图像，直到迭代后期也未能真正获得由九张图像组成的满意集合。前10%参与者的迭代次数均值（$4.11, \sigma = 2.08$）显著低于后10%参与者（$6.69, \sigma = 4.70$）（$p = 0.05^{*}$）。前10%参与者第2次迭代与第1次迭代之间的余弦相似度均值（$0.940, \sigma = 0.0497$）显著较低（$p < 0.01^{**}$），意味着两者之间的距离大于后10%参与者（$0.979, \sigma = 0.0234$）。

7 结论

通过试用不同交互程度的 MBC工具，我们确认了用户反馈的有效性，它能让概念设计师更投入地完成情绪板创作任务。我们通过提供利用二维语义空间的自适应查询更新，为计算创造力工具领域作出了贡献；在该空间中，用户可以重新排列情绪板上的图像。将用户对语义标签的反馈纳入其中的工具版本同样有效。我们对 CSI 和 GTQ问卷因素得分进行的事后分析表明，从用户视角来看，行动与结果之间关系的透明度具有优势，尽管这增加了工作量。我们还观察到，初始口头查询的特征是影响用户是否觉得自己在处理的概念具有创造力的重要因素。对表现最佳者和最差者的分析揭示了语义收敛过程在各次迭代中的差异。表现较好者似乎倾向于较早进行探索，之后再加以利用。

参考文献

[1] Andrés Lucero. 构建框架、对齐、悖论化、抽象化与引导：设计情绪板如何运作。载于《交互系统设计会议论文集》，第 438–447 页，2012 年。doi: http://dx.doi.org/10. 1145/2317956.2318021。[2] Steve Garner 和 Deana McDonagh-Philp。通过视觉刺激阐释和解决问题：情绪板在设计教育中的应用。《艺术与设计教育杂志》，20(1):57–64，2001 年。doi: http: //dx.doi.org/10.1111/1468-5949.00250。[3] A Edwards、SA Fadzli 和 Rossitza Setchi。实体与数字情绪板开发的比较研究。载于 Anais. . . 第五届创新生产机器与系统国际会议。英国卡迪夫，2009 年。[4] Tracy Cassidy。将情绪板流程建模并理解为定性设计研究工具。《时尚实践》，3(2):225–251，2011 年。doi: http://dx.doi.org/10.2752/175693811X13080607764854。[5] Tian-tian Li 和 Dan-hua Zhao。基于高层语义视觉表征、促进创意设计的情绪板工具。载于应用人因与工效学国际会议，第 408–419 页。S pringer，2021 年。doi: http://dx.doi.org/10.1007/978-3-030-79760-7_49。[6] Deana McDonagh 和 Ian Storer。情绪板作为设计催化剂与资源：对一个研究不足领域的探究。《设计期刊》，7(3):16–31，2004 年。doi: http://dx.doi.org/10.2752/146069204789338424。[7] Carole Bouchard、Roger Camous 和 Ameziane Aoussat。中间表征（IR）在设计过程中的性质与作用：汽车设计案例研究。《国际车辆设计杂志》，38(1):1–25，2005 年。doi: http://dx.doi.org/10.1504/IJVD.2005.006602。[8] Rossitza Setchi、Qiao Tang 和 Ivan Stankov。支持概念设计的基于语义的信息检索。《先进工程信息学》，25(2):131–146，2011 年。doi: http://dx.doi.org/10.1016/j.aei.2010.07.006。[9] Arnold WM Smeulders、Marcel Worring、Simone Santini、Amarnath Gupta 和 Ramesh Jain。早期发展阶段结束时的基于内容的图像检索。《IEEE模式分析与机器智能汇刊》，22 (12):1349–1380，2000 年。doi: http://dx.doi.org/10.1109/34.895972。[10] Janin Koch、Andrés Lucero、Lena Hegemann 和 Antti Oulasvirta。人工智能？通过协作式上下文老虎机进行设计构思。载于《2019年计算机系统中的人因国际会议论文集》，第 1–12 页，2019 年。doi: http://dx.doi.org/10.1145/3290605.3300863。[11] Sé bastien Bubeck 和 Nicolò Cesa-Bianchi。随机与非随机多臂老虎机问题的遗憾分析。《机器学习基础与趋势》，5(1):1–122，2012 年。doi: http://dx.doi.org/10.1561/ 2200000024。[12] Janin Koch、Nicolas Taffin、Andrés Lucero 和 Wendy E. Mackay。语义拼贴：通过语义标签丰富数字情绪板设计。载于《2020年ACM交互系统设计会议论文集》，第 407–418 页。计算机协会，2020 年。doi: https://doi.org/10.1145/3357236.3395494。[13] Shin Sano 和 Seiji Yamada。通过字符空间构建进行人工智能辅助设计概念探索。《心理学前沿》，12:6722，2022 年。[14] Cristina Nardin Zabotto、Daniel Capaldo Amaral、Costa Janaina Mascarenhas Hornos、Benedito Galvão Benze 等。借助感性工学连接用户与设计师的自动数字情绪板。《国际工业工效学杂志》，74:102829，2019 年。doi: http://dx.doi.org/10.1016/j.ergon.2019.102829。[15] Dzmitry Aliakseyeu、Sriram Subramanian、Andrés Lucero 和 Carl Gutwin。在数字桌面上与成堆的物件交互。载于《先进视觉界面工作会议论文集》，第 159–162 页，2006 年。doi: http://dx.doi.org/10.1145/1133265.1133298。[16] Michael J Wilber、Chen Fang、Hailin Jin、Aaron Hertzmann、John Collomosse 和 Serge Belongie。Bam! 用于超越摄影的识别的 Behance艺术媒体数据集。载于《IEEE国际计算机视觉会议论文集》，第 1202–1211 页，2017 年。doi: http://dx.doi.org/10.1109/ICCV.2017.136。

[17] Pierre Sachse 和 Marco Furtner。《设计中的具身知识》，载于《认知主义之后》，第 163–179 页。Springer，2009。doi: http://dx.doi.org/10.1007/978-1-4020-9992-2_10.[18] Vincent Rieuf、Carole Bouchard、Vincent Meyrueis 和 Jean-François Omhover。《早期沉浸式设计中的情绪活动：虚拟现实中的草图与情绪板》。Design Studies，48:43–75，2017。[19] Karen Renaud 和 Judy Van Biljon。《最大限度提升知识可视化沟通能力的框架》。载于《2019 年南非计算机科学家与信息技术专家协会会议论文集》，第 1–10 页。计算机协会，美国纽约州纽约市，2019。ISBN 9781450372657。doi: http://dx.doi.org/10.1145/3351108.3351111。[20] Jonathan Baron。《思考与决策》。剑桥大学出版社，2000。doi: http://dx.doi.org/10.1017/ CBO9780511840265。[21] Sangho Yi、Derrick Kondo 和 Artur Andrzejak。《通过检查点机制降低亚马逊弹性计算云竞价实例的成本》。载于《2010 年 IEEE 第三届云计算国际会议》，第 236–243 页。IEEE，2010。doi: http://dx.doi.org/10.1109/CLOUD.2010.35。[22] Shih-Hsin Chen 和 Yi-Hui Chen。《一种基于 Google Cloud Vision API 和 WordNet 的基于内容的图像检索方法》。载于《亚洲智能信息与数据库系统会议论文集》，第 651–662 页。Springer，2017。doi: http://dx.doi.org/10.1007/978-3-319-54472-4_61.[23] Bhargav Srinivasa-Desikan。《自然语言处理与计算语言学：使用 Python、Gensim、spaCy 和 Keras 进行文本分析的实用指南》。Packt 出版社有限公司，2018。ISBN 9781788837033，1788837037。[24] Erin Cherry 和 Celine Latulipe。《通过创造力支持指数量化数字工具对创造力的支持》。ACM《计算机与人类交互汇刊》（TOCHI），21(4):1–25，2014。doi: http://dx.doi.org/ 10.1145/2617588。[25] Samuel S Monfort、John J Graybeal、Amanda E Harwood、Patrick E McKnight 和 Tyler H Shaw。《剩余心理资源的单项评估：油箱问卷（GTQ）的开发与验证》。《人体工效学科学理论问题》，19(5):530–552，2018。doi: http://dx.doi.org/10.1080/1463922X.2017. 1397228。[26] Sandra G Hart 和 Lowell E Staveland。《NASA-TLX（任务负荷指数）的开发：实证与理论研究结果》。载于《心理学进展》，第 52 卷，第 139–183 页。Elsevier，1988。doi: http: //dx.doi.org/10.1016/S0166-4115(08)62386-9。[27] Tomas Mikolov、Kai Chen、Greg Corrado 和 Jeffrey Dean。《向量空间中词表示的高效估计》。arXiv 预印本 arXiv:1301.3781，2013。doi: https://doi.org/10.48550/arXiv.1301.3781。[28] Tomáš Mikolov 、Wen-tau Yih 和 Geoffrey Zweig。《连续空间词表示中的语言规律》。载于《2013 年美国计算语言学协会北美分会会议：人类语言技术论文集》，第 746–751 页，2013。