



# 结合设计经验的图像美学评分机制：基于共识建模的跨学科研究综述

## 引言：图像美学评价的重要性与行业现状

图像美学质量的自动评估是计算机视觉领域的一个重要课题，不仅因为审美评价具有高度主观性和复杂性，还因为其在实际应用中价值巨大<sup>1 2</sup>。随着社交媒体、电子商务和内容生成的蓬勃发展，每天有海量图像涌现，如何从中挑选出视觉上令人愉悦、符合用户偏好的图像变得至关重要<sup>3 4</sup>。研究表明，在电子商务情境下，如果将商品图片按美学评分排序呈现，用户点击率可提高逾50%<sup>2</sup>；又如在社交媒体和照片管理中，具有更高美学品质的图片往往能获得更多的关注和互动。这些现象说明**图像美学评价**对于提升用户体验和商业转化有直接影响。

大型科技公司也日益重视图像的美学质量控制。例如，Amazon在训练其生成式图像模型时，引入了一个模拟人类审美判断的图像质量分类器，对海量候选训练图像进行筛选，只有美学得分高于阈值的高质量图片才被用于模型训练<sup>5</sup>。Google则通过研究提出了算法来自动预测图像的审美得分，并将其应用于照片编辑和内容优化等场景：如Google的NIMA模型能够可靠地为图像打分且与人类评价高度相关，可用于智能照片编辑、提升用户参与度的视觉优化，以及减少成像管线中的主观视觉瑕疵<sup>6</sup>。这些行业实践表明，一个有效的图像美学评分机制在**产品设计、时尚、AI生成内容、广告营销**等诸多领域都有广泛用途。各大研究机构（如Google Research、Adobe Research、Amazon Science等）近年来持续推出相关项目和论文，推动图像美学评价技术从学术走向应用，成为跨计算机视觉、心理学、设计艺术等领域交叉的热门课题<sup>7</sup>。

## 类比引入：参数建模与主观评分机制

要理解“结合设计经验的美学评分”理念，我们可以先用几个日常例子类比说明**参数化评价与主观评分**如何结合。拿奶茶来说，一杯奶茶的口味评价往往因人而异，但其客观参数（如甜度、茶浓度、奶香、温度、配料颗粒感等）可以被量化描述。调配师基于经验为这些参数设定范围和配比，从而创造出多数人喜欢的口味。然而，不同顾客对甜度或茶味的偏好权重不同，因而最终对奶茶好喝与否的主观评分存在差异。又比如香水，其气味由前中后调的香料组成（参数组合），专业调香师会根据美学经验平衡各香调的浓度，使香水散发出和谐的气息。但消费者对香味的喜好非常主观，有人钟爱花香清新，有人偏好木质浓郁，这使得同一款香水会获得不同人的不同评分。再看精酿啤酒，酿酒师通过控制酒花苦度（IBU指标）、麦芽浓度、酒精度、色泽澄清度等参数来实现风味和观感的目标，而最终的品鉴评分仍取决于品酒师和消费者的主观感受。有些啤酒口感复杂、风味独特，可能两极分化——爱的人给出高分，不习惯的人则评价低。这些例子说明：**客观参数建模**提供了质量和风格的基础框架，而**主观评分机制**体现了大众对这些参数组合效果的感知差异。换言之，在奶茶、香水、啤酒等领域，业内专家往往依据多年经验确定影响体验的关键参数和理想范围，然后通过消费者反馈来调整配方，逐步形成“**参数->主观体验->再调优**”的循环，以达成大多数人认可的品质共识。这正类似于图像美学评价中，先由设计师和视觉专家确定构图、色彩等影响美感的要素，再通过大众评分来校准模型，对应地建立起**美感共识模型**。

## 共识美学：大众评分与参数组合形成美感共识

美学评价具有强烈的主观性，不同人对同一图像的审美打分可能存在显著差异。然而，当我们收集了足够多的观众评分后，往往可以提炼出一定程度上的**共识美感评价**。所谓**美感共识**，是指通过集体主观评价汇聚而成的相对客观的审美衡量。常见的做法是邀请一群评分者为图片打分，然后取平均值作为图像的美学分数，或设定阈值将其归为高/低美感类别。但需要注意的是，简单使用平均分或二分类可能掩盖了评分者意见的多样性<sup>8 9</sup>。例如，AVA大型美学数据集中的某两张照片平均得分同为5.9左右，但一张照片的评分集中且大部分人评

价一致，而另一张照片的评分分布从1到10高度分散<sup>10 11</sup>。对于后一种情况，尽管平均分相同，却存在明显的意见分歧——不少人打了高分，同时也有不少人打了低分。如果仅看平均值，这种差异将被忽略，导致模型预测的“共识”与实际观感不符<sup>9</sup>。因此，建模图像美学时有必要考虑评分分布和主观不确定性。近期有研究提出以概率分布来表征美学偏好，例如用Beta分布拟合人类评分直方图，以量化反映“一致喜好”或“意见分歧”的程度<sup>8 9</sup>。通过这种方式，可以得到更直观的主观性指标（如“不确定概率”），帮助模型了解哪些图像是“众口一词”，哪些则是“见仁见智”<sup>8 9</sup>。

心理学与经验审美领域的研究也支持这种观点——美感具有共识与分歧并存的特性<sup>12 13</sup>。Pombo等人（2024）的实证美学研究将图片按审美评分的方差分组，发现有些图像的美丽几乎是“无可争议”的：所有人都打出类似的高分或低分，其评分分布可用单峰高斯拟合（表示高度共识）；而另一些图像的审美判断则存在显著“争议”：喜欢的人给高分，不喜欢的人给低分，整体分布呈双峰（可用两个高斯成分拟合）<sup>12 13</sup>。该研究强调，平均分并不能完全描述美的评价，评分的方差和分布形态同样是审美判断的重要组成<sup>13</sup>。换言之，在美学评价中，大众评分的分布本身就蕴含了共识度的信息：分布越集中，代表审美共识越强；分布越分散，则说明审美偏好存在显著差异甚至阵营分化。

为了在模型中更好地利用“大众智慧”形成美感共识，计算机视觉研究者们开始探索将主观评分转化为客观模型的途径。Google Research在其NIMA模型中引入了一种思路：与其将每张图简单回归一个平均分，不如直接预测整张图的评分分布直方图<sup>14</sup>。具体而言，NIMA模型输出一个1到10分的概率分布，对应不同打分的可能性，从中可计算出期望得分（即平均分）以及反映分歧程度的熵等指标<sup>14</sup>。实验证明，这种预测评分分布的方法比单纯回归平均分更符合人类的评价结果，在与真实用户偏好的相关性上也更高<sup>14</sup>。实际上，这种基于共识建模的理念正是希望既得到一个大众认可的审美评分，又能刻画出评分共识或分歧的程度。有了这样的模型，我们既可以用平均分数代表“大众美感”进行图片筛选和排序，又能够监测模型预测的不确定性，避免对那些存在明显意见分歧的图像给出武断的判断。

简而言之，共识美学评价机制通过集合多人的主观评分，在大量数据中挖掘稳定的审美评价。这为图像美学评分提供了客观性基础：当某种图像特征组合一再被多数人青睐，我们便可视其为客观上更具美感的因素。据此，可以将主观美学偏好“转译”成算法可以学习的目标，从而为后续的图像参数建模和模型优化打下基础。

## 图像参数建模：构图、色彩、景深、清晰度等维度的重要性

如果说大众评分提供了美感评价的标尺，那么图像本身有哪些客观参数影响了这一标尺的刻度？美学质量之所以可预测，在于图像的某些可量化属性与人类的审美感觉存在密切关联。大量跨学科研究表明，构图、色彩、景深、主体清晰度等视觉参数是影响图像美感的关键维度，它们可以被建模并用于解释和预测审美评分<sup>15 16</sup>。

- **构图 (Composition)**：构图反映图像中视觉元素的布局和组织，是摄影与绘画美学的核心要素之一。良好的构图能使画面产生和谐的视觉平衡和节奏感<sup>15</sup>。例如，“三分法”（Rule of Thirds）、对称与平衡（Balancing Elements/Symmetry）等都是构图中的经典规则，这些原则源自艺术设计的经验，被认为有助于提升画面的美感。一项研究指出，图像的构图信息体现了画面视觉元素之间的相互依赖关系，这种关系遵循艺术原理中的和谐原则，为学习图像美学提供了更丰富的信息<sup>15</sup>。Netflix的研究团队提出将图像密集地划分为局部区域，并构建区域组成图（region composition graph），用图卷积网络来显式建模局部区域间的依赖关系<sup>17 18</sup>。结果表明，这种显式编码构图的方法有效提升了美学评分预测的性能，使模型在AVA等美学评价基准上达到当时的最新水平<sup>19 20</sup>。此外，图像的长宽比例也是构图的一个因素：过去CNN模型往往需要固定输入尺寸，可能破坏原始构图比例。为此有研究引入自适应非整数空洞卷积来嵌入图像长宽比信息，保证在不裁切的前提下保持构图完整，实验在AVA数据集上取得了领先表现<sup>21 22</sup>。这些工作都佐证了一个事实：构图之美能够被参数化和算法化地捕捉，当模型“看懂”了画面布局的讲究之处，预测美学评分就更为准确。
- **色彩 (Color)**：色彩对人类情感和审美的影响极为直观。图像的色调搭配是否和谐（Color Harmony）、色彩是否鲜艳生动（Vivid Color）、对比是否适当，都会直接影响观看者的审美体验。在

摄影和设计中，有诸如色轮配色理论、互补色/类似色搭配原则等专业经验总结，它们对应了人眼对色彩舒适度的普遍偏好。在美学数据集中，色彩和谐经常被作为显式标注的评价维度之一。例如AADB数据集就包含“ColorHarmony”和“VividColor”两个属性，分别衡量色彩搭配的和谐程度和色彩的饱和鲜艳程度。实证表明，色彩和谐度高、色调统一且对比适宜的图像更可能获得高的审美评分；而色彩杂乱或灰暗的图像则往往得分偏低。这一点在组合多个因素的案例中尤为明显：当画面主题鲜花与天空呈现出色彩呼应且布局协调时，人们几乎都会认为该照片美感很高<sup>16</sup>。由此可见，将色彩相关的客观参数纳入模型（例如直方图分布、色调多样性指数、对比度指标等），并结合色彩理论知识进行特征设计或学习，有助于提升美学评估的解释性和准确性。

- **景深 (Depth of Field) 与清晰度 (Clarity)**：摄影美学中常常强调主体突出、背景虚化，这涉及景深控制和对焦清晰度。浅景深照片（主体清晰而背景有意模糊）往往呈现出专业而美观的视觉效果，因为虚化背景可以凸显主体、增加画面层次感和艺术氛围。AADB数据集设置了“DoF”（景深）和“Object”这样的属性标签，前者反映背景虚化程度，后者则可理解为主体是否清楚突出<sup>23 24</sup>。当景深运用得当、主要对象清晰聚焦时，观者更容易将注意力集中在主体上，产生良好的审美印象。相反，如果整个画面都过于锐利或背景杂乱无层次，往往显得平淡或凌乱，降低美感。此外，图像的清晰度还包括避免不必要的运动模糊（MotionBlur属性）和确保适当的光影细节（Light属性）等。总体来说，清晰且有层次的图像更容易得到大众青睐，这些因素也可以通过计算视觉算法提取（如边缘锐度、背景复杂度、噪声水平等）并在美学模型中加权考虑。

除了上述主要维度，图像美学评价还涉及其他参数，如**内容元素的趣味性**（Content，有无吸引人的主题）、**重复纹样**（Repetition，图案或元素的重复是否带来节奏美）、**光影与明暗**（Lighting，曝光和对比是否恰当）等。这些都曾在不同研究中被提出并验证对美学评分有影响。比如有研究统计了专业摄影比赛中的高分作品，发现它们往往在构图上符合经典原则、色调上和谐统一、主体清晰且富有趣味<sup>16</sup>。**公开数据集**为这类参数化建模提供了基础：AVA数据集包含25万张由摄影爱好者评分的图片，每张平均有约200人评分<sup>25</sup>；AADB数据集则在数万张图片上标注了审美评分以及上述**八大属性**评分，以方便研究者分析各因素与总体美感的关系<sup>26</sup>。通过对这些数据的建模训练，算法能够逐步“学会”人类在审美打分时看重的图像特征。Google Research的MUSIQ等模型更进一步，利用多尺度视觉Transformer直接处理任意分辨率的全尺寸图片，提取多粒度特征来预测图像质量<sup>27 28</sup>。该模型在多个大型数据集上取得了领先效果，包括在技术画质评价数据集上大幅超越以往方法，并在AVA美学评分上达到与最佳模型相当的性能<sup>29</sup>。这表明：**充分保留图像的构图比例和细节、多方面提取参数特征**，可以显著提高模型对美感的判断力<sup>30 31</sup>。

综上所述，基于**构图、色彩、景深、清晰度等多维参数的建模**，使得图像美学评分机制具有了可解释的客观基础。各参数维度既相互独立又共同作用，最终形成一幅图像的综合美感印象。通过跨学科的研究，我们逐步明确了哪些客观指标与审美主观评分有显著关联，从而能够在算法中融合这些**“设计经验”**，让模型不仅能给出分数，更能从中解释“为何美或不美”。

## 专家经验的角色：设计师对维度定义、权重调整与模型优化的引导

虽然数据驱动的模型可以从海量图片中学到美学评分的规律，但**设计师和美学专家的经验**在这一过程中扮演着不可或缺的角色。专家经验主要体现于两个方面：其一，确定和定义审美评价的**维度体系**；其二，针对具体任务对各维度调整权重并引导模型优化方向。

首先，设计领域的专家（如摄影师、平面设计师、美术指导等）往往拥有系统的美学理论知识和实践经验。他们深知**什么决定了一幅作品的视觉魅力**，因此能够帮助挑选和定义模型应关注的特征维度。例如，设计师知道画面需要在整体与细节之间取得平衡：既要颜色、字体、布局等细节元素协调，又要保证不同元素组合在整体上和谐统一<sup>32</sup>。Adobe Research的相关工作指出，专业设计师在评价版式美感时期望版面达到“双重层次的和谐”，既要细粒度元素（色彩、字体等）的搭配美观，又要全局布局上各元素相互呼应、协同一致<sup>32</sup>。这些经验其实勾勒出了审美评价的**维度蓝图**：色彩是否和谐、元素是否平衡、层次是否分明、风格是否统一等。有了专家提供的这套维度体系，算法才能“按图索骥”地去提取对应的特征进行建模。因此，在建立图像美学评分机制时，设计师往往会参与早期的特征工程或指标体系制定。例如，一个摄影美学评分模型可能在专

家建议下关注“主体是否三分构图”“背景光线是否柔和”“色调是否统一温馨”等等具体维度。这使模型的评价标准与人类专业审美标准对齐，增强了模型判断的可信度和专业性。

其次，专家经验还体现为对不同维度相对重要性的调配。并非所有美学要素在任何场景下都同等重要，设计师能够依据上下文和目的，调节各指标的权重，从而更贴合特定受众的审美取向。举例来说，在人像摄影的美学评估中，构图和用光可能被赋予较高权重，因为它们直接影响人物的呈现和氛围；而在广告平面设计的美学中，色彩搭配和字体选择可能更为关键。人类专家可以通过小规模的实验或多年反馈体会到这些差异，并将其融入模型优化。例如，在模型训练过程中加入**专家权重**：对数据集中的某些维度特征给予更强的监督，或者在损失函数中针对特定属性的误差提高惩罚权重，以此引导模型更关注这些属性的正确预测。专家还可以通过**人工校正回路**参与模型优化：模型初步输出各维度的评分或对最终美学分数的解释后，专家审核这些输出，指出不合理之处并调整权重，然后再反馈给模型重新训练。这样的人机协作循环能够逐步提升模型的性能和可信度，避免模型陷入纯数据驱动可能导致的偏差。近年来兴起的**可解释AI**技术也为专家介入提供了工具，比如一些研究开发了生成型模型，可以针对输入图像给出属性导向的审美意见（critique），指出图像在哪些方面有所不足<sup>33 34</sup>。设计师可以根据这些属性级的反馈来理解模型的判断依据，并进一步调整模型或输入，提高评分一致性。这种互动过程相当于专家在教授模型“审美课”，让模型逐步接近专业水准。

总的来说，**设计师的经验是连接主观审美和客观模型的纽带**。他们帮助模型“看”到数据看不到的美学细节和规则，并通过调整权重确保模型的评价符合专业标准和目标受众的品味。在跨部门合作的美学评分机制开发中，让算法工程师和设计专家协同工作，可以加速建立一套既科学有效又符合业务需求的评价体系。正如艺术与科学的结合能够迸发创意火花一样，融合设计经验的美学模型将比纯粹的数据驱动模型更具解释力和实用价值。

## 模型落地建议：设计师与算法协作的小范围验证与共识拟合

为了将上述美学评分机制有效应用于实际产品或项目中，我们建议采用**设计师与算法协作迭代**的方式，在小范围内进行验证并逐步拟合共识。具体而言，可以考虑以下实施步骤：

- 1. 试点数据集与基线模型：**首先在目标应用领域内收集一批具有代表性的图像，以及这些图像的人类审美评分（可同时包含普通用户和专业设计师的评分，以兼顾大众喜好与专家标准）。利用前述研究中的成熟模型作为起点（例如基于AVA预训练的美学评估模型），在这个小型数据集上进行微调，得到初步的美学评分模型。此阶段相当于建立一个**基线**，提供模型和人工评价之间差异的参考。
- 2. 设计师参与评估：**将模型对试点图像的评分结果与人类评分进行对比分析，重点检查模型在哪些案例上与**设计师判断**出现偏差。邀请资深设计师对这些有分歧的样本进行详细审阅，找出模型可能忽略的美学要素。例如，设计师可能指出：“这张照片构图略有问题，但模特表情和氛围很好，所以整体美感依然不错，人类打分高于模型预测。”通过这类实例，整理出模型评价维度或权重需要**调整的方向**。设计师还可提出新的评价维度（例如发现某类图像的美感与**情感感染力**有关，而这未在模型特征中体现）供模型开发者考虑加入。
- 3. 共识拟合迭代：**根据专家反馈，对模型进行有针对性的调整。例如，增加模型对特定特征的关注（加入新的输入特征或子网络来度量情感感染力），或者直接调整决策层对各维度的权重，使之更符合专家认可的比例。然后在试点数据上重新训练或校准模型。调整后的模型再次与人类（包括专家和普通用户）的评分进行对比。如果仍有较明显偏差的案例，则回到上一步继续分析原因。如此反复迭代，直到模型评分与人类共识评分基本吻合。这个过程实际上是在**拟合“共识函数”**：通过少量高质量的数据和专家修正，让模型学会将多维参数映射到与大众/专家共识相符的分值上。
- 4. 小范围AB测试：**在实际业务中小规模引入模型评分机制进行测试。例如，将电子商务网站上一部分商品的图片按模型美学分排序展示给用户，另一部分保持原有方式，然后比较用户行为指标的变化（类似于前文提到的房地产网站实验<sup>2</sup>）。通过AB测试验证模型评分带来的效益，如用户停留时间增加、点击

率提升等。同时收集用户的主观反馈（如是否觉得推荐的图片确实更吸引人）。这些真实环境的数据将进一步用于微调模型，使其适应业务场景。

**5. 逐步扩大应用：**如果小范围验证显示模型评分机制效果良好、与人类共识高度一致且能带来业务价值，那么可以逐步扩大应用范围。在推广过程中仍需定期监控模型输出，收集新数据来发现可能的问题。例如，不断跟踪模型在新出现的图像类型上的表现，确保共识拟合能力不随应用扩展而劣化。如果发现偏差，再通过增量训练和专家干预进行修正。如此保证模型始终贴合当前受众的审美共识。

通过上述流程，设计师的审美经验和反馈贯穿模型开发和落地的全周期，与算法训练形成闭环。小范围试验的好处在于风险可控、反馈快速，使我们能在投入大量资源部署前，就充分验证模型的可靠性和共识度。最终得到的美学评分机制将在客观参数和主观偏好之间取得平衡，既能利用算法的效率和一致性，又融合了人类审美的精妙判断，适应真实场景的需求。

## 风险与缓解策略：避免个体偏差与动态调整维度结构

在构建和应用图像美学评分机制的过程中，还需关注潜在的风险并制定缓解策略，以确保模型长期有效且公平可靠。

- **避免个体或数据偏差：**共识建模的前提是拥有足够多元、足够大的评价数据。如果采集的评分过于局限（例如主要来自少数专家或单一文化背景的用户），模型学到的可能是带偏的审美观，而非普适的共识美感。这会导致模型评分倾向于某些偏好，无法泛化到更广泛的人群。例如，一个仅依据西方审美偏好训练的模型，可能低估某些东方风格美学元素的价值。缓解这一风险需要在数据收集阶段确保**多样性和代表性**：涵盖不同性别、年龄、文化背景的受众，以及丰富的图像题材和风格。同时，可以在模型中引入**个性化调节**的机制，当应用于特定人群时，允许对评分进行二次校准。这类似于“因地制宜”的原则：在保持总体共识框架的基础上，尊重局部用户群体的特殊品味。例如，系统可根据用户历史偏好对通用美学分做轻微调整，提供**个性化的美学排序**，以避免“一刀切”的偏颇。
- **动态调整维度结构：**审美趋势和视觉风尚并非一成不变。随着时间推移，新兴的艺术风格、流行审美会不断涌现。如果模型固化在既有的维度和权重上，可能无法适应未来的变化。例如，几年前无人机航拍可能还属新奇构图，而如今俯拍视角已相当常见，公众对其美感的评价标准也可能演化。为此，模型的维度结构和参数权重应保持**一定的动态可调性**。一方面，可以定期引入新数据重新训练或微调模型，使其跟上最新的审美趋势；另一方面，在设计模型架构时预留拓展性，当发现某些重要美学因素未被覆盖时，可以较方便地加入新的特征分支或评价维度。不妨将美学评分机制视为一个**不断学习的系统**：就像设计师本身会随着审美潮流更新观念一样，模型也需要持续学习新知识。例如，最近兴起的AI生成艺术、美学滤镜等，如果影响了大众审美取向，我们也应把相关特征纳入模型考量。通过**持续监控和更新**，模型才能始终与动态的美学共识保持同步。
- **防止过度优化与审美多样性的丢失：**还有一个值得注意的风险是，过度追求某一套共识评分可能抑制多样性。如果模型和策略一味引导所有内容朝着同一种高分风格靠拢，长远来看可能导致视觉生态的单一。例如，如果算法偏好明亮高对比的照片而给予高分，用户和内容制作者可能迎合算法口味，产生趋同的影像风格，反而减少了审美上的丰富性。这就要求我们在设计评分机制时保持**包容性**。解决办法包括：在训练数据中涵盖各种不同类型的高分实例，让模型认识到多种风格都可以是美的；在评分输出中不仅给出总体分，还可以提供按风格类别细分的评价（例如一张另类风格的照片在主流审美上得分一般，但在特定亚文化审美中评价很高，这种信息应该能够从模型中体现出来）。通过这种方式，鼓励模型承认**审美的多元共存**，引导应用策略上不过度“一刀切”。同时，可由设计师和业务方共同制定评价结果的使用准则，避免将美学分数当作唯一标准去简单奖惩内容，而是将其作为辅助参考，留有人为复核和创意发挥的空间。
- **透明度与解释性：**最后，确保模型决策的透明和可解释也是降低风险的重要策略。如果模型给出的美学评分无法解释原因，设计师和内容创作者可能无所适从，甚至对其产生不信任，从而妨碍机制的落地。

因此应尽量提供模型评分的可解释输出，例如列出主要贡献的几个参数维度及其权重占比，或给出对图像的简短“美学点评”（哪方面出色，哪方面不足）。这既有助于专业人士校正模型，也能让内容制作者从中学习如何提升作品美感。如今一些学术工作已在朝这个方向努力，例如通过生成属性导向的评价语句来解释模型评分<sup>33</sup><sup>34</sup>。在实际部署中，我们也应加强这方面功能，以打造一个可信赖、能沟通的美学评分机制。

综上，在构建结合设计经验的图像美学评分体系时，我们不仅要关注模型在当前数据上的精度，更要有预见性地规避潜在风险。通过多元数据收集、动态模型更新、保持审美多样性、增强解释透明等措施，我们可以大大缓解这些问题，使模型能够长久稳定地服务于不同场景的审美评估需求。最终，希望该美学评分机制真正做到“以人为本”：既反映群体审美共识，又尊重个体差异和创意，让技术与设计经验融合，持续助力业务决策和用户体验的优化。

## 结语

本报告综述了结合设计经验的图像美学评分机制的研究进展和实践建议。通过引入跨学科视角，我们认识到：图像美感评价虽属主观，但可以通过共识建模凝练出客观指标；借助构图、色彩、景深等参数的量化，我们为审美评分奠定了解释性基础；融合设计师的专业经验，模型得以及时校准和优化；在实际落地中，通过小范围试验和人机协同，我们能有效构建并验证模型，使之贴合业务需求；同时，通过风险管理与动态调整，模型能够与不断演变的审美潮流保持同步。总的来说，图像美学评分机制的建立不是单一学科的任务，而是设计艺术、心理认知与人工智能的交融协作。这一机制在社交媒体内容筛选、电子商务商品展示、广告创意设计、自动化视觉美学反馈等方面都有广阔的应用前景。随着CVPR、TPAMI、SIGGRAPH等顶级会议近年大量相关成果的涌现，我们有理由相信，一个更加通俗易用而又专业可靠的美学评分系统指日可待。在未来的工作中，持续引入更广泛的美学观、多元文化和人文关怀，将使这一机制愈发完善，真正实现技术与艺术的融合，为各行业的视觉决策提供有力支持。

### 参考文献：

1. Pfister J. et al. "Self-Supervised Multi-Task Pretraining Improves Image Aesthetic Assessment." **CVPR Workshops 2021:** 816-825.
2. Jang H. et al. "Modeling, Quantifying, and Predicting Subjectivity of Image Aesthetics." arXiv preprint arXiv:2208.09666 (2022).
3. Talebi H., Milanfar P. "NIMA: Neural Image Assessment." **IEEE Transactions on Image Processing,** **27**(8): 3998-4011, 2018 <sup>14</sup> <sup>25</sup>.
4. Chen Q. et al. "Adaptive Fractional Dilated Convolution Network for Image Aesthetics Assessment." **CVPR 2020:** 14114-14123 <sup>21</sup> <sup>22</sup>.
5. Liu D. et al. "Composition-Aware Image Aesthetics Assessment." **WACV 2020:** 3568-3577 <sup>15</sup> <sup>16</sup>.
6. Kong S. et al. "Photo Aesthetics Ranking Network with Attributes and Content Adaptation." **ECCV 2016:** 662-679 <sup>35</sup> <sup>36</sup>.
7. Pombo M. et al. "Consensus and contention in beauty judgment." **iScience,** **27**(7): 110213, 2024 <sup>12</sup> <sup>13</sup>.
8. Li L. et al. "Towards Explainable Image Aesthetics Assessment with Attribute-Oriented Critiques Generation." **IEEE Trans. on Circuits and Systems for Video Technology,** 2024 <sup>37</sup>.

9. Patnaik S. et al. "AesthetiQ: Enhancing Graphic Layout Design via Aesthetic-Aware Preference Alignment of Multi-modal Large Language Models." arXiv:2503.00591 (2023) <sup>32</sup>.
  10. Hardesty L. "Amazon Web Services releases two new Titan vision-language models." **Amazon Science Blog**, 2025 <sup>5</sup>.
  11. García N. et al. "Validation of an Aesthetic Assessment System for Commercial Tasks." Applied Sciences, **12**(3): 1249, 2022 <sup>2</sup>.
- 

<sup>1</sup> <sup>2</sup> <sup>3</sup> <sup>4</sup> <sup>7</sup> Validation of an Aesthetic Assessment System for Commercial Tasks - PMC  
<https://pmc.ncbi.nlm.nih.gov/articles/PMC8774337/>

<sup>5</sup> Amazon Web Services releases two new Titan vision-language models - Amazon Science  
<https://www.amazon.science/blog/amazon-web-services-releases-two-new-titan-vision-language-models>

<sup>6</sup> <sup>14</sup> <sup>25</sup> Introducing NIMA: Neural Image Assessment  
<https://research.google/blog/introducing-nima-neural-image-assessment/>

<sup>8</sup> <sup>9</sup> <sup>10</sup> <sup>11</sup> arxiv.org  
[https://arxiv.org/pdf/2208.09666](https://arxiv.org/pdf/2208.09666.pdf)

<sup>12</sup> <sup>13</sup> Consensus and contention in beauty judgment - PMC  
<https://pmc.ncbi.nlm.nih.gov/articles/PMC11245996/>

<sup>15</sup> <sup>16</sup> <sup>17</sup> <sup>18</sup> <sup>19</sup> <sup>20</sup> Composition-Aware Image Aesthetics Assessment  
[https://openaccess.thecvf.com/content\\_WACV\\_2020/papers/Liu\\_Composition-Aware\\_Image\\_Aesthetics\\_Assessment\\_WACV\\_2020\\_paper.pdf](https://openaccess.thecvf.com/content_WACV_2020/papers/Liu_Composition-Aware_Image_Aesthetics_Assessment_WACV_2020_paper.pdf)

<sup>21</sup> <sup>22</sup> CVPR 2020 Open Access Repository  
[https://openaccess.thecvf.com/content\\_CVPR\\_2020/html/Chen\\_Adaptive\\_Fractional\\_Dilated\\_Convolution\\_Network\\_for\\_Image\\_Aesthetics\\_Assessment\\_CVPR\\_2020\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2020/html/Chen_Adaptive_Fractional_Dilated_Convolution_Network_for_Image_Aesthetics_Assessment_CVPR_2020_paper.html)

<sup>23</sup> <sup>24</sup> (PDF) Photo Aesthetics Ranking Network with Attributes and Content Adaptation  
[https://www.researchgate.net/publication/303821694\\_Photo\\_Aesthetics\\_Ranking\\_Network\\_with\\_Attributes\\_and\\_Content\\_Adaptation](https://www.researchgate.net/publication/303821694_Photo_Aesthetics_Ranking_Network_with_Attributes_and_Content_Adaptation)

<sup>26</sup> <sup>35</sup> <sup>36</sup> [1606.01621] Photo Aesthetics Ranking Network with Attributes and Content Adaptation  
<https://arxiv.org/abs/1606.01621>

<sup>27</sup> <sup>28</sup> <sup>29</sup> <sup>30</sup> <sup>31</sup> MUSIQ: Assessing Image Aesthetic and Technical Quality with Multi-scale Transformer  
<https://research.google/blog/musiq-assessing-image-aesthetic-and-technical-quality-with-multi-scale-transformers/>

<sup>32</sup> AesthetiQ: Enhancing Graphic Layout Design via Aesthetic-Aware Preference Alignment of Multi-modal Large Language Models  
<https://arxiv.org/html/2503.00591v1>

<sup>33</sup> <sup>37</sup> Towards Explainable Image Aesthetics Assessment With Attribute ...  
[https://www.researchgate.net/publication/384494496\\_Towards\\_Explainable\\_Image\\_Aesthetics\\_Assessment\\_with\\_Attribute-oriented\\_Critiques\\_Generation](https://www.researchgate.net/publication/384494496_Towards_Explainable_Image_Aesthetics_Assessment_with_Attribute-oriented_Critiques_Generation)

<sup>34</sup> Image Aesthetics Assessment With Attribute-Assisted Multimodal ...  
<https://www.semanticscholar.org/paper/Image-Aesthetics-Assessment-With-Attribute-Assisted-Li-Zhu/402baaa1e6510a1382cfa404205a1d2b151c8e39>