照片编辑中的个性化体现在哪里？来自一位摄影师 Lightroom 历史记录的证据

Dongwu Qian 独立研究者
qiandongwu123@qq.com

2026年11月21日

arXiv:2611.29307v1 [cs.CV] 22 Nov 2026

摘要

个性化照片编辑系统必须同时推断摄影师反复表现出的偏好，以及特定拍摄批次的意图，但这两种信号很少被分别评估。我们研究了一个纵向私有数据集，其中包含一位专业摄影师在 41 个目录中编辑的 3,193 张照片，并按批次级划分留出集，以考察可解释的 Lightroom 调整。我们比较了绝对值目标和相对于先验的目标、图像条件化预测器，以及能够接收目标批次少量编辑的估计器。残差预测降低了验证误差，但在留出批次上的优势很小，且在批次聚类自助法下不显著（曝光值曝光平均绝对误差为 0.592 对 0.582）。在批次共享协议中，使用 4、8 或 16 个同批次锚点的闭式收缩，配合各个 k 下表现最佳的估计器，可分别达到 0.466、0.457 或 0.453 曝光值；仅先验基线则为 0.586–0.592 曝光值。配对降幅在批次聚类自助法下具有显著性。学习得到的锚点条件化头未能优于这一闭式估计器；而经过负向训练的先验已能实现较低的冷启动图像误编辑率，因此额外加入门控器几乎没有收益。推荐的估计器是闭式且可审计的，并且只需少量同批次编辑。针对 383 个原生 Lightroom 导出文件进行的渲染空间检查得出了相同的排名：在 $k = 8$ 时，在线经验贝叶斯相较于先验将 ΔE_{00} 降低了 0.41 （ $p = 0.042$ ），而摄影师的真实参数将其降低了 0.78（$p = 0.034$）。这些结果说明，应在拍摄批次级评估个性化，并将参数保真度、冷启动风险和用户偏好视为彼此独立的结果。附录将动作空间从六个扩展至十三个全局控制项：同批次中位数外观使渲染色度更接近导出文件，而逐照片残差头再次未能迁移。本研究仅涉及一位摄影师，并不声称滑块准确度能够衡量审美偏好。

1 引言

一位从一次拍摄中编辑数百帧照片的摄影师，并不会将每一帧单独处理。经过几次校正后，这次拍摄中的其余照片往往也会获得相似的曝光提升、对比度曲线和色彩处理。商业工具宣传称，它们可以从 Adobe Lightroom[1] 目录 [1],但其中的根本问题很少被拆解：预测出的调整有多少来自图像本身，有多少来自摄影师长期的历史记录，又有多少来自当前正在编辑的拍摄?

本文基于真实编辑历史，衡量这种分解。我们采集了一位专业摄影师的私人 Adobe Lightroom 目录集合：41 个目录、3,193 张编辑后的照片、六项可解释的全局调整（曝光、对比度、饱和度、自然饱和度、清晰度、纹理），以及模拟实际操作工作流的批次级留出集——“先编辑几帧，再将结果应用于同一场拍摄的其余照片”。研究范围有意设得较窄。主要实验使用六项全局调整；附录 D 将动作空间扩展至高光、阴影、白色色阶、黑色色阶、八波段 HSL 饱和度和去朦胧，并表明同批次中位数外观可以迁移，而逐照片残差头则不能。我们不对局部蒙版或完整的 Adobe Lightroom 渲染管线进行建模，也不声称滑块准确度等同于审美偏好。正是这种研究范围的限定，使各类信号得以区分。

我们的结果支持四项结论。第一，批次标识是标签变异的重要来源：单向分解表明，批次分别解释了 36% 的曝光方差和 27% 的对比度方差；批次中位数与全局中位数相差 +0。84 至 −0。62 曝光值。因此，忽略拍摄批次的逐图像预测器，评估的对象就错了。

其次，相对于先验的（残差）目标能降低域内验证误差，但无法迁移到不同拍摄中。在验证集上，残差头的误差为 0.171 EV，优于绝对值头的 0.195 EV。但在来自 11 个未见批次的 1,448 张留出照片上，结果顺序逆转（ 0.592 对 0.582 EV）；按批次聚类的自助法得出配对差值为 +0。010 EV，95%区间为 [−0。016， +0。033] (p = 0。51)。此前的照片级检验（p = 0。027）夸大了证据，因为它将同一拍摄中的相关照片视为彼此独立。

第三，少量同批次锚点是可靠的信号。在批次共享协议下——每个批次使用一组锚点，并由该批次中的所有目标共享——闭式收缩可将曝光平均绝对误差从 0.586–0.592 EV 降至 $k = 4$ 时的 0.466 EV、 $k = 8$ 时的 0.457 EV，以及 0.453

曝光值为 $k = 16$ （每个 k 下表现最佳的实验组）。相较于先验的配对降幅为

[1]Adobe 和 Lightroom 是 Adobe Inc. 的商标或注册商标。本作品独立制作，与 Adobe 无关联，也未获得 Adobe 的认可。

0.11–0.14 曝光值，并且在批次聚类自助法下仍具有显著性（$p \leq 0$。003）。学习得到的锚点条件化头，包括在闭式估计器基础上训练的残差头，表现均未优于该估计器。因此，这一提升并不意味着模型“理解了这次拍摄”；它表明拍摄级偏移是低维的，可以通过少量编辑进行估计。

第四，风险与偏好是彼此独立的结果。使用未编辑负样本训练的先验，在未编辑图像上的冷启动误编辑率已达 0.99%（在 302 张留出的未编辑图像上，预测曝光幅度≥ 0. 3 EV）。加入逻辑回归或多层感知机门控器后，误编辑率升至 1.99–7.28%，而编辑召回率和测试平均绝对误差均未改变。在一项包含 50 张图像的探索性盲测比较中，最受偏好的实验组仅在 18.4% 的情况下参数误差最低，在 14.3% 的情况下渲染误差最低；偏好与保真度不应混为一谈。

贡献。 (1) 一种批次共享测量协议，用于区分图像内容、用户先验和当前拍摄上下文；锚点排除在目标之外，推断则按批次级聚类。(2) 一项纵向真实 Lightroom 研究，探讨表示方式的选择（绝对值与相对于先验）、负向训练先验和同批次适应，并涵盖残差表示的跨批次负迁移。(3) 证据表明，闭式收缩估计器是同批次适应的强大且可审计的基线，而在这一数据规模下，学习型锚点条件变体并未胜过它。(4) 明确分析研究边界：一位摄影师，以目录作为批次代理指标；主研究包含六个全局参数（附录 D中为十三个）；针对原生 Lightroom 导出文件进行渲染空间检查；并清楚区分参数保真度、冷启动风险与偏好。

2 相关工作

参数预测与白盒编辑。 基于曝光的系统从一组风格照片中学习一系列白盒滤镜 [2]；WISE通过示例学习白盒风格化 [3]；神经样条变换对个性化色彩曲线建模 [4]；Deep Preset根据参考图像对，预测低级色彩变换（包括 Lightroom 预设）的设置 [13]。深度双边方法学习图像自适应的局部算子 [5]。这些系统证明了可解释的参数空间监督切实可行，但它们学习的是一组风格，而非单个摄影师的纵向历史；同时，它们也没有区分检索、先验和拍摄上下文。

个性化与基于检索的编辑。 Adobe 的专利描述了为相似图像检索已存储的调整参数 [6]。商业服务从用户的目录中学习，而开源的 StyleAI 则将图像嵌入与向量存储以及绝对值显影目标相结合 [7]。近期的检索增强修图会以参考图像为条件来生成模型而无需测试时微调 [8]。PPS 从 767 名用户的成对偏好判断中学习用户的照片风格 [14]；PIF 通过文本反演，利用少量参考图像迁移风格 [15]；商业 AI 配置文件在摄影师数千次编辑的基础上进行训练，并预测每张图像的滑块设置（例如 Imagen AI [17]）。这些系统会个性化最终图像应呈现的样子。我们的问题有所不同：摄影师的调整中哪些部分是特定于某次拍摄的，以及需要多少张同一拍摄中的编辑才能恢复这些调整。因此，我们将它们视为互补方法，而非相互竞争：强大的逐图像风格模型可以建立在我们所测量的批次级偏移量之上。

用户建模。 PieNet根据比较结果对用户品味建模 [9], ，相关研究则将用户嵌入共享偏好空间。这些方法回答的是“这个用户喜欢什么？”；我们提出的问题是“这个用户的哪些调整可以根据当前拍摄恢复？”这两个问题互为补充，而我们的负面结果则对后一个问题作出了限制。

预设与拍摄级一致性。预设是业界针对拍摄级一致性的手动解决方案。我们的批次锚点可以理解为自动估计、带有收缩权重的单次拍摄预设。批量编辑也被描述为将一项已演示的编辑应用于一组图像 [16]；我们研究的是其统计版本——根据少量编辑估计拍摄级偏移量，并在批次级进行评估。区别在于测量方式：我们报告完整的 k 曲线、批次宏平均和最差批次表现，并指出少量锚点就已能带来大部分收益。

评估。参数误差、渲染图像误差和人类偏好是不同的评估目标。近期的奖励模型会根据人类偏好评估指令引导的编辑 [10]。我们的结果还提示了一点：在一项规模较小、仅涉及一位摄影师的研究中，摄影师偏好的实验组往往并非参数误差或渲染误差最低的实验组，因此，指标的选择本身就体现了我们对个性化含义的主张。

3 数据

数据来源与分析单位。数据来自一位专业摄影师的私人 Lightroom 目录集合，以只读方式扫描。目录中有 5,325 张照片，其六个建模字段均包含非默认值。其中，3,196 张照片在档案中有原始文件，并重建了“处理前”预览；其余 2,129 张仅存在于目录中，因为原始文件未存档。分析单位为单张照片，并按其目录级批次分组，以此近似代表一次拍摄或拍摄场次。

每张照片都会从显影设置中读取一个六维标签：曝光（曝光值）、对比度、饱和度、自然饱和度、清晰度和纹理。对比度和饱和度以乘数 $1 + v/100$ 存储；自然饱和度、清晰度和纹理以 $v/100$ 存储。这六个字段是唯一建模的全局调整

表 1： 排除后的数据划分。验证集为域内数据。

数据划分	n
训练集	1,571
验证集（域内）	174
测试集：近期	217
测试集：未见	790
测试集：第二组	441
测试集总数	1,448
排除（边界）	3

项。67.3% 的照片包含局部蒙版，但不会对其进行预测。

数据划分与排除。 数据划分按批次级进行，绝不按随机照片划分。26 个目录级批次被划分为 15 个训练/验证批次和 11 个测试批次。训练集使用这 15 个批次中的 1,571 张照片，验证集使用从其中 14 个批次抽取的 174 张照片；11 个测试批次包含 1,448 张照片，分为三组：四个近期批次（217 张照片）、四个未见批次（790 张）以及三个第二组批次（441 张）。因此，验证集是域内验证集，而不是批次级留出集。所有报告的测试集数值均使用这 11 个留出批次（表 1）。

三张照片（索引 3123–3125）位于批次划分边界：其目录属于测试批次之一，但文件位于一个通常分配给训练/验证集的目录中。它们的预览哈希是唯一的，但它们与测试批次照片的 SigLIP 相似度达到 0.98，与同一次拍摄相符。由于正确的归属存在歧义，我们将其从训练集和验证集中排除；测试集不变。v1 集合与 1,329 张照片重叠；其中 1,867 张是新增照片。由于目录在 v1 扫描后再次编辑，74 个 v1 标签发生了变化；v2 使用当前值。附录 A 单独报告 v1 结果，且绝不与 v2 合并。

处理后图像。 摄影师还保留了 Lightroom 原生导出文件：48 个导出文件夹中共有 6,347 个 JPEG 文件，且保留了原始文件名。通过文件名和内容指纹进行匹配，3,196 条记录中有 2,751 条（86.1%）关联到唯一的导出文件（其中 2,702 个位于同一文件夹，49 个位于不同文件夹，并经感知检查确认）；其余 445 条记录对应的照片未曾导出，或所属目录中的原始文件未归档。在 1,448 张测试照片中，有 1,226 张（84.7%）有导出文件。第 5.8 节使用这些配对中未裁剪的子集，在渲染空间中检验参数空间排序。

批次构成与异质性。表2对 11 个测试批次进行了匿名化处理。批次大小从 38 张到 338 张照片不等；平均曝光值从 −0.32 到 +1.31 曝光值。对全部 3,196 个标注进行的单向分解显示，批次标识可解释曝光方差的 36% 和对比度方差的 27%（饱和度 10%、自然饱和度 13%、清晰度 14%、纹理 15%）；各批次中位数相对于全局中位数的范围为 +0.84 到 −0.62 曝光值。这是描述性结果，并非因果关系：

表 2： 匿名化的留出批次。

批次	组	n	平均曝光值（曝光值）
B01	近期	41	−0。32
B02	近期	57	+0。24
B03	近期	38	+1。00
B04	近期	81	−0.19
B05	未见	338	+1。15
B06	未见	246	+0.46
B07	未见	146	+1。31
B08	未见	60	+0.32
B09	第二组	149	−0。14
B10	第二组	222	−0。17
B11	第二组	70	+0.36

不同目录在主体、光线、相机和摄影师意图方面各不相同，而且目录仅是拍摄的替代指标。

数据不包含的内容。 这 2,751 张带有原生导出的照片提供了真实的处理后图像，但它们是摄影师最终的编辑结果，并非中立的参考；而且我们的渲染器缺少 Adobe 的相机配置文件；因此，第 5.8节仅在未裁剪子集上将这些图像用于相对比较。我们曾测试根据目录参数重建处理后图像，但最终否决了这一做法：预览中的自动增亮基线与 Lightroom 的中性渲染相差 1–3曝光值，因此重建图像不能作为有效参考。本文将参数保真度作为主要评估结果，以渲染结果进行排序核验，并将 50 张图像的盲法研究视为探索性研究。

4 方法

特征与先验。 每张照片均由一个冻结的 SigLIP-large 编码器（1,024 维）编码，并与 84 个手工设计特征拼接，得到 1,108 维表示。该编码器不会进行微调。可部署的先验是一个多层感知机（$1{,}108 \to 192 \to 9$，约 0.22M 个参数），使用 AdamW 在 1,571 张训练集照片上训练，批次大小为 64，训练 60 轮，丢弃率为 0.1，并在 174 张验证集照片上进行模型选择。为兼容生产运行环境，保留了九个输出槽位；其中六个接受监督，而白平衡、色调和去朦胧均设为中性值。

全文使用的先验是 **V25-neg**，其训练中加入了602张未经编辑的冷图像作为负例，权重为0.3（另有一组不重叠的603张冷图像留作测试集；其中301张用于校准门控器，302张用于测试门控器）。负项使先验学会对摄影师未编辑的照片输出接近默认值的调整。在302张冷启动测试图像上，该先验的误编辑率为0.99%，定义为调整幅度为$|\hat{} e| \geq 0$。3曝光值的图像所占的比例。

绝对值头和残差头。 两个头共享同一架构。**绝对值**头直接预测六个标签。**残差**头预测 $\Delta = y - P(x)$，其最终层初始化为零，因此训练从先验值开始；在推断时

预测为 $P(x) + \beta\hat{}\Delta$，主表中使用 $\beta = 1$。在验证集上，β 扫描（0 到 1.5）在 1.0 到 1.5 之间保持平坦。两个头均使用随机种子 0 训练 60 个训练轮次。绝对值头在不使用负样本的情况下进行训练；另一个变体则以 0.6 的权重加入 602 个冷启动负样本。残差头在不使用负样本的情况下进行训练，这遵循了先前的发现：负项会损害残差目标。

批次共享锚点协议。部署问题是：给定当前批次中 k 张已编辑的照片，我们能多好地预测其余照片？对于每个测试批次和每次重复，我们无放回地抽取 k 个锚点，并在该批次的其余照片上进行评估；所有目标共享同一个锚点集。这就是批次共享协议。它不同于逐目标抽样，在逐目标抽样中，每个目标都会抽取自己的锚点集；这种方案不对应批次工作流，并且可能使估计曲线向任一方向偏移，因此我们仅将其作为诊断使用。

主曲线使用 $k \in \{0、1、2、4、8、16、32\}$，重复20次，随机种子 7。包含 $n < k + 1$ 的批次使用 $k = n - 1$；在$k = 0$ 时，每个估计器都会退化为先验。锚点变体实验使用相同的协议，随机种子 21，以及$k \in \{4、8、16\}$。任何锚点都不会同时作为目标，并且训练集批次中的照片不会被用作测试批次的锚点。

估计器。令 P_i 为先验预测，A 为锚点集。先验：P_i。批次中位数：y_a 在$a \in A$上的逐字段中位数。残差中位数：$P_i +$ 中位数$_{a\in A}(y_a - P_a)$。分层经验贝叶斯：$P_i + w \odot$中位数$_{a\in A}(y_a - P_a)$，其中 $w_j = k/(k + \kappa_j)$，κ_j为训练批次噪声信号比：$y - P$ 的平均批次内方差除以逐批次残差均值的方差。*Mix50*：0。5·中位数$_{a\in A}(y_a)$+0。5·P_i。在线经验贝叶斯*(E6d)*：采用相同的收缩形式，其中 $\kappa_j = \mathrm{Var}_A(y_j - P_j)/\mathrm{Var}_A($ $y_j)$根据当前锚点进行估计；无需历史批次统计量，并在 $k = 0$ 处回退至先验。

我们还在相同协议下比较了三种学习得到的锚点条件变体：拼接（图像特征加上 13 维锚点摘要）、median+残差和 E6d+残差。它们用于检验：在明确估计偏移量后，内容特征是否能提供偏移量之外的信息。

冷启动门控器与指标。逻辑回归和多层感知机门控器在 1,571 张已编辑训练照片和 602 个冷启动负样本上进行训练；在 301 张冷启动校准图像上设置阈值，以将假阳性率控制在 $\alpha \in \{0.05,0.10\}$。被拒绝的照片采用默认值（不编辑）。主要指标是以曝光值为单位的曝光平均绝对误差；我们还报告六字段归一化平均绝对误差（逐张照片计算六个字段的 |误差|/范围均值，再取平均）和冷启动误编辑率。对于锚点，我们报告微平均、跨 11 个批次的批次宏平均，以及最差批次。

置信区间采用批次聚类自助法：有放回地对 11 个测试批次进行重采样，共 10,000 次，每次保留所抽取批次中的所有照片，并重新计算配对差值。这种方法考虑了同一场拍摄中照片之间的相关性，也是用于主要结论的唯一置信区间

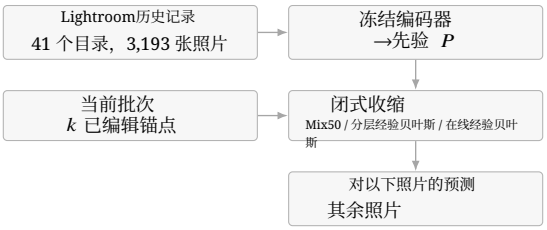

图 1：信号分解。通过闭式收缩将内容先验与少量同批次锚点结合起来；评估单位是批次，锚点绝不会作为目标。

表 3：绝对值头与残差头（V25-neg 先验，排除后）。

数据划分（n）	先验	绝对值	残差
验证集（174）	0.1811	0.1950	**0.1715**
测试集，全部（1,448）	0.5856	**0.5824**	0.5925
测试集，近期（217）	0.4030	0.4246	0.4024
测试集，未见（790）	0.7442	0.7353	0.7614
测试集，第二组（441）	0.3913	0.3860	0.3834

。由于只有 11 个聚类，区间较宽；p 值仅作描述性参考，而非用于验证结论。报告照片级测试仅为展示其对精确度的夸大程度。

5 结果

5.1 批次标识是一阶信号

第 3 节的分解表明，标签方差中有相当大一部分可归因于批次标识（曝光占 36%，对比度占 27%），且批次中位数的跨度超过 1.4 EV。若按照片评估预测器，并随机划分照片，那么训练集和测试集中都会出现来自同一拍摄批次的许多近重复标签。这既说明了采用批次级划分和批次共享锚点协议的必要性，也解释了为什么仅内容模型在域内的表现看起来会比它在新拍摄批次上的实际表现更好。这并非相机效应：机身只能解释 1.2% 的曝光方差，而批次可解释 35.9%；26 个批次中有 20 个同时包含两种机身（批次内机身平均值$\eta^2 = 0.09$）。

5.2 残差目标有助于域内表现，但无法跨批次泛化

在验证集上，与绝对值头相比，残差头将曝光平均绝对误差降低了 0.024 EV；与先验相比，则降低了 0.010 EV。在完整测试集上，结果顺序反转：残差头比绝对值头差 0.010 EV。配对批次聚类差值（残差 − 绝对值）为+0.0101 EV，95%置信区间为 $[-0.0158, +0.0329]$，$p = 0.51$；批次宏平均差值为 +0.0043 EV。照片级配对检验得出 $p = 0.0266$，这正是独立同分布分析会报告的数值，但我们采用的不是这个数值。

分组细分显示了结果反转的原因：残差头在近期组和第二组

表 4： 批次共享锚点曲线（V25-neg，$n = 1,448$，20 次重复，随机种子 7）。在线经验贝叶斯使用随机种子 21 和 $k \in \{4, 8, 16\}$。

k	先验	Mix50	分层经验贝叶斯	在线经验贝叶斯
0	0.5856	0.5856	0.5856	0.5856
1	0.5856	0.5186	0.5343	–
2	0.5872	0.4934	0.5252	–
4	0.5863	**0.4657**	0.4739	0.4805
8	0.5896	0.4717	0.4650	**0.4573**
16	0.5920	0.4712	**0.4525**	0.4559

批次上具有竞争力（0.402 对 0.425，以及 0.383 对 0.386），但在未见批次上表现较差（0.761 对 0.735）；未见批次包含了大部分极端曝光标签（99 张照片中有 80 张带有|曝光| > 2标签）。绝对值头在两个较大的测试组上优于先验（未见组为 0.735 对 0.744；第二组为 0.386 对 0.391），但在近期组上不如先验（0.425 对 0.403）；残差头在近期批次上的表现接近先验，但在未见批次上明显较差。因此，我们将残差表示视为一项描述性、域内结果，而非普遍改进。采用不同数据划分和先验的 v1实验显示了更大的域内增益（附录A）；这一结果无法在 v2 中复现为跨批次结论。

5.3 负向训练先验可应对冷启动

仅使用先验时，在完整测试集上的结果为 0.5856 EV。将负项加入绝对值头后，测试平均绝对误差从 0.5824 EV 变为 0.5890 EV，产生了少量编辑误差代价；部署的负向训练先验达到 0.5856 EV。未使用负样本训练的残差头达到 0.5925 EV。在 302 张冷启动测试图像上，先验的误编辑率为 0.99%——本就很低：对于摄影师未作处理的照片，它很少会建议进行可见的曝光变化。在部署配置下，这种权衡是有利的，但它终究是一种权衡，而非没有代价的改进。

5.4 同批次锚点是稳健信号

曲线从 $k = 0$到 $k = 4$急剧下降（表 4，图 2）。此后，分层经验贝叶斯仍缓慢改善（从 $k = 4$ 到 16，由 0.4739 降至 0.4525），而 mix50 基本持平（0.4657 到 0.4712）。在 $k = 4$时，mix50 表现最佳（0.4657）；在 $k = 8$时，在线经验贝叶斯表现最佳（0.4573）；在 $k = 16$时，分层经验贝叶斯表现最佳（0.4525）。固定 k时，三个闭式方案之间的差异很小（0.003–0.015EV），批次聚类区间也无法将它们区分开来；没有任何一个方案在所有 k条件下都占优。

学习型锚点条件变体并未胜过闭式解（表5）：拼接结果接近先验，中位数+残差接近中位数，而在在线经验贝叶斯之上添加残差头会使结果朝错误方向变化 0.003–0.005曝光值 。在线经验贝叶斯对训练批次构图：排除三张边界照片后，由于训练批次 κ 发生变化，$k = 4$ 处的分层经验贝叶斯从 0.4591 变为 0.4739，而 $k = 4$ 处的在线经验贝叶斯保持在 0.4805。

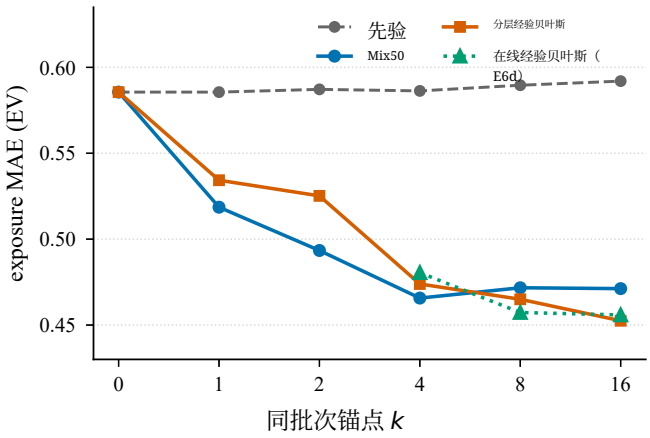

图 2： 曝光 MAE 随同批次锚点数量的变化。效果最佳的闭式解方案会随 k 改变；在固定的 k 下，各方案之间的差异为 0.003–0.015 曝光值。

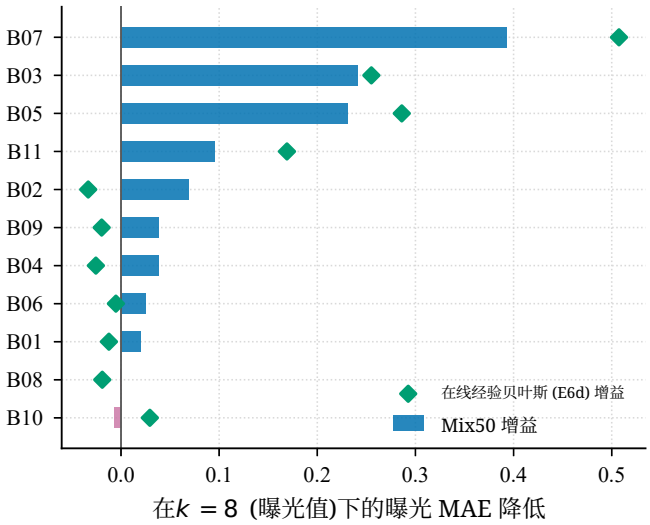

图 3： 在 $k = 8$ 下各批次的曝光 MAE 降幅，按 mix50 增益排序。柱形：mix50；菱形：在线经验贝叶斯。曝光偏移较大的批次（B03、B05、B07）增益最大；其中一个批次（B10）在 mix50 下表现略差。

对于分层经验贝叶斯，$k = 8$处的批次宏平均降至 0.434，最差批次降至 0.796；$k = 16$处则分别降至 0.422 和 0.760。改善不仅体现在微平均上，也体现在宏观平均和最差批次层面。最差批次仍是曝光变化最大的批次（先验下为 1.24 曝光值），这与锚点能够捕捉变化方向、但无法充分反映极端拍摄的变化幅度相一致。

在每个 $k \geq 4$ 下，锚点增益是经批次聚类后仍然显著的唯一效应，对 mix50/分层经验贝叶斯的影响约为 0.11–0.14 曝光值 (表 6)。在线经验贝叶斯在$k = 4$ ($p = 0.070$) 下仅达到边缘显著，在 $k = 8$ 和 16 下显著；其基于锚点估计的 κ 在锚点很少时噪声较大。仅中位数和学习式拼接方案在任何 k ($p = 0.22$–0.85) 下均不显著。

图3 显示，增益因批次而异。在 11 个批次中，Mix50 有 7 个批次优于分层经验贝叶斯，但分层经验贝叶斯的微平均更低，因为它胜出的批次规模较大。这

表5： 锚点变体（V25-neg，$n = 1,448$，随机种子 21）。粗体标示本表中该列的最佳值；分层经验贝叶斯在表 $k = 16$中达到 0.45254。

k	先验	中位数	在线经验贝叶斯	拼接	中位数+残差。	E6d+残差。
4	0.5871	0.5517	**0.4805**	0.5807	0.5783	0.4852
8	0.5896	0.5180	**0.4573**	0.5863	0.5278	0.4607
16	0.5934	0.5059	**0.4559**	0.5903	0.5294	0.4598

表 6： 相对于先验的配对差异，按批次聚类的 95% 置信区间（11 个批次，10,000 次重采样）。

比较	k	Δ EV	95%置信区间	p
Mix50 –先验	4	−0.1200	[−0.221, −0.020]	0.0010
分层经验贝叶斯 – 先验	4	−0。1116	[−0.206, −0.014]	0.0026
混合50 –先验	8	−0.1171	[−0.208, −0.027]	< 0。001
分层经验贝叶斯 – 先验	8	−0。1229	[−0.228, −0.016]	0.0024
混合50 –先验	16	−0。1187	[−0.207, −0.028]	0.0002
分层经验贝叶斯 – 先验	16	−0.1353	[−0.252, −0.018]	0.0024
在线经验贝叶斯 – 先验	4	−0。1057	[−0.211, +0.005]	0.070
在线经验贝叶斯 – 先验	8	−0。1305	[−0.252, −0.007]	0.0246
在线经验贝叶斯 – 先验	16	−0。1340	[−0.254, −0.017]	0.0038

就是我们同时报告微观平均和宏观平均，而不指定唯一最佳实验组的原因。

锚点解决了什么问题。 在 $k = 8$，收益主要集中在曝光：平均绝对误差从 0.589 降至 0.463（21%）；对比度、自然饱和度和清晰度提升了 7–9%，而饱和度和纹理的提升不到 2%。锚点校正的是这次拍摄的色调偏移，而非色彩或细节处理。

哪些锚点？ 我们还探究了锚点的选择是否重要。固定每个测试批次的评估半部分，并从另一半中抽取锚点，我们在 $k \in \{2, 4, 8, 16\}$ 个锚点时比较六种规则：随机选择；该次拍摄的前 k 帧；曝光编辑幅度最大的 k 张照片；特征空间中最接近批次质心的照片；在该空间中进行最远点采样；以及选择最接近批次标签中位数的照片的预言机。在 $k = 4$ 个锚点时，随机选择的曝光平均绝对误差为 0.467，前 k 帧为 0.671，极端编辑为 0.563，特征中心点为 0.550，特征多样性为 0.528，预言机为 0.450（批次聚类 $p = 0.032$，与随机选择相比）。只有预言机的优势具有统计显著性：没有任何可部署的、基于内容的规则优于随机选择，而开头帧的表现明显更差。因此，锚点选择几乎没有可部署的提升空间——该协议采用的随机锚点并未牺牲准确率，用户也无需挑选要编辑哪些帧。

5.5 协议和先验选择会改变答案

有三项选择会实质性地改变报告结果。首先，按目标采样锚点 会为每个目标分别设置锚点集；在 $k = 4/8/16$ 时，其结果分别为 0.4762/0.4603/0.4503，与批次共享实验组的结果接近，有时甚至更高。由于
协议有所不同，因此我们只将其视为诊断。其次，对照片进行留一法更为乐观：相同的

表7： 冷启动错误编辑率和门控。召回率和测试平均绝对误差基于编辑后的测试集。

预测器	冷启动误报率	召回率	测试平均绝对误差
先验（V25-neg）	**0.99%**	100%	0.5856
逻辑回归门控，$\alpha = 0.05$	1.99%	100%	0.5856
逻辑回归门控，$\alpha = 0.10$	6.29%	100%	0.5856
多层感知机门控，$\alpha = 0.05$	4.64%	100%	0.5856
多层感知机门控，$\alpha = 0。10$	7.28%	100%	0.5856

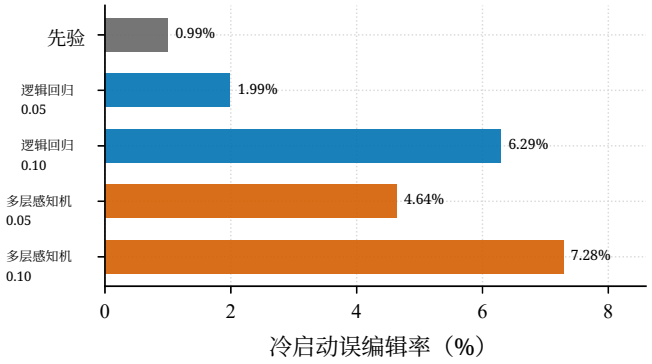

图4： 冷启动误编辑率。负向训练先验已达到0.99%；添加门控器会增加误编辑。

在线经验贝叶斯在留一法下达到约 0.433 曝光值，而在 $k = 8$ 处使用批次共享采样时为 0.457。第三，先验版本也很重要：较早的 V20e 先验得到 0.469/0.455/0.453，而 V25-neg 得到 0.4805/0.4573/0.4559。这些数值均不应跨先验进行比较。实际规则很简单：报告批次共享曲线、批次宏平均、最差批次和批次聚类区间；不要将单个留一法数值或逐目标数值作为可部署结果报告。

5.6 在负向训练先验下，额外的门控器是多余的

在这些阈值下，门控器从未在经过编辑的测试照片上触发（召回率为 1.0），因此对编辑集不起作用，只会在冷图像上增加误编辑（表 7，图 4）。距离门控器的表现更差：它在新批次上的召回率仅为 0.1–8%，会抑制大多数真实编辑。在 V25-neg 先验下，额外的门控器要么多余，要么有害。早期的 v1 结果 —— 门控器将冷图像上的误编辑率从 51% 降至 7%——无法推广到这个先验和数据划分；它只是历史结果，并非关于当前系统的结论。

表8：针对383个原生 Lightroom 导出文件的渲染空间检查。越低 ΔE_{00} 且结构相似性指数越高越好。差异在每张照片内进行配对，并按批次聚类（10,000次重采样）；区间为95%。

Arm	ΔE_{00}	与先验相比	SSIM	与先验相比
未编辑	13.91	−0。03	0.5104	—
先验	13.94	—	0.5033	—
Mix50 (k=8)	13.70	−0。24 [−0。87, +0。08]	0.5149	+0。012
在线经验贝叶斯 (k=8)	13.53	−0。41 [−1。03, −0。02]	0.5241	+0.021
真实参数	13.16	−0.78 [−1.94, −0.08]	0.5342	+0.031

5.7 探索性研究：保真度与偏好各不相同

在一项探索性盲法研究中，我们展示了涵盖测试批次的 50 张照片，每张配有三个经过匿名处理且顺序随机排列的候选结果：生产模型（v7）、基于 CLIP 的先验（V25-CLIP），以及带有同批次锚点的 SigLIP 先验。摄影师在 44% 的案例中偏好 v7，在 34% 的案例中偏好 V25-CLIP，在 20% 的案例中偏好 SigLIP+锚点（有一张图像被所有候选结果拒绝）。然而，只有在 18.4% 的图像中，摄影师偏好的实验组才同时具有最低的绝对曝光误差；最低的渲染 ΔE_{00} 则出现在 14.3% 的图像中；随机情况下的比例为 33%。因此，在这个样本中，参数和渲染保真度都无法预测偏好。我们将此报告为一项边界观察，而非偏好研究：研究仅涉及一位摄影师，使用不带高光恢复功能的简化预览，也没有采用受控的偏好评估方案。正式研究需要多名评审者和盲法 A/B 设计；第5.8 节给出了像素级检查，这是本项探索性研究无法做到的。

5.8 渲染空间验证：与原生导出进行比较

参数误差不等于像素误差。对于带有原生导出的测试照片，我们在每个实验组下渲染相同的处理前图像，并将其与摄影师的输出进行比较。在 1,226 张配对测试照片中，有 777 张的导出图像与原始画幅具有相同的宽高比；由于无法复现裁剪，因此排除裁剪后的导出图像。处理前图像采用固定设置（相机白平衡、不启用自动亮度）从 RAW 解码，在每个实验组下以 512 px 渲染，并使用 ΔE_{00} 和结构相似性指数进行比较。

要使比较具有参考价值，需满足两个条件。渲染器必须遵循 Lightroom 的曝光和对比度语义，因此我们在单独的导出阶梯上对其进行校准（线性光线 2^{EV}曝光；在三个场景、九种设置下拟合的软枢轴对比度 S 曲线，其均方根为 0.001–0.018，而线性拉伸为 0.04–0.23）。此外，它的绝对基线与 Adobe 的基线也不同，存在逐批次曝光偏移 −0。5 至 +2。0 曝光值。我们在每个批次的不重叠一半数据上（仅近中性编辑）校准一次，并在另一半数据上进行评估：共 383 张照片。

表 8 重现了参数空间排序：真实参数 < 在线经验贝叶斯 < 混合 50 < 先验 ≈ 未编辑。在 $k = 8$，在线经验贝叶斯使 ΔE_{00} 降低 0.41 [0.02、1.03]，并使结构相似性指数相较于先验提高 0.021（$p = 0.042$ 和 $p = 0.006$）；真实参数则使 ΔE_{00} 降低 0.78 [0.08、1.94]。其影响幅度很小：约为 13 的绝对值 ΔE_{00} 主要源于我们的渲染器与 Adobe 相机配置文件之间的差异，因此只有相对比较才有意义。在此限制下，以参数空间衡量的拍摄层面增益，确实使渲染图像更接近摄影师导出的图像。当动作空间扩展到六个以上滑块时，也会出现同样的模式：同批次中位数外观使渲染图像更接近导出图像，而逐照片残差头则不会（附录 D）。

6 讨论

个性化体现在哪里。在这个数据集中，可恢复的信号几乎完全是拍摄级偏移。少量同批次编辑可将曝光误差降低 0.11–0.14 曝光值，而且按批次聚类后，这种降低依然存在。通过冻结的通用编码器和一个小型头部捕获的图像内容，在估计偏移量后几乎没有额外贡献：经过学习的锚点条件头部并不优于闭式收缩，而残差表示在域内取得的增益也无法迁移。这并不意味着图像内容与编辑无关；这意味着在当前数据规模下，使用六个全局参数时，内容残差相对于批次偏移很小，而且无法泛化到其他拍摄。

为什么残差目标无法跨批次泛化。残差头经过训练，用于预测 $y - P(x)$。在域内，先验和残差是在相同批次上估计的，因此残差会吸收稳定的批次特定偏移。在新批次上，该偏移会有所不同；残差头学会了重现训练批次的偏移，却没有机制对其进行更新。锚点估计器之所以成功，恰恰是因为它们会在推理时根据目标批次估计偏移量。失败并非源于优化失败，而是表征与评估单位不匹配。

外部压力测试。在 FiveK 上（4 位专家，每位专家 231 张图像；通过反演渲染输出获得伪标签），在我们的重新分析中，专家身份只能解释 1–8% 的目标方差；平均先验或岭回归的表现与记忆或小型头部模型相当，跨用户模型也能匹配自模型 [11]。这是截然相反的数据条件：风格离散度较低。这支持了这样的解读：当拍摄间的离散度较高时，批次锚点才会发挥作用；同时也提醒我们，FiveK 并非个性化基准。这也表明，估计器并非普遍有益；其收益取决于数据条件。

一种可部署的制品，而非全新架构。 我们推荐的估计器是闭式的，可在中央处理器上运行，除先验之外无需训练权重；它已集成到一款生产级编辑器中。在 $k = 0$ 处，负向训练先验的冷启动误编辑率为 0.99%；在 $k \geq 4$ 处，锚点估计器将曝光平均绝对误差降低了 0.11–0.14 曝光值，且批次聚类区间不包含零。尽管

它并非全新的神经网络，但其可审计性对于一款会修改摄影师文件的工具而言是一项优势。

实用方案。在 $k = 0$，使用负向训练先验；在 $k = 1$–2，先验与批次中位数按 50/50 混合较为稳定；在$k \geq 4$，在线经验贝叶斯或分层经验贝叶斯能获得大部分收益，而在闭式方案中应根据批次宏平均和最差批次表现来选择，而非只看单一微平均。不要在负向训练先验上添加逐照片门控器，也不要使用留一法或逐目标抽样来得出可部署的数值。

对“风格学习”产品的启示。声称能够学习摄影师风格的产品，实际上提出了两项主张：长期先验和逐次拍摄适配。在这套工作流程中，后者更容易实现，也更有价值；只需少量锚点，闭式估计器便能实现这一点。前者确实存在，但作用有限：先验优于简单的基线，但其跨批次误差仍然很高，而仅内容模型也无法弥合差距。供应商应说明其评估衡量的是哪种信号。

7 局限性

1. **一位摄影师**。 所有真实数据均来自一位专业摄影师。本文是一项纵向案例研究，而非跨用户研究。FiveK 风格的多专家数据无法替代：其参数是根据渲染输出推断得出的 [11], ，而且在我们的重新分析中，专家风格高度趋同。

2. **以目录作为批次代理指标**。 目录和文件夹可近似代表拍摄，但与拍摄会话并不完全对应。由于目录和文件夹记录不一致，有三张照片不得不排除。批次效应包含主体、光线、相机和编辑会话效应。

3. **十一个测试批次**。 批次聚类自助法仅有 11 个聚类。区间较宽， p 值应作描述性解读。若要加强推断，首要需求是增加拍摄场次。

4. **主研究中的六个全局参数**。 附录 D将动作空间扩展至十三个全局控制项（高光、阴影、白色色阶、黑色色阶、八波段色相、饱和度和明度中的饱和度、去朦胧），并表明同批次中位数外观可以迁移，但局部蒙版（占目录的 67.3%；占成对导出的 75.4%）、色调曲线、白平衡以及完整的渲染流程仍未纳入研究。结论适用于全局滑块预测，而非完整修图。

5. **渲染空间评估是相对的**。 测试集中的 84.7% 具有原生导出文件，经过裁剪筛选后，渲染检查使用了 383 张照片。其绝对值 ΔE_{00} （约 13）主要受缺失的 Adobe 相机配置文件以及必须在不相交子集上进行校准的逐批次基线影响；只有排序具有可解释性。偏好研究仍属于探索性研究，且由单评定者完成。

6. **选择与漂移**。 这 2,129 张仅目录中的照片之所以缺失，是因为其原始文件未被归档；如果它们存在系统性差异，数据集就会产生偏差。v1 扫描后，有 74 个 v1 标签发生了变化；v2 使用当前值。

7. **未开展正式的偏好或采用研究**。本文不对采用率、用户满意度或审美质量作出任何主张。

哪些情况会改变结论。如果另一位摄影师的拍摄层面离散度较低（如 FiveK），锚点增益就会减小；多评审者偏好研究可能会显示，参数保真度与偏好的差异会进一步扩大；而采用不同的批次定义，也可能降低测得的批次效应。因此，这些主张取决于数据条件，并非普遍适用。

8 可复现性与隐私

我们将发布评估代码、估计器实现、数据模式、划分构建规则，以及冻结清单和划分文件的 SHA-256哈希值。我们不会发布照片、Lightroom目录或伴随文件、真实滑块参数、图像嵌入、批次名称，或可能记住训练图像的检查点。我们会随软件包发布足以核验主要表格的聚合统计数据，并提供一个用于测试数据模式的合成示例。私有数据集只能由其所有者重新创建；发布的代码旨在使估计器和协议能够应用于任何兼容的参数编辑数据集。冻结快照哈希值（前 16 位十六进制字符）为：清单 5b57ec3d8abaff5b，划分 b32ae1f9cc42922c，SigLIP特征 42f64327c560626a，CLIP特征 7ba52678e4f1fa70，阶段标签 33a8c2ee73b6edff。三张排除的照片已记录在划分元数据中；排除前的划分则予以保留，以供审计。

9 结论

照片编辑个性化通常被视为一种单一能力。在一位摄影师的 Lightroom历史记录中，它可以分解为：较弱的跨拍摄内容信号、有用的长期先验，以及能够根据少量同批次编辑估计出的强大拍摄级偏移量。拍摄级偏移量是目前可以迁移到可部署工作流中的部分；残差表示和学习到的锚点条件化头则无法迁移。我们希望，这里的批次共享协议和负面结果能让未来的个性化主张更易于比较，并鼓励将参数保真度、冷启动风险和偏好作为彼此独立的结果进行报告。

A 历史v1结果

v1 集合包含 1,387 张编辑后的照片，采用了不同的数据划分和 V20e/V24e先验。在 v1 中，残差头的表现达到 0.1260 曝光值

，而绝对值头为 0.1404 曝光值；50/50 残差融合达到 0.1164 曝光值，而生产模型为 0.1357 曝光值。这些数字促成了 v2 研究，但不能与 v2 表格直接比较：样本、数据划分、先验和评估协议均不相同。v1 门控结果（冷启动误编辑率从 51% 降至 7%）同样只是历史数据；在 V25-neg 先验下，仅先验本身就能达到 0.99%。

B 在 $k = 8$ 时的每批次行为

表 9： $k = 8$ 时的每批次曝光平均绝对误差。online-EB 列使用单独的采样种子。

批次	组	n	先验	Mix50	E6d
B01	近期	41	0.280	0.260	0.292
B02	近期	57	0.330	0.261	0.364
B03	近期	38	0.850	0.608	0.594
B04	近期	81	0.308	0.270	0.333
B05	未见	338	0.868	0.637	0.582
B06	未见	246	0.392	0.367	0.398
B07	未见	146	1.241	0.848	0.733
B08	未见	60	0.290	0.291	0.310
B09	第二	149	0.415	0.376	0.434
B10	第二	222	0.332	0.338	0.303
B11	第二	70	0.532	0.437	0.363

Mix50 在 11 个批次中有 7 个优于分层经验贝叶斯，但分层经验贝叶斯的微观平均值更低，因为它的优势出现在规模较大的批次中。因此，方案排名取决于批次；论文报告了微观平均、宏观平均和最差批次的数值，并未宣布某个实验组为唯一最佳方案。

C 主线之外的负面结果

以下结果记录于技术报告中，不属于论文的主张：（i）逐阶段编辑：人工工作流平均每张照片包含4.9个阶段，但在每个阶段进行带内容残差的预测，效果都不如同批次锚点中位数（曝光 0.463 → 0.515、色调 9.93 → 13.53、色相、饱和度和明度 4.49 → 7.11）；（ii）渲染回看：在预言机曝光校正后重新编码图像，并不能改善后续阶段的预测（色调 12.52 → 12.74；色相、饱和度和明度 7.55 → 7.75）；（iii）不确定性头：使用折外或留一批次训练目标训练的照片级不确定性模型，对 |曝光| > 2的曲线下面积为 0.30–0.48，低于免训练锚点离散度信号（0.878–0.892）；（iv）合成扰动预训练：ATune 式扰动预训练会使测试平均绝对误差从 0.586 上升至 0.607/0.633；（v）锚点条件学习：拼接、中位数+残差和 E6d+残差的表现均不优于闭式在线经验贝叶斯；（vi）裁剪排序：一个 4B 视觉语言模型在 60 张图像中仅有 36.7% 的结果与摄影师一致，低于固定中心85%基线的 40%；（vii）量化：一个 4B 模型采用 NF4 后可装入 8 GB 内存（峰值 2.86 GiB），但其行为相较半精度浮点有所偏移，因此只能作为建议使用。

D 扩展动作空间：批次级完整调色效果

主要研究预测六项全局调整。目录中的调整更多。在41个目录中的5,357张编辑后的照片中（包括2,129张原始文件未归档的照片），99.5%的照片其高光、阴影、白色色阶或黑色色阶的数值并非默认值，93.6%的照片使用了非恒等色调曲线，57.3%的照片进行了HSL色带编辑。在拥有原生导出的2,751张照片中，36.4%的照片平均色带饱和度变化低于 −10，其中三分之二的照片其全局饱和度标签在 ±2以内；全局饱和度标签与平均色带饱和度之间的相关性为 −0。003。因此，六字段目标未能反映大量色彩处理，而处理后图像也无法为这些维度提供可靠标签。

因此，我们构建了一个 13 维扩展：高光、阴影、白色色阶、黑色色阶、八个 HSL 饱和度区间，以及去朦胧（均以滑块单位计，除以 100）。协议保持不变：3,196 张清单照片，按批次级划分（训练集 1,571 张、验证集 174 张、测试集 1,448 张），并使用相同的 1108 维图像特征。表 10 报告了参数平均绝对误差。对八个同批次锚点取批次中位数，可将平均绝对误差从 0.1605（零值）和 0.0916（全局中位数）降至 0.0823；使用十六个锚点时则达到 0.0789。逐照片残差头无法迁移：其验证集误差为 0.0615，但测试集误差为 0.0960（直接预测为 0.1029）；正则化线性残差为 0.0870–0.0973，均差于批次中位数。相对于批次中位数的配对退化为 +0。0137 [0.0094，0。0195]，根据批次聚类自助法得出。这与六字段残差头的情况相同：域内验证具有误导性，只有批次级统计量能够迁移。

表 10： 13维扩展上的参数平均绝对误差（滑块单位，除以 100）。数值越低越好。

估计器	MAE
Zero	0.1605
全局中位数	0.0916
批次中位数（留一法）	0.0765
批次中位数 , $k = 8$	0.0823
批次中位数 , $k = 16$	0.0789
MLP, 直接	0.1029
多层感知机, $k = 8$ +残差	0.0960
岭回归， $k = 8$ +残差	0.0870–0.0973

表11： 对140张配对测试照片进行渲染检查（对照片取平均值）。ΔE_{00} 是相对于原生导出的结果； $u'v'$ 比率具有曝光不变性，1.0表示完美的颜色匹配。

Arm	ΔE_{00}	$u'v'$比率
真实六个参数	17.80	1.555
+批量风格（$k = 8$）	**16.86**	**1.236**
+真实扩展参数 16.92		1.144

表格 11 将 140 张测试照片与其原生导出文件进行对比。我们在每个实验组中使用真实六个参数，并通过逐通道256点色调查找表应用扩展调色，以此单独考察扩展效果；基线仅使用真实六个参数。加入批量风格后，平均 ΔE_{00} 从 17.80 降至 16.86，且不受曝光影响的 $u'v'$ 色度比从 1.555 降至 1.236（140 张照片中有 105 张有所改善）；照片的真实扩展参数则达到 16.92 和 1.144。因此，八个锚点基本捕获了全部全局增益。剩余差距来自局部蒙版（占配对照片的 75.4%）、色调曲线和 Adobe 相机配置文件。

另外两项控制也未带来正面结果。批量中位数色调曲线呈现出明显的参数信号（9点平均绝对误差从 0.0139 降至 0.0065， −53%），但没有渲染增益：将其加入批量风格后，平均 ΔE_{00} 仅改变了 −0。004，真实曲线也同样没有帮助。曲线算子本身是准确的——逐通道曲线能够复现 Lightroom 的对比度阶梯

：仅亮度应用时为 0.54–1.05 ΔE_{76}，而逐通道曲线为 1.45–2.77——因此，零结果源于与基础色调路径存在冗余，而非样条曲线不佳。色彩分级没有批次信号（平均绝对误差为 0.0350，对比 0.0349）。加入批次中位数 HSL 色相和亮度也会降低效果（+0。30 ΔE_{00}，140 张中仅 33 张有所改善），因此最终采用的风格只保留饱和度。

该产品的实现采用闭式且可审计：项目外观命令会对同一文件夹中已编辑照片的四个色调字段、八个饱和度区间和去朦胧取中位数（仅读取应用侧车文件或 Lightroom XMP），并将结果应用于所选照片，而不触及几何、蒙版或包含六个字段的自动色调快照。色调曲线和色彩分级则明确排除在外。

E 数值溯源

主表中的所有数值均来自冻结的 v2 快照中排除后的报告。发布包中的 aggregates/main_results.json 包含相同数值；dataset_summary.json 包含计数和标签极值；hashes.json 包含指纹。对应关系如下：E5 来自 E5-v2 报告及其批次聚类置信区间文件；E6b 来自批次共享锚点曲线及其聚类置信区间文件；锚点变体来自锚点变体评估及其聚类置信区间文件；E7 来自 E7-v2 报告；盲法研究来自 50 张图像审阅日志。附录中的扩展数据来自单独的全设置扫描（grade_full_2026_11_21.jsonl）和 E8/E8b/E8c/E8d 报告；这些数据是六字段台账的扩展，绝不会与主表数据合并。论文中的所有数值均直接取自这些文件，且论文与汇总文件会同步更新。

参考文献

[1] Adobe Inc.，《Adobe Lightroom》 [计算机软件]。[2] Y. Hu、H. He、C. Xu、B. Wang 和 S. Lin，《曝光：一种白盒照片后期处理框架》，ACM Transactions on Graphics，第 37 卷第 2 期，2018 年。[3] W. Lötzsch、M. Reimann、M. Büssemeyer、A. Semmo、J. Döllner 和 M. Trapp，《WISE：基于示例学习的白盒图像风格化》，载于 ECCV，2022 年，第 135–152 页。[4] S. Bianco、C. Cusano、F. Piccoli 和 R. Schettini，《利用神经样条颜色变换实现个性化图像增强》，IEEE Transactions on Image Processing，第 29 卷，2020 年，第 6223–6236 页。[5] M. Gharbi、J. Chen、J. T. Barron、S. W. Hasinoff 和 F. Durand，《用于实时图像增强的深度双边学习》，ACM Transactions on Graphics，第 36 卷第 4 期，2017 年。[6] Adobe Inc.，《学习用户对照片调整的偏好》，美国专利 9,361,666，2016 年。[7] RareLight，《StyleAI》 [计算机软件], GitHub 代码仓库，https://github.com/RareLight/StyleAI。[8] T. M. Weldengus、B. Liu、F. Kou、Y. Lyu、J. Chen、Q. Fan 和 C. Zou，《RefRetouch：无需测试时微调的个性化图像修图》，arXiv:2602.17044，2026 年。[9] H.-U. Kim、Y. J. Koh 和 C.-S. Kim，《PieNet：个性化图像增强网络》，载于 ECCV，2020 年，第 374–390 页。[10] K. Wu、S. Jiang、M. Ku、P. Nie、M. Liu 和 W. Chen，《EditReward：与人类偏好对齐的指令引导图像编辑奖励模型》，载于 ICLR，2026 年。[11] V. Bychkovsky、S. Paris、E. Chan 和 F. Durand，《利用输入/输出图像对数据库学习摄影全局色调调整》，载于 CVPR，2011 年。[12]D. Goyal、G. Ruiz 和 A. Meisami，《ATune：通过图像编辑器控件进行无监督照片修图》，arXiv:2611.16236，2026 年。[13] M. M. Ho 和 J. Zhou，《Deep Preset：通过色彩风格迁移混合与修饰照片》，载于 WACV，2021 年。[14] J. Kim、J. Yoo 和 S. J. Kim，《根据成对用户偏好学习个性化摄影风格》，载于 CVPR，2026 年，第 1134–1144 页。[15] C. Zhu、C. Xu、S. Weng、J. Fang、P. Zhang、S. Li 和 B. Shi，《个性化图像滤镜：掌握你的摄影风格》，arXiv:2510.16791，2025 年。[16] T. Nguyen、U. Ojha、Y. Li、H. Liu 和 Y. J. Lee，《Edit one for all：交互式批量图像编辑》，载于 CVPR，2024 年，第 8271–8280 页。[17] Imagen AI，《Personal AI profile》 [计算机软件], https: //imagen-ai.com。