

SAGA

解决3D提示性分割中数据稀缺、标注成本高、多粒度歧义等问题

三大核心设计：

高斯亲和特征 软尺度门机制 尺度感知对比训练
(D=32)

尺度条件3D特征：

→ 将2D掩码的视觉大小转化为3D空间物理尺度

基于3D物理尺度的特征适配机制，通过量化2D掩码对应的3D目标物理大小，为不同粒度的3D分割提供统一判断标准

计算逻辑：

(1) input:

2D掩码 M_i : 由SAM提取2D多视角掩码

相机内参：用于2D投影到3D

3D深度信息：由预训练3DGs预测

(2) 公式：

$$s_M = 2\sqrt{\text{std}(\mathcal{X}(\mathcal{P}))^2 + \text{std}(\mathcal{Y}(\mathcal{P}))^2 + \text{std}(\mathcal{Z}(\mathcal{P}))^2},$$

s_M — 2D掩码对应的3D尺度 \mathcal{P} — 将2D掩码 M_i 投影到3D空间后得到的点云

$X(\mathcal{P}), Y(\mathcal{P}), Z(\mathcal{P})$ — 点云 \mathcal{P} 在 X, Y, Z 三个坐标轴上的集合

$\text{std}(\cdot)$ — 计算集合标准差 (反映点云在该轴分布范围，即目标的尺寸跨度)

乘2：标准差范围 → 目标直径级尺度

整体流程：

特征嵌入 → 尺度匹配 → 训练蒸馏 → 推理分割

输入：

预训练的3DGs模型，多视角2D图像，2D视觉提示，指定3D尺度 S

核心组件：

高斯亲和特征：为每个3D高斯附加高维特征 $f_g \in \mathbb{R}^D$ ($D=32$)，作为分割能力载体

软尺度门：根据尺度 S 调整特征通道权重，解决多粒度歧义

训练阶段：

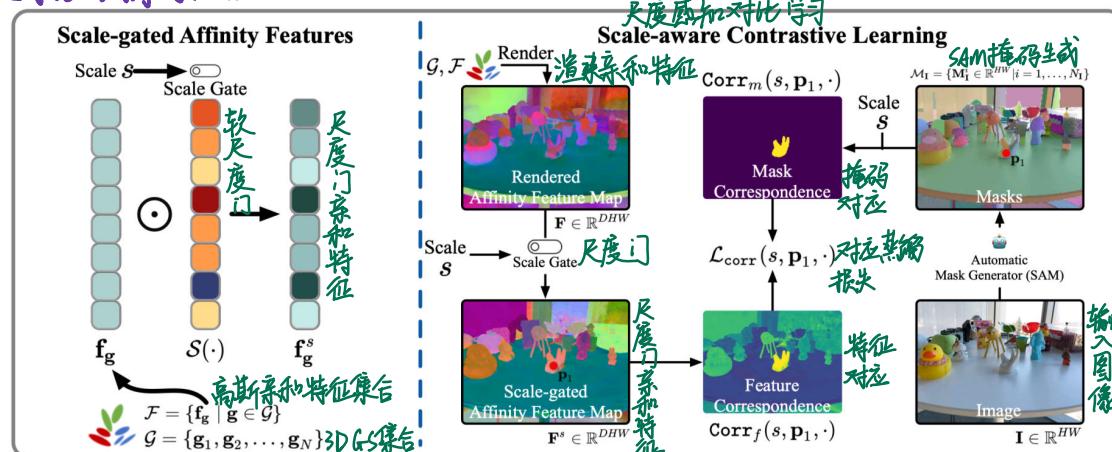
用SAM 提取多视角2D掩码，计算每个掩码3D物理尺度 s_M

通过“尺度感知对比学习”，将2D掩码的分割能力蒸馏到3D高斯的亲和特征中

加入局部特征平滑和特征范数正则化，提升特征稳定性和平齐度。

输出：

训练后，给定2D提示和尺度 S ，输出对应3D目标的分割结果 (由3D高斯表示)，支持场景自动分解，开放词汇分割等任务



高斯亲和特征(Gaussian Affinity Feature)

1. 尺度门亲和特征 (Scale-Gated Affinity Feature)

让同一亲和特征在不同尺度下适配不同粒度的分割需求，解决多粒度歧义

尺度门结构：量化化，由单线性层 + Sigmoid 函数组成，映射关系为

$$S : [0, 1] \rightarrow [0, 1]^D \text{ 将尺度 } S \text{ 转换为门向量 } S(s)$$

公式：尺度门与亲和特征的哈达玛积（逐元素相乘），得到尺度适配后的特征

$$f_g^s = S(s) \odot f_g$$

训练：先将 3D 亲和特征渲染为 2D 特征图 $F(p)$ ，再应用尺度门 $F^s(p) = S(s) \odot F(p)$

推理：直接对 3D GS 用尺度门

2. 局部特征平滑：

3D 空间中存在噪声高斯，会产生假阳性分割结果

利用 3D 高斯空间局部性，用 K 近邻 ($K=16$) 的亲和特征均值替换

$$f_g \leftarrow \frac{1}{K} \sum_{g'} e^{-\gamma \|g-g'\|_2^2} f_g'$$

消除噪声干扰

尺度感知对比学习 (Scale-Aware Contrastive Learning) 训练亲和特征的核心策略

通过 2D 掩码的“尺度一像素归属关系”监督 3D 特征学习，利用可微分光栅化实现反向传播

尺度感知像素身份向量 (Scale-Aware Pixel Identity Vector)

通过 2D 掩码转换为“尺度依赖的监督信号”，明确不同尺度下像素的掩码归属

构建步骤：

(1) 按 3D 尺度降序排序图像 I 的掩码集合 M_I ，得有序列表 $O_I = (M_I^{(1)}, \dots, M_I^{(N)}) (S_{M_I^{(1)}} > \dots > S_{M_I^{(N)}})$

(2) 对每个像素 P 和尺度 S ，定义向量 $V(S, P) \in \{0, 1\}^{N_I}$

计掩码强度 $S_{M_I^{(i)}} < S$ ：

向量第 i 位为掩码在 P 点的值 $M_I^{(i)}(P)$ ；

else：

向量第 i 位为 1 当且仅当 $M_I^{(i)}(P)=1$ ，且所有更小尺度掩码在 P 点为 0

两个像素若在尺度 S 下共享至少一个掩码 ($V(S, p_1), V(S, p_2) > 0$)，则其尺度门特征应相似

假设我们有一张桌子的多视角图像 I，SAM 自动提取了 3 个多粒度 2D 掩码（对应 3D 目标的不同粒度），每个掩码的 3D 物理尺度已计算完成：

- 掩码 M1：苹果（最小粒度，3D 尺度 $s_1=0.2$ ，仅覆盖苹果区域）
- 掩码 M2：桌面（中粒度，3D 尺度 $s_2=0.8$ ，仅覆盖桌面区域，不含苹果）
- 掩码 M3：整张桌子（粗粒度，3D 尺度 $s_3=1.5$ ，覆盖桌面 + 苹果 + 桌腿）

按论文要求，先将掩码按 3D 尺度降序排序，得到有序列表 $O_I = [M_3 (s=1.5), M_2 (s=0.8), M_1 (s=0.2)]$ ，此时 $N_I=3$ （向量维度为 3）。

我们选两个典型像素：

- p1：苹果表面的像素（属于 M1、M3，不属于 M2）
- p2：桌面边缘的像素（属于 M2、M3，不属于 M1）

例子 1：指定尺度 $s=0.5$ （中粒度，关注“桌面 / 苹果”级别）

按构建步骤计算 $V(s=0.5, p1)$ 和 $V(s=0.5, p2)$ ：

1. 对每个掩码判断“尺度是否 $\geq s=0.5$ ”：

- M3 ($s=1.5 \geq 0.5$)：属于“尺度 $\geq s$ ”的掩码
- M2 ($s=0.8 \geq 0.5$)：属于“尺度 $\geq s$ ”的掩码
- M1 ($s=0.2 < 0.5$)：属于“尺度 $< s$ ”的掩码

(1) $V(s=0.5, p1)$ (苹果上的像素)：

- 第 1 位 (对应 M3)： $M_3(p1)=1$ (苹果在整张桌子里)；且所有 “ $s \leq s_{Mj} < s_{M3}$ ” 的掩码是 M2 ($s=0.8$, 满足 $0.5 \leq 0.8 < 1.5$)， $M_2(p1)=0$ (苹果不在桌面里) → 满足条件，取 1
- 第 2 位 (对应 M2)： $M_2(p1)=0$ (苹果不在桌面里) → 取 0
- 第 3 位 (对应 M1)： $M_1(p1)=1$ (像素在苹果掩码里) → 取 1
- 最终 $V(s=0.5, p1) = [1, 0, 1]$

(2) $V(s=0.5, p2)$ (桌面边缘的像素)：

- 第 1 位 (对应 M3)： $M_3(p2)=1$ (桌面在整张桌子里)；但 “ $s \leq s_{Mj} < s_{M3}$ ” 的掩码是 M2 ($s=0.8$)， $M_2(p2)=1$ (像素在桌面里) → 不满足“所有更小尺度掩码在 $p2$ 为 0”，取 0
- 第 2 位 (对应 M2)： $M_2(p2)=1$ (像素在桌面里)；且所有 “ $s \leq s_{Mj} < s_{M2}$ ” 的掩码不存在 (M1 的 $s=0.2 < 0.5$ ，不满足 $s \leq 0.2 < 0.8$) → 取 1
- 第 3 位 (对应 M1)： $M_1(p2)=0$ (像素不在苹果掩码里) → 取 0
- 最终 $V(s=0.5, p2) = [0, 1, 0]$

3. 核心意义验证：

计算两个向量的点积： $V(p1) \cdot V(p2) = (1 \times 0) + (0 \times 1) + (1 \times 0) = 0 \rightarrow$ 不共享任何掩码，因此 p1 和 p2 的“ $s=0.5$ 尺度门特征”应不相似（符合直觉：中粒度下，苹果和桌面是不同目标，不应被归为同一类）。

Loss function：对应蒸馏损失 + 特征范数正则化

(1) 对应蒸馏损失：

掩码对应性： $\text{Corr}_m(S, P_1, P_2) = |V(S, P_1), V(S, P_2)|$ 指示函数，1 表示同属一个掩码

特征对应性： $\text{Corrf}(S, P_1, P_2) = \langle F^s(P_1), F^s(P_2) \rangle$ 尺度门余弦特征相似度

损失： $L_{\text{corr}}(S, P_1, P_2) = C_1 - 2 \cdot \text{Corrm}(S, P_1, P_2) \cdot \max(\text{Corrf}(S, P_1, P_2), 0)$

例子 2：指定尺度 $s=1.2$ (粗粒度，关注“整张桌子”级别)

同样计算 $V(s=1.2, p1)$ 和 $V(s=1.2, p2)$:

1. 对每个掩码判断“尺度是否 $\geq s=1.2$ ”：

- M3 ($s=1.5 \geq 1.2$) : 属于“尺度 $\geq s$ ”的掩码
- M2 ($s=0.8 < 1.2$) : 属于“尺度 $< s$ ”的掩码
- M1 ($s=0.2 < 1.2$) : 属于“尺度 $< s$ ”的掩码

2. 逐位计算向量值：

(1) $V(s=1.2, p1)$:

- 第 1 位 (对应 M3) : M3 ($p1=1$)；且所有 “ $s \leq s_{Mj} < s_{M3}$ ” 的掩码是 M2 ($s=0.8 < 1.2$, 不满足 $s \leq 0.8 < 1.5$ 且 $0.8 \geq 1.2$) → 无符合条件的更小尺度掩码, 取 1

- 第 2 位 (对应 M2) : M2 ($p1=0$) → 取 0

- 第 3 位 (对应 M1) : M1 ($p1=1$) → 取 1

最终 $V(s=1.2, p1) = [1, 0, 1]$

(2) $V(s=1.2, p2)$:

- 第 1 位 (对应 M3) : M3 ($p2=1$)；且所有 “ $s \leq s_{Mj} < s_{M3}$ ” 的掩码是 M2 ($s=0.8 < 1.2$, 不满足 $0.8 \geq 1.2$) → 无符合条件的更小尺度掩码, 取 1
- 第 2 位 (对应 M2) : M2 ($p2=1$) → 取 1
- 第 3 位 (对应 M1) : M1 ($p2=0$) → 取 0

最终 $V(s=1.2, p2) = [1, 1, 0]$

3. 核心意义验证：

计算点积： $V(p1) \cdot V(p2) = (1 \times 1) + (0 \times 1) + (1 \times 0) = 1 > 0 \rightarrow$ 共享 M3 (整张桌子) 掩码, 因此 p1 和 p2 的“ $s=1.2$ 尺度门特征”应相似 (符合直觉：粗粒度下，苹果和桌面都属于整张桌子，应归为同一类)。

(2) 特征范数正则化：

2D 特征是多个 3D 特征的线性组合，可能导致 2D 分割好，3D 分割差 (特征方向不一致)

渲染 2D 特征时，先将 3D 特征归一化为单位向量，再约束 2D 特征 L_2 范数逼近 1 (表示特征方向一致)

$$L_{norm}(p) = 1 - \|F(p)\|_2$$

$$L = \sum_{(p1, p2)} L_{corr} + \sum_p L_{norm}$$

额外训练策略 (解决数据不平衡)

(1) 尺度敏感不平衡：多数像素对尺度变化不敏感 → 重采样“尺度敏感像素对”

(2) 正负样本不平衡：负样本远多于正样本 → 按 1:1 比例采样正负样本

(3) 目标大小不平衡：大目标主导优化 → 按像素所属掩码的平均大小加权 $W(p1, p2) = 1/(mp1, mp2)$

Inference

1. 提示性分割：

输入特定视角 2D 点提示 + 指定 3D 尺度 s

2D 提示映射为 3D 尺度门查询特征，计算其所有 3D 高斯亲和特征的相似度，相似度 > 阈值的高斯构成目标分割结果

2. 3D 场景自动分解：

利用训练好的亲和特征，直接用 HDBSCAN 聚类

3. 开放词汇分割：

CLIP + 投票机制

聚类多视角 2D 掩码 → 提取每个掩码的 CLIP 特征 → 输入文本提示，计算 mask 与文本相关性 → 聚类分割

Limitation:

对未见过的目标的分割能力缺失 (尤其是小目标)

SAGA 的高斯亲和特征完全依赖 SAM 提取的多视角 2D 掩码进行训练，只有出现在这些掩码中的目标/部件，其对应的 3DGs 才能学到有效的分割特征。

算法流程：

1. 提取多粒度 2D 掩码 (SAM)

2. 计算 3D 物理尺度 (预测深度、相机内参)

3. 初始化高斯亲和特征

4. 尺度感知对比训练 (蒸馏 2D \rightarrow 3D)

(1) 构建尺度感知像素身份向量

(2) α -blending 渲染亲和特征图

(3) 尺度门适配

(4) 损失计算与优化： $L_{corr} + L_{norm}$

5. 推理