



Idea

- 潜在拡散モデルを用いた**レイアウト制御可能な画像生成手法**を提案
- レイアウトを決定づける重要な要因となる**初期ノイズを最適化**

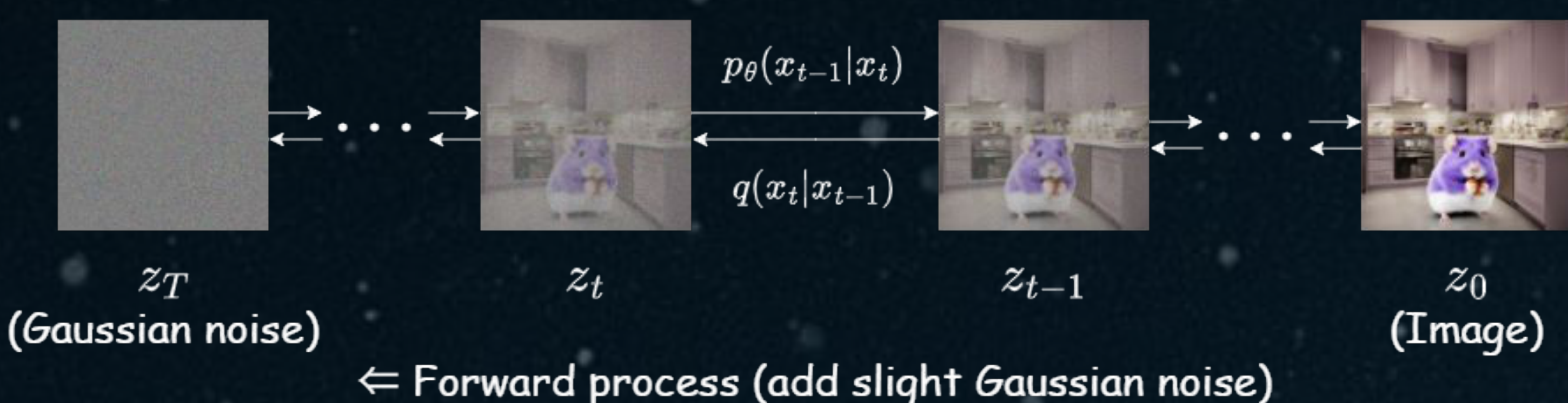


Background

潜在拡散モデル [Rombach+ CVPR'22]

- 初期ノイズから逐次的にノイズを除去し画像を生成
- テキスト等の条件付けにより任意の画像を生成可能

Reverse process (estimate and eliminate slightly added Gaussian noise) \Rightarrow



初期ノイズ z_T の性質

- ガウス分布に従う乱数であり, 構造決定の一因
- 特定の画素 Block が特定の概念と癒着 [Mao+ ECCV'24]

Layout to Image の現状

- 高コストな fine-tuning による手法 [Li+ CVPR'23]
- attention の編集による training-free の手法は初期ノイズによる影響を受け低品質 [Chen+ WACV'24]

目的

- 追加学習なしで, 自然なレイアウト制御を実現
- 目的の attention に一致するよう初期ノイズを最適化
- 初期ノイズ最適化が及ぼす生成画像への影響を分析

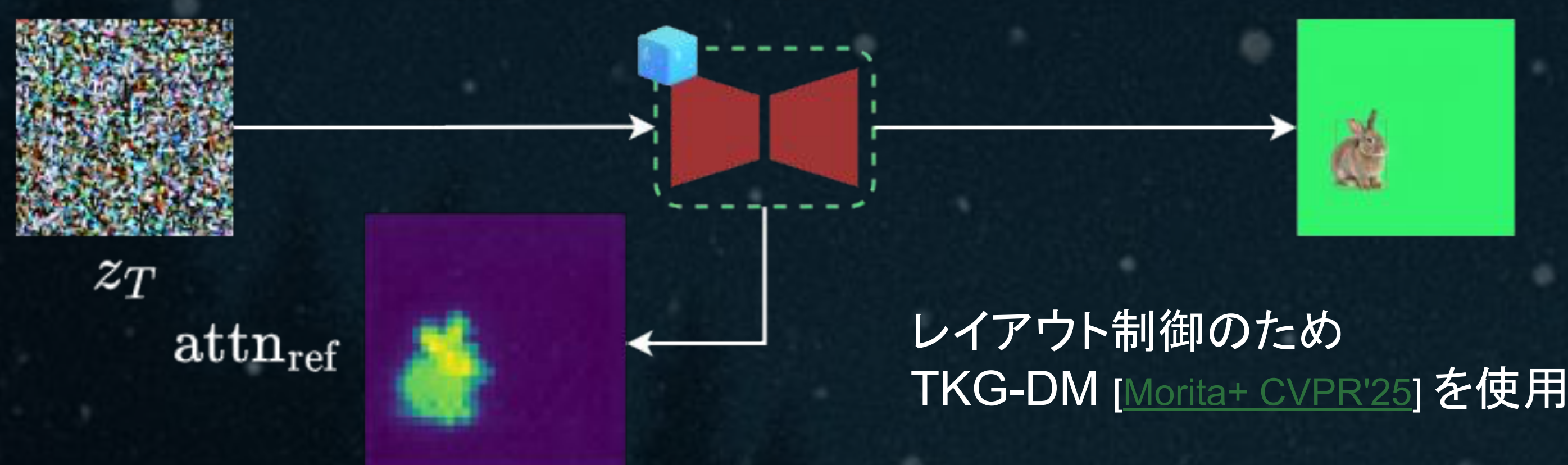
NOEL-Diffusion Noise Optimization for Enhanced Layout in Diffusion Model

アイデア

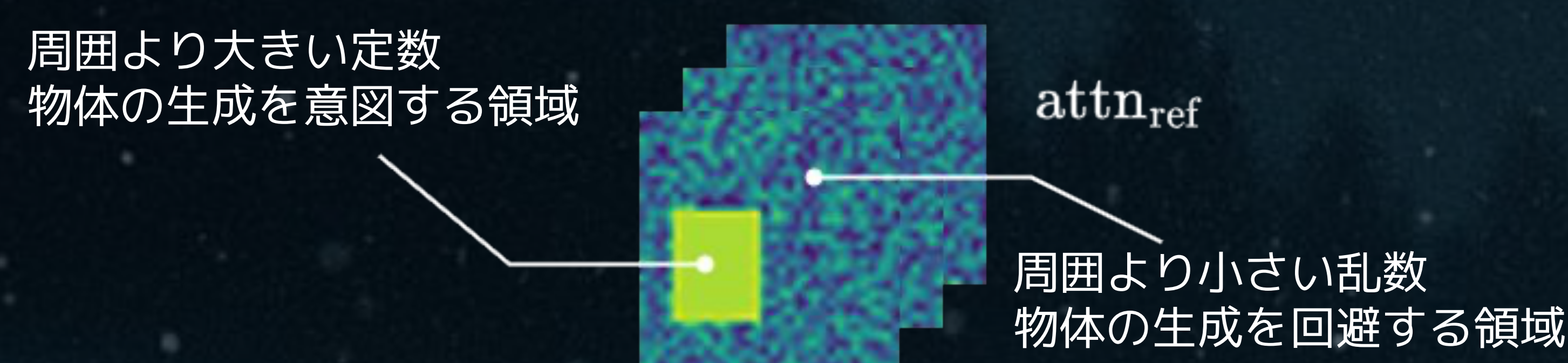
目的の attention (attn_{ref}) をもとに z_T を最適化

2種の attn_{ref} の取得方法を検討

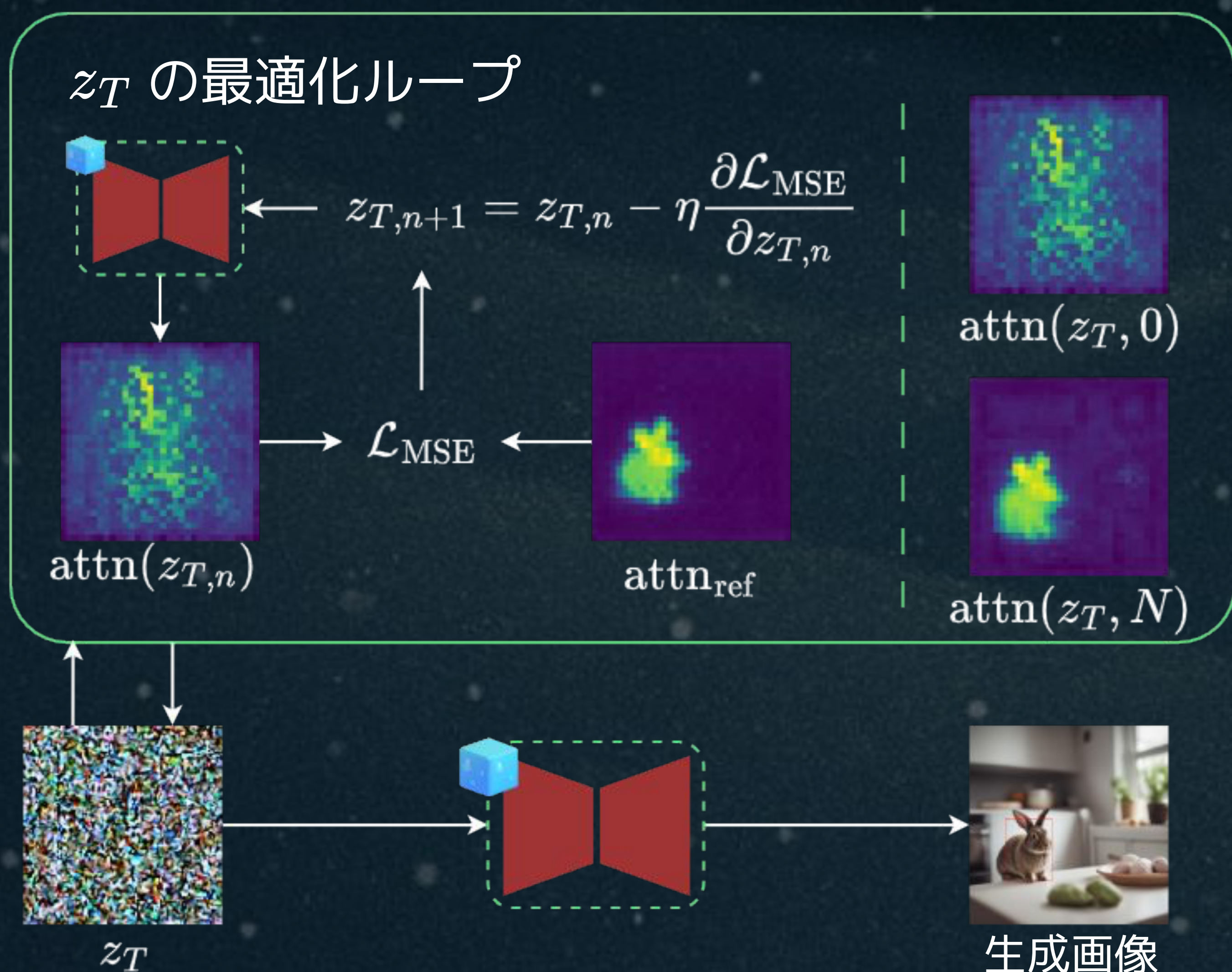
方法1: 中間生成物として取得



方法2: 物体の位置を考慮して手動で構築



画像生成 with NOEL-Diffusion

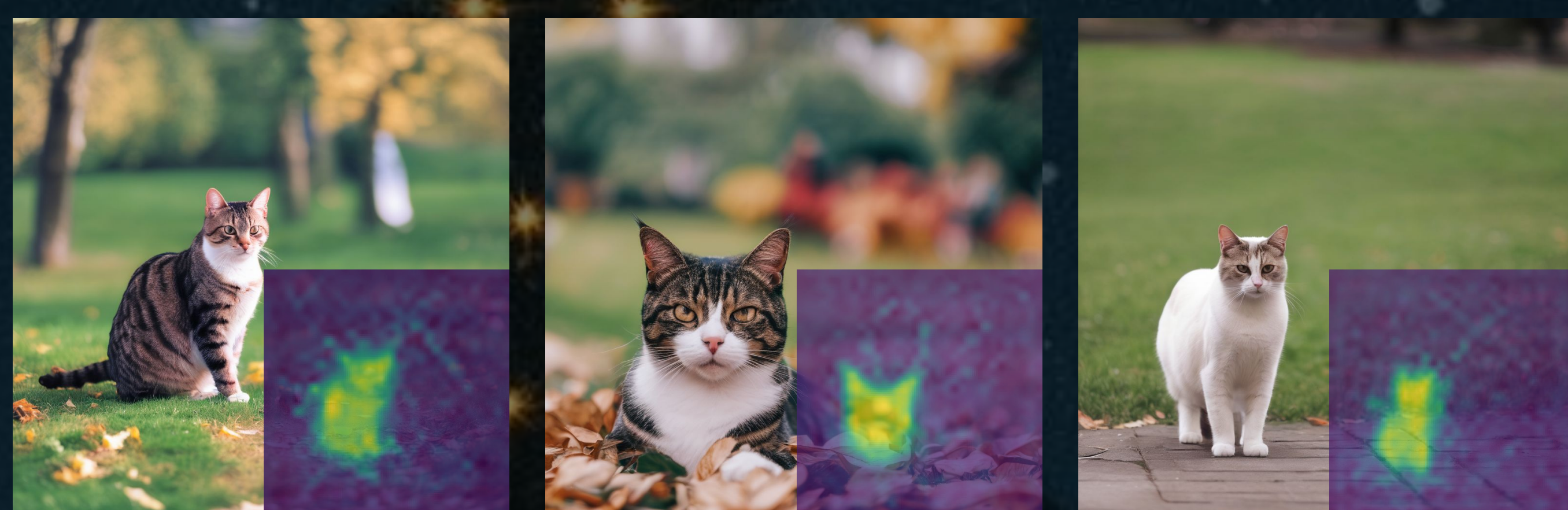


Generation Results & Discussion

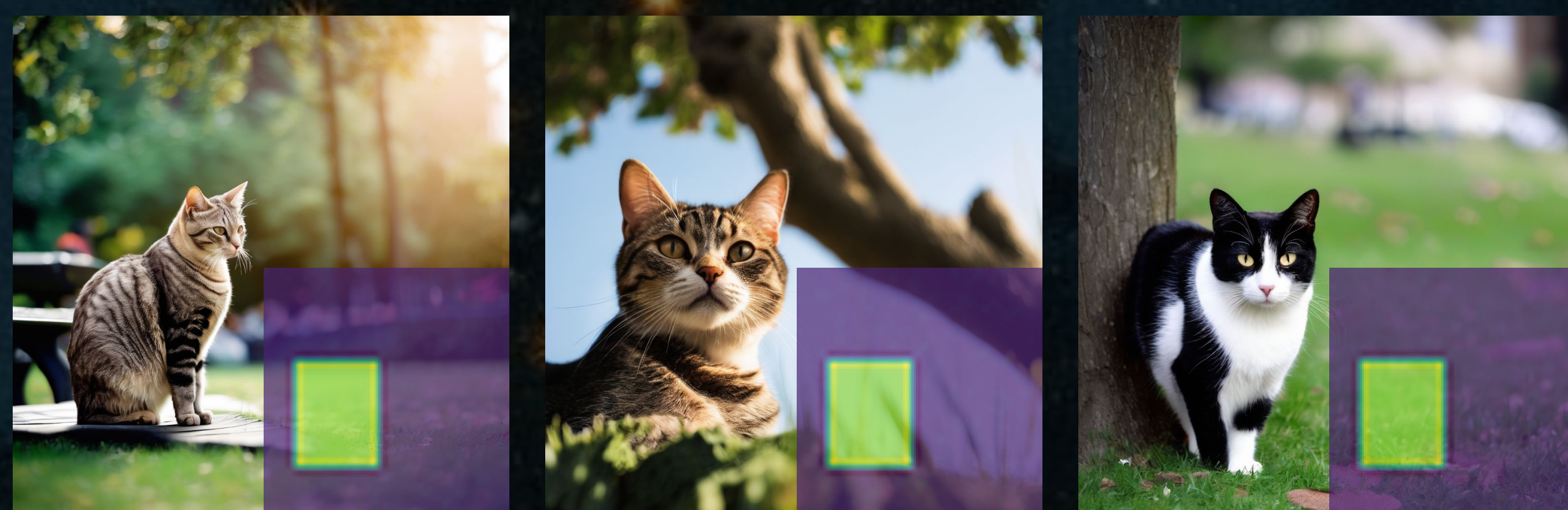
model: SDXL [Podell+ ICLR'24]

prompt: 'Photo of a cat in the park'

方法1



方法2



結果

- 方法1では物体が attn_{ref} の形状を概ね反映し, 指定位置に対象物体が生成された
- 簡易的な矩形の attention を与えた方法2でも, 指定位置に対象物体が生成された

考察

- 方法2から z_T の最適化において U-Net から得られる厳密な attention を必要としないことが示された
 - デノイズ初期は低周波情報が優位であり, 方法1の attn_{ref} のように詳細な形状情報は不要
- 方法2では対象物体周辺が無作為な attention となるように最適化を行う
 - 特定概念と癒着しない初期ノイズが得られる
 - 対象物体が指定位置外に生成されることを抑制

今後の展望

- 広域的な実験と定量評価を行う
 - 画質・テキストとの一致度・レイアウト忠実度 etc.