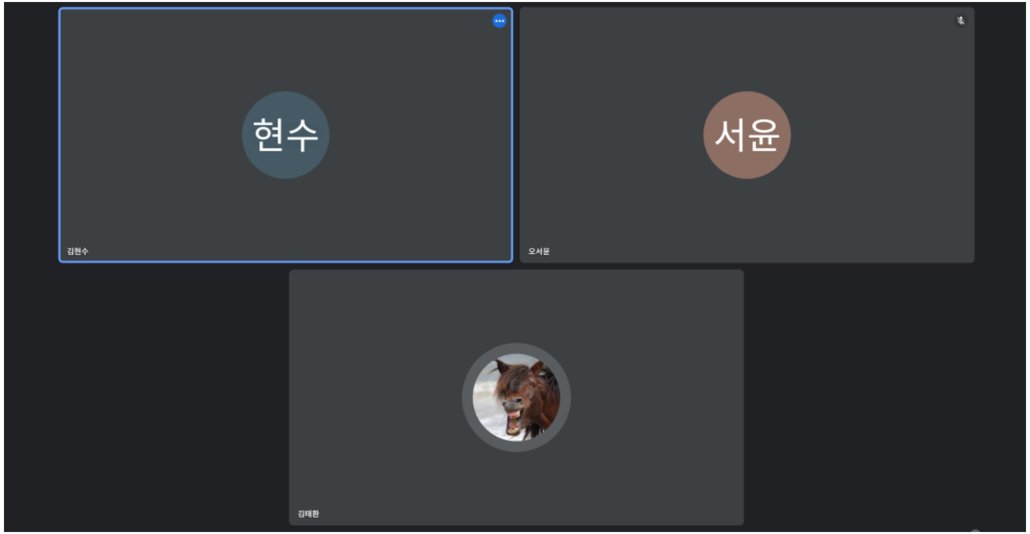
CUAI 딥러닝 논문 리뷰 스터디 CV 1팀

2024.11.12

발표자 : 김태환

목차

스터디원 소개 및 만남 인증



스터디원 1 : 김태환

스터디원 2 : 김현수

스터디원 3 : 오서윤

논문 리스트

Long Short Term Memory

Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling

EfficientNet

DragDiffusion

DragonDiffusion

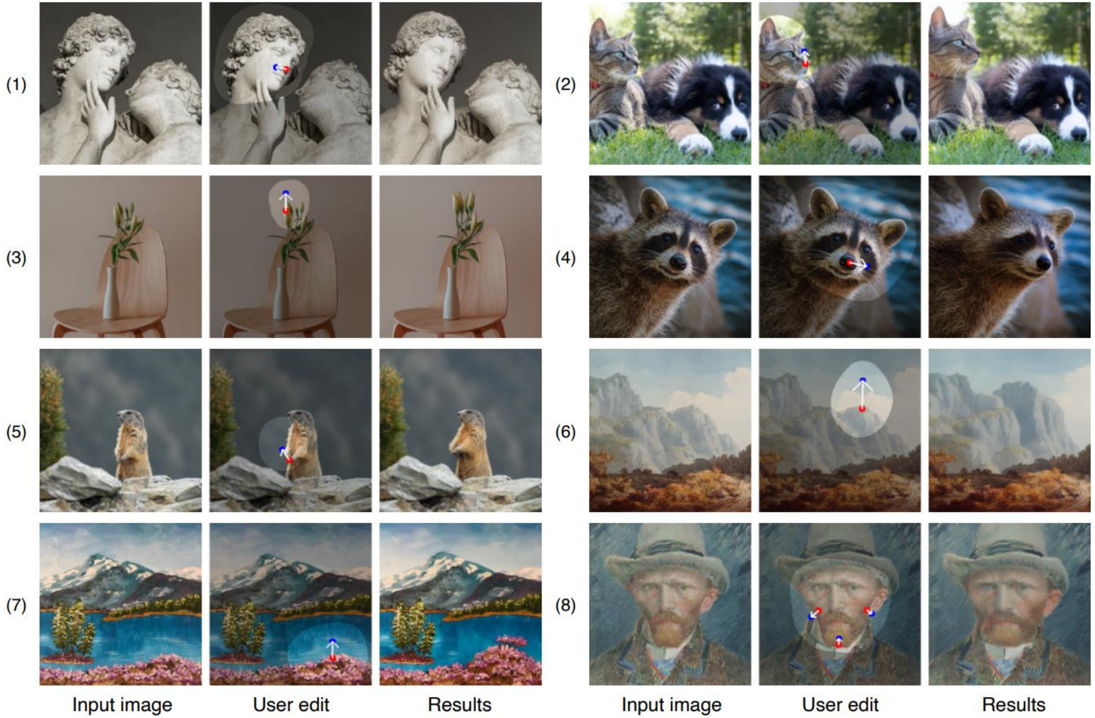
CLIP

Mask R-CNN

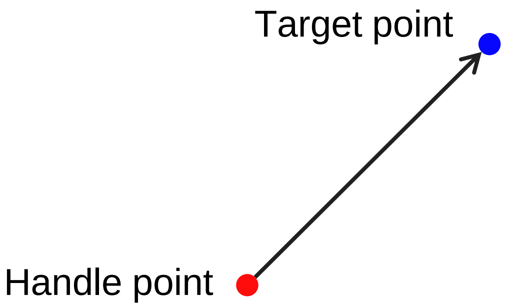
DETR

Task

Point-based image editing (Drag-based image)

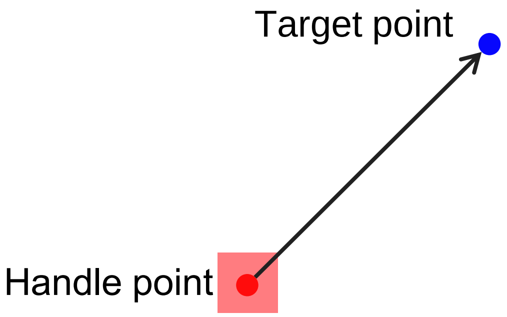


기호 및 용어 정리



기호 및 용어 정리

$\Omega(h_i^k, r)$: k번 업데이트된 i번째 handle point 주변 한 변이 2r인 정사각형 패치



기호 및 용어 정리

ϵ_θ: Unet

z_t: t번 노이즈가 들어간 latent

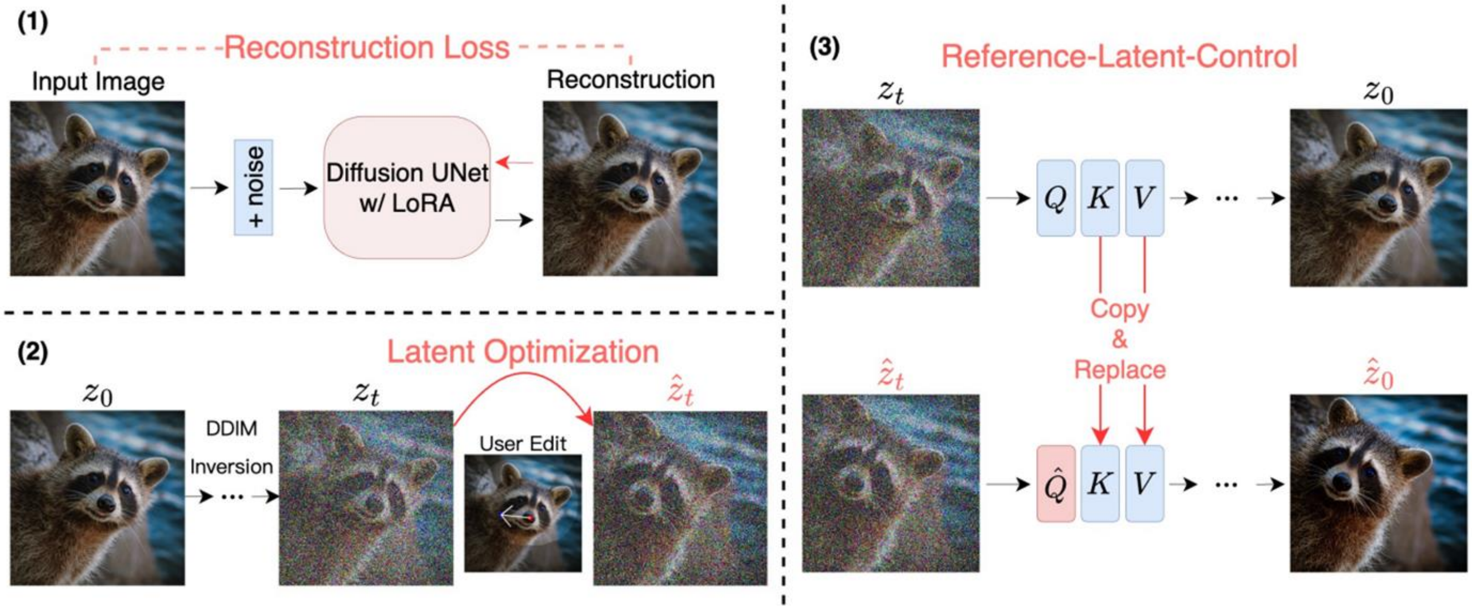
$\hat{z}_t^k$: t번 노이즈가 들어가고, k번 optimize된 latent

$F_q(\cdot)$: input의 위치 q에서의 feature

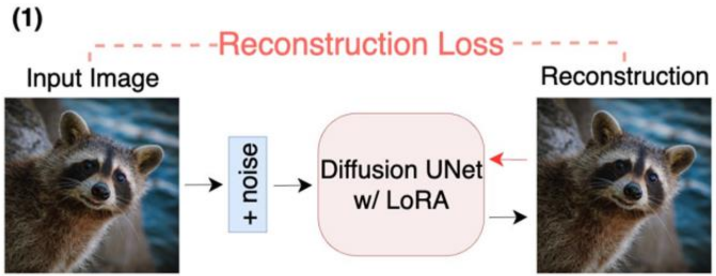
$sg(\cdot)$: stop gradient

M: 편집할 영역의 binary mask

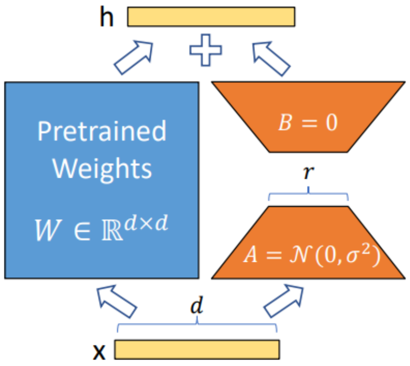
DragDiffusion: 전체 프로세스



DragDiffusion: Identity-preserving Fine-tuning

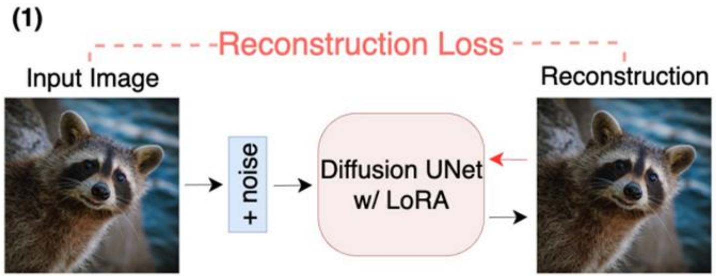


LoRA

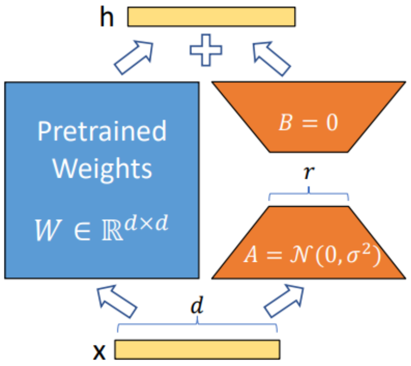


$$\mathcal{L}_{\text{ft}}(z, \Delta\theta) = \mathbb{E}_{\epsilon,t}\left[\|\epsilon - \epsilon_{\theta+\Delta\theta}(\alpha_t z + \sigma_t \epsilon)\|_2^2\right],$$

DragDiffusion: Identity-preserving Fine-tuning



LoRA



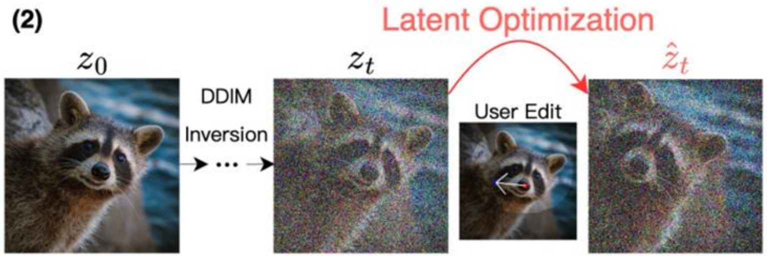
$$\mathcal{L}_{\text{ft}}(z, \boxed{\Delta\theta}) = \mathbb{E}_{\epsilon,t}\left[\|\epsilon - \epsilon_{\theta\boxed{+\Delta\theta}}(\alpha_t z + \sigma_t \epsilon)\|_2^2\right],$$

DragDiffusion: Motion tracking

Motion supervision

$$\mathcal{L}_{\text{ms}}(\hat{z}_t^k) = \sum_{i=1}^{n} \sum_{q \in \Omega(h_i^k, r_1)} \left\| F_{q+d_i}(\hat{z}_t^k) - \text{sg}(F_q(\hat{z}_t^k)) \right\|_1 + \lambda \left\| (\hat{z}_{t-1}^k - \text{sg}(\hat{z}_{t-1}^0)) \odot (\mathbb{1} - M) \right\|_1$$

$$\hat{z}_t^{k+1} = \hat{z}_t^k - \eta \cdot \frac{\partial \mathcal{L}_{\text{ms}}(\hat{z}_t^k)}{\partial \hat{z}_t^k}$$



Point tracking

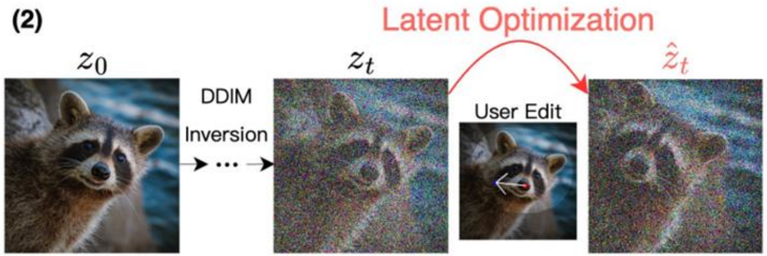
$$h_i^{k+1} = \underset{q \in \Omega(h_i^k, r_2)}{\arg\min} \left\| F_q(\hat{z}_t^{k+1}) - F_{h_i^0}(z_t) \right\|_1$$

DragDiffusion: Motion tracking

Motion supervision

$$\mathcal{L}_{\text{ms}}(\hat{z}_t^k) = \boxed{\sum_{i=1}^{n} \sum_{q \in \Omega(h_i^k, r_1)} \left\| F_{q+d_i}(\hat{z}_t^k) - \text{sg}(F_q(\hat{z}_t^k)) \right\|_1} + \lambda \left\| (\hat{z}_{t-1}^k - \text{sg}(\hat{z}_{t-1}^0)) \odot (\mathbb{1} - M) \right\|_1$$

$$\hat{z}_t^{k+1} = \hat{z}_t^k - \eta \cdot \frac{\partial \mathcal{L}_{\text{ms}}(\hat{z}_t^k)}{\partial \hat{z}_t^k}$$



Point tracking

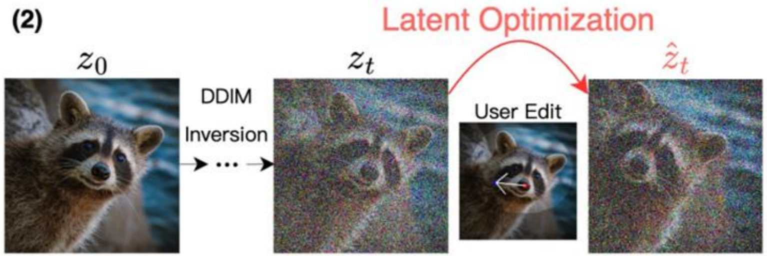
$$h_i^{k+1} = \underset{q \in \Omega(h_i^k, r_2)}{\arg\min} \left\| F_q(\hat{z}_t^{k+1}) - F_{h_i^0}(z_t) \right\|_1$$

DragDiffusion: Motion tracking

Motion supervision

$$\mathcal{L}_{\mathrm{ms}}(\hat{z}_t^k) = \sum_{i=1}^{n} \sum_{q \in \Omega(h_i^k, r_1)} \left\| F_{q+d_i}(\hat{z}_t^k) - \mathrm{sg}(F_q(\hat{z}_t^k)) \right\|_1 \boxed{+ \lambda \left\| (\hat{z}_{t-1}^k - \mathrm{sg}(\hat{z}_{t-1}^0)) \odot (\mathbb{1} - M) \right\|_1}$$

$$\hat{z}_t^{k+1} = \hat{z}_t^k - \eta \cdot \frac{\partial \mathcal{L}_{\mathrm{ms}}(\hat{z}_t^k)}{\partial \hat{z}_t^k}$$



Point tracking

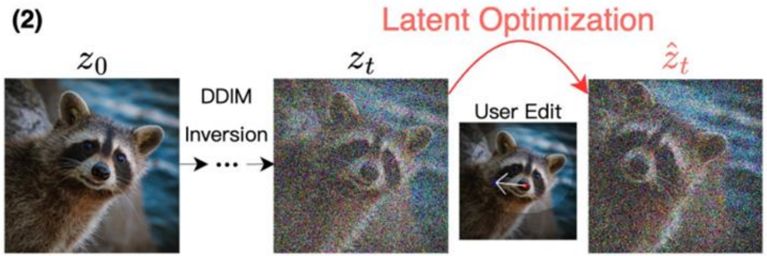
$$h_i^{k+1} = \underset{q \in \Omega(h_i^k, r_2)}{\arg\min} \left\| F_q(\hat{z}_t^{k+1}) - F_{h_i^0}(z_t) \right\|_1$$

DragDiffusion: Motion tracking

Motion supervision

$$\mathcal{L}_{\text{ms}}(\hat{z}_t^k) = \sum_{i=1}^{n} \sum_{q \in \Omega(h_i^k, r_1)} \left\| F_{q+d_i}(\hat{z}_t^k) - \text{sg}(F_q(\hat{z}_t^k)) \right\|_1 + \lambda \left\| (\hat{z}_{t-1}^k - \text{sg}(\hat{z}_{t-1}^0)) \odot (\mathbb{1} - M) \right\|_1$$

$$\boxed{\hat{z}_t^{k+1} = \hat{z}_t^k - \eta \cdot \frac{\partial \mathcal{L}_{\text{ms}}(\hat{z}_t^k)}{\partial \hat{z}_t^k}}$$



Point tracking

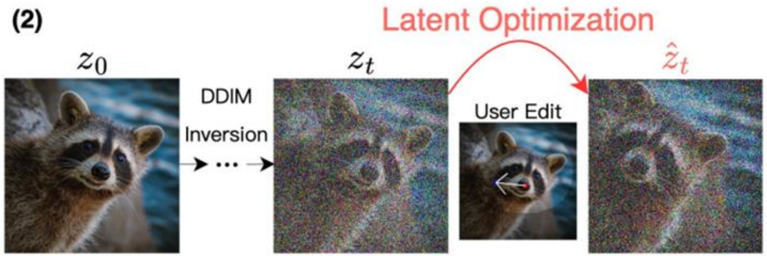
$$h_i^{k+1} = \underset{q \in \Omega(h_i^k, r_2)}{\arg\min} \left\| F_q(\hat{z}_t^{k+1}) - F_{h_i^0}(z_t) \right\|_1$$

DragDiffusion: Motion tracking

Motion supervision

$$\mathcal{L}_{\text{ms}}(\hat{z}_t^k) = \sum_{i=1}^{n} \sum_{q \in \Omega(h_i^k, r_1)} \left\| F_{q+d_i}(\hat{z}_t^k) - \text{sg}(F_q(\hat{z}_t^k)) \right\|_1 + \lambda \left\| (\hat{z}_{t-1}^k - \text{sg}(\hat{z}_{t-1}^0)) \odot (\mathbb{1} - M) \right\|_1$$

$$\hat{z}_t^{k+1} = \hat{z}_t^k - \eta \cdot \frac{\partial \mathcal{L}_{\text{ms}}(\hat{z}_t^k)}{\partial \hat{z}_t^k}$$

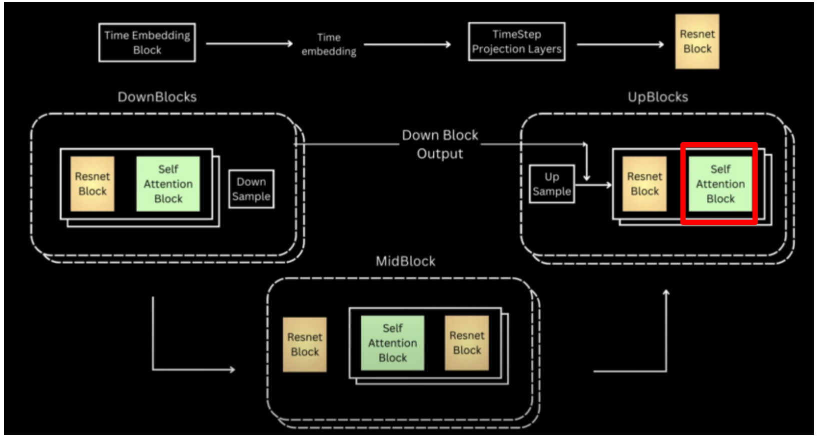


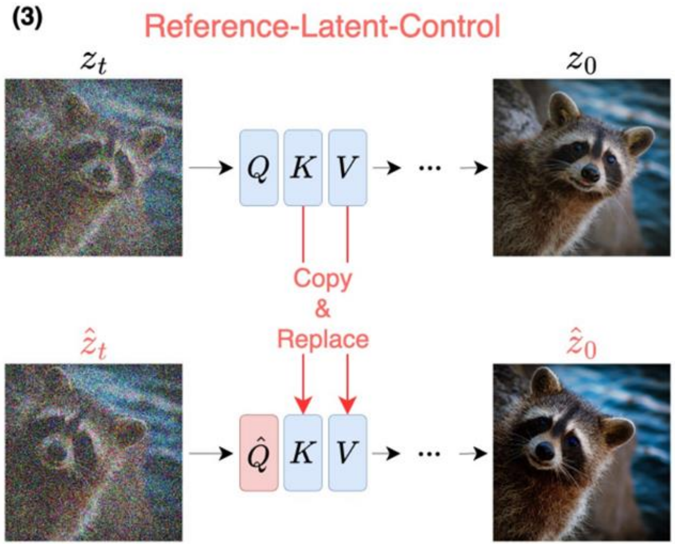
Point tracking

$$h_i^{k+1} = \boxed{\arg\min_{q \in \Omega(h_i^k, r_2)} \left\| F_q(\hat{z}_t^{k+1}) - F_{h_i^0}(z_t) \right\|_1}$$

DragDiffusion: Reference-latent control

Diffusion UNet





$$Attention(\hat{Q},\hat{K},\hat{V}) = softmax\left(\frac{\hat{Q}\hat{K}^T}{\sqrt{d_k}}\right)\hat{V} \quad \rightarrow \quad Attention(\hat{Q},K,V) = softmax\left(\frac{\hat{Q}K^T}{\sqrt{d_k}}\right)V$$

DragGAN과의 비교

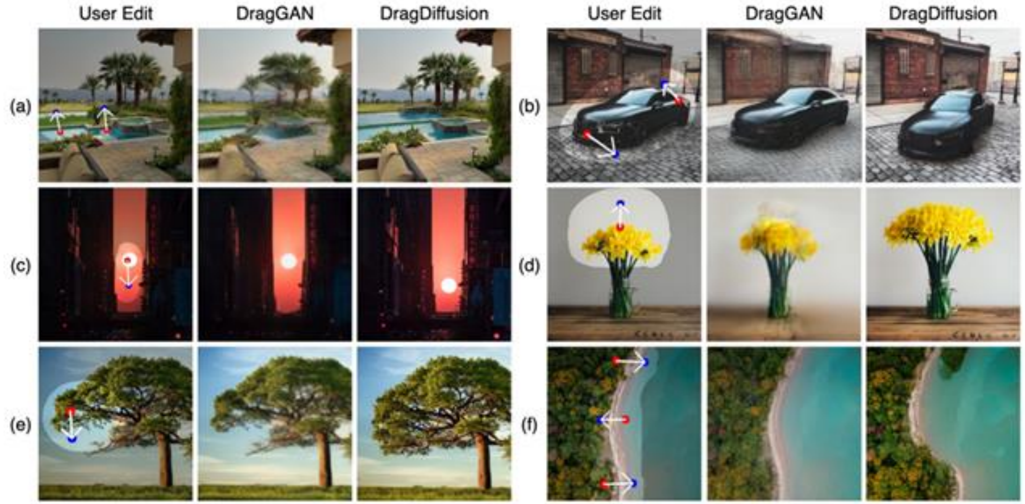


Figure 4. Comparisons between DRAGGAN and DRAGDIFFUSION. All results are obtained under the same user edit for fair comparisons.

Ablation Study

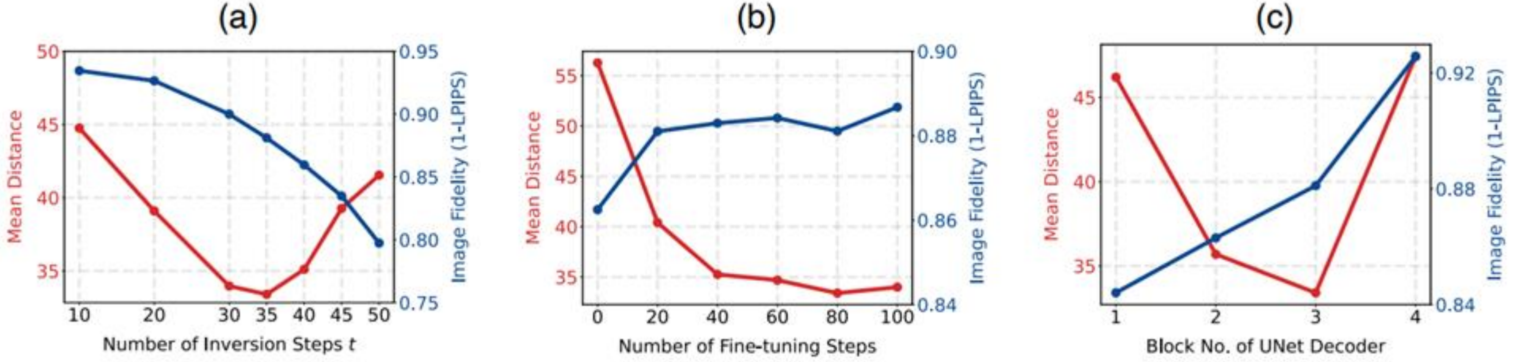


Figure 7. Ablation study on **(a)** the number of inversion step t of the diffusion latent; **(b)** the number of identity-preserving fine-tuning steps; **(c)** Block No. of UNet feature maps. Mean Distance (↓) and Image Fidelity (↑) are reported. Results are produced on DRAGBENCH.

Figure 6. Ablating the number of inversion step t. Effective results are obtained when $t \in [30, 40]$.

Ablation Study

Figure 9. Qualitative validation on effectiveness of identity-preserving fine-tuning and reference-latent-control.

감사합니다