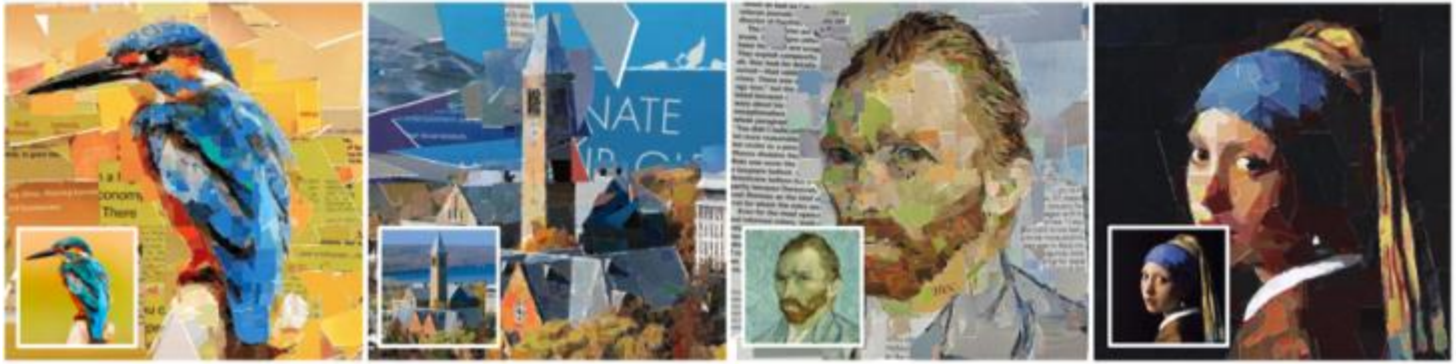
Neural Collage Transfer : Artistic Reconstruction via Material Manipulation

Ganghun Lee, Minji Kim, Yunsu Lee, Minsu Lee*, Byoung-Tak Zhang*
Seoul National University

CVPR 2023

TABLE OF CONTENTS

INTRODUCTION



Given an image and materials, each output collage was generated based on the proposed complexity-aware multi-scale collage method.

RELATED WORK

Neural Style Transfer (NST)

1. NST has been a prominent technique in the field of artistic style transfer.
2. The goal of NST is to transform a target image while preserving the content of the target style.
3. Conventional NST methods employ pixel-wise gradient descent or trained models to approximate the distribution of the target style.
4. Advanced NST models, despite covering various styles, are limited in their applicability to collage styles.
5. Pixel-wise style extraction in NST primarily focuses on common patterns for various styles, while collage styles require a different approach.

RELATED WORK

Stroke-based Rendering (SBR)

1. SBR is an automated method using discrete elements like strokes to generate non-photorealistic images.
2. Training involves human sketch demonstrations, facing challenges due to data collection.
3. Training the painting agent without supervision using RL, overcoming challenges of supervised methods.
4. Utilizing fully-differentiable painting designs for optimization.
5. Primarily concentrated on environments where stroke structures are pre-modeled, such as sketches and paintings.

RELATED WORK

Collage Generation

1. Previous research has explored artistic collage generation but hasn't specifically focused on collage transfer.
2. CLIP-CLOP generates collage artworks from text prompts using predefined strokes with modifiable properties.
3. The approach in this paper uses non-predefined materials for collage, generating images.

METHOD

Collage MDP

1. Preliminary
2. State and Transition
3. Action Design
4. Differentiable Collage
5. Reward Function

$$\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$$

Value function

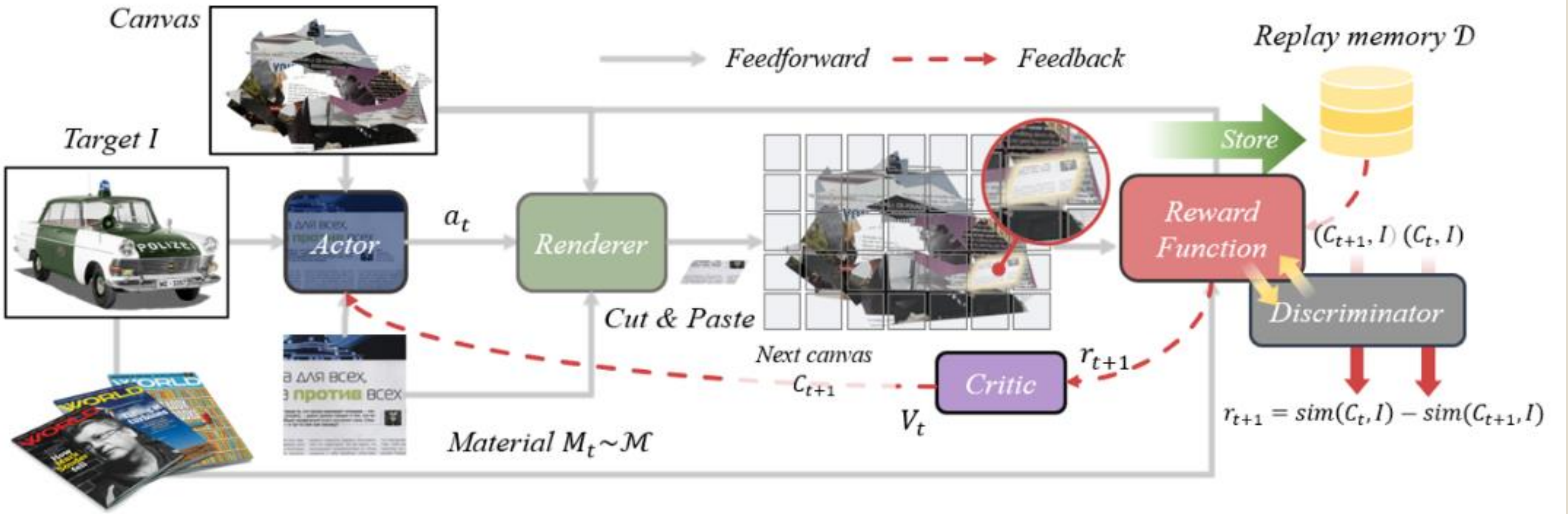
$$V_\pi(s_t) = r_{t+1} + \gamma r_{t+2} + \cdots + \gamma^{T-t-1} r_T$$

$$\pi^* = \underset{\pi}{\operatorname{argmax}}\, V_\pi(s)$$

METHOD

$$s_{t+1} = \mathcal{P}(s_t, a_t) = (\delta(C_t, M_t, a_t), I, M_{t+1}, (T_M - t_M)/T_M, c)$$

Collage MDP



Remaining time $l_t = (T_M - t_M)/T_M$

METHOD

Collage MDP

1. Preliminary
2. State and Transition
3. Action Design
4. Differentiable Collage
5. Reward Function

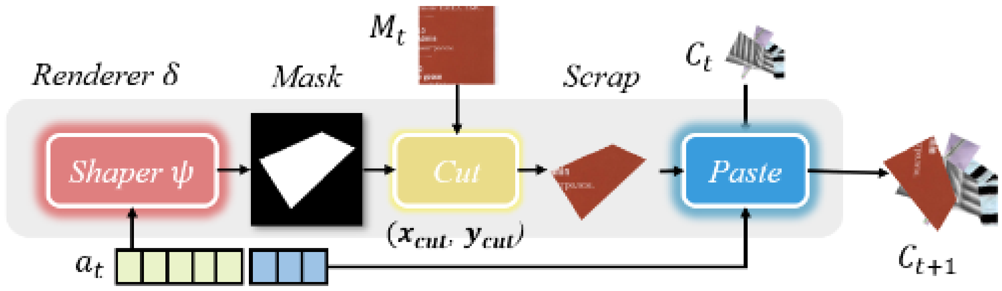
$$a = \langle x_{cut}, y_{cut}, w, h, p_1, p_2, p_3, p_4, x_{glue}, y_{glue}, \theta, v \rangle$$

Material acceptor $v \in [0, 1]$
If $v < 0.5$ => agent can deny a poor given material and request another one

METHOD

Collage MDP

1. Preliminary
2. State and Transition
3. Action Design
4. Differentiable Collage
5. Reward Function



The differentiable rendering process in collage MDP. The actions determining the cutting shapes are input into the pre-trained shaper network ψ. The resulting mask is then used to cut the material, generating a scrap to be pasted onto the canvas.

METHOD

Collage MDP

1. Preliminary
2. State and Transition
3. Action Design
4. Differentiable Collage
5. Reward Function

$$r_t = sim(C_{t-1}, I) - sim(C_t, I)$$

METHOD

Training

1. Model-based SAC
2. Training Scheme

Value function of traditional RL and SAC

$$V_{\pi}^{RL}(s_t) = \mathbb{E}_{a_t \sim \pi}\left[Q(s_t, a_t)\right],$$

$$V_{\pi}^{SAC}(s_t) = \mathbb{E}_{a_t \sim \pi}\left[Q(s_t, a_t) - \log \pi(a_t|s_t)\right].$$

METHOD

Advanced Techniques

1. Active Material Selection
2. Multi-Scale Collage
3. Complexity-Aware Multi-Scale Collage

$$Q(s_t, a_t) = r(s_t, a_t) + \mathbb{E}_{s_{t+1} \sim \mathcal{P}} \left[V(s_{t+1}) \right]$$

$$\begin{aligned} m_t^* = \underset{m}{\operatorname{argmax}} \left(r(s_t, a_t) + \gamma V(s_{t+1}) \right), \; m \in \mathcal{M}, \\ a_t = \mathbb{E}_{s_t \sim \mathcal{P}} \left[\pi(s_t) \right], \; s_{t+1} = \mathcal{P}(s_t, a_t). \end{aligned}$$

METHOD

Advanced Techniques

1. Active Material Selection
2. Multi-Scale Collage
3. Complexity-Aware Multi-Scale Collage

$$\mathcal{U} = (u_1, u_2, \ldots, u_n), \text{ where } u_1 > u_2 > \cdots > u_n \text{ and } u, n \in \mathbb{N}$$

$$k(u) = (\lceil (W - u)/\rho \rceil + 1)(\lceil (H - u)/\rho \rceil + 1)$$

The sequence of grapes shows our collage generation process.

METHOD

Advanced Techniques

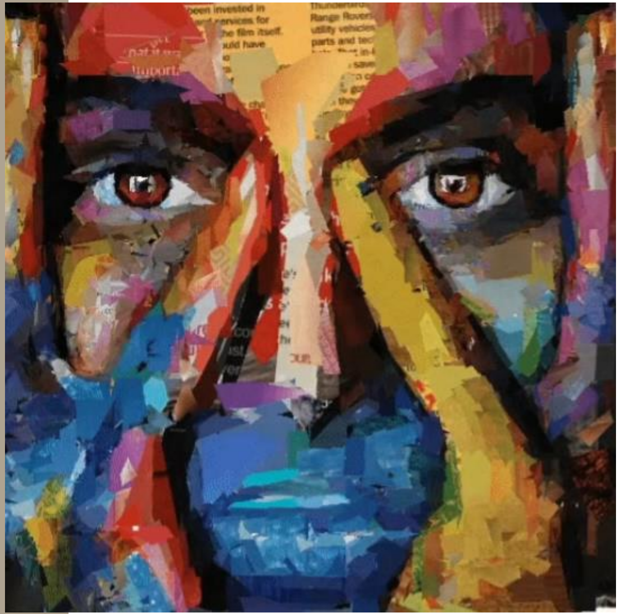
1. Active Material Selection
2. Multi-Scale Collage
3. Complexity-Aware Multi-Scale Collage

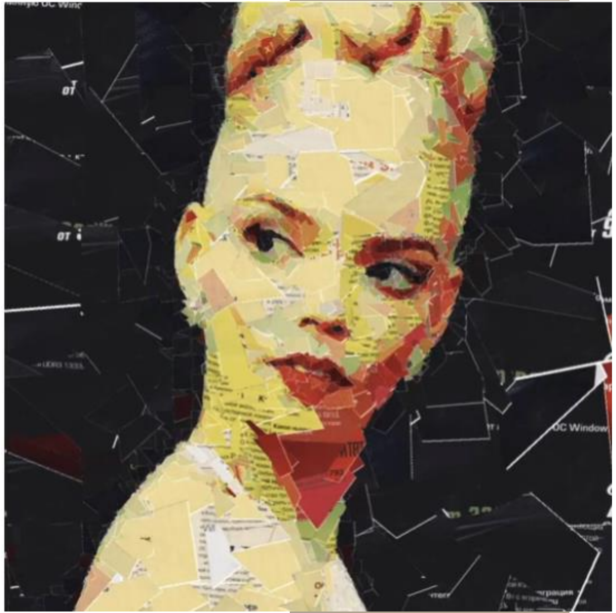
$$\mathcal{U} = (u_1, u_2, \ldots, u_n), \text{ where } u_1 > u_2 > \cdots > u_n \text{ and } u, n \in \mathbb{N}$$

$$k(u) = (\lceil (W - u)/\rho \rceil + 1)(\lceil (H - u)/\rho \rceil + 1)$$

The sequence of grapes shows our collage generation process.

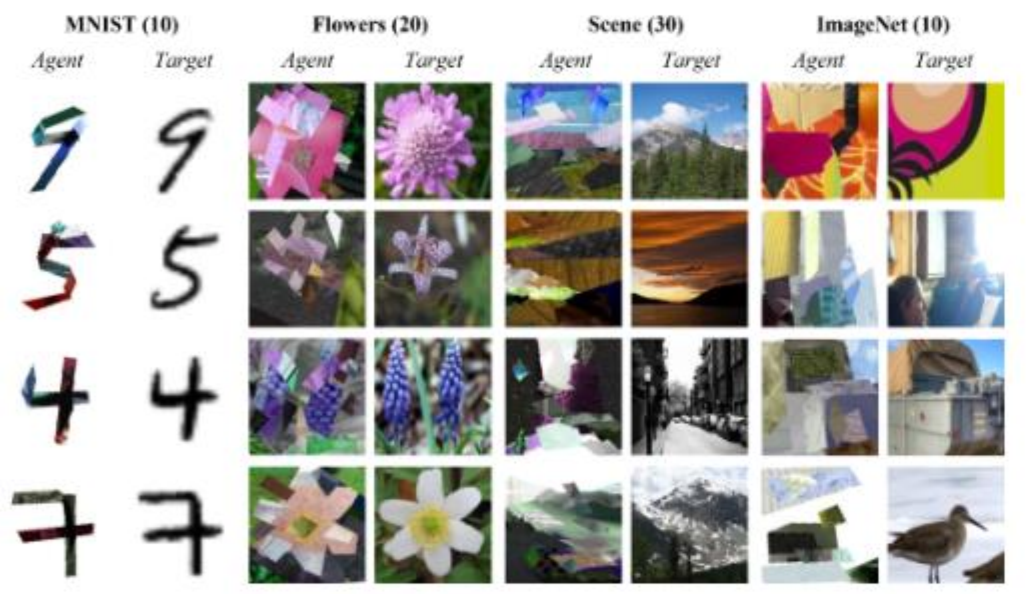
RESULT





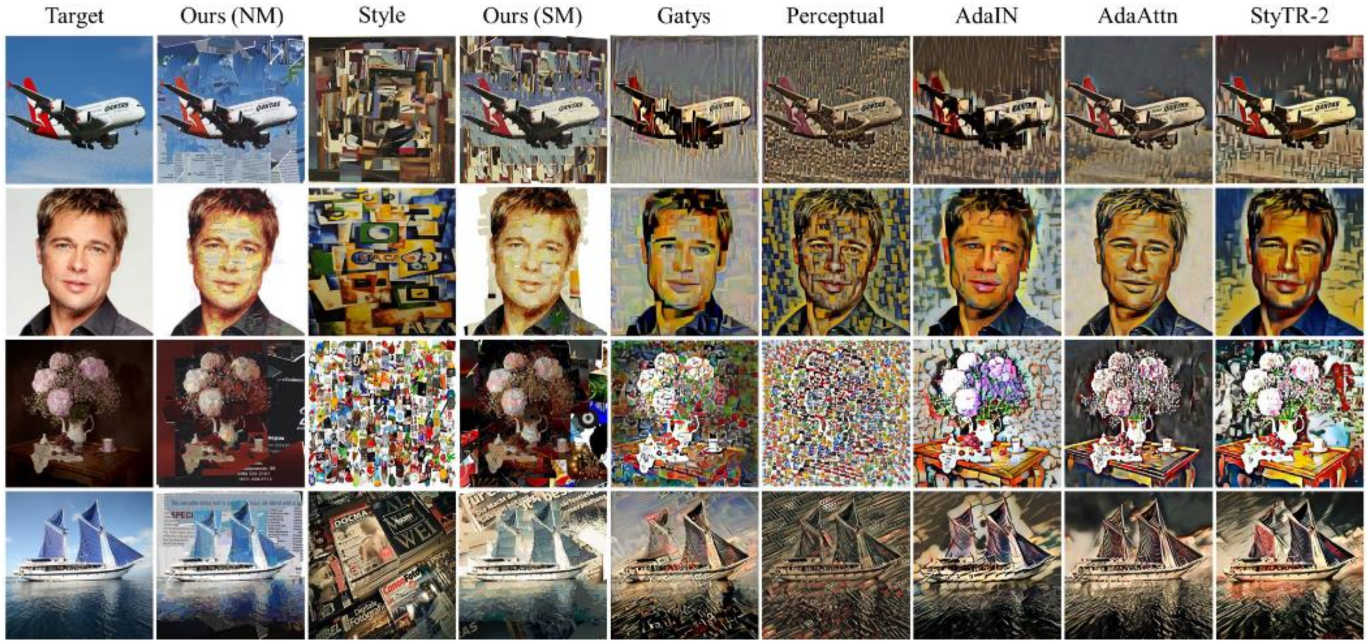
RESULT

Comparison with Single-Scale Collage



RESULT

Comparison with NST



RESULT

Comparison with NST

Methods	*CLIP score* ↑ [42]			*CLIP vote* ↑	*LPIPS* [55] ↓
	content	*human*	*collage*	*collage*	VGG
Target	0.276 ± 0.027	0.213 ± 0.018	0.200 ± 0.017	0.633	-
AdaAttn [34]	0.278 ± 0.021	0.247 ± 0.018	0.241 ± 0.010	0.027	0.597 ± 0.103
Adain [16]	0.251 ± 0.019	0.239 ± 0.010	0.236 ± 0.008	0.017	0.662 ± 0.103
Gatys [8]	0.226 ± 0.013	0.260 ± 0.006	0.250 ± 0.006	0.290	0.708 ± 0.098
Perceptual [22]	0.239 ± 0.019	0.246 ± 0.006	0.234 ± 0.007	0.307	0.722 ± 0.117
StyTR-2 [4]	0.261 ± 0.023	0.238 ± 0.010	0.235 ± 0.009	0.027	0.613 ± 0.115
Ours (32)	**0.280 ± 0.026**	0.262 ± 0.017	0.281 ± 0.020	0.100	**0.510 ± 0.111**
Ours (64)	0.262 ± 0.028	0.272 ± 0.020	0.259 ± 0.015	0.667	0.565 ± 0.112
Ours (128)	0.225 ± 0.023	**0.288 ± 0.015**	**0.272 ± 0.016**	**1.000**	0.610 ± 0.115

CONCLUSION AND FUTURE WORK

1. Novel RL-based training architecture (MB-SAC) for stroke-based collage transfer.
2. Complexity-aware multi-scale techniques enhance the agent's ability to handle different target image sizes.
3. Autonomous learning, producing aesthetically pleasing collages without demonstration data.
4. Limitations include the constraint to quadrilateral stroke shapes, suggesting potential future extensions to more unconstrained shapes.
5. Custom reward factors reflecting intentional distortions or style variations could be added for further improvement.