Q-Align: Alleviating Attention Leakage in Zero-Shot Appearance Transfer via Query-Query Alignment

Namu Kim[1], Wonbin Kweon[2], Minsoo Kim[3], Hwanjo Yu[3]

[1]KT Corporation, South Korea

[2]University of Illinois Urbana-Champaign, USA

[3]Pohang University of Science and Technology (POSTECH), South Korea

1. Task – Zero-Shot Appearance Transfer

Appearance Transfer

- Transfer the *appearance* (color, texture, style) of one image onto the *structure* (shape, layout) of another.

Goal

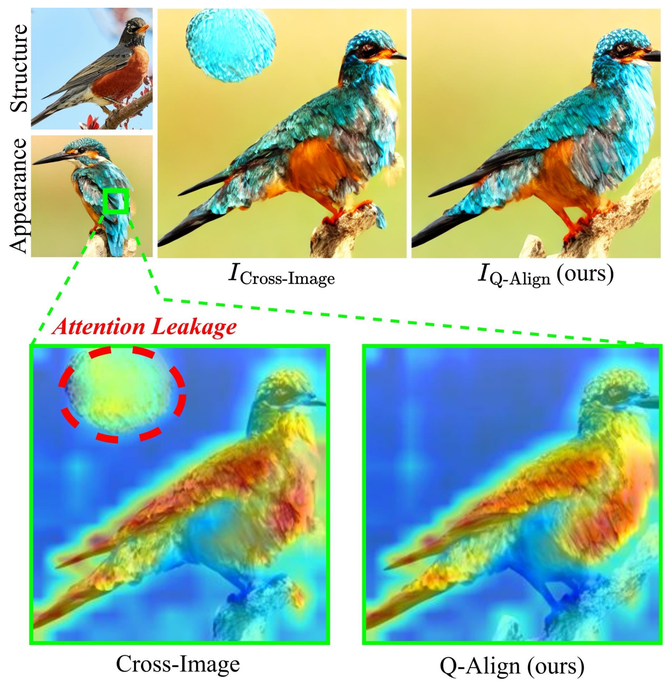
- I_out that preserves the structure of I_str while adopting the appearance of I_app.

2. Problem – Attention Leakage

Attention Leakage

- The unintended spread of attention to semantically irrelevant regions, causing structural distortion and appearance inconsistency.

- Attention leakage occurs when the semantic focus spills to irrelevant regions, misaligning structure and appearance.



3. Method – Q-Align replaces query–key alignment with query–query alignment.

Q-Align Steps

Step1. Compute Query-Query alignment

$$S = Q_{\text{app}} Q_{\text{str}}^{\top}$$

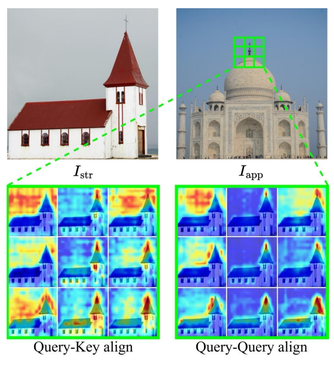
Step2. Construct the rearrangement matrix

$$(P^{\top})_{r,c} = \begin{cases} \frac{1}{k}, & \text{if } S_{r,c} \in \text{Top-}k(S_r) \\ 0, & \text{otherwise} \end{cases} \qquad P_{r,r} = 1, \text{ If } \sum_{c=1}^{n} P_{r,c} = 0$$

$$P'_{r,c} = \begin{cases} \frac{e^{P_{r,c}}}{\sum_{c' \in [n], P_{r,c'} \neq 0} e^{P_{r,c'}}}, & \text{if } P_{r,c} \neq 0 \\ 0, & \text{if } P_{r,c} = 0 \end{cases}$$

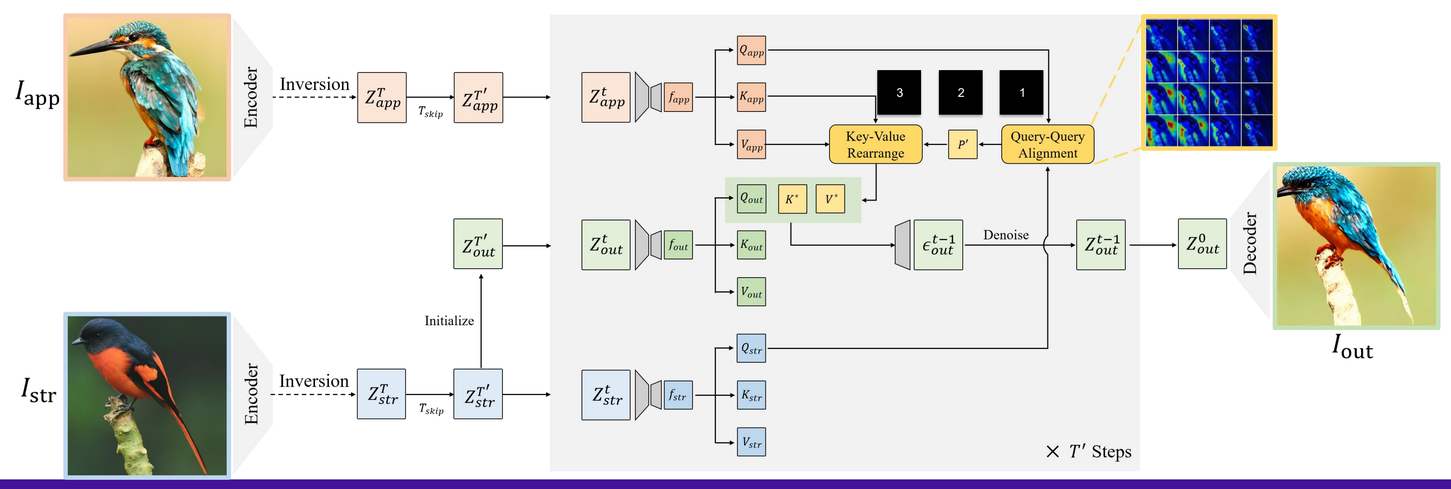
Step3. Apply rearranged Keys and Values

$$K^* = P' K_{\text{app}},$$
$$V^* = P' V_{\text{app}}.$$



- Query-Query alignment provides more sophisticated spatial attention, while Query-Key distributes attention broadly across related regions

- Query-Key align : $softmax(Q_{str} K_{app}^{T})$
- Query-Query align : $softmax(Q_{str} Q_{app}^{T})$



4. Results

Visually Coherent Results

- Q-Align effectively alleviates attention leakage, generating images with more faithful appearance transfer and better structural consistency.

- Aligning features in the query space leads to more stable and semantically consistent attention behavior.