Interactive Object Placement with Reinforcement Learning

Shengping Zhang [1] **Quanling Meng** [1] **Qinglin Liu** [1] **Liqiang Nie** [2]
Bineng Zhong [3] **Xiaopeng Fan** [4] **Rongrong Ji** [5]

Abstract

Object placement aims to insert a foreground object into a background image with a suitable location and size to create a natural composition. To predict a diverse distribution of placements, existing methods usually establish a one-to-one mapping from random vectors to the placements. However, these random vectors are not interpretable, which prevents users from interacting with the object placement process. To address this problem, we propose an Interactive Object Placement method with Reinforcement Learning, dubbed IOPRE, to make sequential decisions for producing a reasonable placement given an initial location and size of the foreground. We first design a novel action space to flexibly and stably adjust the location and size of the foreground while preserving its aspect ratio. Then, we propose a multi-factor state representation learning method, which integrates composition image features and sinusoidal positional embeddings of the foreground to make decisions for selecting actions. Finally, we design a hybrid reward function that combines placement assessment and the number of steps to ensure that the agent learns to place objects in the most visually pleasing and semantically appropriate location. Experimental results on the OPA dataset demonstrate that the proposed method achieves state-of-the-art performance in terms of plausibility and diversity.

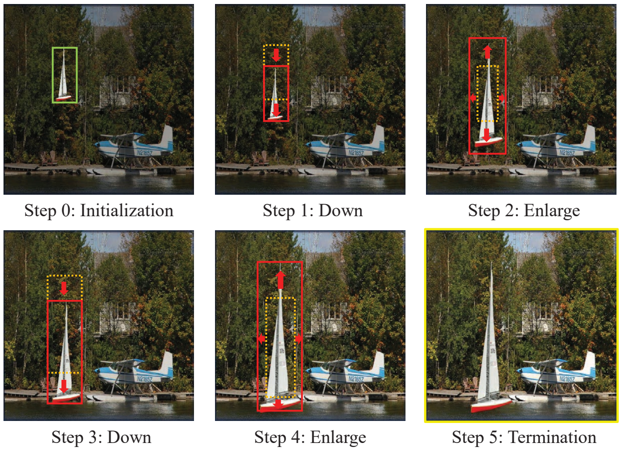

Figure 1. Illustration of the decision-making process. We first give an initial location and size of the foreground object (depicted as a boat in a green rectangle). Then, the agent continuously adjusts the location and size (indicated by red arrows) of the boat until the termination action is selected. Red (*resp.*, Orange) rectangles indicate the current (*resp.*, last) location and size of the boat.

1. Introduction

Image composition (Niu et al., 2021) aims to paste a foreground object from one image on another background image to produce a realistic composite image, which has broad application prospects in the realm of art, entertainment, commerce (Chen & Kae, 2019; Weng et al., 2020; Zhang et al., 2020b), and data augmentation (Dwibedi et al., 2017; Remez et al., 2018; Ouyang et al., 2018). Object placement is a crucial aspect of image composition that focuses on estimating the location and size of the foreground object to make the composition image natural. Learning object placement is a very challenging task, as it requires taking into account various factors such as location, size, and occlusion within the background image (Niu et al., 2021).

Early methods (Georgakis et al., 2017; Remez et al., 2018; Wang et al., 2019; Fang et al., 2019; Zhang et al., 2020b) attempt to find reasonable locations and scales for foreground objects by designing explicit rules. However, these methods struggle to cope with diverse and complex scenarios.

[1] School of Computer Science and Technology, Harbin Institute of Technology, Weihai, China [2] School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China [3] Guangxi Key Laboratory of Multi-Source Information Mining and Security, Guangxi Normal University, Guilin, China [4] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China [5] Department of Artificial Intelligence, School of Informatics, Xiamen University, Xiamen, China. Correspondence to: Liqiang Nie <nieliqiang@gmail.com>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

Recent work (Tripathi et al., 2019; Zhang et al., 2020a; Zhou et al., 2022) employs deep learning-based methods, specifically Generative Adversarial Networks (GANs), to generate placements given a pair of foreground and background. TERSE (Tripathi et al., 2019) produces a single placement, which ignores the diversity of object placement. PlaceNet (Zhang et al., 2020a) and GracoNet (Zhang et al., 2020a) combine foreground and background features with sampled random vectors to predict a diverse distribution of placements, which establishes a one-to-one mapping from random vectors to the placements. However, these random vectors are not interpretable, which prevents users from interacting with the object placement process. In some scenarios, such as artistic creation and automatic advertising, users often have specific expectations for the location and size of the foreground in the background image. Unfortunately, the above two methods usually first produce a large number of reasonable candidate placements and then use a post-processing procedure to search for placements based on user expectations, which is inefficient and impractical.

In this paper, we propose an Interactive Object Placement method with Reinforcement Learning, dubbed IOPRE, to make sequential decisions for producing a reasonable placement given an initial location and size of the foreground. Our method mimics the manual compositing process, where users typically first define an initial location and size of the foreground object and then make small adjustments until they are satisfied with the composite result. Specifically, the proposed method takes the user-annotated initial location and size of the foreground object as auxiliary inputs, which are continuously adjusted during the decision-making process to ensure that the final composite image is both realistic and meets the specific expectations of users. In our method, we first design a novel action space to flexibly and stably adjust the location and size of the foreground while preserving its aspect ratio. Then, we propose a multi-factor state representation learning method, which integrates composition image features and sinusoidal positional embeddings of the foreground to make decisions for selecting actions. Finally, we design a hybrid reward function that combines placement assessment and the number of steps to ensure that the agent learns to place objects in the most visually pleasing and semantically appropriate locations. As shown in Figure 1, the agent adjusts the location and size of the foreground object until a reasonable object placement is obtained. Our main contributions are as follows:

- We propose an interactive object placement method with reinforcement learning to make sequential decisions for producing a reasonable placement. To our best knowledge, we are the first to take into account the user-annotated initial location and size of the foreground object, making it highly interactive and efficient for image composition.
- We design an aspect ratio preserved action space, a multi-factor state representation learning method, and a hybrid reward function to make the agent learn to place objects in the most visually pleasing and semantically appropriate location.
- Experimental results on the OPA dataset demonstrate that our method achieves interactive object placements. The proposed method produces plausible and diverse composite images and achieves state-of-the-art performances in terms of various metrics.

2. Related Work

2.1. Object Placement

Traditional object placement methods (Georgakis et al., 2017; Remez et al., 2018; Wang et al., 2019; Fang et al., 2019; Zhang et al., 2020b) design explicit rules to find reasonable placements for foreground objects. However, these explicit rules have limited application scenarios.

Deep learning based object placement methods employ neural networks to predict reasonable placements, which can be divided into category-specific and instance-specific methods (Niu et al., 2021). The category-specific methods (Tan et al., 2018; Lee et al., 2018) aim to predict plausible bounding boxes given a background image and a foreground category. Some methods (Dvornik et al., 2018; 2019; Volokitin et al., 2020) predict whether a bounding box is suitable for certain foreground categories by modeling context. The category-specific methods assume that any instance belonging to the same foreground category can be placed in the predicted bounding box, which is not reasonable.

The instance-specific methods (Lin et al., 2018; Tripathi et al., 2019; Zhan et al., 2019; Kikuchi et al., 2019; Azadi et al., 2020; Zhang et al., 2020a; Zhou et al., 2022) aim to predict plausible spatial transformations given pairs of foreground objects and background images. Tripathi et al. (2019) propose a framework consisting of a generator, discriminator, and target network to produce composite images. Zhang et al. (2020a) propose a PlaceNet that samples random variables to predict multiple reasonable placements. Zhou et al. (2022) treat object placement as a graph completion problem and design a dual-path framework to generate plausible object placements. Liu et al. (2021a) and Niu et al. (2022) focus on the object placement assessment (OPA) task, which aims to verify whether a composite image is plausible in terms of object placement.

2.2. Reinforcement Learning

Reinforcement learning aims to train an agent to learn the optimal policy by interacting with an environment, which

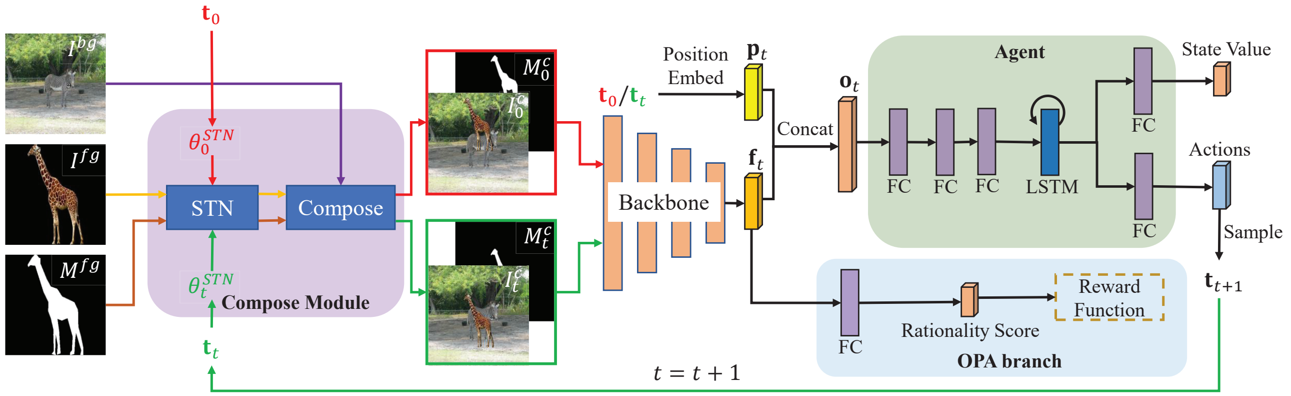

Figure 2. Architecture of the proposed IOPRE method. Given a background image I^{bg}, foreground image I^{fg}, foreground object mask M^{fg}, and initial transformation vector $\mathbf{t}_0$, the composite module produces a composite image I_0^c and composite mask M_0^c. At each time step t, the backbone processes I_t^c and M_t^c to generate a feature vector $\mathbf{f}_t$, and $\mathbf{t}_t$ is mapped to the positional embedding $\mathbf{p}_t$. The observation $\mathbf{o}_t$ is the concatenation of $\mathbf{f}_t$ and $\mathbf{p}_t$. $\mathbf{f}_t$ is fed into the object placement assessment (OPA) branch to output a rationality score for computing the reward. The agent processes $\mathbf{o}_t$ to generate the value and action output. The actor output is used to sample an action from the action space to generate $\mathbf{t}_{t+1}$. The process ends when the terminal state is reached.

has been introduced in many computer vision tasks, including object detection (Jie et al., 2016; Pirinen & Sminchisescu, 2018), image cropping (Li et al., 2018), and person re-identification (Wu et al., 2022). Jie et al. (2016) propose a Tree-structured Reinforcement Learning approach to incorporate global interdependency between objects into object detection. Compared with typical region proposal networks, drl-RPN (Pirinen & Sminchisescu, 2018) optimizes an objective closer to the final detection task by using reinforcement learning. Li et al. (2018) propose a weakly supervised aesthetics aware reinforcement learning framework for automatic image cropping. Wu et al. (2022) propose a novel Temporal Complementarity-Guided Reinforcement Learning approach for image-to-video person re-identification.

In this paper, we formulate object placement as a sequential decision-making process and utilize reinforcement learning to achieve interactive object placement.

3. Method

3.1. Problem Formulation

Given a background image I^{bg}, foreground image I^{fg}, and foreground object mask M^{fg}, the general object placement model $\mathcal{F}$ aims to produce a composite image I^c and composite mask M^c, which is formulated as

$$< I^c, M^c >= \mathcal{F}(I^{bg}, I^{fg}, M^{fg}; W) \tag{1}$$

where W represents the parameters of $\mathcal{F}$. To realize interactive object placement, an initial transformation vector $\mathbf{t}_0$ is required as an additional input to generate an object placement, which is formulated as

$$< I^c, M^c >= \mathcal{F}(I^{bg}, I^{fg}, M^{fg}, \mathbf{t}_0; W) \tag{2}$$

where $\mathbf{t}_0$ indicates the initial location and size of I^{fg} in I^{bg} and the detailed definition of the transformation vector is introduced in Section 3.2.

Inspired by the manual compositing image process, we formulate object placement as a sequential decision-making process and propose an Interactive Object Placement method with Reinforcement Learning, dubbed IOPRE, as shown in Figure 2. Given the initial transformation vector $\mathbf{t}_0$, an agent interacts with an environment $\mathcal{E}$ and takes actions step by step to adjust the location and size of the foreground to obtain a reasonable placement. At the beginning of this process, the composite module processes I^{bg}, I^{fg}, M^{fg}, and $\mathbf{t}_0$ to produce a composite image I_0^c and composite mask M_0^c. At each time step t, the agent receives an observation $\mathbf{o}_t$ according to the composite image I_t^c and transformation vector $\mathbf{t}_t$, and combines it with historical observations $\{\mathbf{o}_0, \mathbf{o}_1, ..., \mathbf{o}_{t-1}\}$ to form the current state s_t. Then, the agent selects an action a_t from the action space $\mathcal{A}$ according to the policy π and executes a_t to get a new transformation vector $\mathbf{t}_{t+1}$ for producing a new composite image I_{t+1}^c and composite mask M_{t+1}^c. After the chosen action a_t is executed, the agent obtains a reward r_t according to the rationality scores of I_t^c and I_{t+1}^c, and receives a new state s_{t+1}. The agent and environment interact until the terminal state is reached. During this process, the goal of the agent is to find a reasonable object placement by maximizing the expectation of the long-term accumulated reward $R_t = \sum_{i=0}^{\infty} \gamma^i r_{t+i}$ for each time step t, where γ

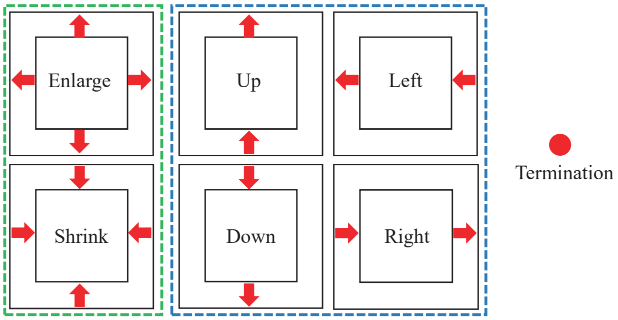

Figure 3. Illustration of the action space $\mathcal{A}$ in IOPRE. There are 7 pre-defined actions in $\mathcal{A}$, which can be divided into three groups: scaling actions, location translation actions, and a termination action. The arrow indicates the moving direction of the foreground image I^{fg}.

is the discount factor. In our method, we design an aspect ratio preserved action space, a multi-factor state representation learning method, and a hybrid reward function to make the agent learn to place objects in background images with suitable locations and sizes.

3.2. Aspect Ratio Preserved Action Space

To ensure the realism of the composite image, the aspect ratio of the foreground needs to be constant during the compositing process. Following (Zhou et al., 2022), we use a transformation vector $\mathbf{t} = [t^r, t^x, t^y]$ to indicate the location and size of the foreground in the background. Specifically, we use $t^r \in (0, 1)$ to represent the scaling ratio of the whole foreground and W (*resp.*, H) to denote the width (*resp.*, height) of the foreground image and background image. Consequently, the width w and height h of the scaled foreground become $t^r W$ and $t^r H$, respectively. Suppose (x, y) indicates the background coordinate for the left top pixel point of the scaled foreground, we then use $t_x = \frac{x}{W-w} \in (0, 1)$ and $t_y = \frac{y}{H-h} \in (0, 1)$ to indicate the relative horizontal and vertical location that the foreground should be placed over the background image.

In our method, we design a set of 7 pre-defined actions to construct the action space $\mathcal{A}$, as illustrated in Figure 3. This action space differs from previous methods (Jie et al., 2016; Li et al., 2018) and is divided into three groups: scaling actions, location translation actions, and a termination action. The first two groups aim to adjust the size and location of the foreground, including 2 and 4 actions, respectively. The agent stops the decision-making process and outputs the current object placement as the final result when the termination action is executed. In addition, we limit the variation range of each element in $\mathbf{t}$ to $(0, 1)$ during the placement process. The proposed action space has two key advantages. Firstly, it preserves the aspect ratio of the foreground while still providing flexibility in adjusting its size and location. Secondly, it allows for the reliable placement of the entire foreground object within the background, preventing any parts of the object from being removed. In IOPRE, the first two groups of actions adjust the size and location by 0.05 at each time step.

3.3. Position-aware State Space

At each time step t, the state s_t in our method consists of current and historical observations, which mimics the human decision-making process. Specifically, s_t can be represented as $s_t = \{\mathbf{o}_0, \mathbf{o}_1, ..., \mathbf{o}_{t-1}, \mathbf{o}_t\}$, where $\mathbf{o}_t$ denotes the current observation of the agent. To provide multi-factor information for the agent, we combine the feature of I^c_t and the positional embedding of $\mathbf{t}_t$ as $\mathbf{o}_t$. Since the goal of the agent is to find a reasonable object placement, the current composite image I^c_t is essential for it to make a decision. Given I^{bg}, I^{fg}, M^{fg}, and $\mathbf{t}_t$, the composite module produces I^c_t and M^c_t. Our backbone takes the concatenation of I^c_t and M^c_t as input and outputs a feature vector $\mathbf{f}_t$. To enable the agent to be aware of the size and location of the foreground object, we map $\mathbf{t} = [t^r, t^x, t^y]$ to a 384-dimensional sinusoidal positional embedding (Vaswani et al., 2017) as the positional embedding $\mathbf{p}$ by

$$\mathbf{p} = \mathrm{Cat}(\mathrm{PE}(t^r), \mathrm{PE}(t^x), \mathrm{PE}(t^y)) \tag{3}$$

where $\mathrm{Cat}(\cdot, \cdot, \cdot)$ denotes the concatenation function and $\mathrm{PE}(\cdot)$ denotes the positional encoding function that maps a float to a vector. In IOPRE, the current observation $\mathbf{o}_t$ is the concatenation of the feature vector $\mathbf{f}_t$ and the positional embedding $\mathbf{p}_t$. Since historical experience is usually valuable for future decision-making, we employ an LSTM unit to memorize historical observations $\{\mathbf{o}_0, \mathbf{o}_1, ..., \mathbf{o}_{t-1}\}$ and combine them with the current observation $\mathbf{o}_t$ to form the current state s_t.

3.4. Object Placement Assessment Reward

The realism of the current composite image is crucial for the agent to make a decision, so we use its rationality score to design a reward function. Inspired by (Liu et al., 2021a), we design an object placement assessment network to evaluate the rationality degree of generated composite images. Specifically, this network takes the concatenation of I^c and M^c as input and predicts a rationality score $s_a(I^c, M^c)$, where the variation range of $s_a(I^c, M^c)$ is $(0, 1)$. $s_a(I^c, M^c)$ is compared to a threshold of 0.5 to produce a binary rationality label. In our method, we discard the binary rationality label and retain $s_a(I^c, M^c)$ to compare the rationality between I^c_t and I^c_{t+1}. The agent receives a positive reward if the rationality score of the new composite image is higher than the previous one, otherwise, it receives a zero or negative reward. Inspired by (Li et al., 2018), we adopt an additional negative reward $-\lambda_s \times (t+1)$ at each time step to accelerate the object placement process, where λ_s is a hyper-parameter that controls the significance of this negative reward, $t+1$ represents the number of steps and t starts from 0. The agent takes an action a_t under the state s_t and then receives a reward $r_t(s_t, a_t)$ by

$$\begin{aligned} r_t(s_t, a_t) =& \operatorname{sign}(s_a(I^c_{t+1}, M^c_{t+1}) - s_a(I^c_t, M^c_t)) \\ & - \lambda_s \times (t+1) \end{aligned} \tag{4}$$

where $\operatorname{sign}(\cdot)$ denotes the sign function. To stabilize the training process of IOPRE, we limit the variation range of $r_t(s_t, a_t)$ to $[-1, 1]$.

3.5. Network Architecture

IOPRE consists of a composite module, backbone, actor-critic branch, and object placement assessment (OPA) branch, which is illustrated in Figure 2. The composite module first converts a transformation vector $\mathbf{t}$ to the affine transformation parameter θ^{STN}, which is defined as

$$\theta^{STN} = \begin{bmatrix} 1 \setminus t^r & 0 & (1-2t^x)(1 \setminus t^r - 1) \\ 0 & 1 \setminus t^r & (1-2t^y)(1 \setminus t^r - 1) \end{bmatrix} \tag{5}$$

Then, it employs the Spatial Transformer Network (STN) (Jaderberg et al., 2015) to produce a new foreground image $\tilde{I}^{fg}$ and a composite mask M^c. Finally, a composite image I^c is obtained by

$$I^c = \tilde{I}^{fg} \odot M^c + I^{bg} \odot (1 - M^c) \tag{6}$$

We adopt Swin Transformer (Liu et al., 2021b) as the backbone, which takes the concatenation of I^c_t and M^c_t as input and outputs a 768-dimensional feature vector $\mathbf{f}_t$ at each time step t. The transformation vector $\mathbf{t}_t$ is mapped to the positional embedding $\mathbf{p}_t$, and the current observation $\mathbf{o}_t$ is the concatenation of $\mathbf{f}_t$ and $\mathbf{p}_t$. IOPRE has two branches, the first one is the actor-critic branch, and the other is the OPA branch. The actor-critic branch consists of 5 fully-connected layers and an LSTM unit. $\mathbf{o}_t$ is fed into the first three fully-connected layers and the LSTM unit to produce the current state s_t. The last two fully connected layers are used to get an output respectively, the first one is the policy output (**Actor**), and the other output is the value output (**Critic**). The policy output is a probability distribution of 7 possible outcomes, which corresponds to 7 pre-defined actions of the action space $\mathcal{A}$. The value output is the expected accumulated reward for the agent starting from state s_t. The OPA branch consists of a single fully-connected layer and outputs rationality scores to compute rewards.

The policy output provides the agent with the probability of each action under the current state. For exploration during the training process, the agent samples an action from the multinomial distribution, with actions having higher probabilities being more likely to be selected. During the testing process, the agent chooses the action with the highest probability to find a reasonable placement according to the learned policy. Note that the OPA branch is only used in the training process.

3.6. Optimization

Inspired by (Li et al., 2018), we modify the asynchronous advantage actor-critic (A3C) algorithm (Mnih et al., 2016) to train the agent for learning the object placement policy. Specifically, we use the mini-batch to replace the asynchronous mechanism to add diversity to exploration during the training process. At each time step t, the accumulated reward R_t is obtained by

$$R_t = \sum_{i=0}^{k-1} \gamma^i r_{t+i} + \gamma^k V\left(s_{t+k}; \theta_v\right) \tag{7}$$

where $\gamma \in (0, 1]$ is the discount factor, r_t is the object placement assessment reward, $V\left(s_t; \theta_v\right)$ is the value output under state s_t, θ_v denotes the network parameters of the **Critic** branch and k ranges from 0 to t_{max}. t_{max} means the maximum number of steps the agent takes before updating. For the policy output, the optimization objective is to maximize both the advantage function (Mnih et al., 2016) $R_t - V\left(s_t; \theta_v\right)$ and the entropy of the policy output (Williams & Peng, 1991) $H\left(\pi\left(s_t; \theta\right)\right)$, where $\pi\left(s_t; \theta\right)$ is the probability distribution of policy output, θ denotes the network parameters of the **Actor** branch, and $H(\cdot)$ is the entropy function. The entropy in the optimization objective improves exploration by discouraging premature convergence to suboptimal deterministic policies. Further, θ can be updated by

$$\begin{aligned} d\theta \leftarrow & d\theta + \nabla_\theta \log \pi\left(a_t \mid s_t; \theta\right)\left(R_t - V\left(s_t; \theta_v\right)\right) \\ & + \beta \nabla_\theta H\left(\pi\left(s_t; \theta\right)\right) \end{aligned} \tag{8}$$

where $\pi\left(a_t \mid s_t; \theta\right)$ is the probability of the sampled action a_t under the state s_t and β is a hyper-parameter to control the importance of $H(\cdot)$. We optimize the value output by minimizing the distance between R_t and $V\left(s_t; \theta_v\right)$. So θ_v can be updated by

$$d\theta_v \leftarrow d\theta_v + \nabla_{\theta_v}\left(R_t - V\left(s_t; \theta_v\right)\right)^2 / 2 \tag{9}$$

The detailed procedure for training our IOPRE is shown in Algorithm 1.

4. Experiment

4.1. Experimental Setting

4.1.1. Dataset and Evaluation Metrics

We conduct all experiments on the Object Placement Assessment (OPA) dataset (Liu et al., 2021a), which contains

Algorithm 1 Training procedure of IOPRE

```
Input: background image $I^{bg}$, foreground image $I^{fg}$, object mask $M^{fg}$, initial transformation vector $\mathbf{t}_0$
Output: network parameters of Actor branch $\theta$ and network parameters of Critic branch $\theta_v$
$< I_0^c, M_0^c >= \text{Composite}(I^{bg}, I^{fg}, M^{fg}, \mathbf{t}_0)$
$t \leftarrow 0, d\theta \leftarrow 0, d\theta_v \leftarrow 0$
repeat
    $\mathbf{f}_t = \text{FeatureExtractor}(I_t^c, M_t^c)$
    $\mathbf{p}_t = \text{PositionalEmbedding}(\mathbf{t}_t)$
    $\mathbf{o}_t = \text{Concat}(\mathbf{f}_t, \mathbf{p}_t)$
    $< V(s_t; \theta_v), \pi(a_t \mid s_t; \theta) >= \text{Actor-Critic}(\mathbf{o}_t)$
    perform $a_t$ to get $\mathbf{t}_{t+1}$ according to $\pi(a_t \mid s_t; \theta)$
    $< I_{t+1}^c, M_{t+1}^c >= \text{Composite}(I^{bg}, I^{fg}, M^{fg}, \mathbf{t}_{t+1})$
    $r_t = \text{Reward}(I_t^c, M_t^c, I_{t+1}^c, M_{t+1}^c, t)$
    $t = t + 1$
until $t == t_{max}$ or $a_{t-1}$ is termination action;
$R = \begin{cases} 0 & \text{if } a_{t-1} \text{ is termination action} \\ V(s_t; \theta_v) & \text{for other actions} \end{cases}$
for $i = t - 1$ to $0$ do
    $R \leftarrow r_i + \lambda R$
    $d\theta \leftarrow d\theta + \nabla_\theta \log \pi(a_i \mid s_i; \theta)(R - V(s_i; \theta_v))$
    $+\beta \nabla_\theta H(\pi(s_i; \theta))$
    $d\theta_v \leftarrow d\theta_v + \nabla_{\theta_v}(R - V(s_i; \theta_v))^2 / 2$
end for
Update $\theta$ with $d\theta$ and $\theta_v$ with $d\theta_v$
```

73,470 composite images with binary rationality labels. This dataset also provides foreground images, foreground object masks, and background images for image composition. Specifically, these composite images are divided into 62,074 training images and 11,396 test images. The training (*resp.*, test) set contains 2,701 (*resp.*, 1,436) unrepeated foreground objects and 1,236 (*resp.*, 153) unrepeated background images. Besides, the training (*resp.*, test) set consists of 21,376 (*resp.*, 3,588) positive samples and 40,698 (*resp.*, 7,808) negative samples. In this paper, we first train our object placement assessment network on the training set and evaluate it on the test set. Then, we train IOPRE on the positive samples of the training set and evaluate it on the positive samples of the test set.

Following (Zhou et al., 2022), we adopt user study, accuracy, and Fréchet Inception Distance (FID) metric (Heusel et al., 2017) to evaluate generation plausibility. Specifically, IOPRE generates a composite image for a pair of foreground and background given a random initial transformation vector. The user study is conducted to compare the composite images generated by various methods. For each pair of foreground and background, 20 voluntary participants are invited to choose the most reasonable composite image. Then, each method is scored by the proportion of participants who choose its composite images. We obtain the final score of each method by averaging its scores over all samples. Accuracy is the proportion of generated composite images that are classified as positive by a binary classifier SimOPA (Liu et al., 2021a) in terms of object placement. We compute FID between the generated composite images and the positive composite images in the test set to measure the similarity of two groups of images. Generally, lower FID indicates that generated composite images are more realistic and plausible. We use the Learned Perceptual Image Patch Similarity (LPIPS) metric (Zhang et al., 2018) to evaluate generation diversity. IOPRE generates 10 composite images for a pair of foreground and background given 10 random initial transformation vectors. LPIPS is used to evaluate the perceptual similarity between two composite images, and higher LPIPS means they are more different. We first compute LPIPS for all pairs of composite images among 10 generation results. Then, the averaged LPIPS is adopted to evaluate diversity.

4.1.2. Implementation Details

We employ Swin-Tiny (Liu et al., 2021b) pre-trained on ImageNet (Deng et al., 2009) to build our object placement assessment network, which is trained with batch size 64 for 30 epochs. The initial learning rate is 1e-4, and we reduce it by a factor of 0.1 every 10 epochs. This network achieves an F1-score of 0.813 and a balanced accuracy of 0.863 on the test set of the OPA dataset (Liu et al., 2021a). All parameters of this network are frozen when training IOPRE. IOPRE takes this network as the backbone, except that the last fully-connected layer of this network is used as the OPA branch of IOPRE. We set the learning rate as 2e-4 and use AdamW (Loshchilov & Hutter, 2017) to train IOPRE with batch size 64 for 15 epochs. For the assessment reward, we set the hyper-parameter λ_s as 0.01. To train IOPRE, we set the discount factor γ as 0.99, the weight of entropy loss β as 0.08, and the maximum number of steps t_{max} as 20. In the training phase, t_0 is obtained by a random initialization. In the test phase, the maximum number of steps is set as 100. All images are resized to 256×256 before being fed into all networks.

4.2. Comparison with the State-of-the-arts

We compare the proposed method with the state-of-the-art methods: TERSE (Tripathi et al., 2019), PlaceNet (Zhang et al., 2020a) and GracoNet (Zhou et al., 2022). Following (Zhou et al., 2022), we remove the target network of TERSE (Tripathi et al., 2019) and maintain the synthesizer and the discriminator to generate composite images. Because we do not need to use the target network for downstream tasks, and the discriminator is enough to help the synthesizer to generate the composite images that we need. PlaceNet (Zhang et al., 2020a) and GracoNet (Zhou et al., 2022) can be directly applied to predict object placements

Table 1. Comparison with the state-of-the-art methods on the OPA dataset. Best results are denoted in boldface.

Method	Plausibility			Diversity
	User Study↑	Acc.↑	FID↓	LPIPS↑
TERSE (Tripathi et al., 2019)	0.168	0.679	46.94	0
PlaceNet (Zhang et al., 2020a)	0.230	0.683	36.69	0.160
GracoNet (Zhou et al., 2022)	0.284	0.847	27.75	0.206
Our IOPRE	**0.317**	**0.895**	**21.59**	**0.214**

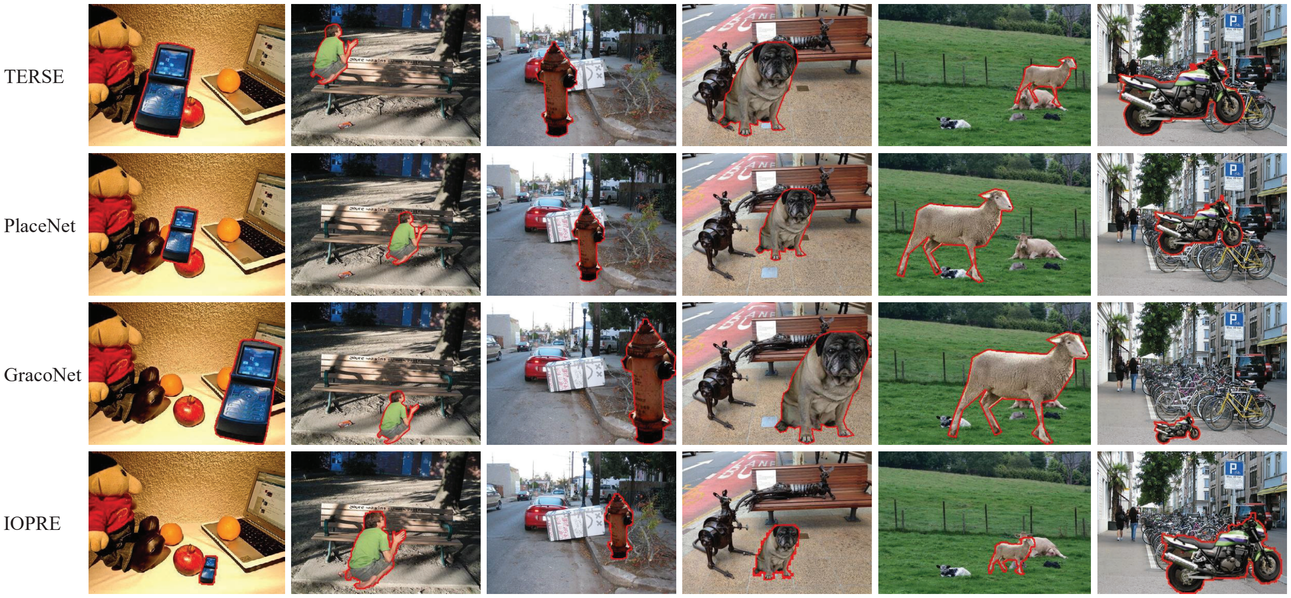

Figure 4. Qualitative comparison results against state-of-the-art methods. The foreground is outlined in red.

without further adjustment. For a fair comparison, we use all training samples of the OPA dataset (Liu et al., 2021a) to train the above methods. Additionally, since the above methods are non-interactive, our IOPRE takes randomly initialized locations and sizes as inputs instead of user-specified initial locations and sizes. As shown in Table 1, the results show that our method performs favorably against state-of-the-art methods in terms of plausibility and diversity. Note that TERSE (Tripathi et al., 2019) outputs only one object placement given a pair of foreground and background, so its diversity is zero.

We also present some qualitative comparison results against state-of-the-art methods on test images of the OPA dataset (Liu et al., 2021a) in Figure 4. Our method also takes randomly initialized locations and sizes as inputs to predict placements for a fair comparison. Compared with other methods, the locations and sizes of the foreground objects are more reasonable in the composite images generated by IOPRE, which verifies the effectiveness of our method in an intuitive way. Additionally, we have provided more qualitative results in the appendix.

4.3. Qualitative Results

As shown in Figure 5, we display the intermediate results, final results, and selected actions in the interactive object placement process. Given a user-specified initial location and size, the agent takes actions step by step to adjust the location and size of the foreground and decides when to stop this process to obtain a reasonable placement. The first two rows of Figure 5 show that IOPRE can produce different object placements when given different initial locations and sizes for a foreground object in the same background, demonstrating its ability to enable flexible and diverse placements based on user preferences. Additionally, if the initial size (*resp.*, location) of the foreground object is reasonable, only its location (*resp.*, size) will be adjusted, as shown in the second (*resp.*, third) row of Figure 5. Moreover, we discover that if the initial location (*resp.*, size) of the fore-

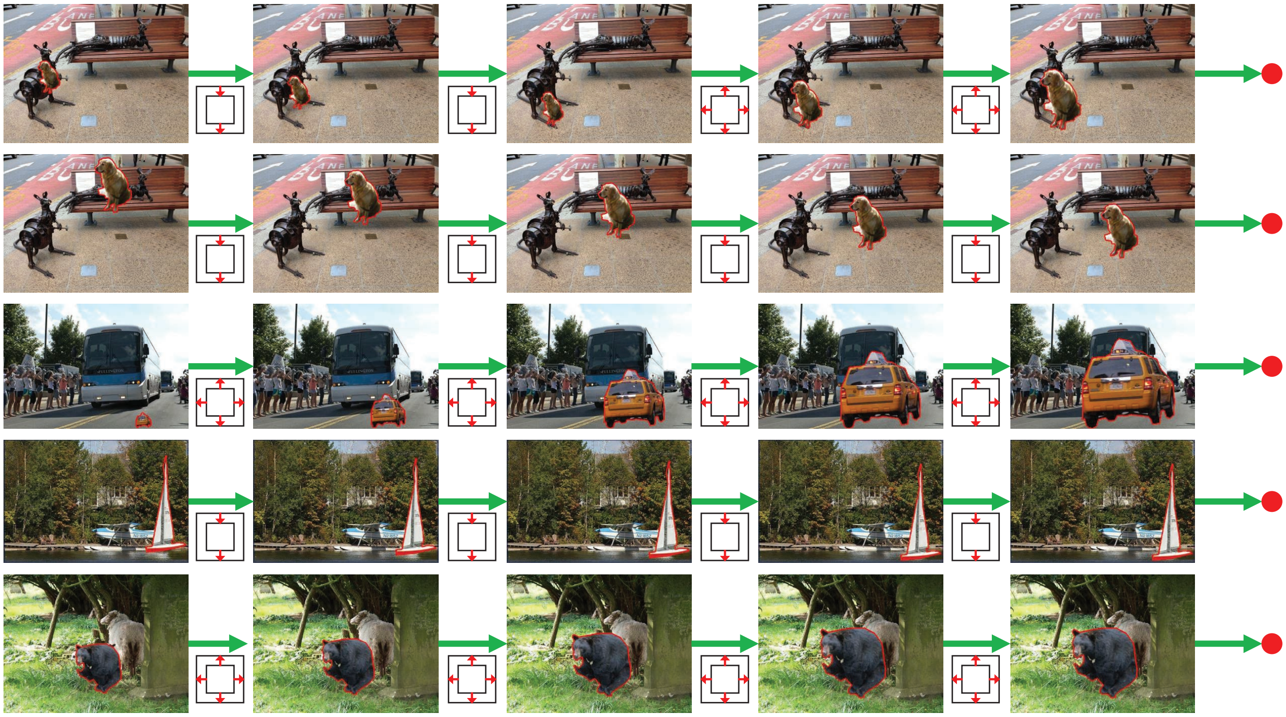

Figure 5. Qualitative results of the proposed method. Each row displays the intermediate results, final result, and selected actions. For convenience, some repetitive actions are omitted in the process. The foreground is outlined in red.

ground object is close to a reasonable location (*resp.*, size), its location (*resp.*, size) will be slightly adjusted, as shown in the fourth (*resp.*, fifth) row of Figure 5. These results indicate that our method can finely perceive the size and location of the foreground object. We also observe that there are no opposite actions in the object placement process, which indicates that the agent has learned an efficient policy. These findings demonstrate the effectiveness of our method for interactive object placement. More results are available in the appendix.

4.4. Ablation Study

We first study the influence of positional embedding. In our IOPRE, we map the transformation vector to the sinusoidal positional embedding, which is combined with the feature of the current composite image to form the current observation. Firstly, we remove the sinusoidal positional embedding and only utilize the feature of the current composite image as the current observation, which is denoted as ‘w/o PE’ in Table 2. Then, we replace the sinusoidal positional embedding with a learnable positional embedding, which is denoted as ‘learnable PE’. The results presented in Table 2 demonstrate that 1) modeling the location and size of the foreground significantly impacts generation plausibility, and 2) the sinusoidal positional embedding better describes the spatial information of the foreground compared to the learnable positional embedding. Note that there is generally a trade-off between plausibility and diversity, and accuracy is more important than other metrics because it directly reflects the proportion of generated reasonable placements.

Table 2. Ablation studies on the positional embedding, aspect ratio preserved action space, usage of LSTM unit, and design of reward function. Best results are denoted in boldface.

Method	Plausibility		Diversity
	Acc.↑	FID↓	LPIPS↑
w/o PE	0.864	23.66	0.227
learnable PE	0.872	22.37	0.218
ratio free	0.558	21.57	0.229
w/o LSTM	0.588	**20.25**	0.244
w/o step	0.858	24.85	0.233
w/o sign	0.672	20.34	**0.247**
Our IOPRE	**0.895**	21.59	0.214

We then study the effect of preserving the aspect ratio of the foreground during the process of composite. In our IOPRE, we utilize t^r to represent the scaling ratio of the whole foreground to keep its aspect ratio constant. To ex-

Table 3. Hyper-parameter analyses on the discount factor of the reward γ, the coefficient of the entropy of loss β, and the hyper-parameter λ_s in the reward function. Best results are denoted in boldface.

Method	Plausibility		Diversity
	Acc.↑	FID↓	LPIPS↑
$\gamma = 0$	0.771	20.48	0.217
$\gamma = 0.5$	0.754	**19.62**	0.186
$\beta = 0.06$	0.885	22.33	0.208
$\beta = 0.1$	0.885	23.70	0.214
$\lambda_s = 0.05$	0.794	21.66	0.239
$\lambda_s = 0.001$	0.842	32.29	**0.262**
Our IOPRE	**0.895**	21.59	0.214

plore the effect of disregarding the aspect ratio, we design an aspect ratio free action space that utilizes two variables to represent the scaling ratios for the width and height of the foreground, respectively, which is denoted as 'ratio free' in Table 2. The experimental results show a decrease in accuracy when using the aspect ratio free action space, which demonstrates the importance of maintaining the aspect ratio of the foreground.

Next, we study the impact of the LSTM unit in our IOPRE. The LSTM unit is responsible for memorizing historical observations and combining them with the current observation to form the current state. We replace the LSTM unit with a fully-connected layer to only focus on the current observation without considering historical observations, which is denoted as 'w/o LSTM' in Table 2. Experimental results show that only focusing on the current observation leads to a notable decrease in accuracy, which indicates the importance of employing an LSTM unit to retain and utilize information from historical observations.

Finally, we analyze the impact of different reward function designs. In IOPRE, our hybrid reward function combines object placement assessment and the number of steps. First, we remove the negative reward in our reward function to ignore the influence of the number of steps, which is denoted as 'w/o step' in Table 2. Then, we remove the sign function to consider the difference in rationality scores between composite images, which is denoted as 'w/o sign'. The results demonstrate that 1) considering the number of steps is significantly important for generation plausibility, and 2) the sign function makes the reward more stable, which is beneficial for model convergence.

4.5. Hyper-parameter Analyses

We conduct ablation studies on major parameters, including the discount factor of the reward γ, the coefficient of the entropy of loss β, and the hyper-parameter λ_s in the reward function, as shown in Table 3. We set γ to 0, 0.5, and 0.99 to analyze the influence of future rewards during the training process, where '$\gamma = 0$' means that the agent only considers the current reward and ignores future rewards. The results demonstrate that our IOPRE achieves the highest accuracy when γ is set to 0.99, indicating the significance of future rewards in learning the placement policy. To investigate the effect of the entropy in the optimization objective, we set β to 0.06, 0.08, and 0.1. When β is set to 0.08, our IOPRE achieves the best performance in both plausibility and diversity because increasing the value of β improves exploration. However, when β is too large, the probability distribution of actions becomes close to the uniform distribution, which is not conducive to learning the optimal policy. For the assessment reward, we set λ_s as 0.05, 0.01, and 0.001 to study the effect of the negative reward based on the number of steps. When λ_s is set to 0.01, IOPRE achieves the best performance in plausibility, which shows that too high or too low negative rewards will lead agents to learn sub-optimal policies.

5. Limitation

One of the main limitations of our method is that it does not take into account aesthetic evaluation during the placement, which is crucial in real-world applications such as artistic creation and automated advertising. We will overcome this limitation in future work.

6. Conclusion

In this paper, we formulate object placement as a sequential decision-making process and propose an interactive object placement method with reinforcement learning to produce a reasonable placement given an initial location and size of the foreground. Specifically, we have designed an aspect ratio preserved action space, a multi-factor state representation learning method, and a hybrid reward function to make the agent learn to place objects in background images with suitable locations and sizes. Experimental results on the OPA dataset demonstrate that our method has achieved interactive object placement and obtained state-of-the-art performance in terms of plausibility and diversity.

Acknowledgements

This work was supported in part by the National Key R&D Program of China (Nos. 2021YFF0900500, 2021ZD0110901, 2022ZD0118202), the National Natural Science Foundation of China (Nos. 62272134 and 62236003), the Taishan Scholars Program of Shandong Province (No. tsqn201812106), and the Shenzhen College Stability Support Plan (No. GXWD20220817144428005).

References

Azadi, S., Pathak, D., Ebrahimi, S., and Darrell, T. Compositional GAN: Learning image-conditional binary composition. *International Journal of Computer Vision*, 128 (10):2570–2585, 2020.

Chen, B.-C. and Kae, A. Toward realistic image compositing with adversarial learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8415–8424, 2019.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Dvornik, N., Mairal, J., and Schmid, C. Modeling visual context is key to augmenting object detection datasets. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 364–380, 2018.

Dvornik, N., Mairal, J., and Schmid, C. On the importance of visual context for data augmentation in scene understanding. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):2014–2028, 2019.

Dwibedi, D., Misra, I., and Hebert, M. Cut, paste and learn: Surprisingly easy synthesis for instance detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1301–1310, 2017.

Fang, H.-S., Sun, J., Wang, R., Gou, M., Li, Y.-L., and Lu, C. InstaBoost: Boosting instance segmentation via probability map guided copy-pasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 682–691, 2019.

Georgakis, G., Mousavian, A., Berg, A. C., and Kosecka, J. Synthesizing training data for object detection in indoor scenes. *arXiv preprint arXiv:1702.07836*, 2017.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, pp. 6626–6637, 2017.

Jaderberg, M., Simonyan, K., Zisserman, A., and kavukcuoglu, k. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, volume 28, pp. 2017–2025, 2015.

Jie, Z., Liang, X., Feng, J., Jin, X., Lu, W., and Yan, S. Tree-structured reinforcement learning for sequential object localization. In *Advances in Neural Information Processing Systems*, volume 29, pp. 127–135, 2016.

Kikuchi, K., Yamaguchi, K., Simo-Serra, E., and Kobayashi, T. Regularized adversarial training for single-shot virtual try-on. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 3149–3152, 2019.

Lee, D., Liu, S., Gu, J., Liu, M.-Y., Yang, M.-H., and Kautz, J. Context-aware synthesis and placement of object instances. In *Advances in Neural Information Processing Systems*, volume 31, pp. 10414–10424, 2018.

Li, D., Wu, H., Zhang, J., and Huang, K. A2-RL: Aesthetics aware reinforcement learning for image cropping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8193–8201, 2018.

Lin, C.-H., Yumer, E., Wang, O., Shechtman, E., and Lucey, S. ST-GAN: Spatial transformer generative adversarial networks for image compositing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9455–9464, 2018.

Liu, L., Liu, Z., Zhang, B., Li, J., Niu, L., Liu, Q., and Zhang, L. OPA: Object placement assessment dataset. *arXiv preprint arXiv:2107.01889*, 2021a.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. Swin Transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10012–10022, 2021b.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., and Kavukcuoglu, K. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, pp. 1928–1937. PMLR, 2016.

Niu, L., Cong, W., Liu, L., Hong, Y., Zhang, B., Liang, J., and Zhang, L. Making images real again: A comprehensive survey on deep image composition. *arXiv preprint arXiv:2106.14490*, 2021.

Niu, L., Liu, Q., Liu, Z., and Li, J. Fast object placement assessment. *arXiv preprint arXiv:2205.14280*, 2022.

Ouyang, X., Cheng, Y., Jiang, Y., Li, C.-L., and Zhou, P. Pedestrian-Synthesis-GAN: Generating pedestrian data in real scene and beyond. *arXiv preprint arXiv:1804.02047*, 2018.

Pirinen, A. and Sminchisescu, C. Deep reinforcement learning of region proposal networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6945–6954, 2018.

Remez, T., Huang, J., and Brown, M. Learning to segment via cut-and-paste. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 37–52, 2018.

Tan, F., Bernier, C., Cohen, B., Ordonez, V., and Barnes, C. Where and who? automatic semantic-aware person composition. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1519–1528. IEEE, 2018.

Tripathi, S., Chandra, S., Agrawal, A., Tyagi, A., Rehg, J. M., and Chari, V. Learning to generate synthetic data via compositing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 461–470, 2019.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017.

Volokitin, A., Susmelj, I., Agustsson, E., Van Gool, L., and Timofte, R. Efficiently detecting plausible locations for object placement using masked convolutions. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, pp. 252–266. Springer, 2020.

Wang, H., Wang, Q., Yang, F., Zhang, W., and Zuo, W. Data augmentation for object detection via progressive and selective instance-switching. *arXiv preprint arXiv:1906.00358*, 2019.

Weng, S., Li, W., Li, D., Jin, H., and Shi, B. MISC: Multi-condition injection and spatially-adaptive compositing for conditional person image synthesis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7741–7749, 2020.

Williams, R. J. and Peng, J. Function optimization using connectionist reinforcement learning algorithms. *Connection Science*, 3(3):241–268, 1991.

Wu, W., Liu, J., Zheng, K., Sun, Q., and Zha, Z.-J. Temporal complementarity-guided reinforcement learning for image-to-video person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7319–7328, 2022.

Zhan, F., Zhu, H., and Lu, S. Spatial fusion GAN for image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3653–3662, 2019.

Zhang, L., Wen, T., Min, J., Wang, J., Han, D., and Shi, J. Learning object placement by inpainting for compositional data augmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, pp. 566–581. Springer, 2020a.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595, 2018.

Zhang, S.-H., Zhou, Z.-P., Liu, B., Dong, X., and Hall, P. What and where: A context-based recommendation system for object insertion. *Computational Visual Media*, 6(1):79–93, 2020b.

Zhou, S., Liu, L., Niu, L., and Zhang, L. Learning object placement via dual-path graph completion. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII*, pp. 373–389. Springer, 2022.

A. More Qualitative Comparison Results against State-of-the-art Methods

We present more qualitative comparison results against state-of-the-art methods on the test images of the OPA dataset (Liu et al., 2021a) in Figure 6 and Figure 7. Specifically, we compare the proposed method, IOPRE, with TERSE (Tripathi et al., 2019), PlaceNet (Zhang et al., 2020a), and GracoNet (Zhou et al., 2022). Our method demonstrates superior performance in placing foreground objects with more reasonable locations and sizes compared to other methods.

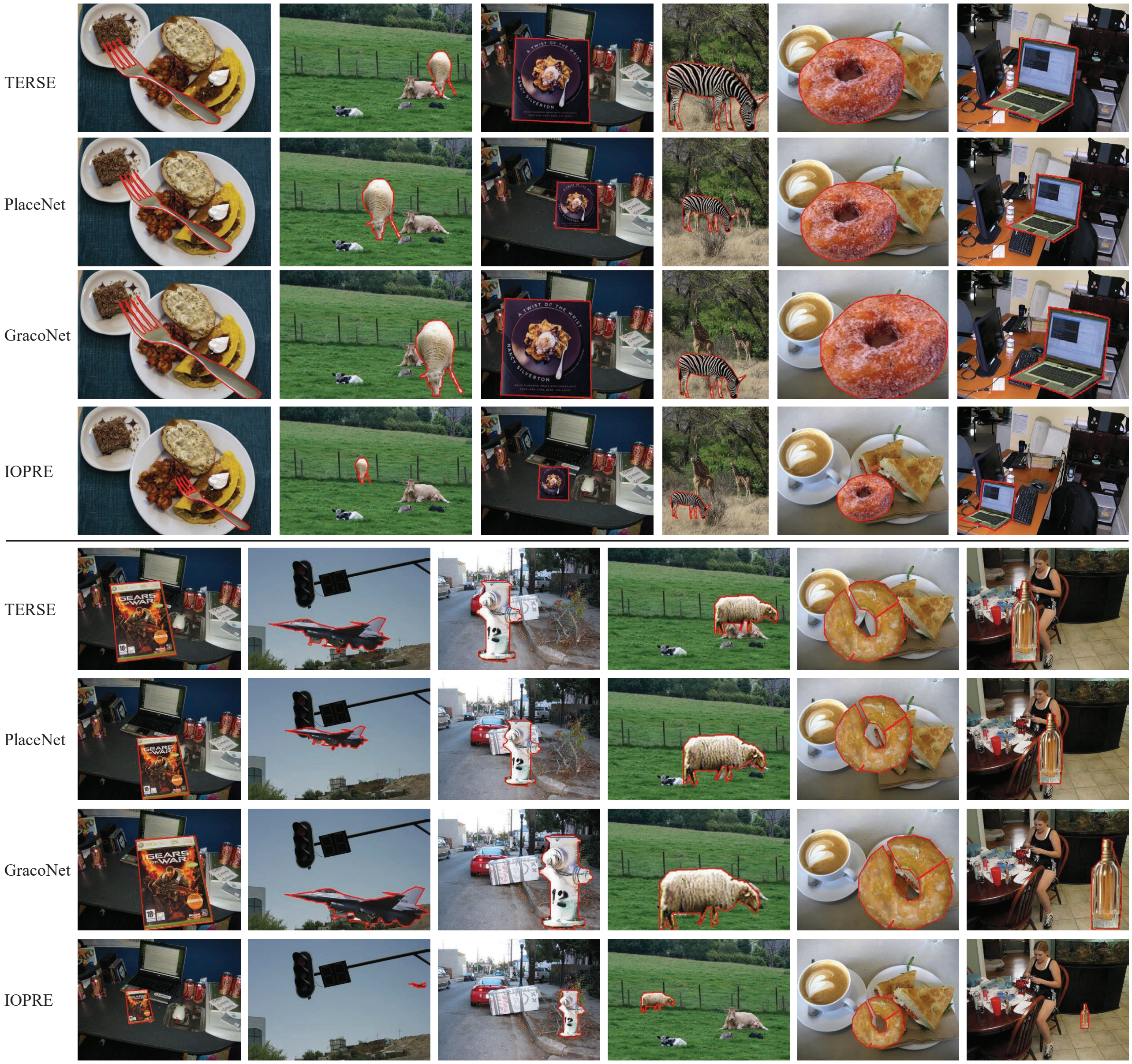

Figure 6. Qualitative comparison results against state-of-the-art methods. The foreground is outlined in red.

B. More Qualitative Results of the Proposed Method

We show more intermediate results, final results, and selected actions during the interactive object placement in Figure 8. Based on the user-specified initial location and size, the agent takes a series of actions according to the learned policy to place an object with a suitable location and size.

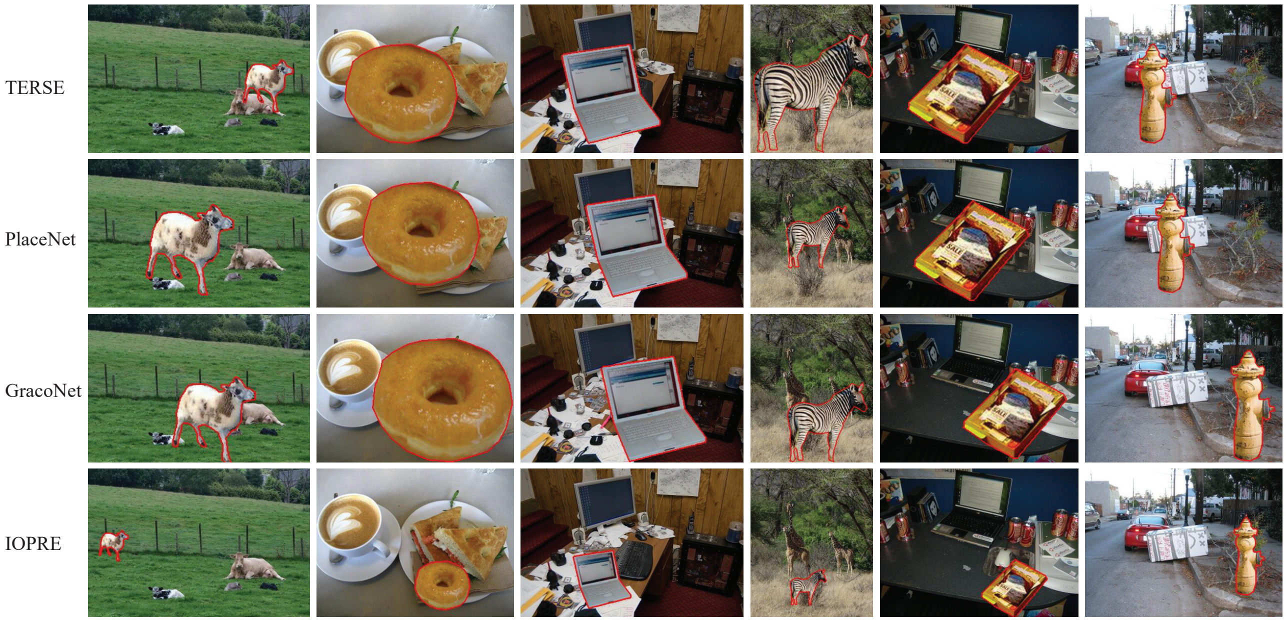

Figure 7. Qualitative comparison results against state-of-the-art methods. The foreground is outlined in red.

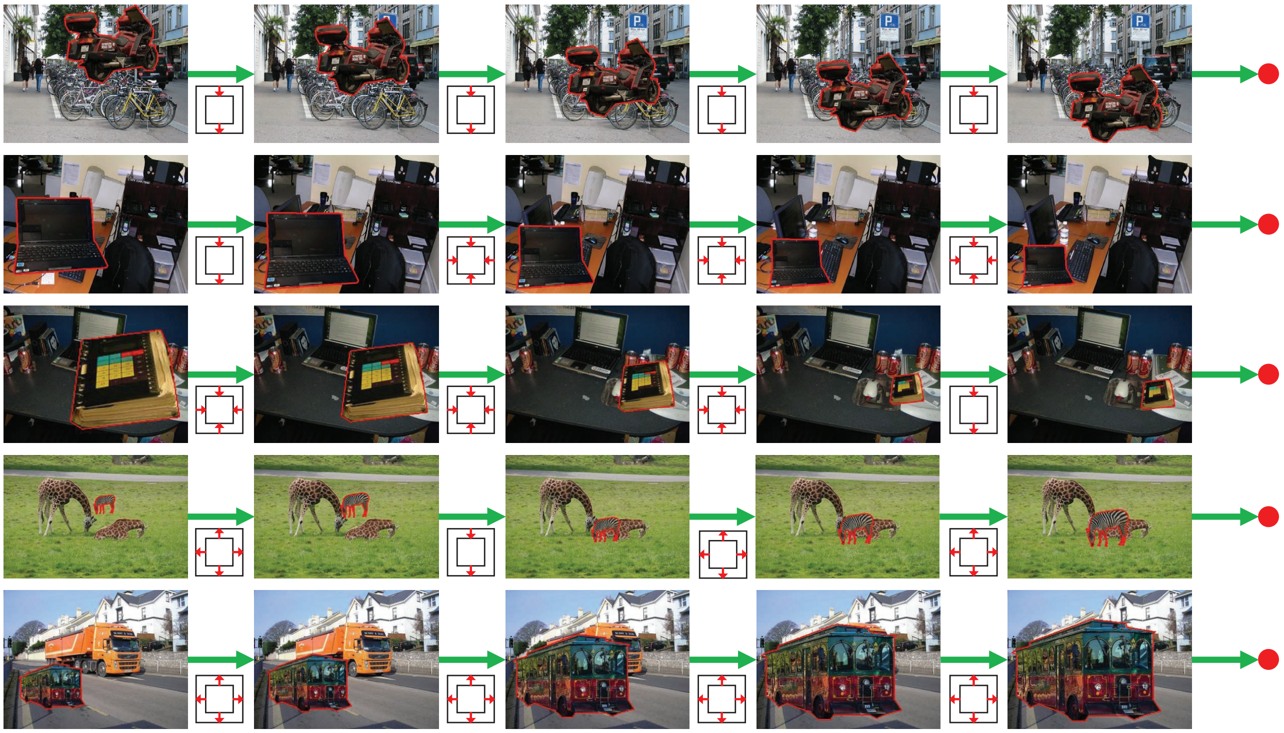

Figure 8. Qualitative results of the proposed method. Each row displays the intermediate results, final result, and selected actions. For convenience, some repetitive actions are omitted in the process. The foreground is outlined in red.