We present ReDO (Redrawing Objects): an unsupervised, data-driven, object segmentation method for real images.

We assume natural images generation is a composite process in which each object is generated independently. Object segmentation is then the discovery of regions that can be redrawn without seeing the rest of the image.

### Image Composition Model

We consider the following underlying generative process $G$ that produce images in three steps.

1. Define the position of the different regions: global structure of the image, by sampling $N$ region masks.

   $M \sim p(M), M \in \{0, 1\}^{N \times W \times H}$, $\sum_{k=1}^{N} M_{k,x,y} = 1$

2. Generate the contents of each region independently.

   $V^k \leftarrow G_i(M^k, z_i), z_i \sim p(z)$ for $k \in \{1, \ldots, n\}$

3. Aggreate the resulting regions into the final image.

   $G(M, z_1, \ldots, z_n) = \sum_{k=1}^{N} M^k \odot V^k$

---

### Towards Learning To Segment

We replace step 1 by a segmentation function $F$ that produce a mask given an image input $I \in \mathbb{R}^{C \times W \times H}$.

1. Obtain the mask using a segmentation function $F$.

   $M \leftarrow F(I), M \in \{0, 1\}^{N \times W \times H}$, $\sum_{k=1}^{N} M_{k,x,y}^z = 1$

2. Generate the contents of each region independently.

   $V^k \leftarrow G_i(M^k, z_i), z_i \sim p(z)$ for $k \in \{1, \ldots, n\}$

3. Aggreate the resulting regions into the final image.

   $G_r(I, z_1, \ldots, z_n) = \sum_{k=1}^{N} M^k \odot V^k$

---

### Conservation of Information

**Problem 1:** Mapping all pixels to one region is a trivial but valid solution.

![Conservation of Information](image)

**Solution:** We add a learned function $\delta$ that tries to reconstruct the noise vectors $z_i$ from the generated image.

**Problem 2:** The segmentation function can ignore the input.

![Conservation of Information](image)

**Solution:** We tie the output to the input by only regenerating one region at a time, keeping the rest of the image the same.

---

### Learning the full model for object segmentation

We can train this model end-to-end using an adversarial loss to match the distribution of generated images to the dataset distribution. But it would naturally converge to trivial and uninformative solutions.

**Objective functions:** We use the hinge version of the adversarial loss.

$$
\max_{\alpha, \delta} \mathcal{L}_G = \mathbb{E}_{I \sim p_{data}}[D(G_r(I, z_i, i)) - \lambda \delta(G_r(I, z_i, i) - z_i)]
$$

$$
\max_{\alpha, \delta} \mathcal{L}_D = \mathbb{E}_{I \sim p_{data}}[\min(0, -1 + D(I))] + \mathbb{E}_{z \sim p_z}[\min(0, -1 - D(G_r(I, z_i, i)))]
$$

---

**Preprint**


**Code and Pretrained**

https://github.com/mickaelChen/ReDO
Experiments

We evaluate ReDO on 3 datasets of real images:
- Without supervision, ReDO discovers meaningful object masks and noise vectors z codes for specific texture.
- ReDO’s performance is comparable to supervised baselines trained with about 50-100 labelled datapoints.
- Preliminary experiments indicates that ReDO can work on datasets with multiple objects or multiple classes without using labels.

ReDO and supervised baselines

Generated Samples

Dataset with 2 Categories

Datasets with 2 Objects

Additional masks