

# Supplementary Material for “Semantic Image Synthesis via Adversarial Learning”

## 1 Additional Results on Oxford-102 Flower Dataset

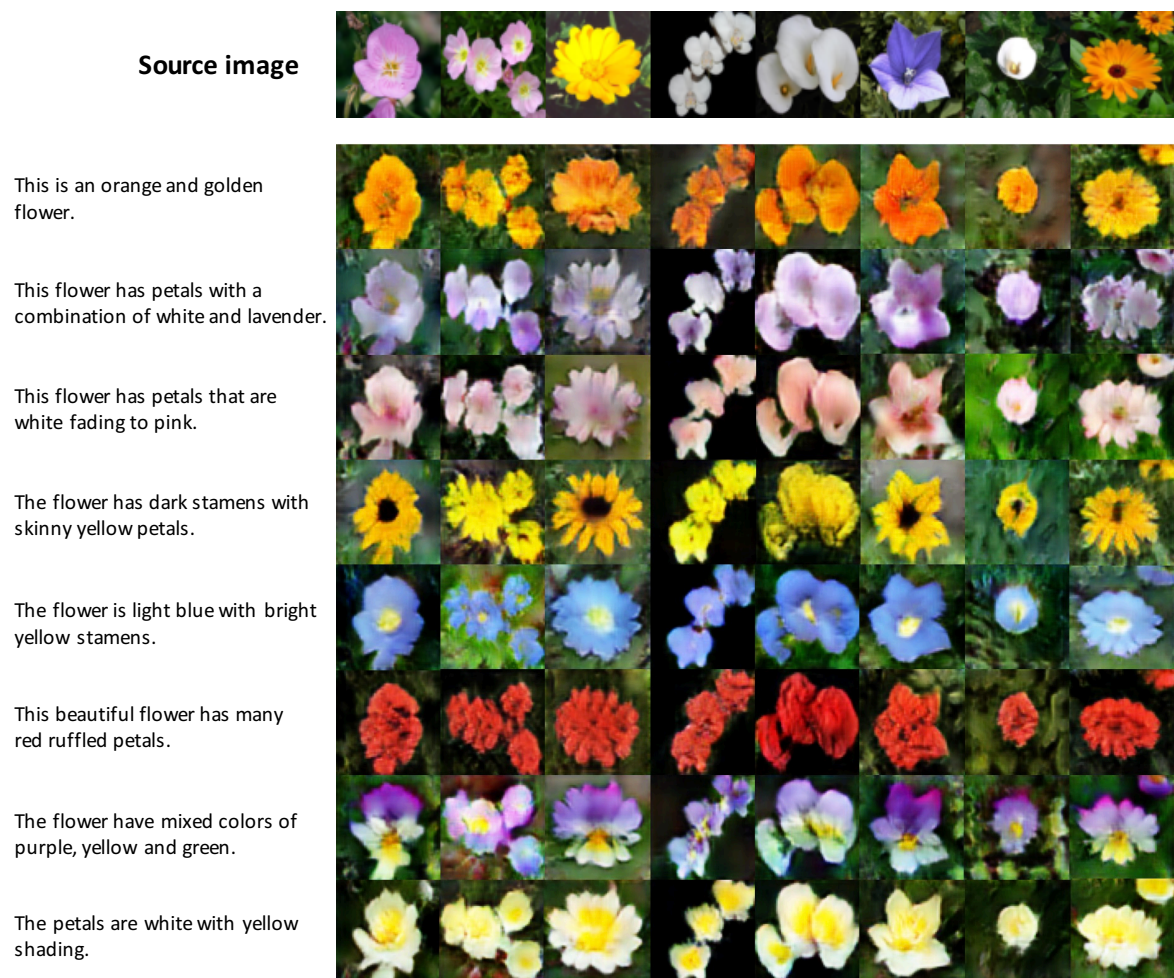


Figure 1: Additional zero-shot results of our method without pretrained VGG encoder on Oxford-102 flower dataset.

## 2 Additional Results on Caltech-200 Bird Dataset



Figure 2: Additional zero-shot results of our method without pretrained VGG encoder on Caltech-200 bird dataset.

### 3 Failure Cases

The main reasons for failure cases are: 1) complex and unrecognized background information; 2) diverse and complicated shape of objects.

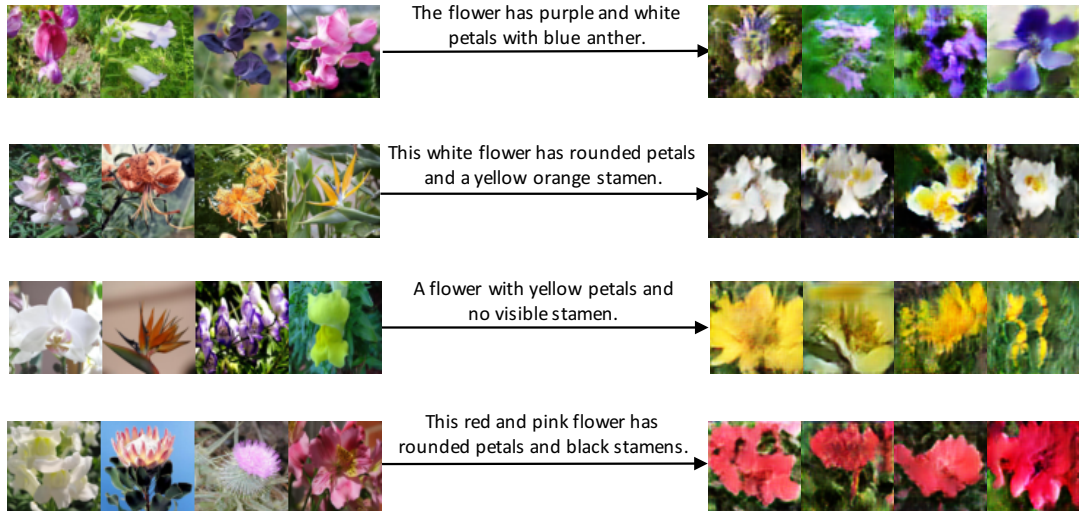


Figure 3: Failure cases of zero-shot results of our method without pretrained VGG encoder on Oxford-102 flower dataset.

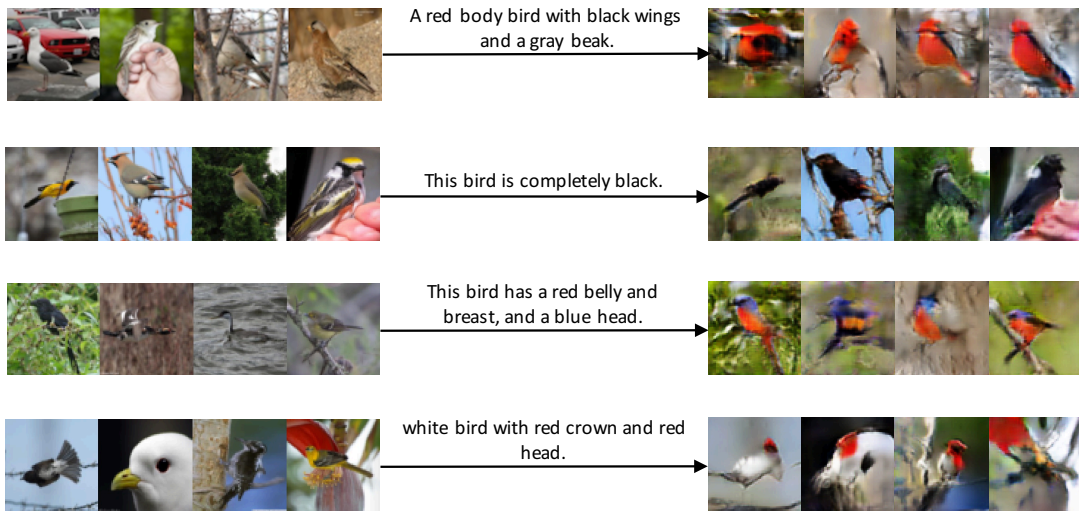


Figure 4: Failure cases of zero-shot results of our method without pretrained VGG encoder on Caltech-200 bird dataset.

## 4 Training Algorithm

**Input:** batches of image  $x$ , matching text  $t$ , mismatching text  $\hat{t}$ , semantically relevant text  $\bar{t}$ , number of training steps  $n$  and learning rate  $\alpha$ .

- 1: **for**  $i = 1$  **to**  $n$  **do**
- 2:    $h \leftarrow \varphi(t)$  encode matching text
- 3:    $\hat{h} \leftarrow \varphi(\hat{t})$  encode mismatching text
- 4:    $\bar{h} \leftarrow \varphi(\bar{t})$  encode semantically relevant text
- 5:    $\bar{x} \leftarrow G(x, \bar{h})$  forward generator with real image and semantically relevant text embedding
- 6:    $s_r \leftarrow D(x, h)$  real image, matching text
- 7:    $s_w \leftarrow D(x, \hat{h})$  real image, mismatching text
- 8:    $s_s \leftarrow D(\bar{x}, \bar{h})$  synthesized image, semantically relevant text
- 9:    $\mathcal{L}_D \leftarrow \log(s_r) + (\log(1 - s_w) + \log(1 - s_s))/2$
- 10:    $D \leftarrow D - \alpha \delta \mathcal{L}_D / \delta D$  update discriminator
- 11:    $\mathcal{L}_G \leftarrow \log(s_s)$
- 12:    $G \leftarrow G - \alpha \delta \mathcal{L}_G / \delta G$  update generator
- 13: **end for**

**Algorithm 1:** Training algorithm for semantic image synthesis via adversarial learning.