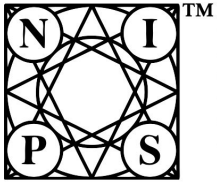


Neural Information Processing Systems Foundation

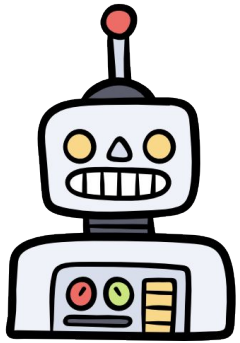
Text-Adaptive Generative Adversarial Networks: Manipulating Images with Natural Language

Seonghyeon Nam, Yunji Kim, Seon Joo Kim

Dept. of Computer Science, Yonsei University

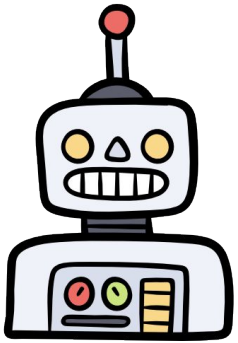
Seoul, South Korea

Manipulating Images with Natural Language



Manipulating Images with Natural Language

This small bird has a *blue crown* and *white belly*.



Icons made by Freepik from www.flaticon.com

Manipulating Images with Natural Language

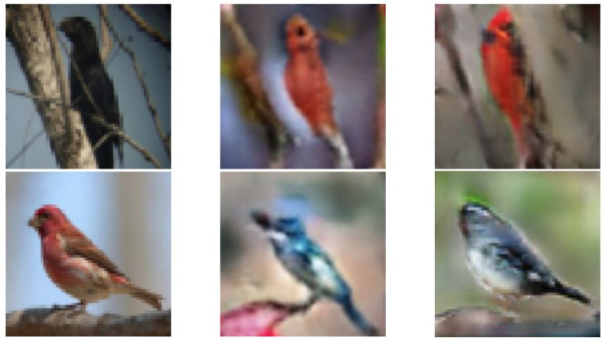
Icons made by Freepik from www.flaticon.com

Related Work

- Existing methods rely heavily on sentence embedding vectors
- They fail to preserve text-irrelevant contents (e.g. background)
- Coarse multi-modal modeling is not enough for the disentanglement

This particular bird with a **red head and breast** and features **grey wings**.

This small bird has a **blue crown** and **white belly**.



Original | [Reed et al., 2016] | [Dong et al., 2017]

Contribution

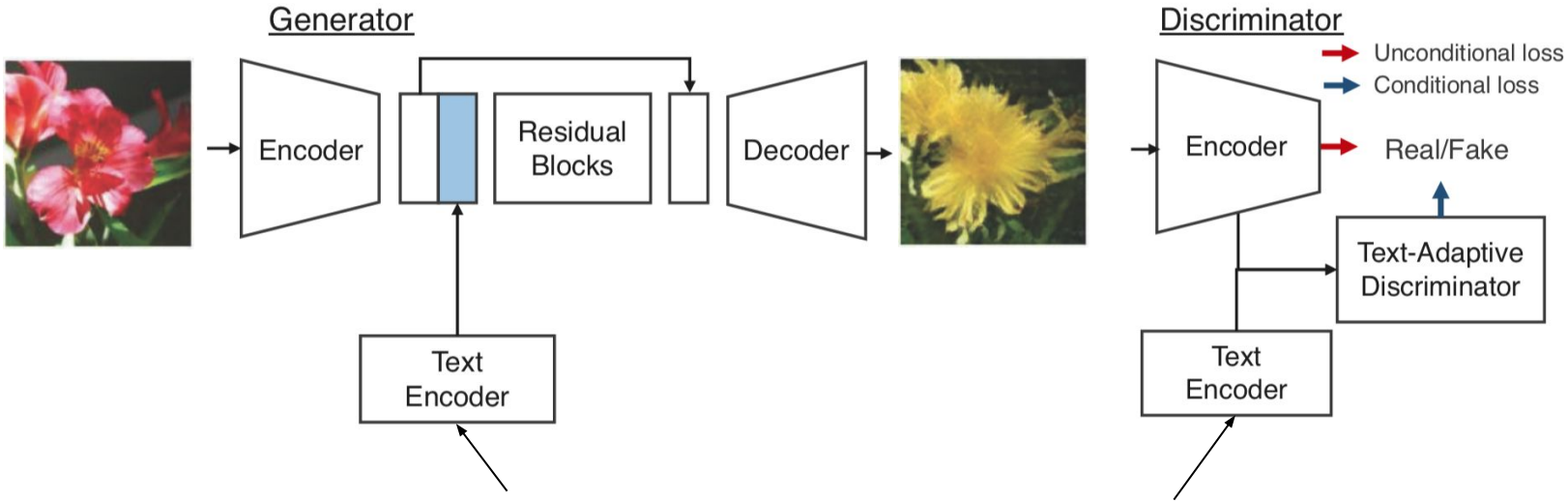
- Our key idea is word-level local discriminators for fine-grained training
- Our method effectively changes visual attributes while preserving text-irrelevant contents

This particular bird with a **red head and breast** and features **grey wings**.

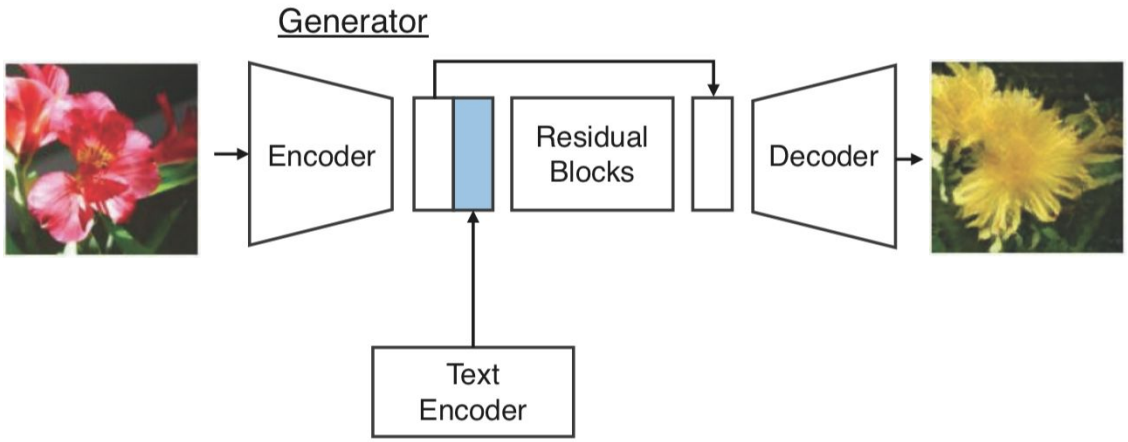
This small bird has a **blue crown** and **white belly**.

Original | [Reed et al., 2016] | [Dong et al., 2017] | Ours

Overview of TAGAN



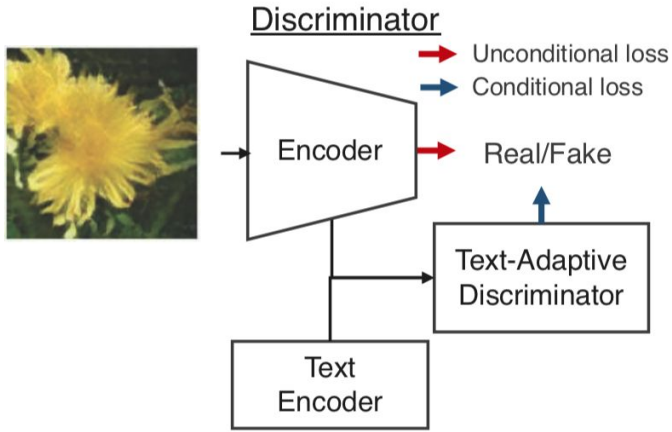
Generator



This flower has petals that are *yellow* and are *very stringy*.

To preserve original contents,
we add a reconstruction loss: $L_{rec} = \|\mathbf{x} - G(\mathbf{x}, \mathbf{t})\|$

Discriminator



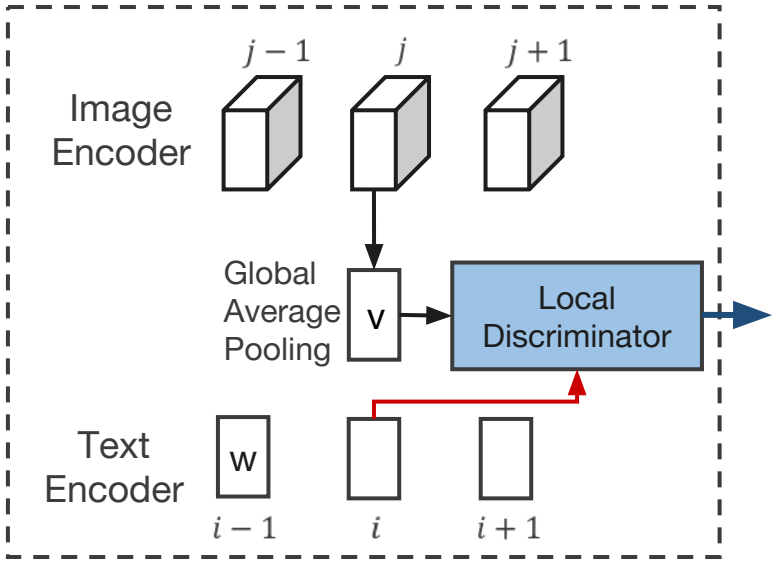
This flower has petals that are *yellow* and are *very stringy*.

The discriminator consists of

1. Unconditional discriminator
   → Make image realistic

2. Text-adaptive discriminator
   → Make image match the text

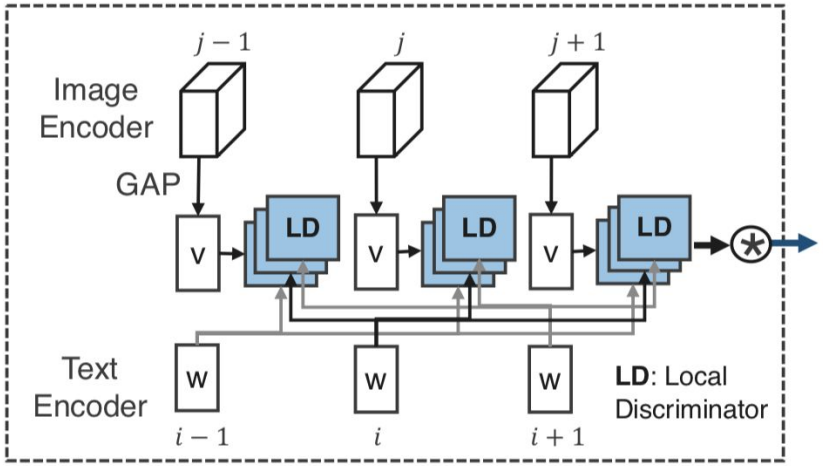
Text-Adaptive Discriminator



1. Compute local discriminator scores

$$f_{\mathbf{w}_i}(\mathbf{v}) = \sigma(\underbrace{\mathbf{W}(\mathbf{w}_i)}_{\text{text}} \cdot \underbrace{\mathbf{v}}_{\text{image}} + \underbrace{\mathbf{b}(\mathbf{w}_i)}_{\text{text}})$$

Text-Adaptive Discriminator



1. Compute local discriminator scores

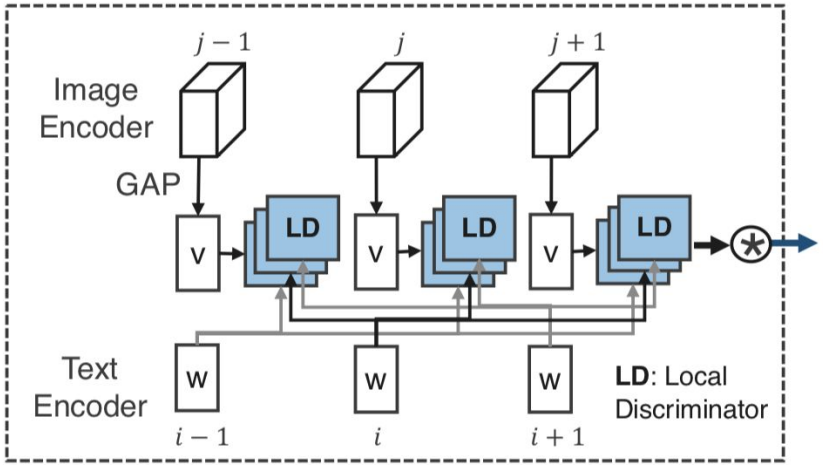
$$f_{\mathbf{w}_i}(\mathbf{v}) = \sigma(\mathbf{W}(\mathbf{w}_i) \cdot \mathbf{v} + \mathbf{b}(\mathbf{w}_i))$$

2. Compute text/image attentions

α_i : softmax weight for word *i*

β_{ij} : softmax weight for word *i*, and image feature level *j*

Text-Adaptive Discriminator



1. Compute local discriminator scores

$$f_{\mathbf{w}_i}(\mathbf{v}) = \sigma(\mathbf{W}(\mathbf{w}_i) \cdot \mathbf{v} + \mathbf{b}(\mathbf{w}_i))$$

2. Compute text/image attentions

α_i : softmax weight for word *i*

β_{ij} : softmax weight for word *i*, and image feature level *j*

3. Aggregate the scores with attentions

$$D(\mathbf{x}, \mathbf{t}) = \prod_{i=1}^{T} [\sum_{j} \beta_{ij} f_{\mathbf{w}_i, j}(\mathbf{v}_j)]^{\alpha_i}$$

Manipulation Results on CUB-200

Original

This particular bird with a **red head and breast** and features **grey wings**.

This bird has **wings that are blue** and has a **white belly**.

A small bird with **white base** and **black stripes** throughout its belly, head, and feathers.

Manipulation Results on Oxford-102

Original

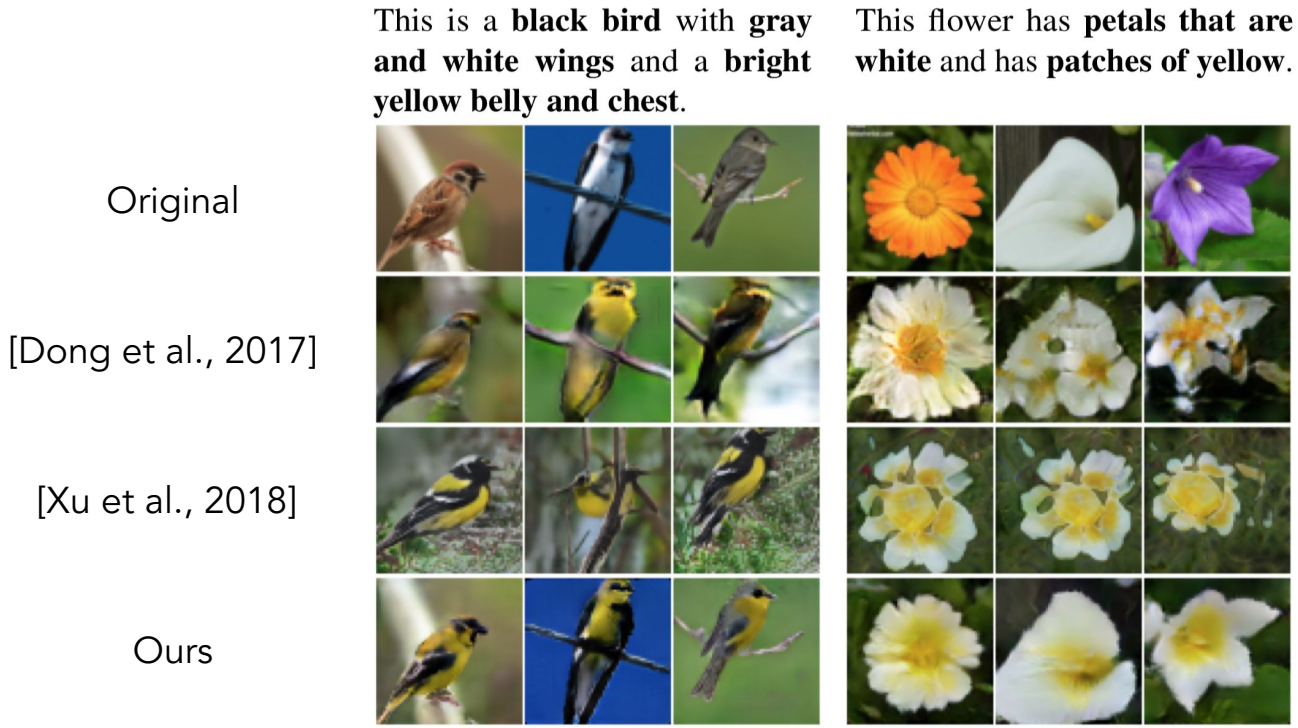
The petals are **white** and the stamens are **light yellow**.

The petals of the flower have **yellow and red stripes**.

This flower has petals of **pink and white color** with **yellow stamens**.

Gazania
Wikipedia

Qualitative Comparison



Conclusion

- We propose a Text-Adaptive Generative Adversarial Network (TAGAN)
- Our method disentangles and manipulates fine-grained visual attributes
- Our method outperforms existing methods on CUB-200 and Oxford-102

Please visit our poster (#126) for more information

https://github.com/woozzu/tagan