Visual Search by Brushing

Baldo Faieta, Mingyang Ling, Yifu Wang | Adobe

Bē
Jon Noorlander

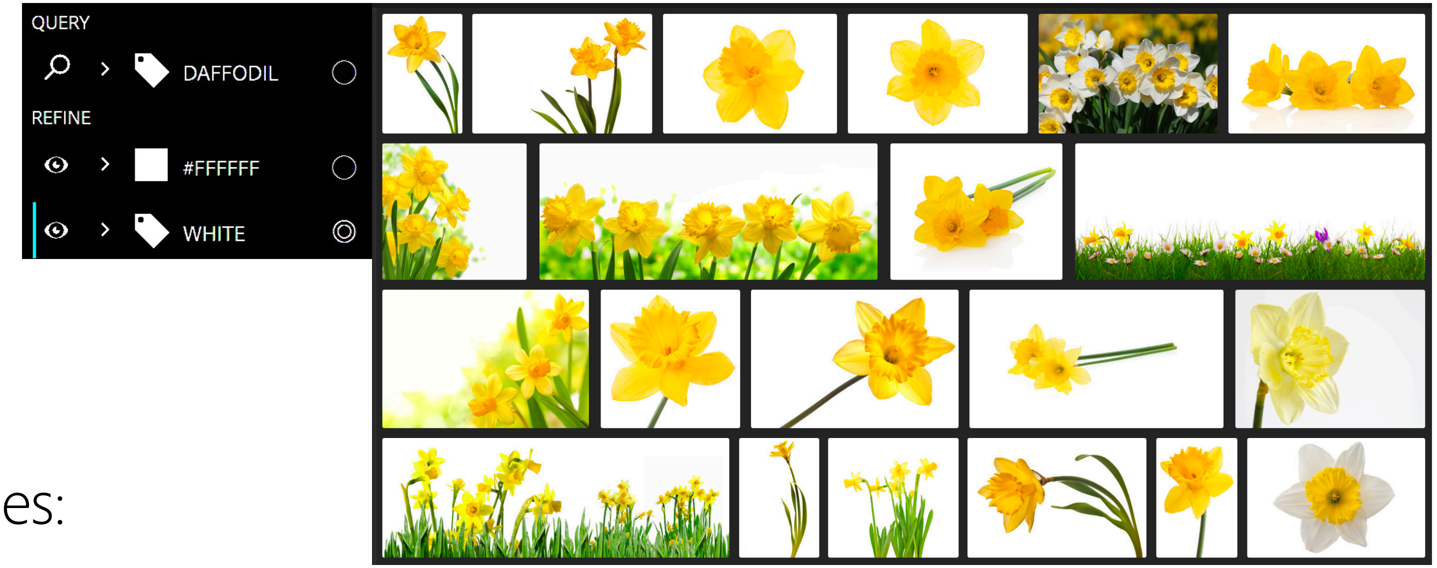
Introduction

- 2B people carry phones with cameras
  - 1 trillion photos / year
  - 10k-100k personal photo collections
- Don't want keyword search on mobiles
- Visual search
  - Needs starting image
  - Hard to refine
- Idea: Generative visual query
- Can one model generate
  "cars", "bridges", "flowers", "dogs" … ?
- Can use reformulate visual (generated) query?

Problem

- Difficult to describe textually image details
- Even harder in <u>mobile</u> and without words

- Query Example: "**White Daffodil**"
  - Fails because:
    - White background
    - Yellow Daffodil most common



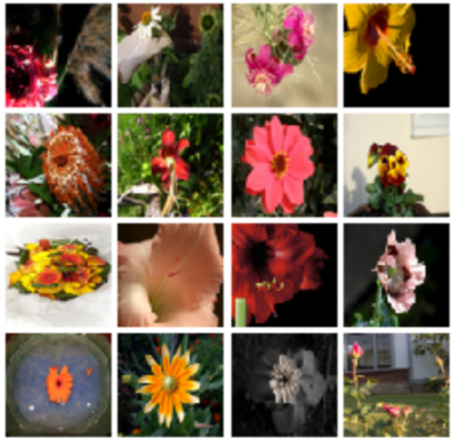
- Solution for mobile touch interfaces:
  - Generative, interactive image query
  - <u>Narrow</u> to text category (e.g., "Daffodil")

Data

- Oxford-102 Flower dataset
  - 8189 flower images, 102 classes

- Adobe Stock 160k (internal) dataset
  - 160k ”squares” samples from 63M images
  - Pre-trained **w2v** for every image
  - w2v trained based on original tags

- Adobe Stock 10k ”flowers” dataset
  - 10k sample filtered to “flower” query

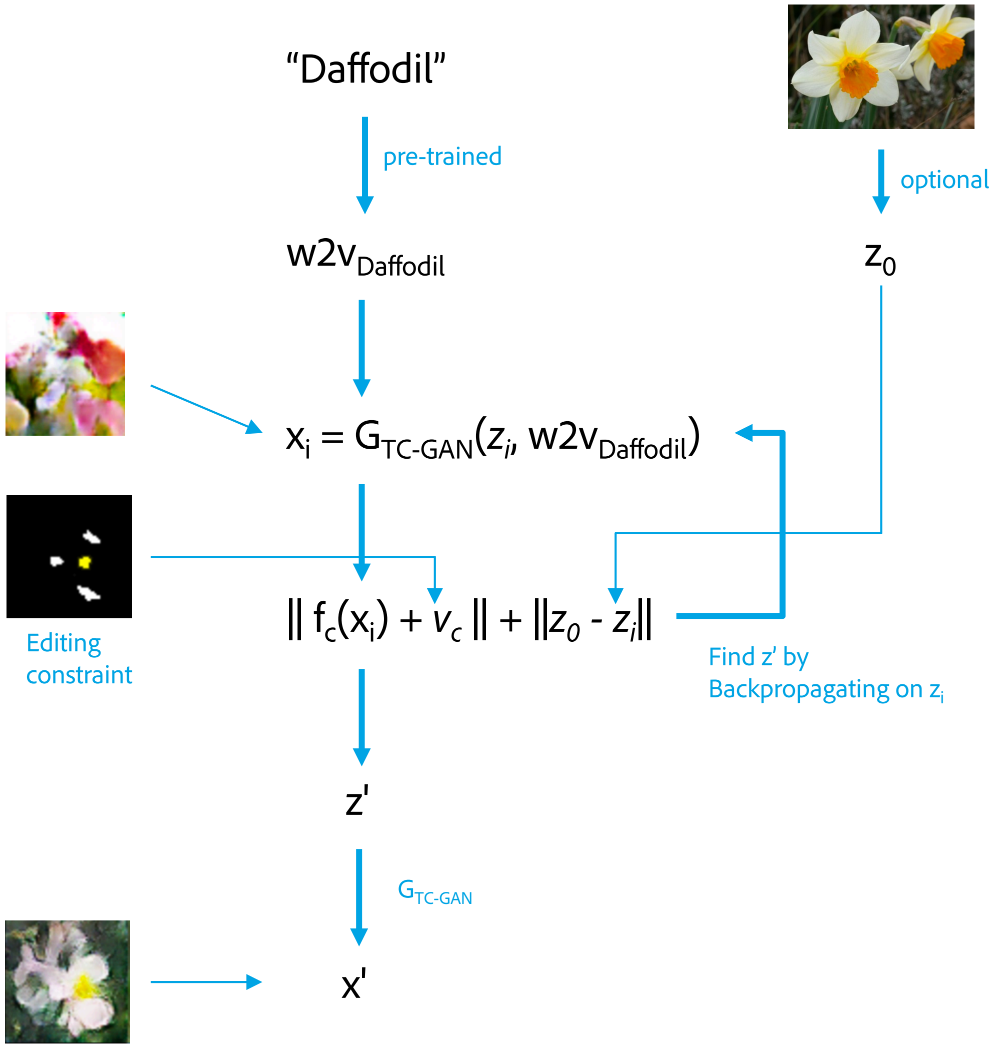


Approach

- 3 components:
  - **TC-GAN**: Text-conditional GAN generates images conditioned on **w2v** (or category)
  - **iGAN**: Apply editing constraints in latent space to generate image that conforms with constraints
  - **G_{XZ}**: Inverse GAN to infer a ***z*** given a image ***x***

1. iGAN as baseline
   - Train with DCGAN on Oxford-flower-102
2. Concurrently, train TC-GAN based on AC-GAN, Text-to-Image Synthesis, TAC-GAN
   - Oxford-flower-102 for classes
   - Adobe Stock 10k flowers with w2v for Text-to-Image Synthesis
3. Train G_{XZ} based on BiGAN, BEGAN, AEGAN, …
4. Integrate to work as one model

Model



"Daffodil"

pre-trained

$w2v_{Daffodil}$

optional

z_0

$x_i = G_{TC\text{-}GAN}(z_i, w2v_{Daffodil})$

$\| f_c(x_i) + v_c \| + \|z_0 - z_i\|$

Editing constraint

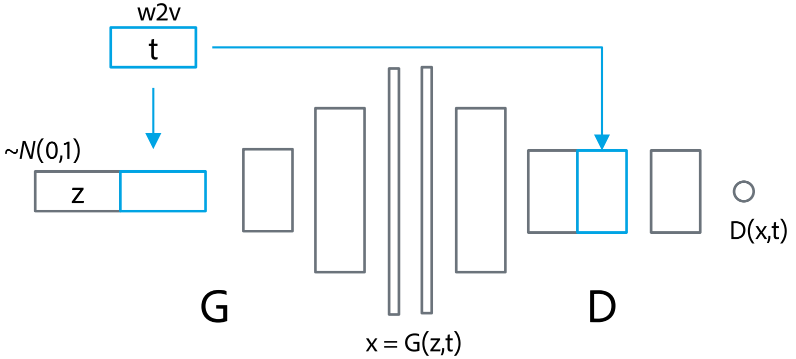
Find z' by Backpropagating on z_i

z'

$G_{TC\text{-}GAN}$

x'

Text Conditional GAN



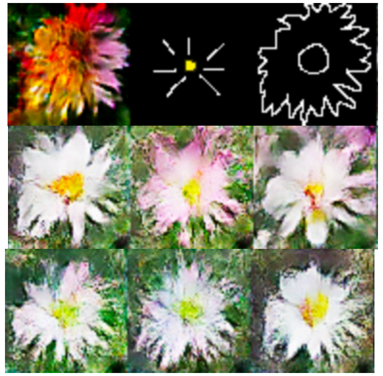
Results

Interactive GAN (iGAN)

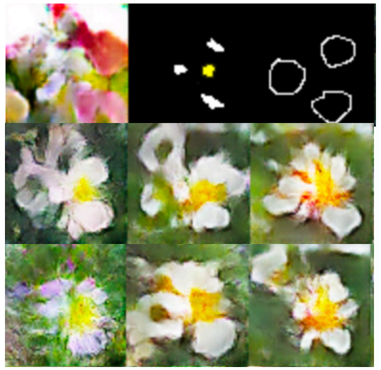
brushes

Outline (optional)

Chrysanthemum Query



Daffodil Query

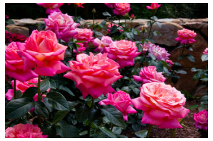


Text-Conditional GAN

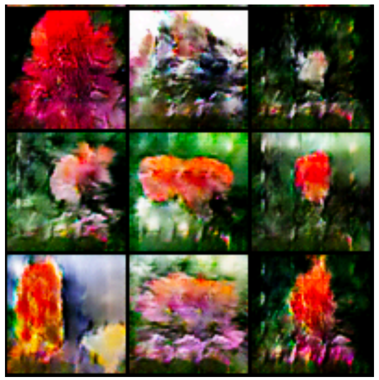
Epoch 74, 11400 iterations fake examples

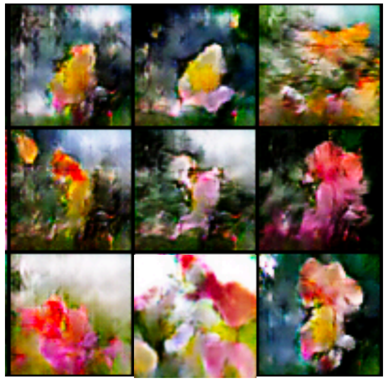
Chrysan-themum

Rose



Daffodil

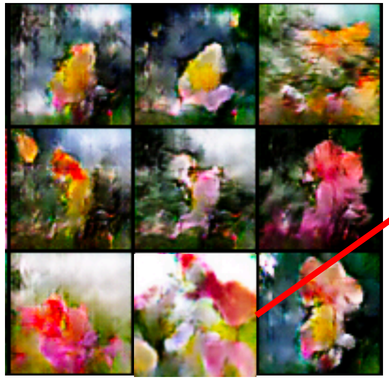




Demo

- Query: “Daffodil”

TC-GAN



iGAN

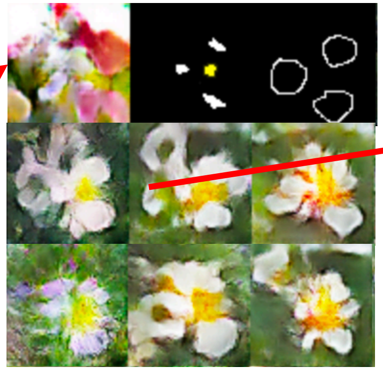
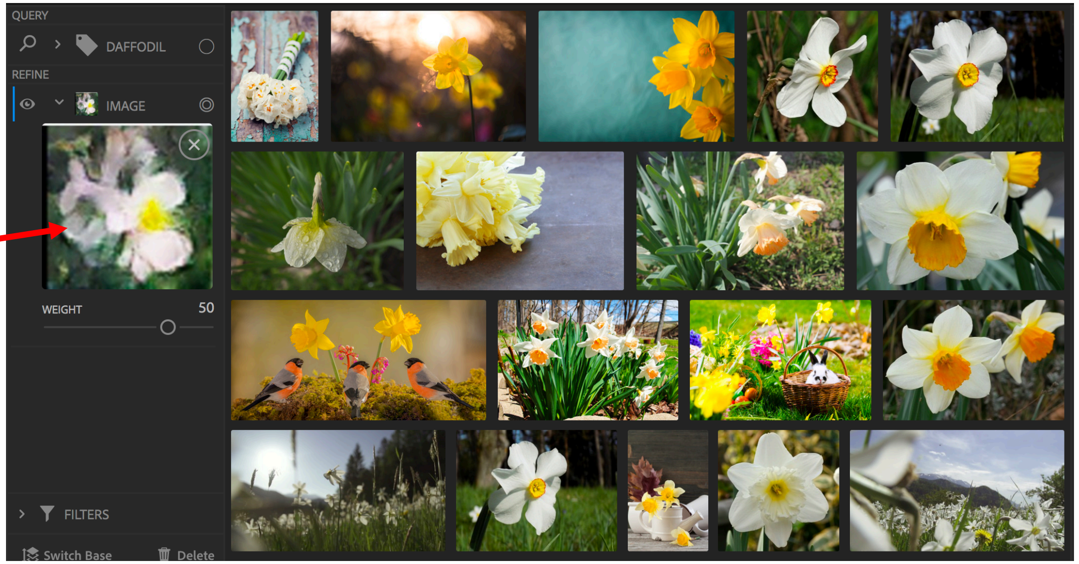


Image Similarity



WEIGHT 50

Conclusion

- Shown (manually) it's possible to connect iGAN with text-conditional GAN (TC-GAN)
- <u>However</u>:
  - TC-GAN still very **<u>poor</u>** discriminating using text (maybe overfitting/model collapse?)
  - Pretrained w2v used in TC-GAN seem very noisy
  - Work-in-progress integrating iGAN and TC-GAN
- Further Analyses:
  - Ablation study whether TC-GAN narrows iGAN choices within category or w2v
  - Measure diversity of generated images
- Future Steps:
  - Improve TC-GAN: Focus on categories rather than text (w2v)
  - TC-GAN and iGAN as one single model