# **Style Transfer**

In this notebook we will implement the style transfer technique from "Image Style Transfer Using Convolutional Neural Networks" (Gatys et al., CVPR 2015).

The general idea is to take two images, and produce a new image that reflects the content of one but the artistic "style" of the other. We will do this by first formulating a loss function that matches the content and style of each respective image in the feature space of a deep network, and then performing gradient descent on the pixels of the image itself.

The deep network we use as a feature extractor is <u>SqueezeNet</u>, a small model that has been trained on ImageNet. You could use any network, but we chose SqueezeNet here for its small size and efficiency.

Here's an example of the images you'll be able to produce by the end of this notebook:

## Setup

#### In [1]:

```
%load ext autoreload
%autoreload 2
from scipy.misc import imread, imresize
import numpy as np
from scipy.misc import imread
import matplotlib.pyplot as plt
# Helper functions to deal with image preprocessing
from cs231n.image utils import load image, preprocess image,
deprocess image
%matplotlib inline
def get session():
    """Create a session that dynamically allocates memory."""
https://www.tensorflow.org/tutorials/using gpu#allowing gpu memory growth
    config = tf.ConfigProto()
    config.gpu options.allow growth = True
    session = tf.Session(config=config)
   return session
def rel error(x,y):
   return np.max(np.abs(x - y) / (np.maximum(1e-8, np.abs(x) + np.abs(y))))
# Older versions of scipy.misc.imresize yield different results
# from newer versions, so we check to make sure scipy is up to date.
def check scipy():
    import scipy
    version = scipy. version .split('.')
```

```
if int(version[0]) < 1:
    assert int(version[1]) >= 16, "You must install SciPy >= 0.16.0 to
complete this notebook."

check_scipy()
```

Load the pretrained SqueezeNet model. This model has been ported from PyTorch, see cs231n/classifiers/squeezenet.py for the model architecture.

To use SqueezeNet, you will need to first download the weights by descending into the cs231n/datasets directory and running get\_squeezenet\_tf.sh. Note that if you ran get assignment3 data.sh then SqueezeNet will already be downloaded.

```
In [2]:
from cs231n.classifiers.squeezenet import SqueezeNet
import tensorflow as tf
import os
tf.reset default graph() # remove all existing variables in the graph
sess = get session() # start a new Session
# Load pretrained SqueezeNet model
SAVE PATH = 'cs231n/datasets/squeezenet.ckpt'
if not os.path.exists(SAVE PATH + ".index"):
    raise ValueError("You need to download SqueezeNet!")
model = SqueezeNet(save path=SAVE PATH, sess=sess)
# Load data for testing
content img test = preprocess image(load image('styles/tubingen.jpg', size=
192))[None]
style img test = preprocess image(load image('styles/starry night.jpg', siz
e=192))[None]
answers = np.load('style-transfer-checks-tf.npz')
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/h5py/__init__.py:34: FutureWa
rning: Conversion of the second argument of issubdtype from `float` to `np.
floating` is deprecated. In future, it will be treated as `np.float64 == np
.dtype(float).type`.
  from . conv import register converters as register converters
INFO:tensorflow:Restoring parameters from cs231n/datasets/squeezenet.ckpt
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:482: Fu
tureWarning: Conversion of the second argument of issubdtype from `int` to
`np.signedinteger` is deprecated. In future, it will be treated as `np.int6
4 == np.dtype(int).type`.
 if issubdtype(ts, int):
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:485: Fu
tureWarning: Conversion of the second argument of issubdtype from `float` t
o `np.floating` is deprecated. In future, it will be treated as `np.float64
== np.dtype(float).type`.
  elif issubdtype(type(size), float):
```

## **Computing Loss**

We're going to compute the three components of our loss function now. The loss function is a weighted sum of three terms: content loss + style loss + total variation loss. You'll fill in the functions that compute these weighted terms below.

## **Content loss**

We can generate an image that reflects the content of one image and the style of another by incorporating both in our loss function. We want to penalize deviations from the content of the content image and deviations from the style of the style image. We can then use this hybrid loss function to perform gradient descent **not on the parameters** of the model, but instead **on the pixel values** of our original image.

Let's first write the content loss function. Content loss measures how much the feature map of the generated image differs from the feature map of the source image. We only care about the content representation of one layer of the network (say, layer  $\ell$ ), that has feature maps  $A^{\ell} \in \mathbb{R}^{1 \times H_{\ell} \times W_{\ell} \times C_{\ell}}$ .  $C_{\ell}$  is the number of filters/channels in layer  $\ell$ ,  $H_{\ell}$  and  $W_{\ell}$  are the height and width. We will work with reshaped versions of these feature maps that combine all spatial positions into one dimension. Let  $F^{\ell} \in \mathbb{R}^{M_{\ell} \times C_{\ell}}$  be the feature map for the current image and  $P^{\ell} \in \mathbb{R}^{M_{\ell} \times C_{\ell}}$  be the feature map for the content source image where  $M_{\ell} = H_{\ell} \times W_{\ell}$  is the number of elements in each feature map. Each row of  $F^{\ell}$  or  $P^{\ell}$  represents the vectorized activations of a particular filter, convolved over all positions of the image. Finally, let  $w_{\ell}$  be the weight of the content loss term in the loss function.

Then the content loss is given by:

```
L_c = w_c \times \sum_{i,j} (F_{ij}^{\ell} - P_{ij}^{\ell})^2
```

In [3]:

```
def content_loss(content_weight, content_current, content_original):
    """
    Compute the content loss for style transfer.

Inputs:
    - content_weight: scalar constant we multiply the content_loss by.
    - content_current: features of the current image, Tensor with shape [1, height, width, channels]
    - content_target: features of the content image, Tensor with shape [1, height, width, channels]

    Returns:
    - scalar content loss
    """
    #pass
    content_loss =
content_weight*tf.reduce_sum((content_current-content_original)**2)
    return content_loss
```

Test your content loss. You should see errors less than 0.0001.

```
In [4]:
```

```
def content_loss_test(correct):
    content_layer = 3
    content_weight = 6e-2
```

```
content_weight = Ge 2
    c_feats = sess.run(model.extract_features()[content_layer], {model.imag}
e: content_img_test})
    bad_img = tf.zeros(content_img_test.shape)
    feats = model.extract_features(bad_img)[content_layer]
    student_output = sess.run(content_loss(content_weight, c_feats, feats))
    error = rel_error(correct, student_output)
    print('Maximum error is {:.3f}'.format(error))
```

Maximum error is 0.000

## Style loss

Now we can tackle the style loss. For a given layer  $\ell$ , the style loss is defined as follows:

First, compute the Gram matrix G which represents the correlations between the responses of each filter, where F is as above. The Gram matrix is an approximation to the covariance matrix -- we want the activation statistics of our generated image to match the activation statistics of our style image, and matching the (approximate) covariance is one way to do that. There are a variety of ways you could do this, but the Gram matrix is nice because it's easy to compute and in practice shows good results.

Given a feature map  $F^{\ell}$  of shape  $(M_{\ell}, C_{\ell})$ , the Gram matrix has shape  $(C_{\ell}, C_{\ell})$  and its elements are given by:

$$G_{ij}^{\ell} = \sum_{k} F_{ki}^{\ell} F_{kj}^{\ell}$$

Assuming  $G^{\ell}$  is the Gram matrix from the feature map of the current image,  $A^{\ell}$  is the Gram Matrix from the feature map of the source style image, and  $w_{\ell}$  a scalar weight term, then the style loss for the layer  $\ell$  is simply the weighted Euclidean distance between the two Gram matrices:

$$\sum_{L_s^{\ell} = w_{\ell}^{i,j} \left( G_{ij}^{\ell} - A_{ij}^{\ell} \right)^2}$$

In practice we usually compute the style loss at a set of layers L rather than just a single layer  $\ell$ ; then the total style loss is the sum of style losses at each layer:

$$\sum_{L_s = \ell \in \mathcal{L}L_s^\ell}$$

Begin by implementing the Gram matrix computation below:

## In [9]:

```
Returns:
    - gram: Tensor of shape (C, C) giving the (optionally normalized)
    Gram matrices for the input image.
"""

#pass
shapes = tf.shape(features)
#Reshape the features tensor from [1, H, W, C] to [H*W, C]
features_reshaped = tf.reshape(features,[shapes[1]*shapes[2],shapes[3]])
gram = tf.matmul(tf.transpose(features_reshaped), features_reshaped)

if normalize == True:
    gram/=tf.cast(shapes[1]*shapes[2]*shapes[3],tf.float32)
return gram
```

Test your Gram matrix code. You should see errors less than 0.0001.

#### In [10]:

```
def gram_matrix_test(correct):
    gram = gram_matrix(model.extract_features()[5])
    student_output = sess.run(gram, {model.image: style_img_test})
    error = rel_error(correct, student_output)
    print('Maximum error is {:.3f}'.format(error))
gram_matrix_test(answers['gm_out'])
```

Maximum error is 0.000

### Next, implement the style loss:

### In [11]:

```
def style loss(feats, style layers, style targets, style weights):
    Computes the style loss at a set of layers.
   Inputs:
    - feats: list of the features at every layer of the current image, as p
roduced by
      the extract features function.
    - style layers: List of layer indices into feats giving the layers to i
nclude in the
     style loss.
    - style targets: List of the same length as style layers, where style t
      a Tensor giving the Gram matrix of the source style image computed at
      layer style layers[i].
    - style weights: List of the same length as style layers, where style w
eights[i]
     is a scalar giving the weight for the style loss at layer style layer
s[i].
   Returns:
    - style loss: A Tensor containing the scalar style loss.
    # Hint: you can do this with one for loop over the style layers, and sh
```

```
ould
    # not be very much code (~5 lines). You will need to use your gram_matr
ix function.
    #pass
    Loss = 0
    for index, layer in enumerate(style_layers):
        Loss +=
style_weights[index]*2*tf.nn.l2_loss(gram_matrix(feats[layer]) -
style_targets[index])
    return Loss
```

Test your style loss implementation. The error should be less than 0.0001.

#### In [12]:

Error is 0.000

## **Total-variation regularization**

It turns out that it's helpful to also encourage smoothness in the image. We can do this by adding another term to our loss that penalizes wiggles or "total variation" in the pixel values.

You can compute the "total variation" as the sum of the squares of differences in the pixel values for all pairs of pixels that are next to each other (horizontally or vertically). Here we sum the total-variation regualarization for each of the 3 input channels (RGB), and weight the total summed loss by the total variation weight,  $w_r$ :

$$L_{tv} = w_t \times \left( \sum_{c=1}^{3} \sum_{i=1}^{H-1} \sum_{j=1}^{W} (x_{i+1,j,c} - x_{i,j,c})^2 + \sum_{c=1}^{3} \sum_{i=1}^{H} \sum_{j=1}^{W-1} (x_{i,j+1,c} - x_{i,j,c})^2 \right)$$

In the next cell, fill in the definition for the TV loss term. To receive full credit, your implementation should not have any loops.

#### In [17]:

```
def tv_loss(img, tv_weight):
    """

Compute total variation loss.
```

```
Inputs:
    - img: Tensor of shape (1, H, W, 3) holding an input image.
    - tv_weight: Scalar giving the weight w_t to use for the TV loss.

Returns:
    - loss: Tensor holding a scalar giving the total variation loss for img weighted by tv_weight.

"""

# Your implementation should be vectorized and not require any loops!

#pass

H_dim_variation = tf.reduce_sum((img[:,1:,:,:] - img[:,:-1,:,:])**2)

W_dim_variation = tf.reduce_sum((img[:,1:,:,:] - img[:,:,:-1,:])**2)

total_loss = tv_weight*(H_dim_variation + W_dim_variation)

return total_loss
```

Test your TV loss implementation. Error should be less than 0.0001.

```
In [18]:
```

```
def tv_loss_test(correct):
    tv_weight = 2e-2
    t_loss = tv_loss(model.image, tv_weight)
    student_output = sess.run(t_loss, {model.image: content_img_test})
    error = rel_error(correct, student_output)
    print('Error is {:.3f}'.format(error))
tv_loss_test(answers['tv_out'])
```

Error is 0.000

## **Style Transfer**

Lets put it all together and make some beautiful images! The style\_transfer function below combines all the losses you coded up above and optimizes for an image that minimizes the total loss.

### In [19]:

```
def style transfer(content image, style image, image size, style size, cont
ent layer, content weight,
                   style layers, style weights, tv weight, init random = Fal
se):
    """Run style transfer!
    Inputs:
    - content image: filename of content image
    - style image: filename of style image
    - image size: size of smallest image dimension (used for content loss a
nd generated image)
    - style_size: size of smallest style image dimension
    - content layer: layer to use for content loss
    - content weight: weighting on content loss
    - style layers: list of layers to use for style loss
    - style weights: list of weights to use for each layer in style layers
    - tv weight: weight of total variation regularization term
    - init random: initialize the starting image to uniform random noise
```

```
# Extract features from the content image
    content img = preprocess image(load image(content image,
size=image size))
    feats = model.extract features(model.image)
    content target = sess.run(feats[content layer],
                              {model.image: content img[None]})
    # Extract features from the style image
    style imq = preprocess image(load image(style image, size=style size))
    style feat vars = [feats[idx] for idx in style layers]
    style target vars = []
    # Compute list of TensorFlow Gram matrices
    for style feat var in style feat vars:
        style target vars.append(gram matrix(style feat var))
    # Compute list of NumPy Gram matrices by evaluating the TensorFlow grap
h on the style image
    style_targets = sess.run(style_target_vars, {model.image: style_img[Non
e]})
    # Initialize generated image to content image
    if init random:
        img var = tf.Variable(tf.random uniform(content img[None].shape, 0,
1), name="image")
    else:
        img var = tf.Variable(content img[None], name="image")
    # Extract features on generated image
    feats = model.extract features(img var)
    # Compute loss
    c loss = content loss(content weight, feats[content layer],
content target)
    s loss = style loss(feats, style layers, style targets, style weights)
    t_loss = tv_loss(img_var, tv_weight)
    loss = c loss + s loss + t loss
    # Set up optimization hyperparameters
    initial_lr = 3.0
    decayed lr = 0.1
    decay lr at = 180
    max iter = 200
    # Create and initialize the Adam optimizer
    lr var = tf.Variable(initial lr, name="lr")
    # Create train op that updates the generated image when run
    with tf.variable scope("optimizer") as opt scope:
        train op = tf.train.AdamOptimizer(lr var).minimize(loss, var list=[i
mg var])
    # Initialize the generated image and optimization variables
    opt vars = tf.get collection(tf.GraphKeys.GLOBAL VARIABLES,
scope=opt scope.name)
    sess.run(tf.variables initializer([lr var, img var] + opt vars))
    # Create an op that will clamp the image values when run
    clamp image op = tf.assign(img var, tf.clip by value(img var, -1.5, 1.5)
    f, axarr = plt.subplots(1,2)
    axarr[0].axis('off')
    axarr[1].axis('off')
    axarr[0].set_title('Content Source Img.')
```

```
axarr[1].set title('Style Source Img.')
axarr[0].imshow(deprocess image(content img))
axarr[1].imshow(deprocess image(style img))
plt.show()
plt.figure()
# Hardcoded handcrafted
for t in range(max iter):
    # Take an optimization step to update img var
    sess.run(train op)
    if t < decay lr at:</pre>
        sess.run(clamp image op)
    if t == decay lr at:
        sess.run(tf.assign(lr var, decayed lr))
    if t % 100 == 0:
        print('Iteration {}'.format(t))
        img = sess.run(img var)
        plt.imshow(deprocess image(img[0], rescale=True))
        plt.axis('off')
        plt.show()
print('Iteration {}'.format(t))
img = sess.run(img var)
plt.imshow(deprocess image(img[0], rescale=True))
plt.axis('off')
plt.show()
```

## Generate some pretty pictures!

Try out style\_transfer on the three different parameter sets below. Make sure to run all three cells. Feel free to add your own, but make sure to include the results of style transfer on the third parameter set (starry night) in your submitted notebook.

- The content image is the filename of content image.
- The style image is the filename of style image.
- The image\_size is the size of smallest image dimension of the content image (used for content loss and generated image).
- The style size is the size of smallest style image dimension.
- The content layer specifies which layer to use for content loss.
- The content\_weight gives weighting on content loss in the overall loss function.

  Increasing the value of this parameter will make the final image look more realistic (closer to the original content).
- style layers specifies a list of which layers to use for style loss.
- style\_weights specifies a list of weights to use for each layer in style\_layers (each of
  which will contribute a term to the overall style loss). We generally use higher weights for the
  earlier style layers because they describe more local/smaller scale features, which are more
  important to texture than features over larger receptive fields. In general, increasing these
  weights will make the resulting image look less like the original content and more distorted
  towards the appearance of the style image.
- tv\_weight specifies the weighting of total variation regularization in the overall loss function. Increasing this value makes the resulting image look smoother and less jagged, at the cost of lower fidelity to style and content.

Below the next three cells of code (in which you shouldn't change the hyperparameters), feel free to

copy and paste the parameters to play around them and see how the resulting image changes.

#### In [20]:

```
# Composition VII + Tubingen
params1 = {
    'content image' : 'styles/tubingen.jpg',
    'style image' : 'styles/composition vii.jpg',
    'image_size' : 192,
    'style size' : 512,
    'content layer' : 3,
    'content_weight' : 5e-2,
    'style layers' : (1, 4, 6, 7),
    'style weights' : (20000, 500, 12, 1),
    'tv weight' : 5e-2
style transfer(**params1)
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:482: Fu
tureWarning: Conversion of the second argument of issubdtype from `int` to
`np.signedinteger` is deprecated. In future, it will be treated as `np.int6
4 == np.dtype(int).type`.
 if issubdtype(ts, int):
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:485: Fu
tureWarning: Conversion of the second argument of issubdtype from `float` t
o `np.floating` is deprecated. In future, it will be treated as `np.float64
== np.dtype(float).type`.
  elif issubdtype(type(size), float):
```

### Content Source Img.



Style Source Img.



### Iteration 0





#### Iteration 199



### In [21]:

```
# Scream + Tubingen
params2 = {
    'content image':'styles/tubingen.jpg',
    'style image':'styles/the scream.jpg',
    'image size':192,
    'style size':224,
    'content layer':3,
    'content weight':3e-2,
    'style layers':[1, 4, 6, 7],
    'style weights':[200000, 800, 12, 1],
    'tv weight':2e-2
style transfer(**params2)
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:482: Fu
tureWarning: Conversion of the second argument of issubdtype from `int` to
`np.signedinteger` is deprecated. In future, it will be treated as `np.int6
4 == np.dtype(int).type`.
 if issubdtype(ts, int):
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:485: Fu
tureWarning: Conversion of the second argument of issubdtype from `float` t
o `np.floating` is deprecated. In future, it will be treated as `np.float64
== np.dtype(float).type`.
```

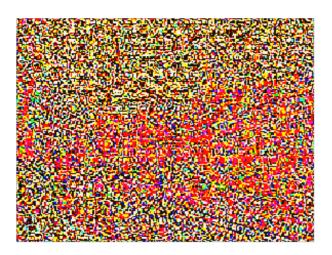
# elif issubdtype(type(size), float):

Content Source Img.





Iteration 0



Iteration 100



Iteration 199





#### In [22]:

```
# Starry Night + Tubingen
params3 = {
    'content_image' : 'styles/tubingen.jpg',
    'style image' : 'styles/starry night.jpg',
    'image size' : 192,
    'style size' : 192,
    'content_layer' : 3,
    'content weight': 6e-2,
    'style layers' : [1, 4, 6, 7],
    'style weights' : [300000, 1000, 15, 3],
    'tv weight' : 2e-2
style_transfer(**params3)
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:482: Fu
tureWarning: Conversion of the second argument of issubdtype from `int` to
`np.signedinteger` is deprecated. In future, it will be treated as `np.int6
```

4 == np.dtype(int).type`.

if issubdtype(ts, int):

/Users/adele/.pyenv/versions/anaconda3-

4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:485: Fu tureWarning: Conversion of the second argument of issubdtype from `float` t o `np.floating` is deprecated. In future, it will be treated as `np.float64 == np.dtype(float).type`.

elif issubdtype(type(size), float):

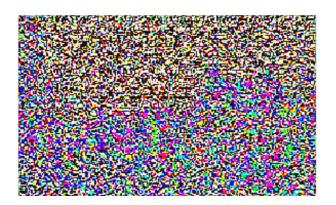
Content Source Img.



Style Source Img.



#### Iteration 0





Iteration 100



Iteration 199



## **Feature Inversion**

The code you've written can do another cool thing. In an attempt to understand the types of features that convolutional networks learn to recognize, a recent paper [1] attempts to reconstruct an image from its feature representation. We can easily implement this idea using image gradients from the pretrained network, which is exactly what we did above (but with two different feature representations).

Now, if you set the style weights to all be 0 and initialize the starting image to random noise instead of the content source image, you'll reconstruct an image from the feature representation of the content source image. You're starting with total noise, but you should end up with something that looks quite a bit like your original image.

(Similarly, you could do "texture synthesis" from scratch if you set the content weight to 0 and initialize the starting image to random noise, but we won't ask you to do that here.)

Run the following cell to try out feature inversion.

[1] Aravinon Manendran, Andrea Vedaldi, Understanding Deep image Representations by inverting them", CVPR 2015

### In [23]:

```
# Feature Inversion -- Starry Night + Tubingen
params inv = {
    'content image' : 'styles/tubingen.jpg',
    'style image' : 'styles/starry night.jpg',
    'image size' : 192,
    'style size' : 192,
    'content layer' : 3,
    'content_weight' : 6e-2,
    'style layers' : [1, 4, 6, 7],
    'style weights' : [0, 0, 0, 0], # we discard any contributions from styl
e to the loss
    'tv weight': 2e-2,
    'init random': True # we want to initialize our image to be random
style transfer(**params inv)
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:482: Fu
tureWarning: Conversion of the second argument of issubdtype from `int` to
`np.signedinteger` is deprecated. In future, it will be treated as `np.int6
4 == np.dtype(int).type`.
 if issubdtype(ts, int):
/Users/adele/.pyenv/versions/anaconda3-
4.3.1/envs/cs231n/lib/python3.6/site-packages/scipy/misc/pilutil.py:485: Fu
tureWarning: Conversion of the second argument of issubdtype from `float` t
o `np.floating` is deprecated. In future, it will be treated as `np.float64
== np.dtype(float).type`.
  elif issubdtype(type(size), float):
```

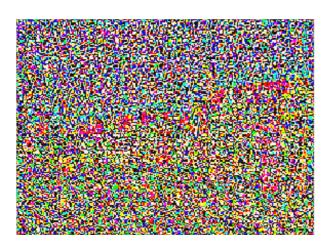
### Content Source Img.



Style Source Img.



Iteration 0



### 

Iteration 100



Iteration 199



In [ ]: