Generating Images from Noise

CS194: Intro to Comp. Vision and Comp. Photo
Alexei Efros, UC Berkeley, Fall 2022
Texture as samples from distribution
Parametric Texture Synthesis

Goal: parametric generative model of the “infinite texture”
Early Vision Texture Models

Heeger & Bergen (1995)
Portilla & Simoncelli (2000)
Multi-scale filter decomposition (steerable pyramid)
Step 1: Convolve with filterbank
Step 2: match per-channel histograms
Step 3: collapse pyramid and repeat!

input

Noise image
Texture Synthesis

Images with equal model response

Portilla & Simoncelli (2000)
Figure 7: (Left pair) Inhomogeneous input texture produces blotchy synthetic texture. (Right pair) Homogeneous input.

Figure 8: Examples of failures: wood grain and red coral.

Figure 9: More failures: hay and marble.
Marginal statistics are not enough
Neighbor filter responses are highly correlated
  - an edge at low-res will cause an edge at high-res
Let's match 2nd order statistics too!

• Match joint histograms of pairs of filter responses at adjacent spatial locations, orientations, and scales.
• Optimize using repeated projections onto statistical constraint surfaces
Convolutional Neural Network Texture Model

Gatys et al. (NIPS 2015)
Texture Synthesis

Image Space

Model Space

Images with equal model response

Portilla & Simoncelli (2000)
CNN - Multiscale Filter Bank

- **conv1_1**
- **pool1**
- **pool2**
- **pool3**
- **pool4**

# features
- 64
- 128
- 256
- 512
CNN - Texture Features

\[ F = [\tilde{f}_1, \tilde{f}_2, \tilde{f}_3, \ldots, \tilde{f}_N]^T \]

\[ G = FF^T \]

\[ \begin{pmatrix} 
\langle \tilde{f}_1, \tilde{f}_1 \rangle & \cdots & \langle \tilde{f}_1, \tilde{f}_N \rangle \\
\langle \tilde{f}_2, \tilde{f}_1 \rangle & \ddots & \vdots \\
\vdots & \ddots & \ddots \\
\langle \tilde{f}_N, \tilde{f}_1 \rangle & \cdots & \langle \tilde{f}_N, \tilde{f}_N \rangle 
\end{pmatrix} \]

\[ \langle \tilde{f}_i, \tilde{f}_j \rangle = \sum_k F_{ik}F_{jk} \]
CNN-Texture Features

Gram Matrices

# features
512
256
128
64
64
Texture Synthesis
Texture Synthesis

$E_L = \sum (G^L - \hat{G}^L)^2$

$G_0 = \sum \hat{G}_0^L \hat{E}_L^k$

$F_L = \sum_{l=0}^{L} \frac{\hat{E}_L^k}{\hat{E}_L^k}$

$F_{L-1}$

$L(\overline{x}, \overline{\hat{x}}) = \sum_{l=0}^{L} w_l E_l$
Texture Synthesis

\[ E_L = \sum (G^L - \tilde{G}^L)^2 \]

\[ E_0 = \sum_{l=0}^L E_0^L E_0^L \]

\[ \frac{\partial E_L}{\partial F^L} \]

\[ E^{L-1} \]

\[ L(\tilde{x}, \hat{x}) = \sum_{i=0}^L w_i E_i \]
Texture Synthesis
Texture Synthesis
Texture Synthesis

\[
E_L = \sum (x^n - x)^2
\]

\[
E_L = \sum L_{x^n x^n}
\]

\[
\hat{t} = \hat{t} - \alpha \frac{\partial \mathcal{L}}{\partial \hat{t}}
\]
Texture Synthesis
Test Julesz’ Conjecture
Test Julesz’ Conjecture
ImageNet Recognition is just Texture Recognition
ImageNet Recognition is just Texture Recognition

Gatys et al., 2017

"Shetland Sheepdog"
A Neural Algorithm of Artistic Style

Gatys, Ecker, Bethge (arXiv 2015)
Van Gogh (1889)
CNN - Texture Synthesis

Gatys et al. (NIPS 2015)
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer
Artistic Style Transfer

\[ \mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{content}} + \beta \mathcal{L}_{\text{style}} \]

\[ E_c = \sum (G^t - G^c)^2 \]

\[ G^t = \sum \hat{p} \hat{p}^t \]

\[ \mathcal{L}_{\text{content}} = \sum (\hat{p} - p)^2 \]

\[ \mathcal{L}_{\text{style}} = \sum w_l E_l \]
Artistic Style Transfer

\[ \mathcal{L}_{total} = \alpha \mathcal{L}_{content} + \beta \mathcal{L}_{style} \]

\[ \mathcal{L}_{content} = \sum (\mathbf{F}^i - \mathbf{F}^c)^2 \]

\[ \mathcal{L}_{style} = \sum_l w_l E_l \]

Artwork images: "Starry Night" by Vincent van Gogh and "A View of le Harve" by Claude Monet.
Artistic Style Transfer

\[ \mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{content}} + \beta \mathcal{L}_{\text{style}} \]

\[ E_C = \sum (\mathbf{G}^c - \mathbf{G}^t)^2 \]

\[ \mathcal{L}_{\text{style}} = \sum_l w_l E_l \]

\[ \mathcal{L}_{\text{content}} = \sum (\mathbf{F}^c - \mathbf{F}^t)^2 \]
Artistic Style Transfer

\[ E_C = \sum (G^c - G^t)^2 \]

\[ E_S = \sum \frac{G^s}{F^s_c F^s_j} \]

\[ \mathcal{L}_{total} = \alpha \mathcal{L}_{content} + \beta \mathcal{L}_{style} \]

\[ \mathcal{L}_{content} = \sum (\hat{I} - I)^2 \]

\[ \mathcal{L}_{style} = \sum_l w_l E_{l} \]

\[ \frac{\partial \mathcal{L}}{\partial \hat{I}} \] Gradient descent

\[ \hat{I} := \hat{I} - \alpha \frac{\partial \mathcal{L}}{\partial \hat{I}} \]
Artistic Style Transfer

\[ L_{total} = \alpha L_{content} + \beta L_{style} \]

\[ E_c = \sum (G^c - G^t)^2 \]

\[ G^c = \sum \hat{F}_{i}^c \hat{F}_{i}^t \]

\[ \hat{F}_{i}^t = \frac{\partial E_c}{\partial F_i} \]

\[ \hat{F}_{i}^{t-1} = \frac{\partial E_c}{\partial F_i^{t-1}} \]

\[ L_{content} = \sum (\hat{F}^c - \hat{F}^t)^2 \]

\[ \hat{L} = \frac{\partial C}{\partial \hat{L}} \]

\[ \hat{L} = \hat{L} - \alpha \frac{\partial C}{\partial \hat{L}} \]

\[ \hat{C} = \hat{C} - \alpha \frac{\partial C}{\partial \hat{C}} \]
Artistic Style Transfer

\[ \mathcal{L}_{total} = \alpha \mathcal{L}_{content} + \beta \mathcal{L}_{style} \]

\[ \mathcal{L}_{content} = \sum (\| \hat{F}^i - F^i \|^2) \]

\[ \mathcal{L}_{style} = \sum_l w_l F_l \]

Gradient descent:

\[ \hat{\theta} := \hat{\theta} - \alpha \frac{\partial \mathcal{L}}{\partial \theta} \]
Relative Weighting of Content and Style

1e-4

1e-3

1e-2

1e-1
Different Reconstruction Layers

Conv2_2

Conv4_2
Different Reconstruction Layers

Conv2_2

Conv4_2
Different Reconstruction Layers

Original  Conv2_2  Conv4_2
General Style Transfer
General Style Transfer
GANs as Texture Synthesis?

- **Conjecture:** GANs might be learning the “right” features to match for natural images.
Projecting onto the “Image Manifold”

Image Space

Model Space

Images with equal model response

Portilla & Simoncelli (2000)
Imagen

diffusion

DALL-E 2

By Steve Seitz
Generate 100 images
Generate 100 images

slide from Steve Seitz's video
Generate 100 images of raspberries
easy random images

hard raspberry images

slide from Steve Seitz's video
random images → hard → raspberry images

slide from Steve Seitz’s video
easy

random images

raspberry images
random images

raspberry images

slide from Steve Seitz’s video
random images → raspberry images
random images

raspberry images

slide from Steve Seitz's video
random images

raspberry images

slide from Steve Seitz’s video
random images → raspberry images

slide from Steve Seitz’s video
random images

raspberry images

random images

slide from Steve Seitz's video
random images → diffusion neural network → raspberry images
random images

diffusion neural network

raspberry images

Training

slide from Steve Seitz's video
Training
slide from Steve Seitz's video
random images — diffusion neural network — raspberry images

Training

slide from Steve Seitz's video
random images \rightarrow \text{diffusion neural network} \rightarrow \text{raspberry images}
random images -> diffusion neural network -> raspberry images

Training slide from Steve Seitz's video
random images → diffusion neural network → raspberry images

slide from Steve Seitz's video
random images

raspberry images

diffusion neural network

slide from Steve Seitz's video
random images

diffusion neural network

raspberry images

slide from Steve Seitz’s video
random images

diffusion neural network

raspberry images

slide from Steve Seitz's video
random images \rightarrow diffusion neural network \rightarrow raspberry images
random images → diffusion neural network → raspberry images

slide from Steve Seitz’s video
random images → diffusion neural network → raspberry images

slide from Steve Seitz’s video
random images → diffusion neural network → raspberry images

slide from Steve Seitz's video
random images \rightarrow \text{diffusion neural network} \rightarrow \text{raspberry images}
random images -> diffusion neural network -> raspberry images
random images → diffusion neural network → raspberry images
random images

diffusion neural network

raspberry images

slide from Steve Seitz's video
raspberry
descriptions of these images:

- peach
- apple
- mango

random images

slide from Steve Seitz's video
Slide from Steve Seitz’s video

- Peach images
- Apple images
- Raspberry images
- Mango images
- Random images
- Diffusion neural network

“Apple”
raspberry beret

diffusion neural network

slide from Steve Seitz's video
Large Language Model -> diffusion neural network

raspberry beret

slide from Steve Seitz's video
raspberry beret
raspberry beret

slide from Steve Seitz’s video
beret of raspberries
beret of raspberries
beret of raspberries
slide from Steve Seitz’s video
chocolate guacamole pancakes

slide from Steve Seitz’s video
squirrel inside a nutshell
An astronaut riding a horse in a photorealistic style (Dall-E 2)
slide from Steve Seitz’s video
Language Generator (+ pixels)

Diffusion (+ language)

Parti

Imagen

slide from Steve Seitz’s video
Language Generator + Diffusion

Dall-E 2

slide from Steve Seitz's video
Parti
slide from Steve Seitz’s video
Squirrel reaching for a nut. Latte art slide from Steve Seitz's video.
A teddy bear making chocolate guacamole pancakes

slide from Steve Seitz’s video
A dog looks curious in the mirror, seeing a cat.
A bowl of soup that looks like a monster knitted out of wool
Impressive compositionality:

DALL-E + Danielle Baskin
“Person holding a heavy box”
“Person holding a laptop”
“Person holding birthday cake”
“Person holding a watermelon”
“Person holding a self-portrait”
“Person holding a photo of himself”
“Person holding a painting of himself”
Language seems the weakest part...
“Bag of words” seems to work as well
five grapes
three grapes
chair with three legs
photo of bart train with wings
fish that has human hands
person with hair of snakes
whatever you do, don't draw a purple elephant
Luxembourg palace with triangular windows
Luxembourg palace with circle windows