

Style and Content, Texture Synthesis

Jun-Yan Zhu

16-726, Spring 2023

Collection Style Transfer



Photograph ©Alexei Efros



Monet



Van Gogh



Cezanne



Ukiyo-e

Style and Content Separation

A

Classification

A	B	C	D	E
A	B	C	D	E
A	B	C	D	E
A	B	C	D	E
A	B	C	D	E
B	C	A	E	D

Domain Adaptation

B

Extrapolation

A	B	C	D	E
A	B	C	D	E
A	B	C	D	E
A	B	C	D	E
A	B	C	D	E
?	?	C	D	E

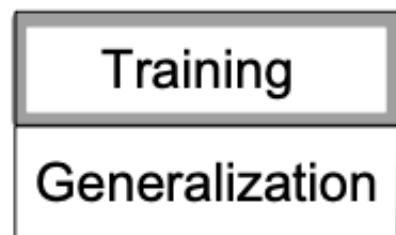
Paired Image-to-Image Translation

C

Translation

A	B	C	D	E	?	?	?
A	B	C	D	E			
A	B	C	D	E			
A	B	C	D	E			
A	B	C	D	E	?	?	?
?				?	F	G	H

Unpaired Image-to-Image Translation



Separating Style and Content
[Tenenbaum and Freeman 1996]

$$y_k^{sc} = \sum_{i=1}^I \sum_{j=1}^J w_{ijk} a_i^s b_j^c.$$

Style and Content

Adversarial loss

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$



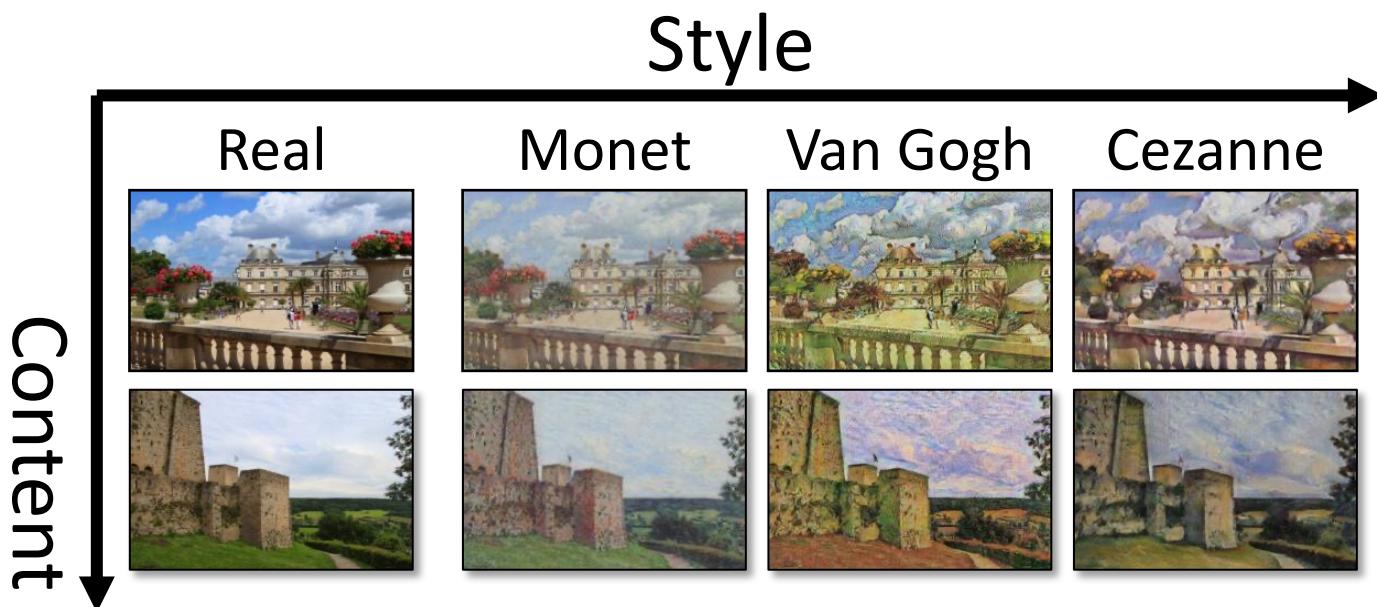
$p(x) \rightarrow p(y)$ change style

Cycle-consistency loss

$$\mathbb{E}_x \|F(G(x)) - x\|_1$$

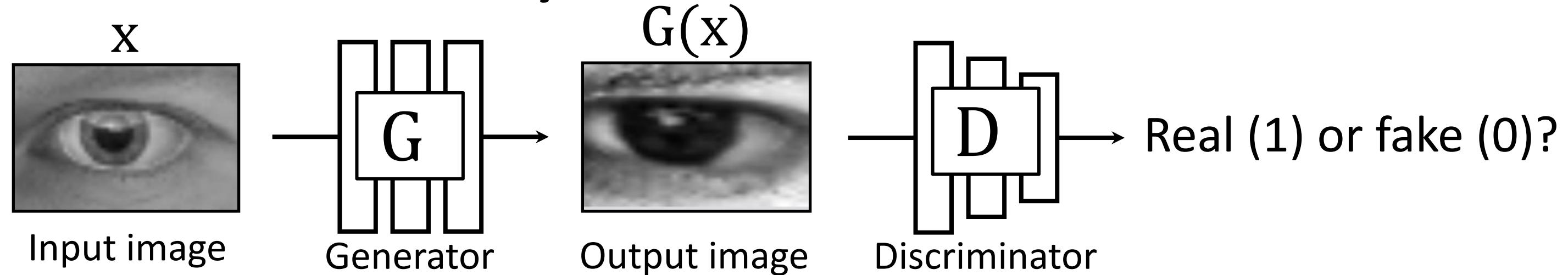


Bidirectional: preserve content



Separating Style and Content
[Tenenbaum and Freeman 1996]

Style and Content

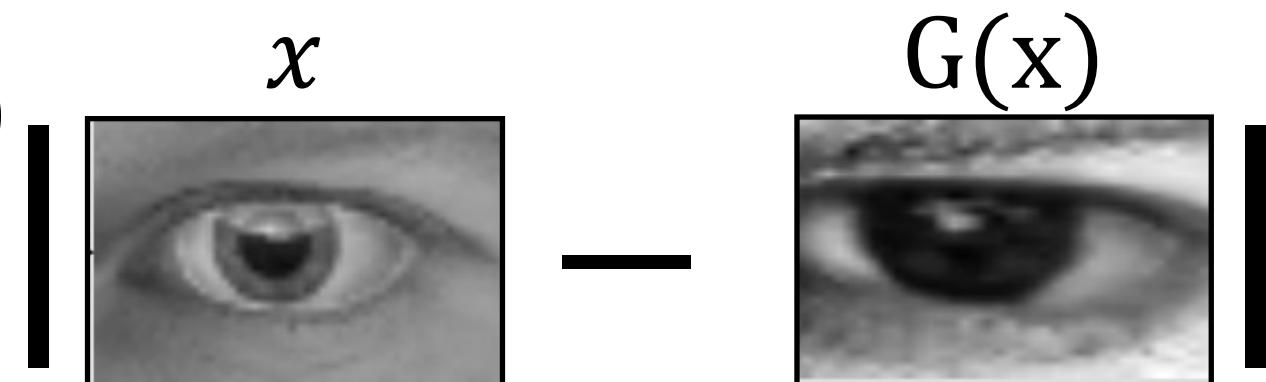


Adversarial loss (change style)

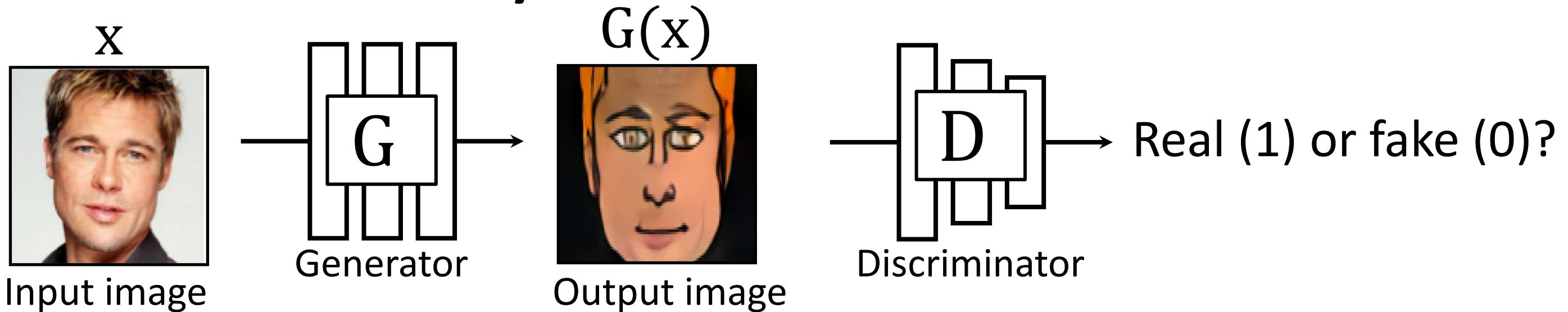
$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

L1 loss (preserve content in pixel space)

$$\mathbb{E}_x \|G(x) - x\|_1$$



Style and Content



Adversarial loss (change style)

$$\mathbb{E}_x \log(1 - D_Y(G(x))) + \mathbb{E}_y \log D_Y(y)$$

Feature loss (Preserve content in feature space)

$$\mathbb{E}_x \|F(G(x)) - F(x)\|$$

$$\|F(\text{Input}) - F(\text{Output})\|$$

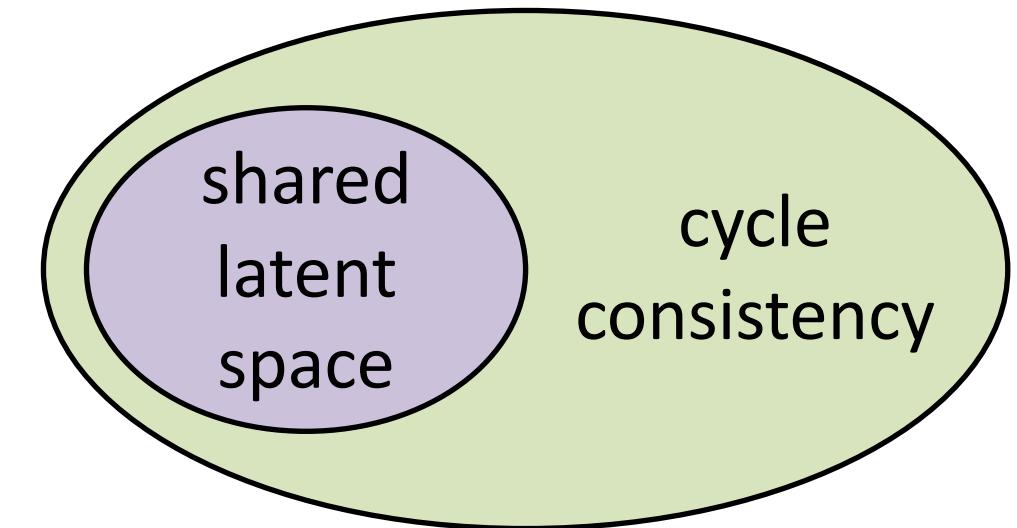
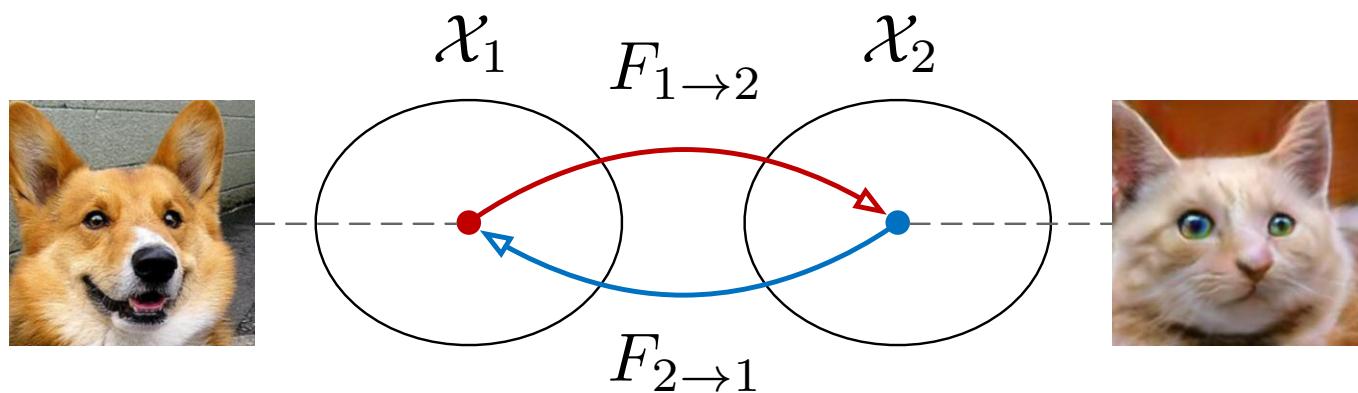
DTN [Taigman et al., 2017]

Style and Content

- Style: domain-specific features
(horse vs. zebra)
- Content: features shared across two domains

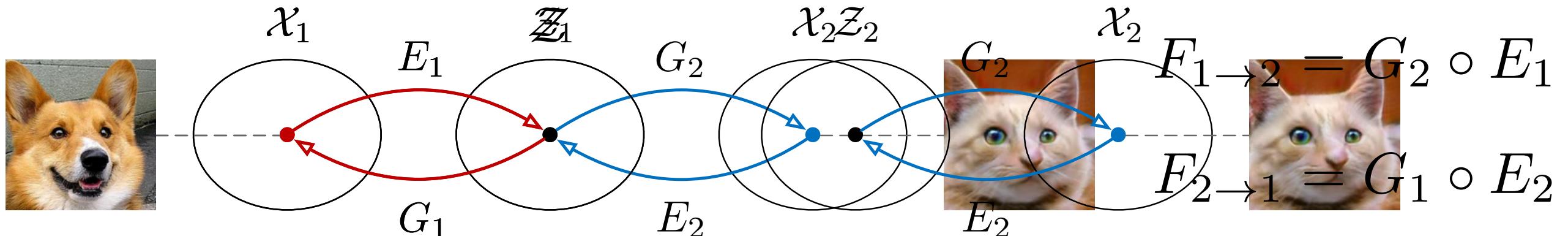
CycleGAN and UNIT

- CycleGAN (**cycle consistency**)



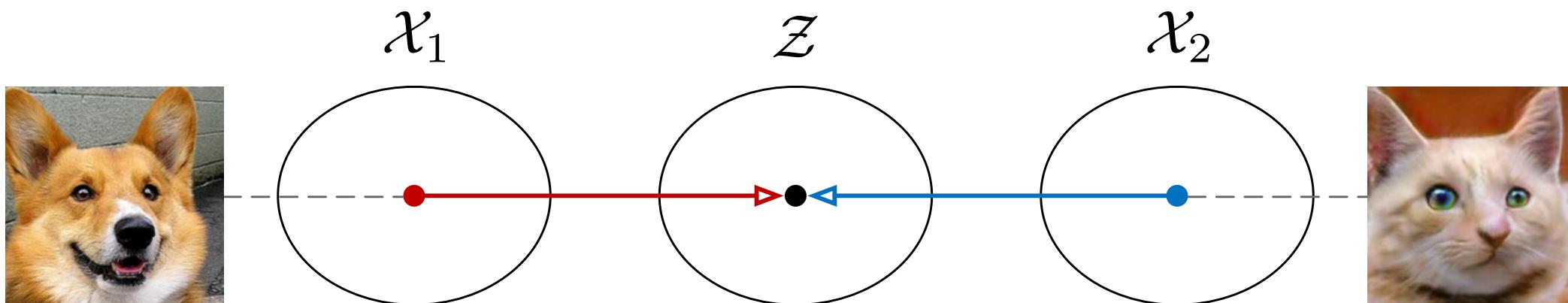
- UNIT (**shared latent space**) [Liu et al. 2017]

shared latent space \Rightarrow cycle consistency



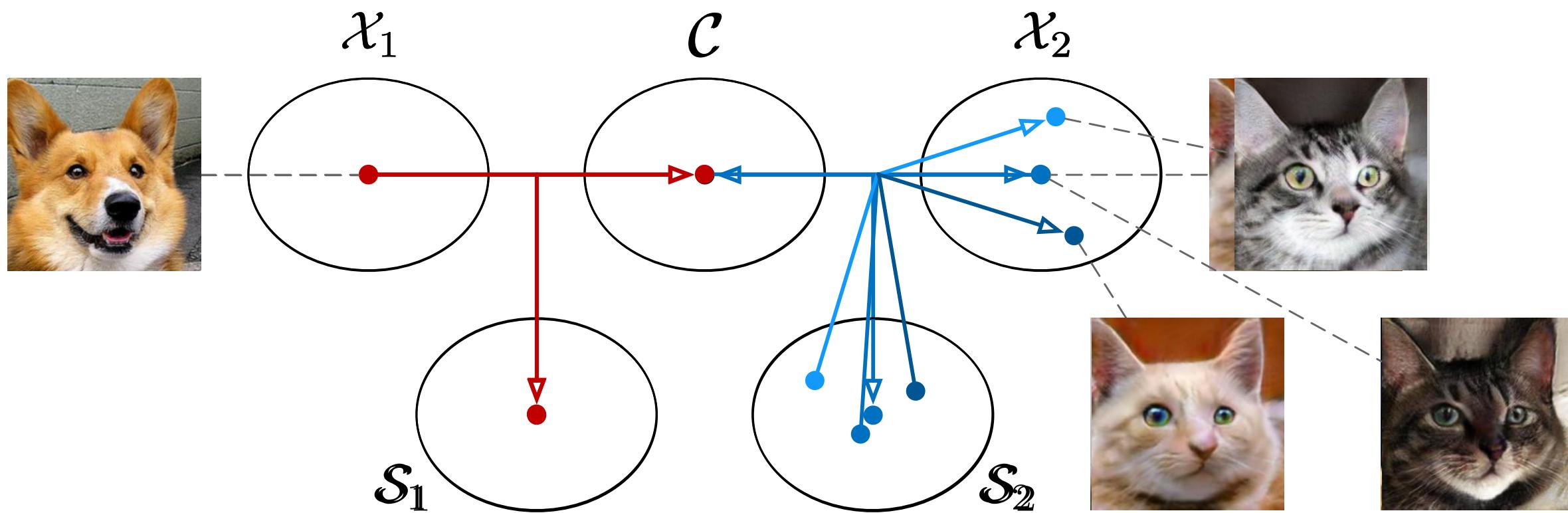
Disentangling the Latent Space

- UNIT
 - A single **shared, domain-invariant** latent space \mathcal{Z}



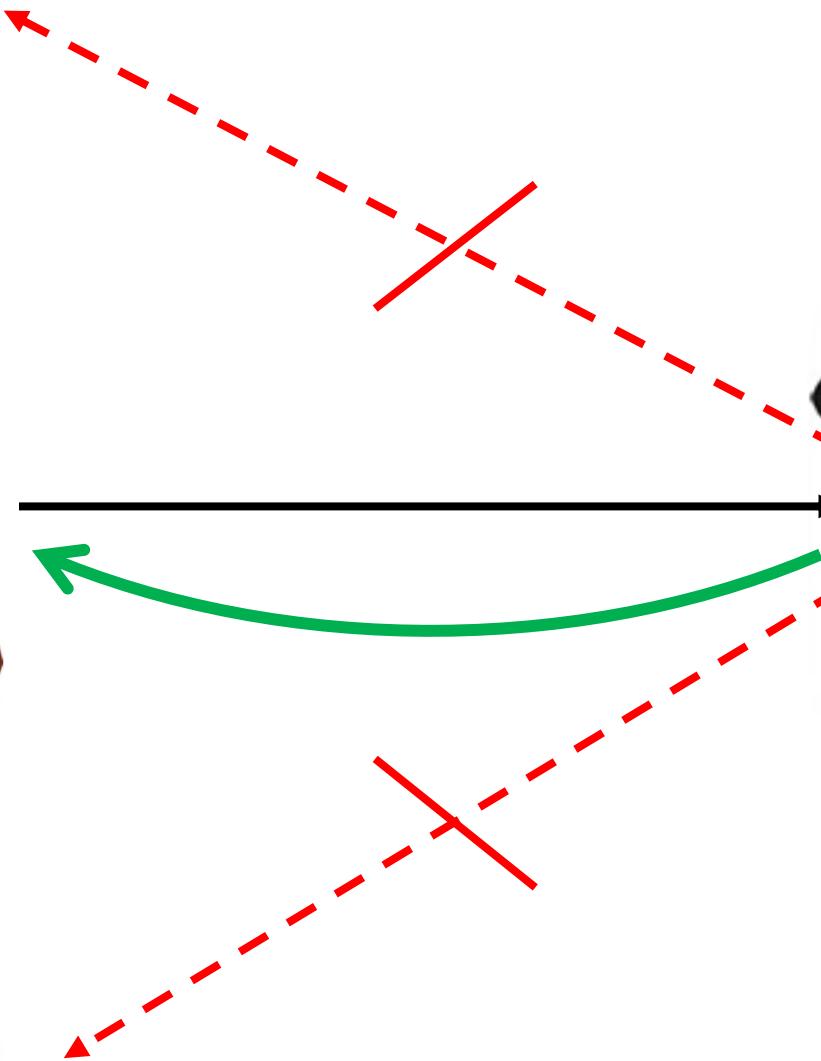
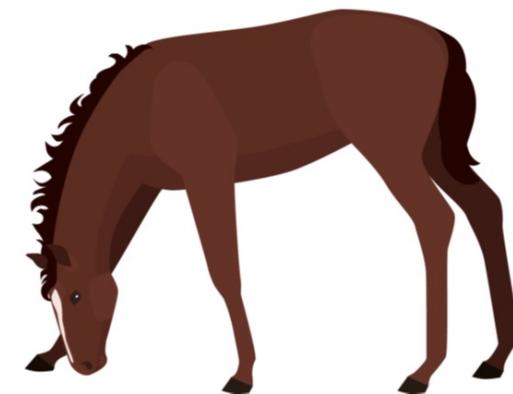
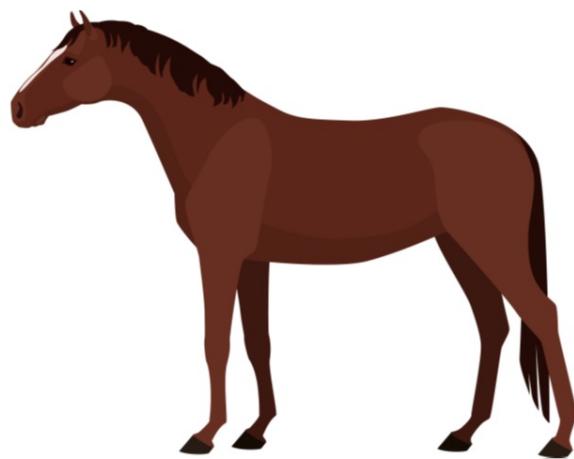
Disentangling the Latent Space

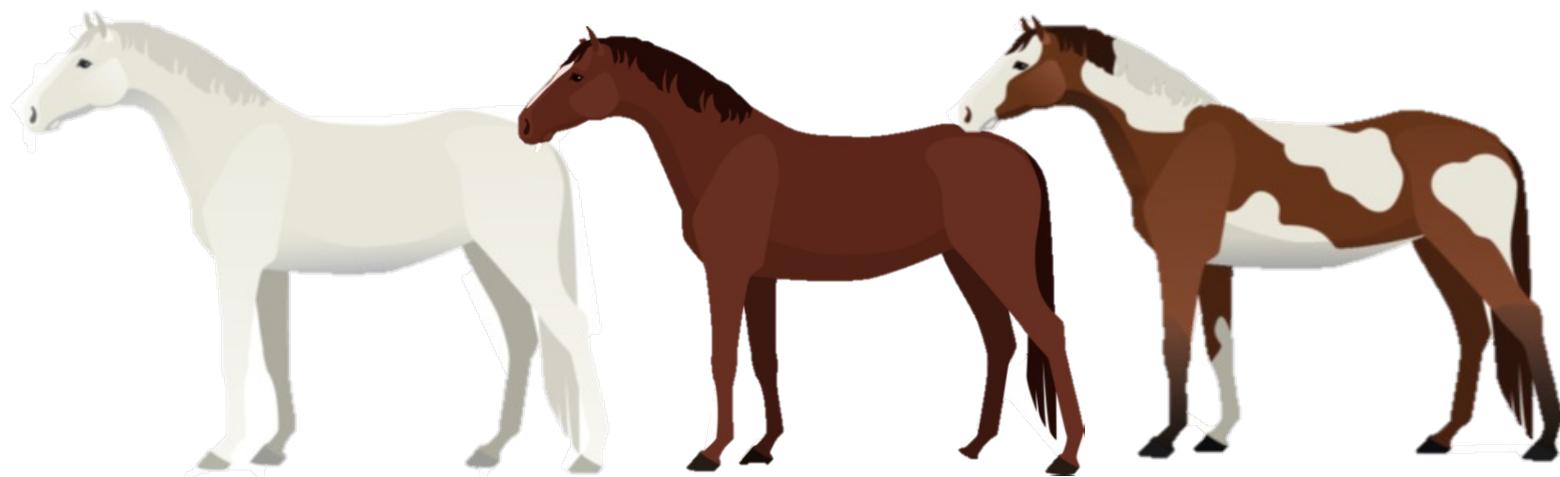
- Multimodal UNIT (MUNIT)
 - A **content** space \mathcal{C} that is **shared, domain-invariant**
 - Two **style** spaces $\mathcal{S}_1, \mathcal{S}_2$ that are **unshared, domain-specific**



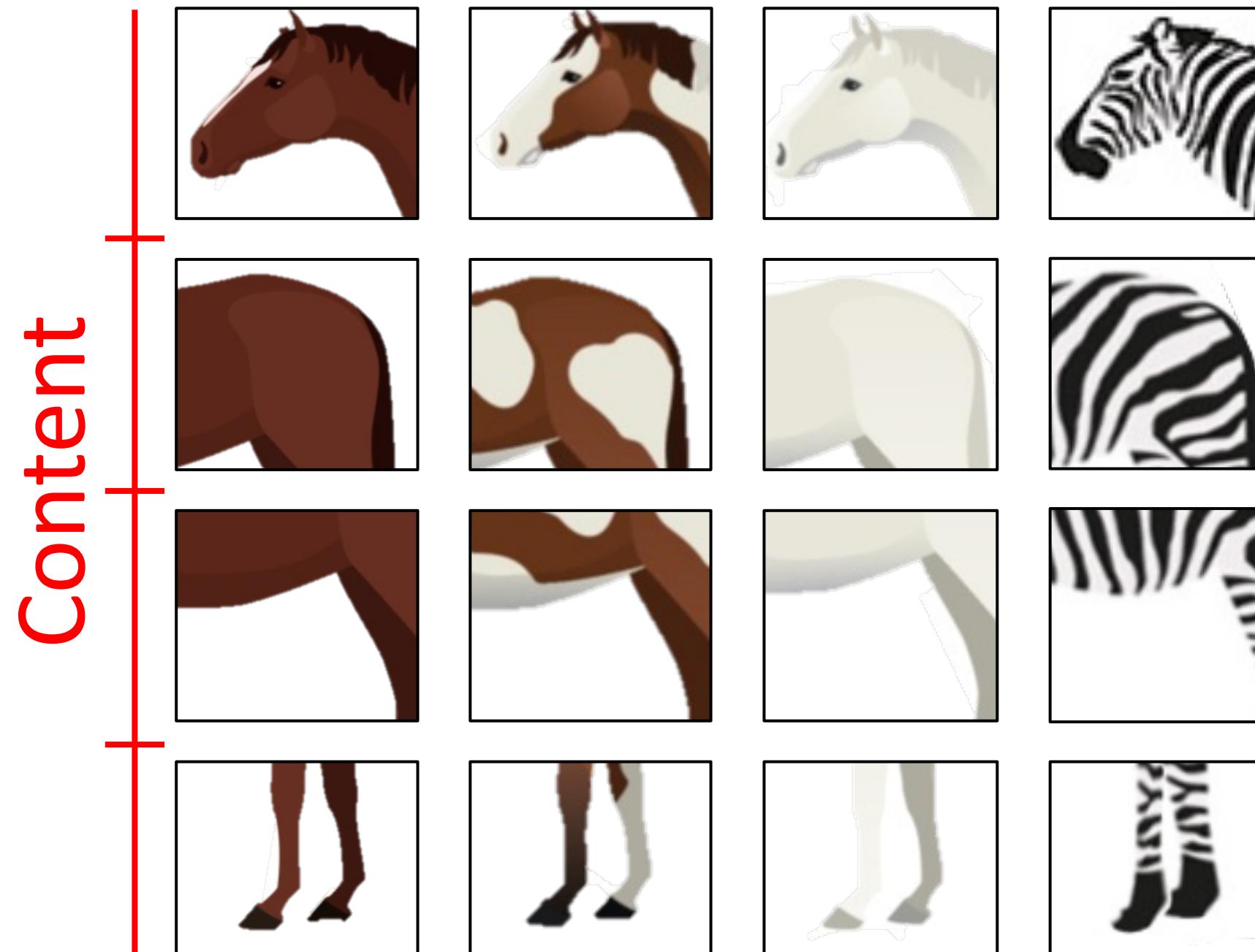
Style and Content

- Style: variations within the same domain
(different colors, textures, etc.)
- Content: features shared across two domains





Style

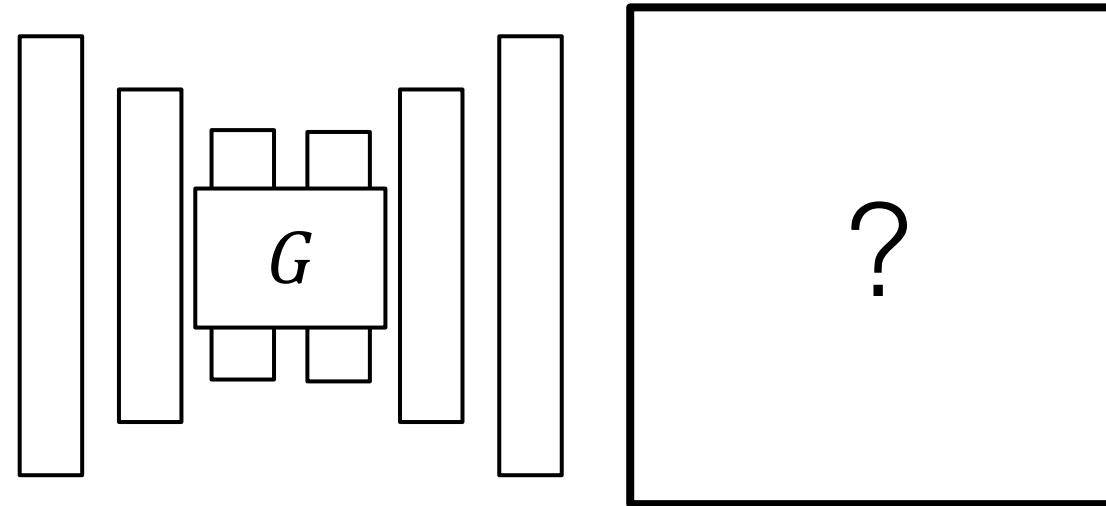


What makes for a good output?

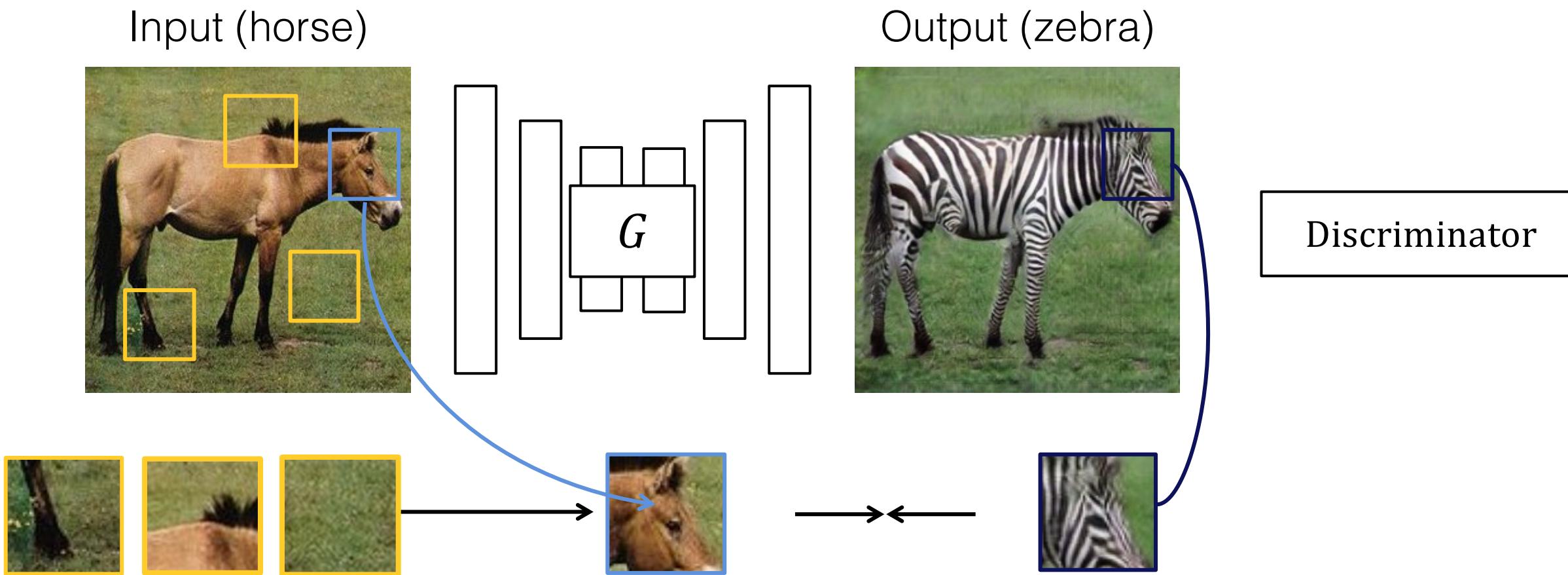
Input (horse)



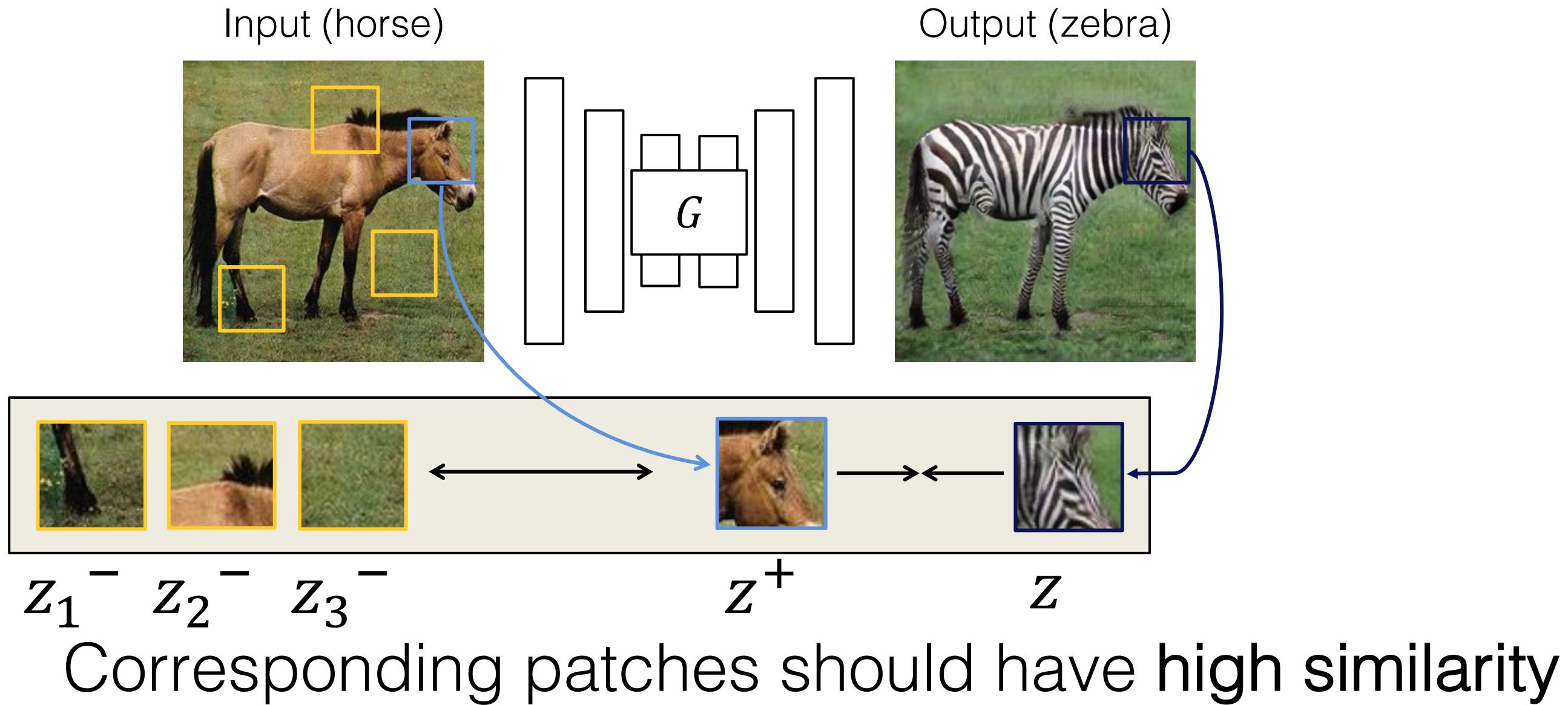
Output (zebra)



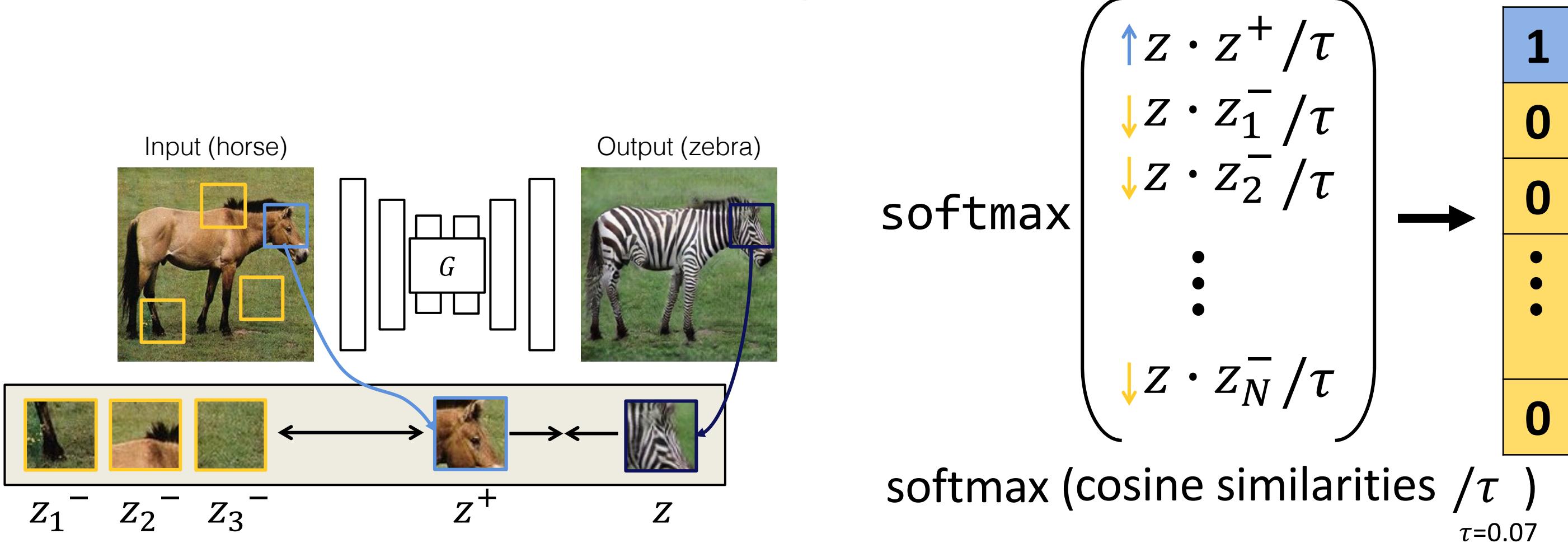
Retaining input content



Retaining input content

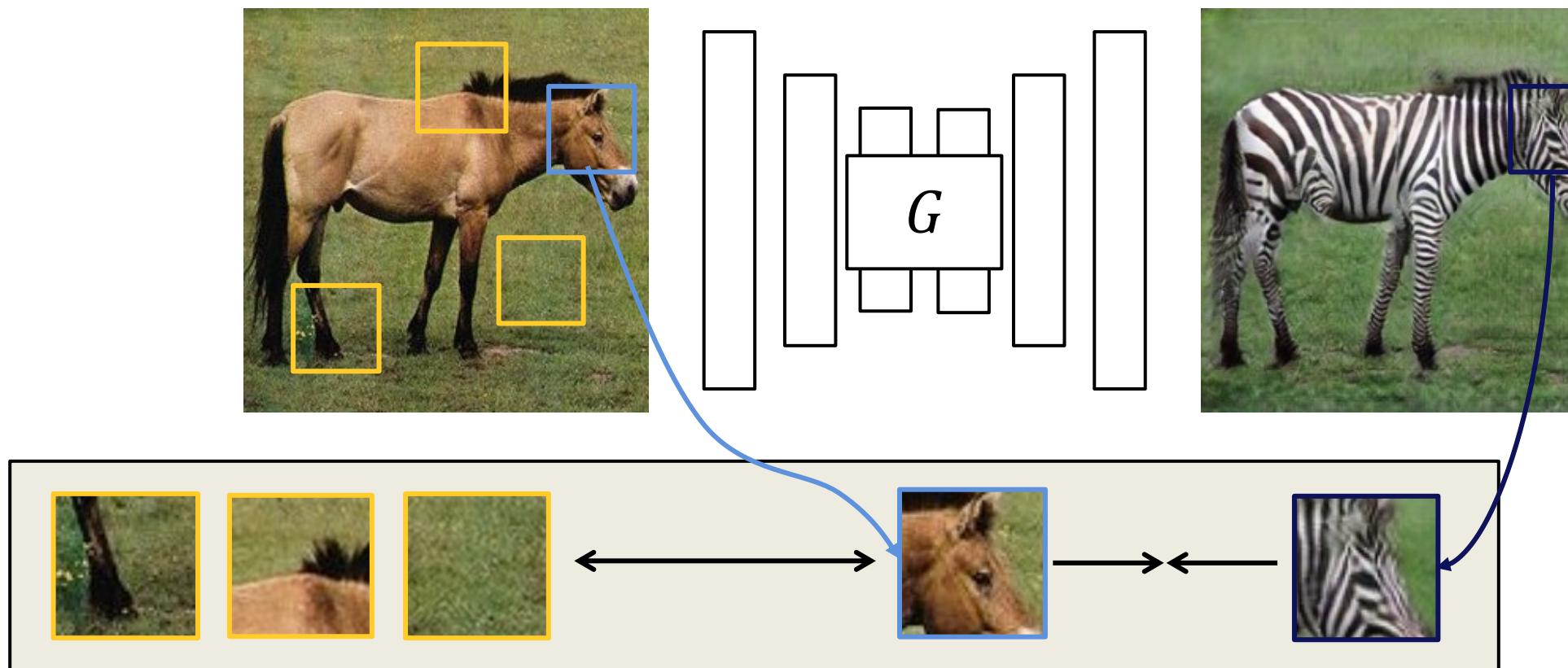


Patch-based Contrastive Loss

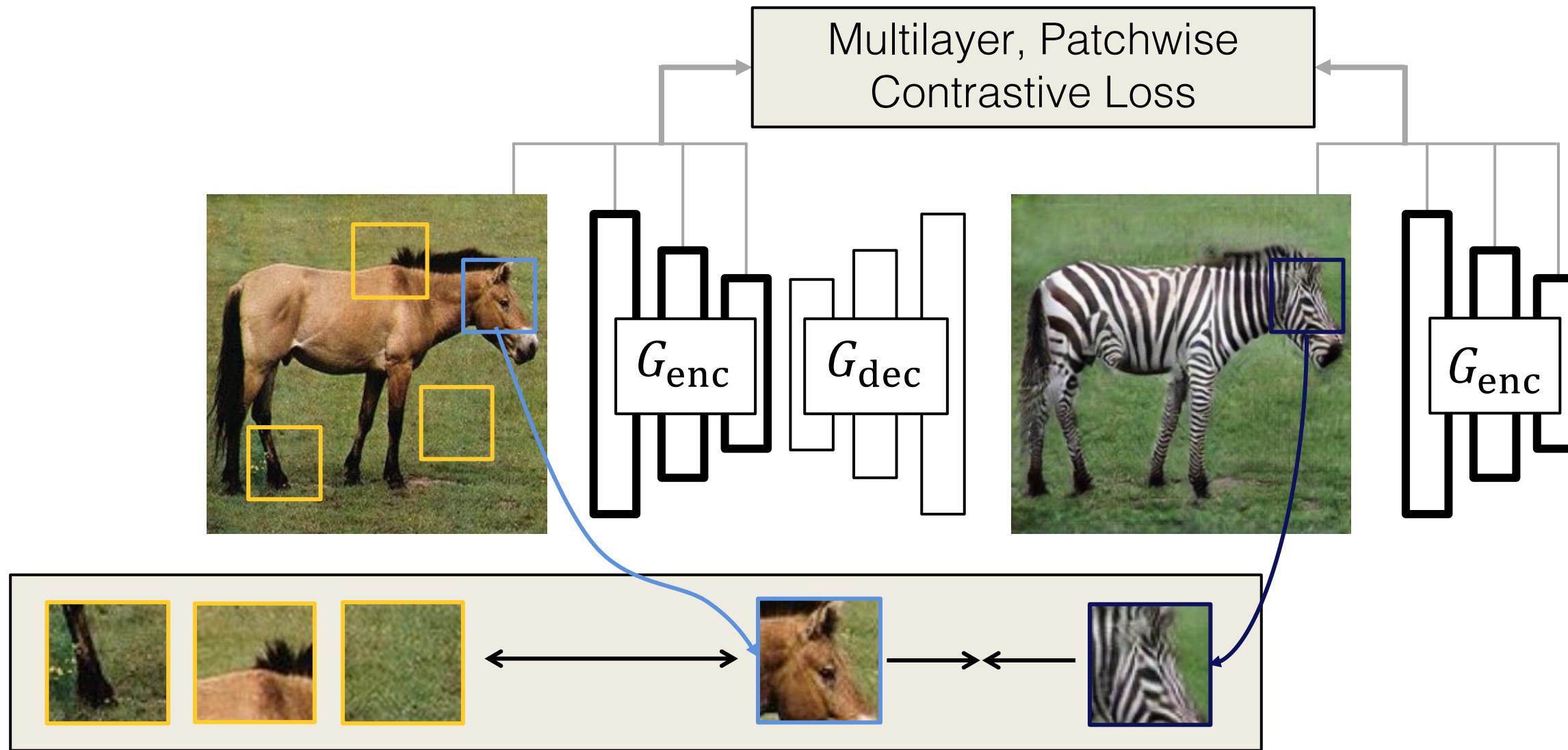


- InfoNCE loss (Gutmann et al., AISTATS18 , van den Oord et al., 2018) used in MoCo and SimCLR
- To produce positive pairs:
 - Handcrafted data augmentation (MoCo, SimCLR, etc.)
 - Input and synthesized image (ours)

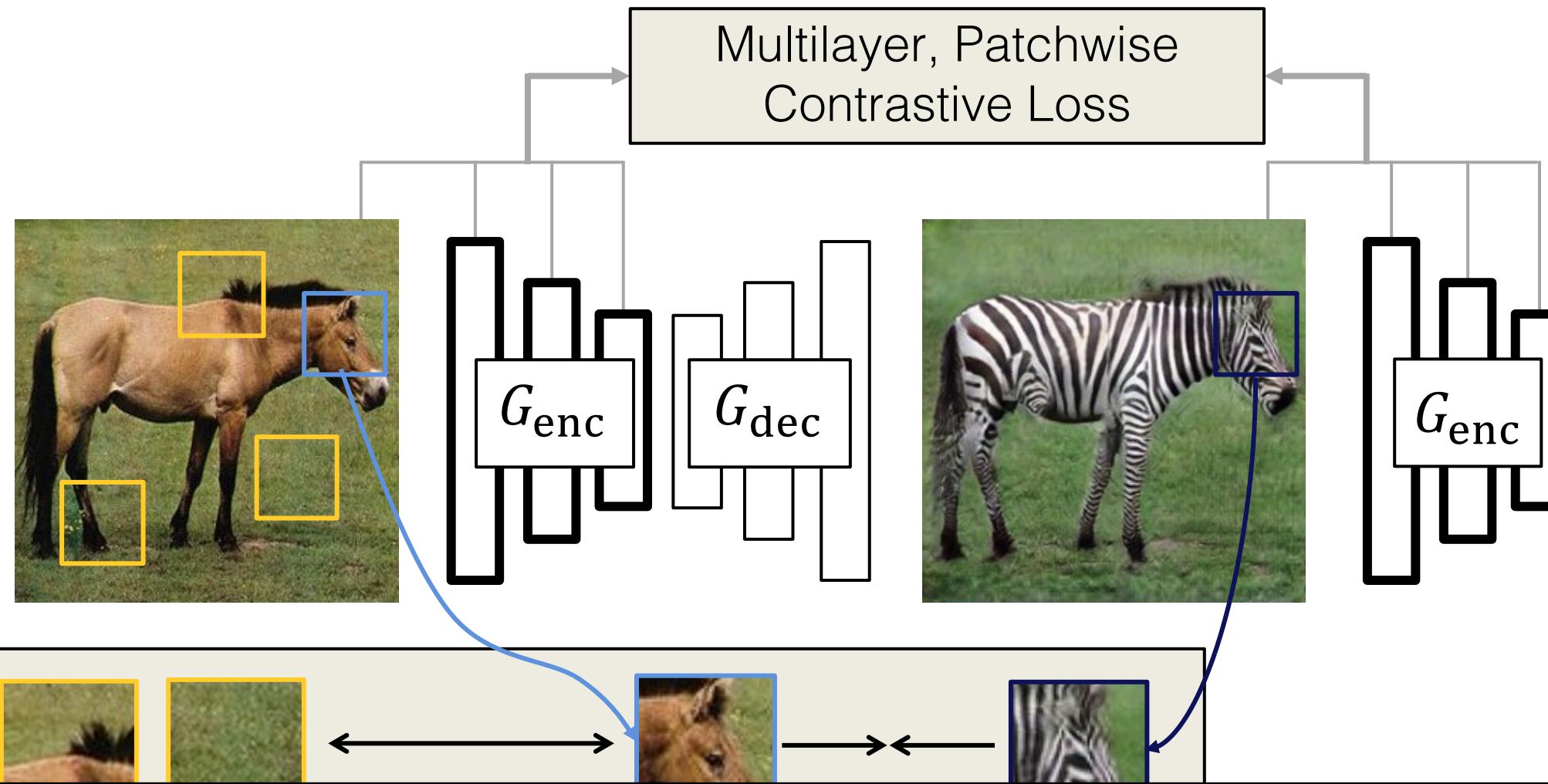
Patchwise contrastive loss



Patchwise contrastive loss

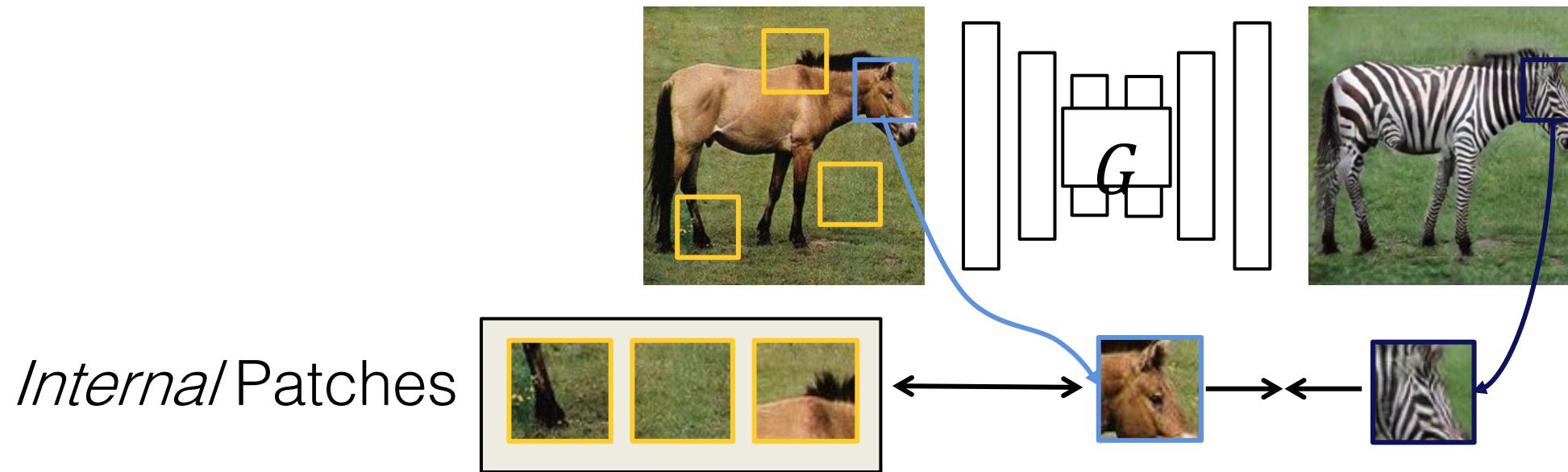


Patchwise contrastive loss

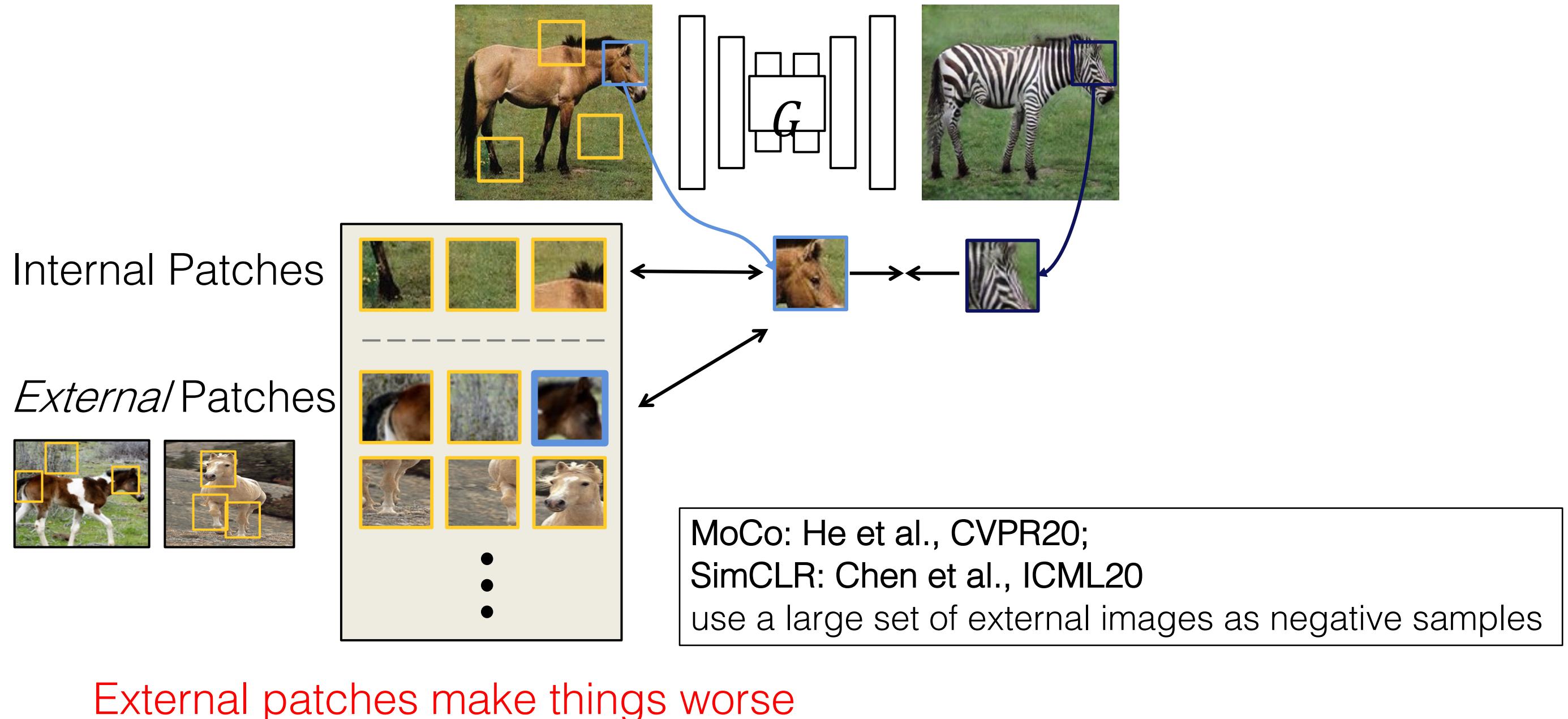


- + No fixed similarity metric (e.g., L1 or perceptual loss)
- + One-sided (no inverse mapping needed)

Internal vs External Patches

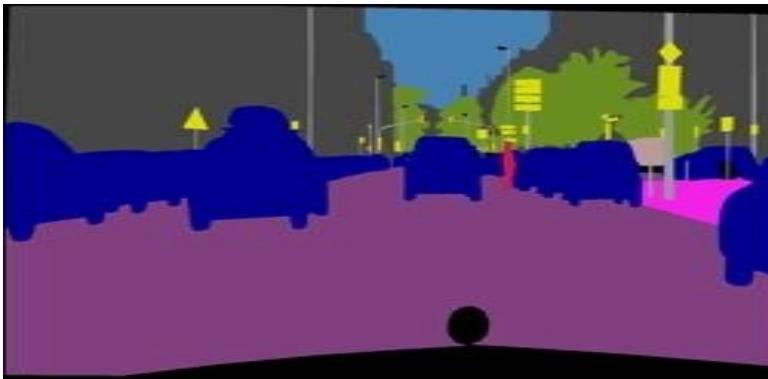
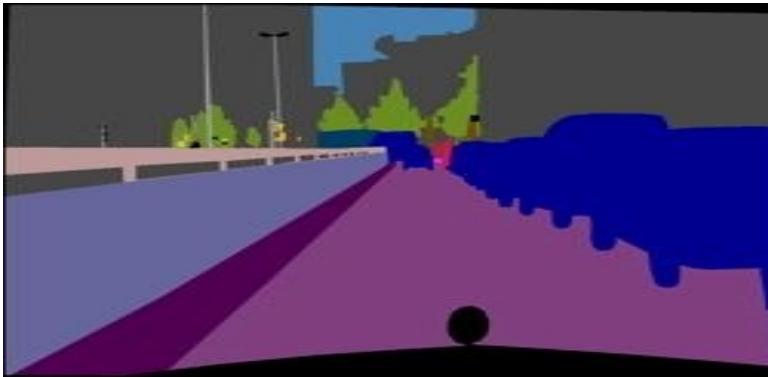


Internal vs External Patches



Internal vs External Patches

input



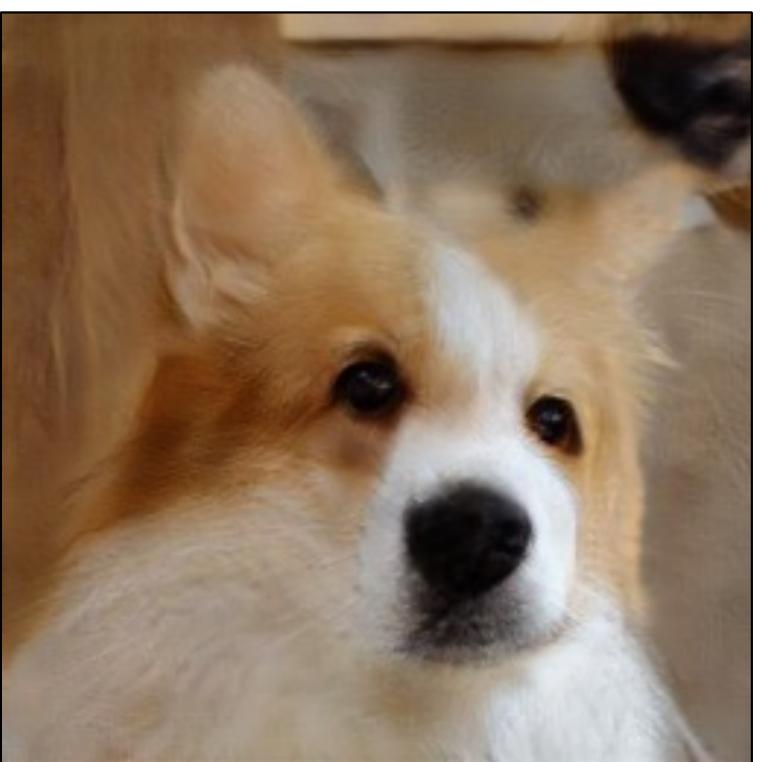
internal patches



external patches



Mode
Collapse!



Cat



Yosemite Summer



Apple



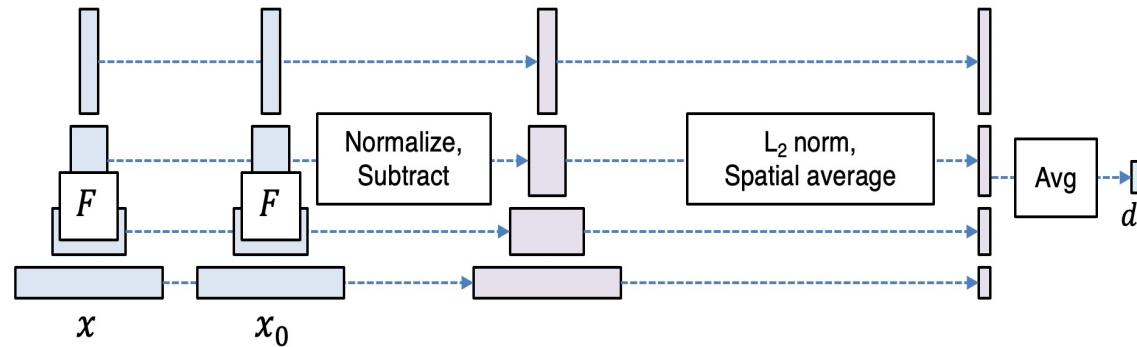
Paris



GTA

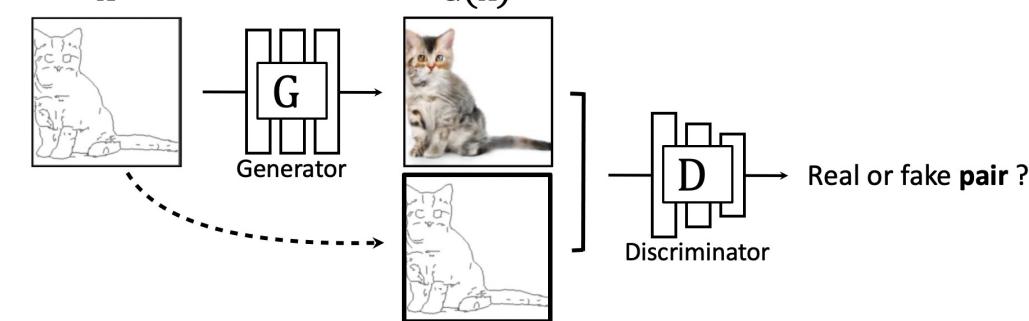
Summary

Perceptual/Feature Loss



How well do “perceptual losses” describe perception?

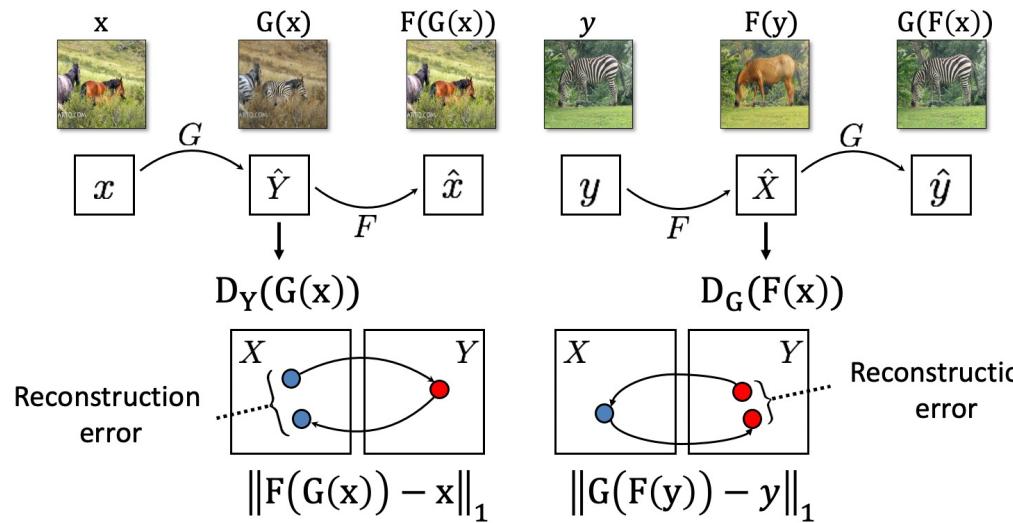
(Conditional) GAN Loss



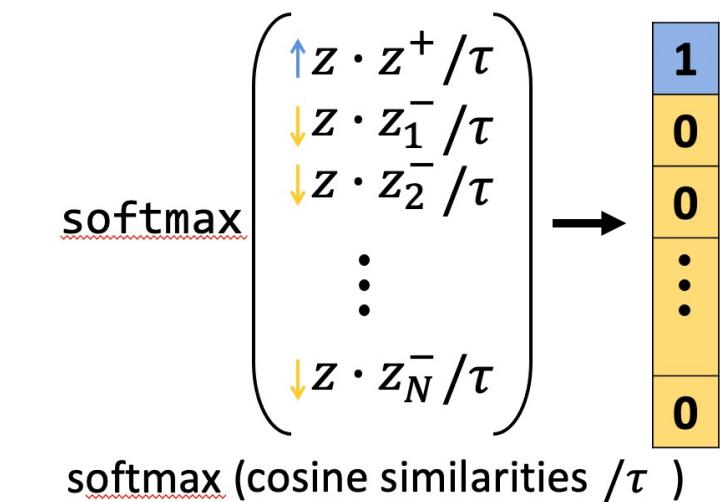
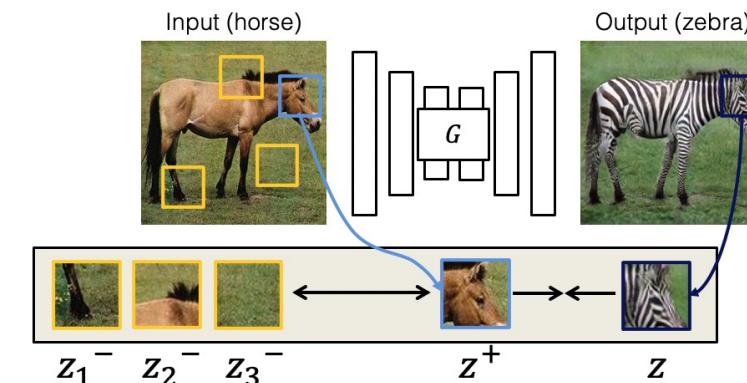
Learning objective

$$\min_G \max_D \mathbb{E}_x[\log(1 - D(\hat{x}, G(x)))] + \mathbb{E}_{x,y}[\log D(\hat{x}, y)]$$

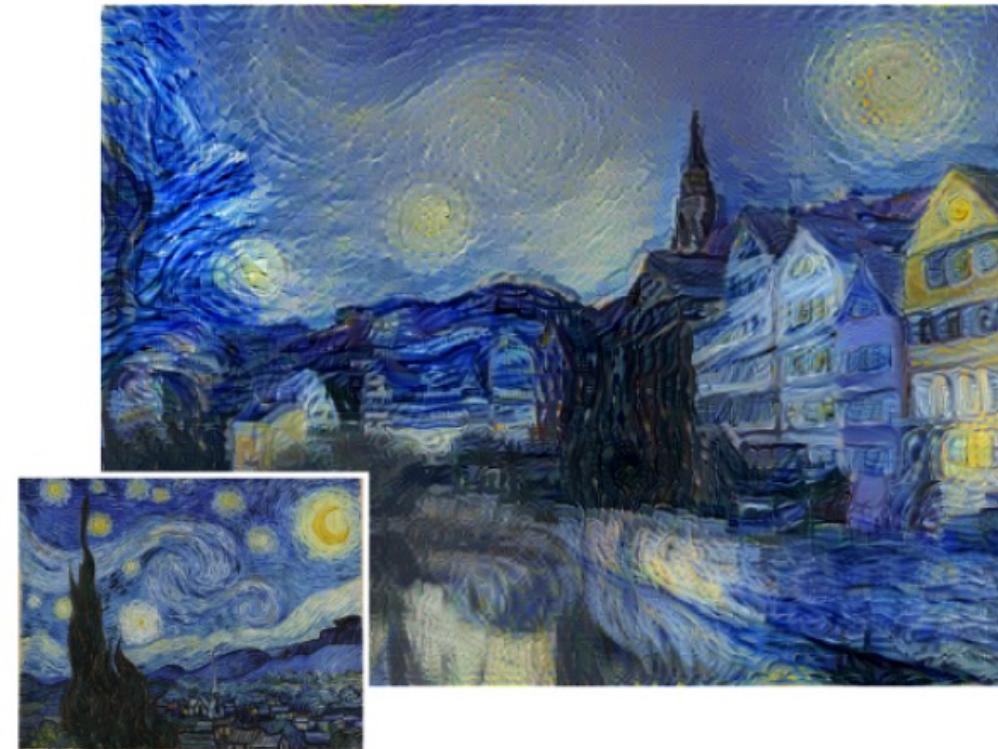
Cycle-Consistency Loss



Patch-wise Contrastive Loss



Other loss functions: Style Loss [Gatys et al.], Contextual Loss [Mechrez et al.], Domain-specific Loss (e.g., face), 3D-aware Loss (for geometric data)



Style and Content, Texture Synthesis

Jun-Yan Zhu

16-726, Spring 2023

Texture

- Texture depicts spatially repeating patterns
- Many natural phenomena are textures



radishes



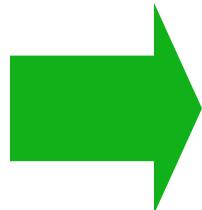
rocks



yogurt

Texture Synthesis

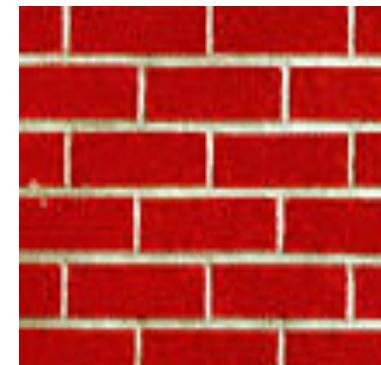
- Goal: create new samples of a given texture
- Applications: virtual environments, inpainting, texturing surfaces



Non-parametric Texture Synthesis

The Challenge

- Need to model the whole spectrum: from repeated to stochastic texture



repeated

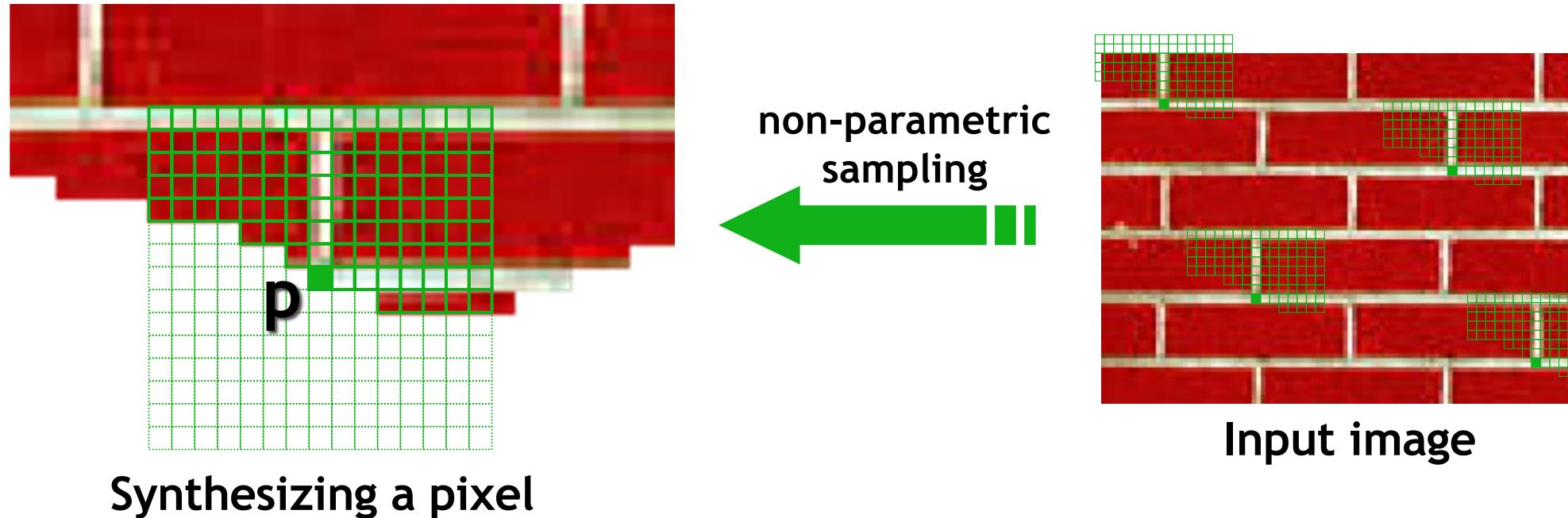


stochastic



Both?

Efros & Leung Algorithm

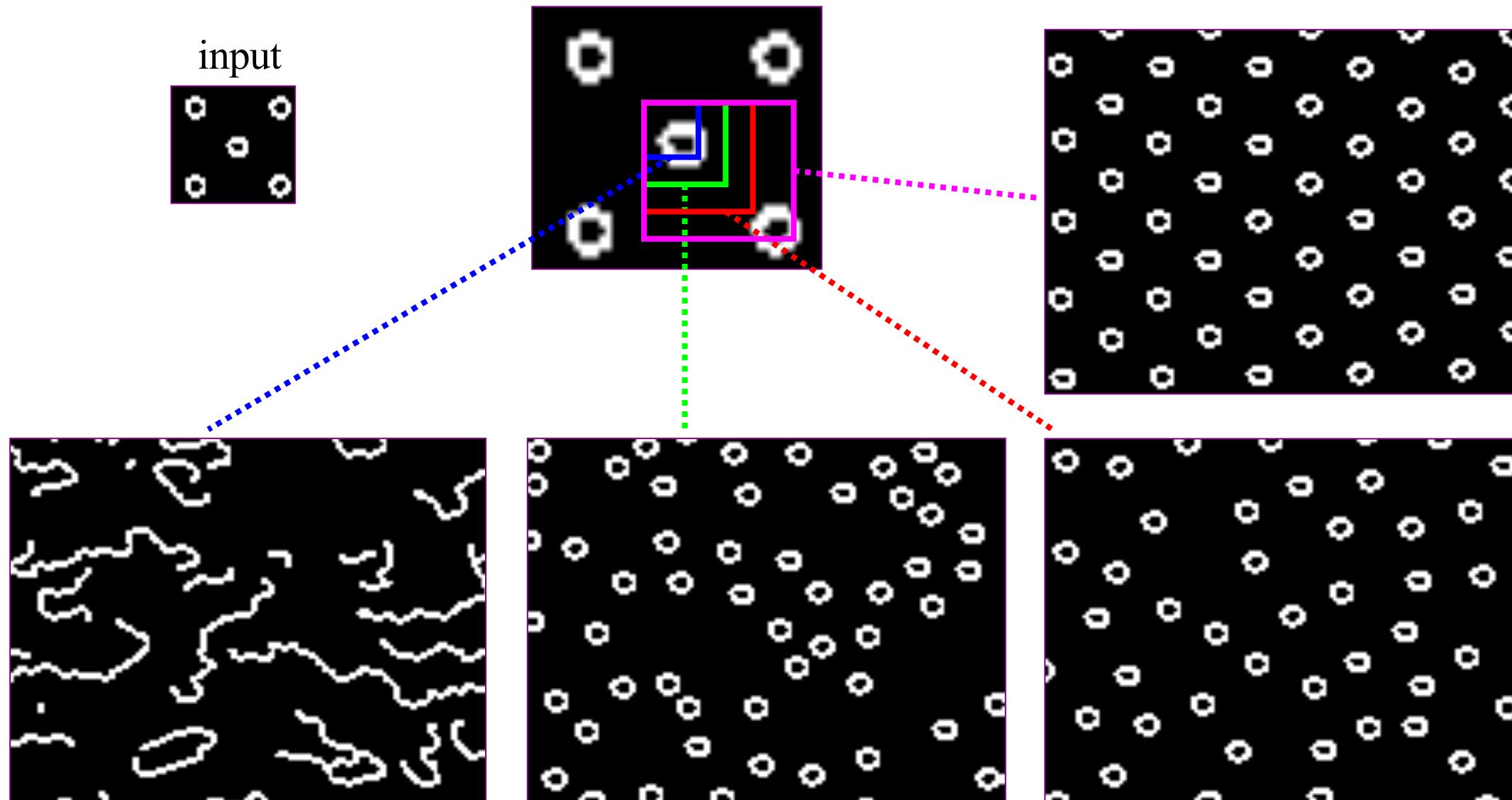


- Assuming Markov property, compute $P(p | N(p))$
 - Building explicit probability tables infeasible
 - Instead, we *search the input image* for all similar neighbourhoods — that's our pdf for p
 - To sample from this pdf, just pick one match at random

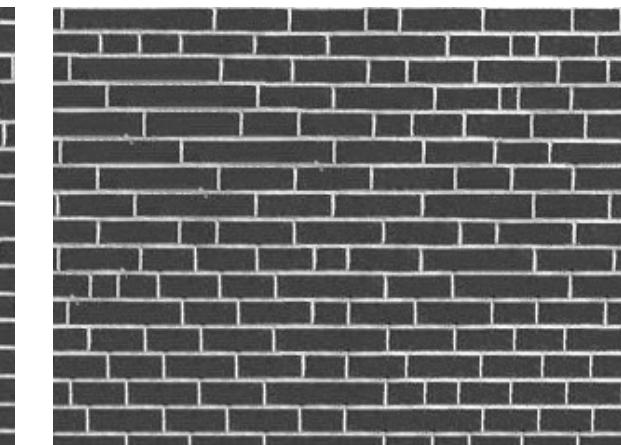
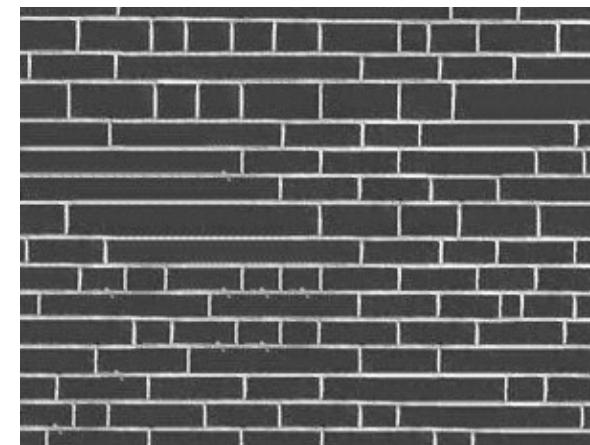
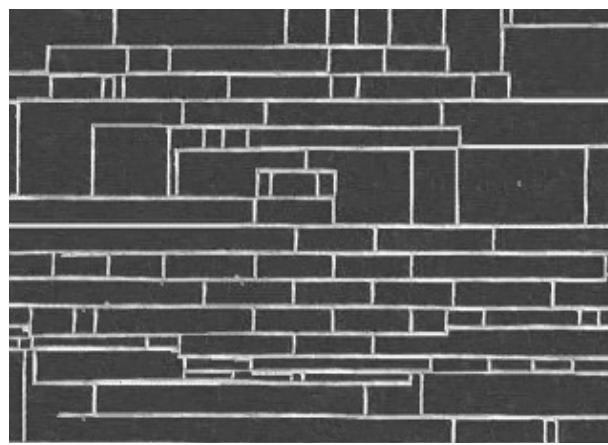
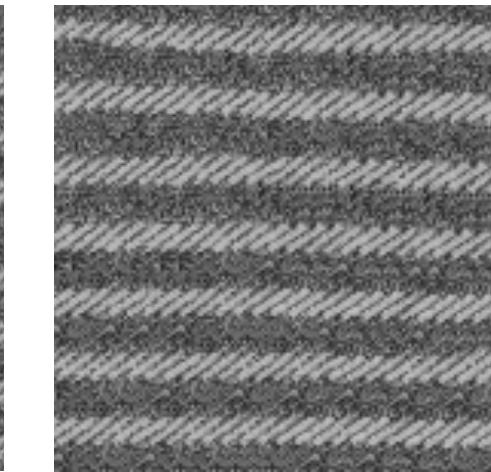
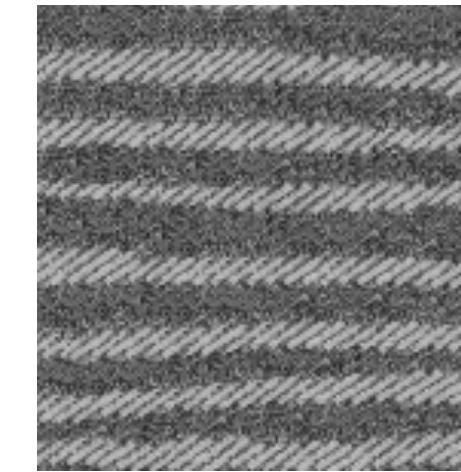
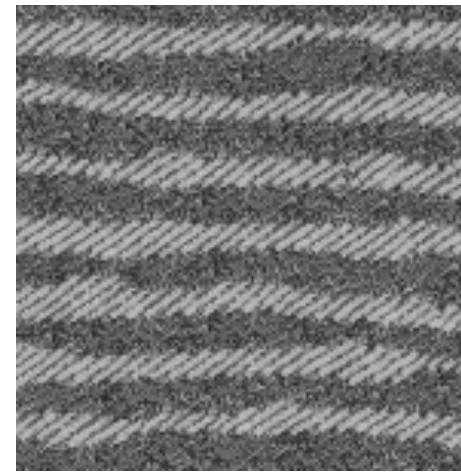
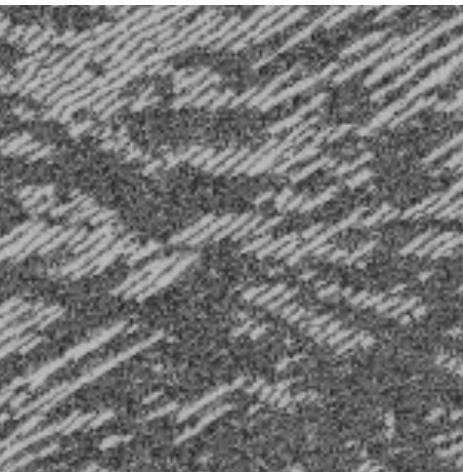
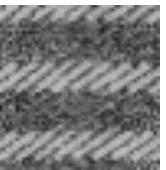
Some Details

- Growing is in “onion skin” order
 - Within each “layer”, pixels with most neighbors are synthesized first
 - If no close match can be found, the pixel is not synthesized until the end
- Using *Gaussian-weighted* SSD is very important
 - to make sure the new pixel agrees with its closest neighbors
 - Approximates reduction to a smaller neighborhood window if data is too sparse

Neighborhood Window



Varying Window Size

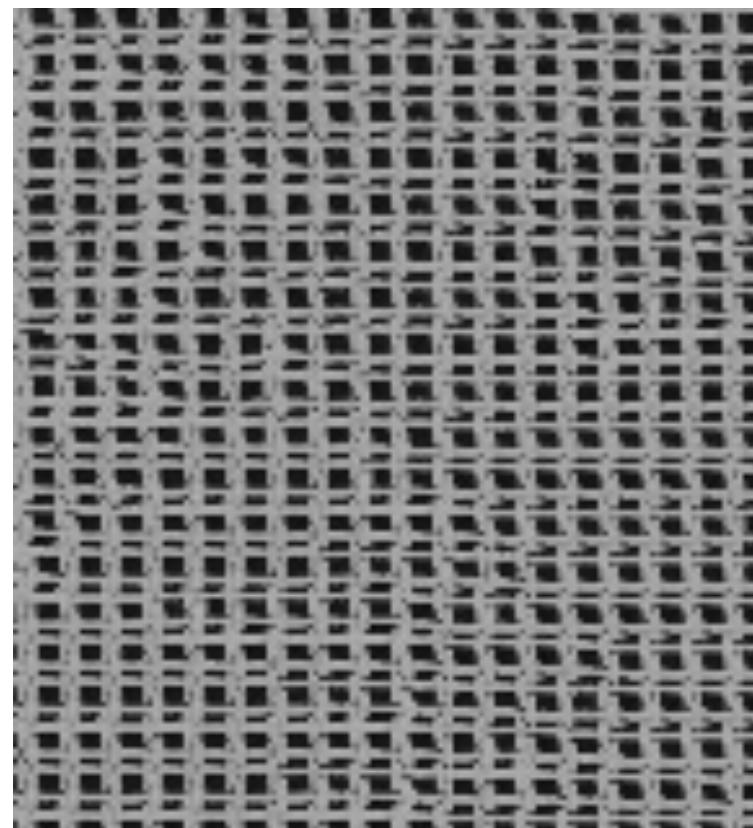
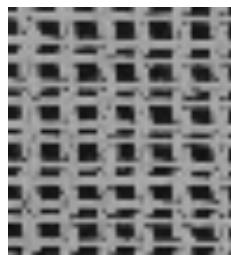


Increasing window size

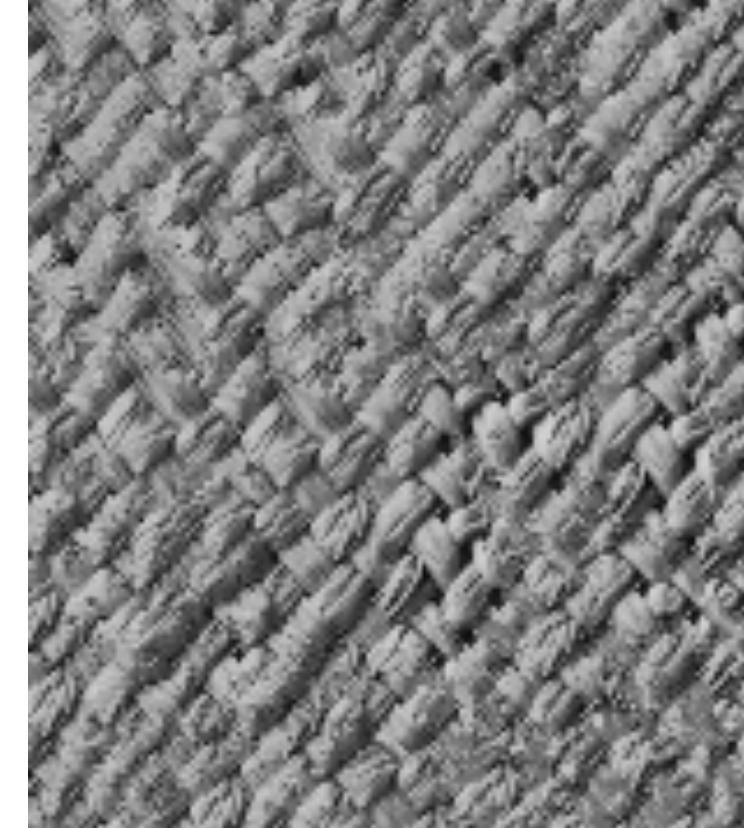
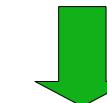
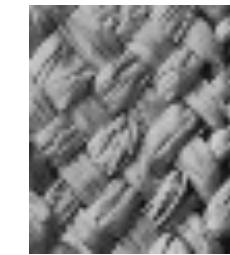


Synthesis Results

french canvas

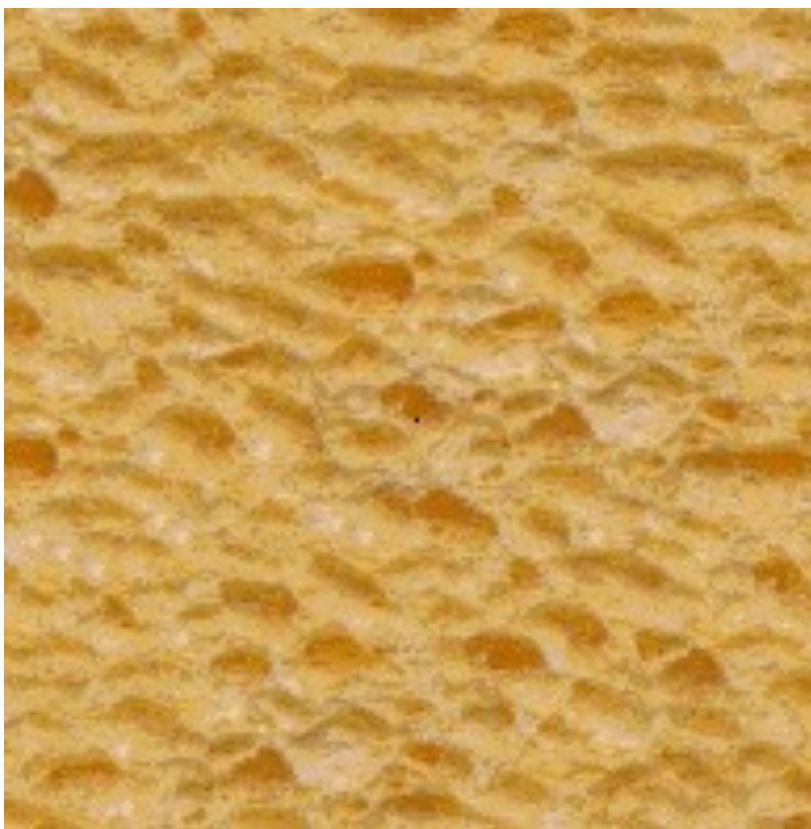


rafia weave

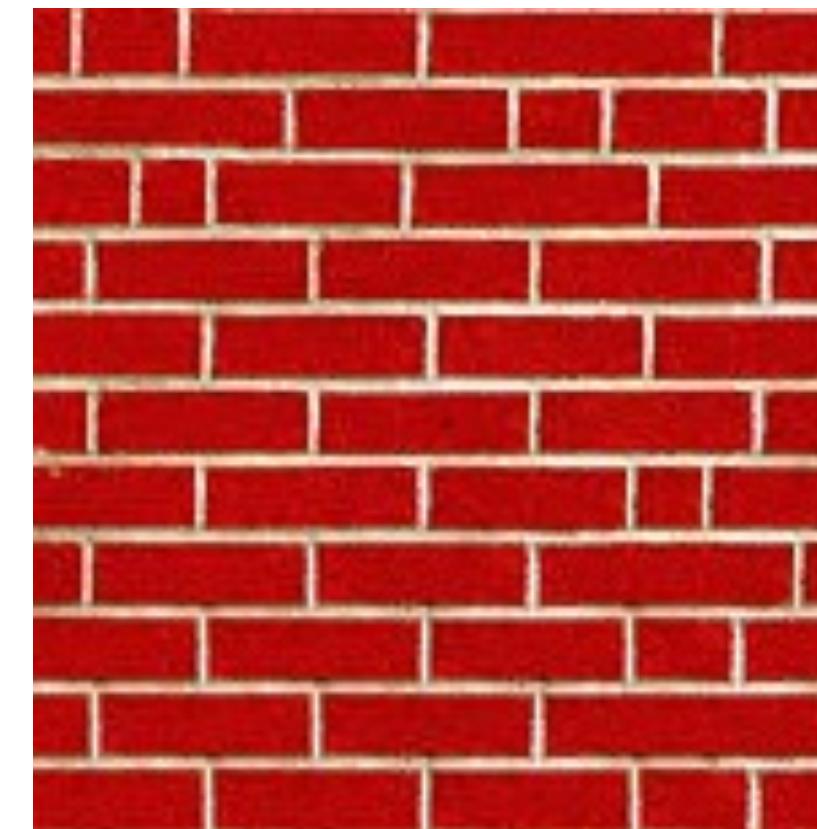
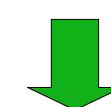
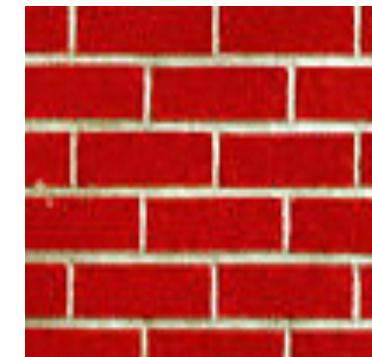


More Results

white bread



brick wall



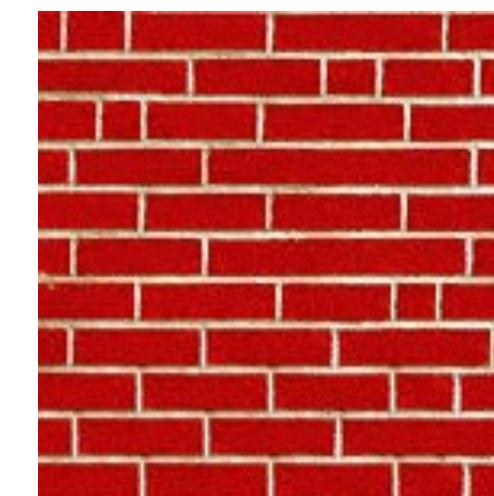
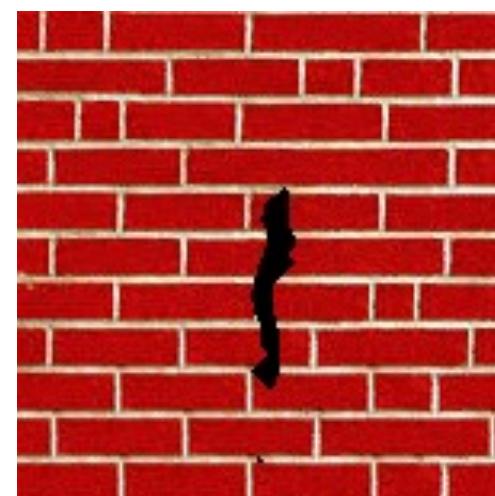
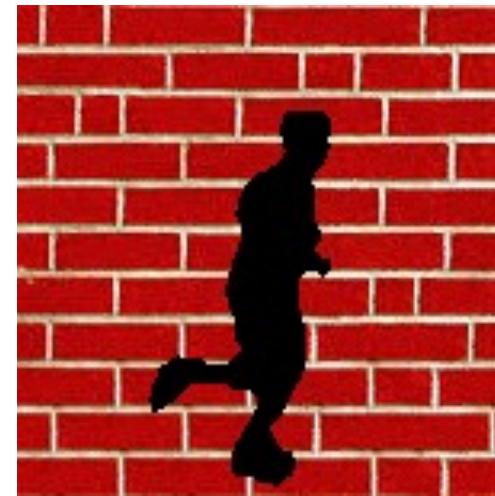
Homage to Shannon

uring his unsensational interview Dick Gephardt was fearful riff on the looming crisis. He only asked, "What's your position?" A heartfelt sigh followed. "It's a story about the emergency of the day," he said, referring to the legal challenges against Clinton. "Boycotting people about continuing the investigation," Gephardt began, patiently observing that the legal system was dealing with this latest tangent.

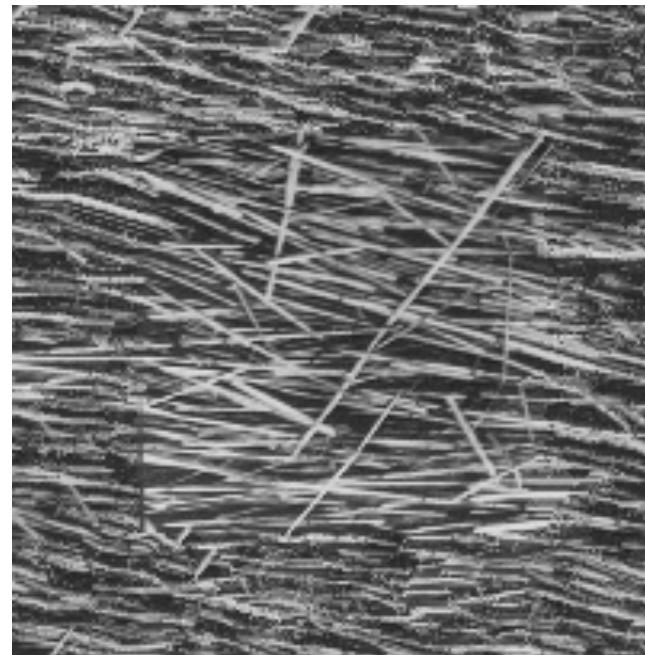
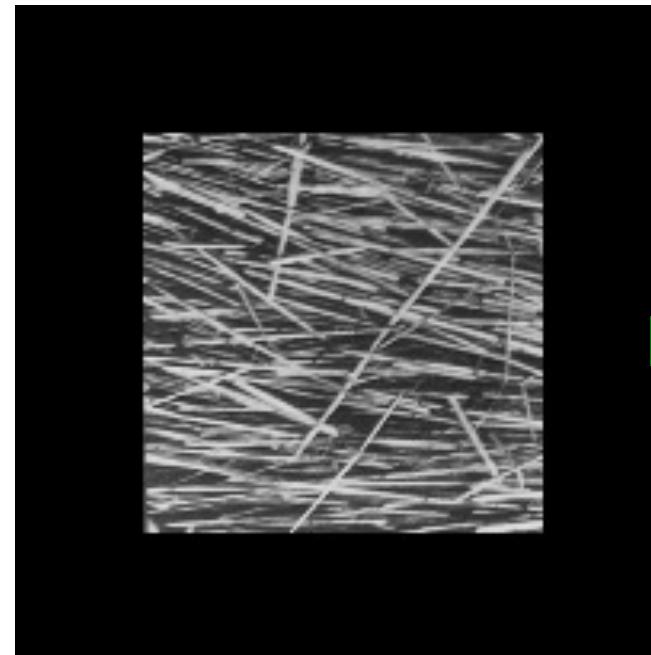
1 b' vnoþiowr. reh. A. v. s. ai thunne; E
egle fer s otot " la ttinu hhtt, t imr ai thunne; E
as " he dCil. eior feori tis u unt a
hC. yf te fit a us? firtfoear, cat h
tiae " t diab. l. n. h. ui hmr te opm t. F. o
e fit seu " vio bntu rn hrist dh pnr ti h
n id tif v. ie, v. dt g. slag l. u. entra mu abry s
utonuc. fu hes, it. , na lenr
ithenly n. Anf er Eloaeuoh, n. B.
dthf p. di b. l. in e o. i. ha muuny, ro. ab
ir fa. if. J. hooegahmtt. syoke
t. l. b. s. l. hsk as ke k. ys. m. C. u. e. t. f. r. l.
e. t. h. t. nit. " " tff afe c. d. t. ell. r. h. C. a. r. s.
l. t. in t. r. l. t. n. l. e. t. y. a. v. i. s. A. ^ p. o. f. r. y. p.
s. C. n. h. r. u. t. n. l. e. t. y. a. v. i. s. A. ^ h. v. e. w. t.
t. h. i. n. g. e. o. r. r. t. e. s. p. o. r. i. u. r. t. w. e. n. n. p. C. n. u. y.
j. h. o. e. e. t. v. e. n. l. i. r. y. l. i. j. u. r. t. w. e. n. n. p. C. n. u. y.
(d. v. t. l. " v. o. e. v. o. e. l. p. " u. l. t. " b. t. e. l.
o. e. [ps. the] l. " v. e. r. n. r. " t. l. f. f. b. l. l. r. r. n. " >. b. t. e. l.
at. t. g. e. l. " v. o. l. f. u. H. s. v. p. n. a. r. p. a. d.
ten. A. " a. r. e. n. l. o. u. m. t. e. n. t. f. r. C. n. t. t. o. s. t. h. p. n. i. s. t. a.
m. r. i. m. p. l. r. t. e. l. u. t. f. u. r. i. t. t. o. s. e. h. f. a. n.
3. m. r. i. m. p. l. r. t. e. s. y. h. p. s. a. l. u. b. t. C. a. i. t. i. a. w. w. f. r. t. t.

thaim. them . "Whnephartfe lartifelintomimer
fel ck Clirtioout omaim thartfelins. f ait s aneit
the ry onst warfē lck Gephntoomimeationl sigal
Chiooufit Clinut Cll riff on. hat's yordn, parut tly
ons yoontonsteht wasked, paim t sahe loo' riff on
nskoneploourtfeas leil A nst Clit, "Wleontongal
k Cirtioouirtfepe ong pme abegal fartfenstemem
tiensteneltorydt telemeplninsverdt was agemer
ff ons artientont Cling peme asurtfe atish, "Boui
hal s fartfelt sig pedr tlrdt ske abounutie aboutioo
tfeaonewwas yous abowonthardt thatins fain, ped,
ains. them, pabout wasy arfiut coutly d, In A h
ole emthrdingbooreme agas fa bontinsyst Clinut
ory about continst Clipeouinst Cloke agatiff out
stome zinemen tly ardt beoraboul n, thenly as t C
cons faimeme Diantont wat coutlyohgans as far
ien, phrtfaul, "Wbaut cout congagal cōmininga
mifmst Cliy abon al coountha.emungairt tf oun
The looorysten loontieph. Intly on, theoplegatick
aul tatiesontly atie Diantomi wal s f tbegae ener
mthahgat's enenhmas fan, "intchthory ahons v

Hole Filling



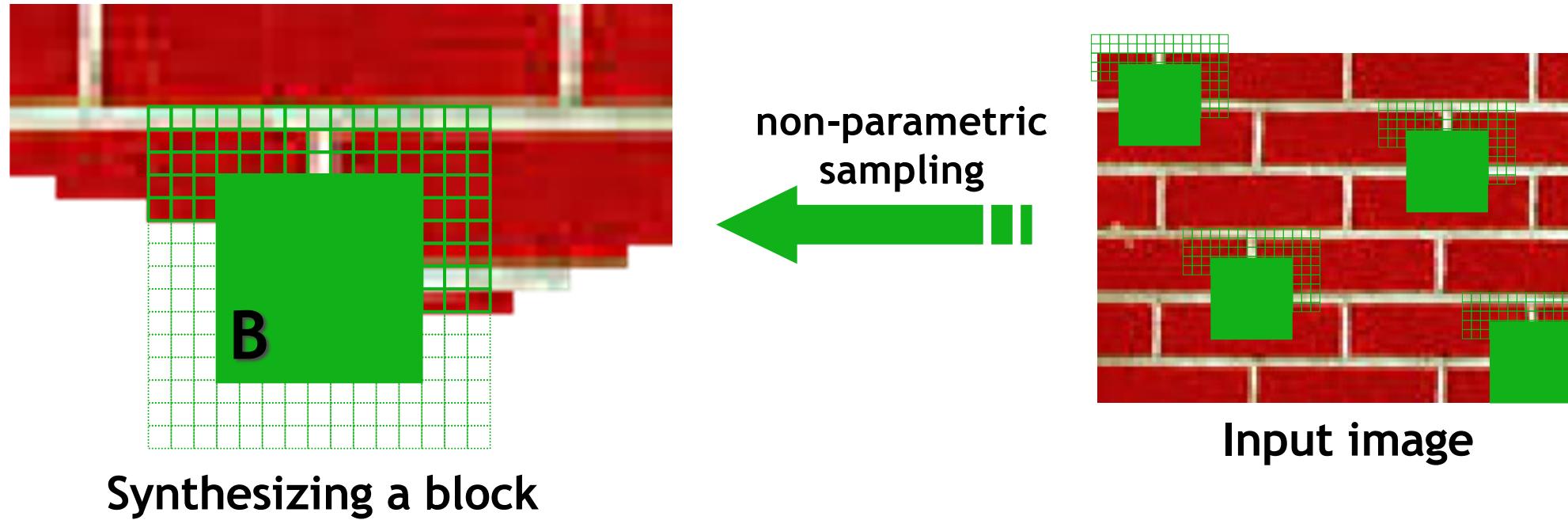
Extrapolation



Summary

- The Efros & Leung algorithm
 - + Very simple
 - + Surprisingly good results
 - + Synthesis is easier than analysis!
 - ...but very slow

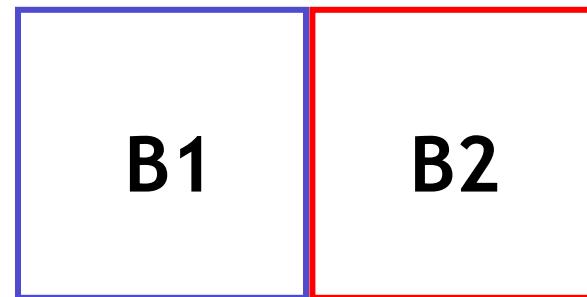
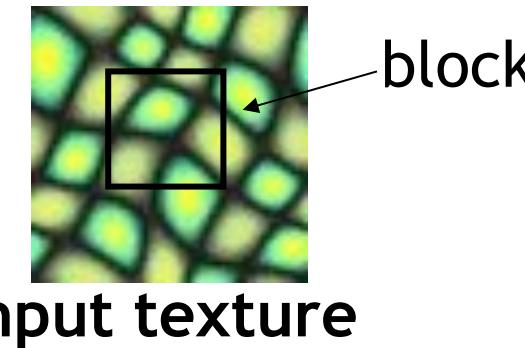
Image Quilting [Efros & Freeman]



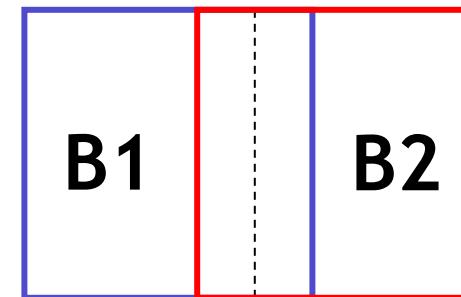
- Observation: neighbor pixels are highly correlated

Idea: unit of synthesis = block

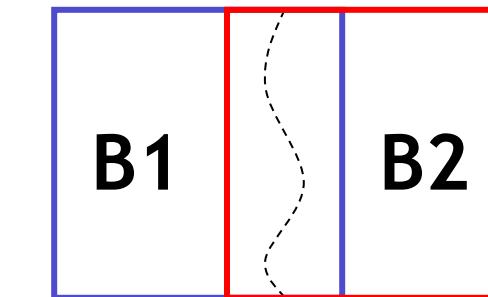
- Exactly the same but now we want $P(B | N(B))$
- Much faster: synthesize all pixels in a block at once
- Not the same as multi-scale!



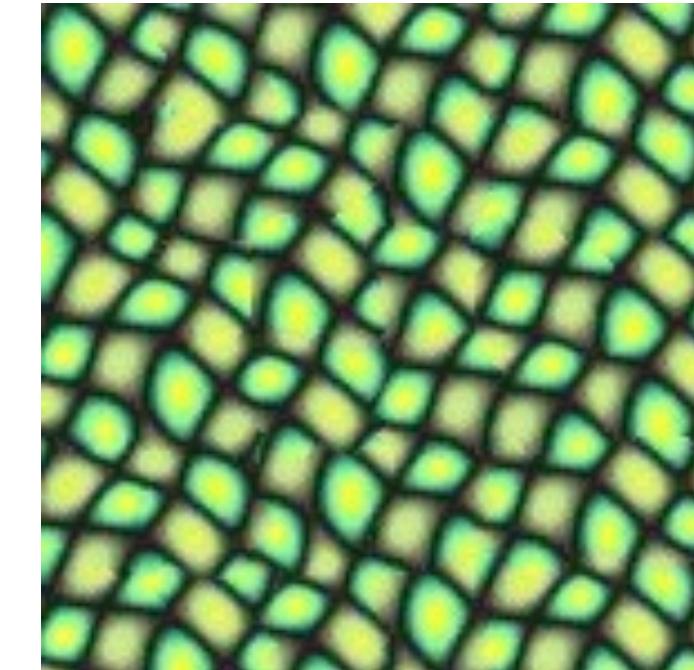
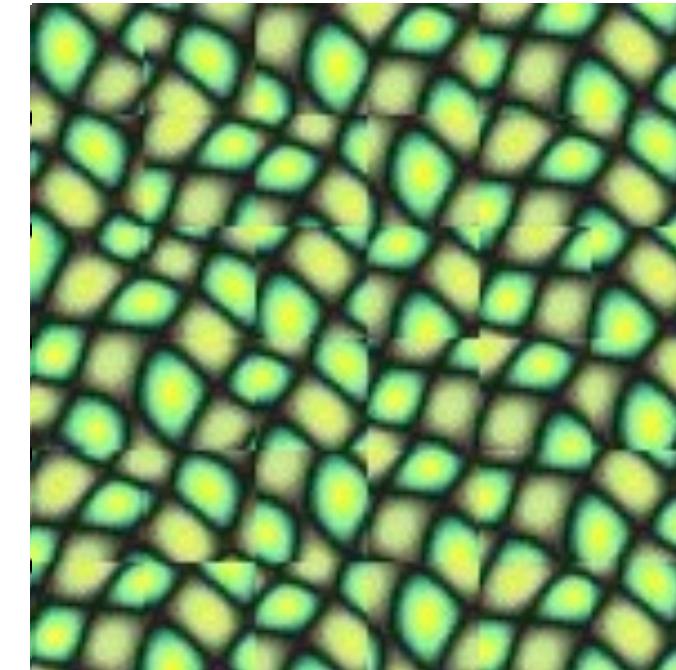
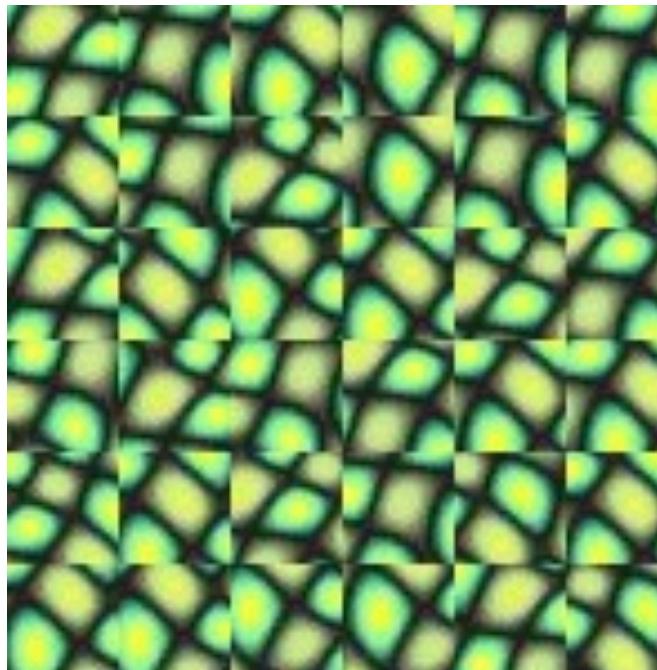
Random placement
of blocks



Neighboring blocks
constrained by overlap

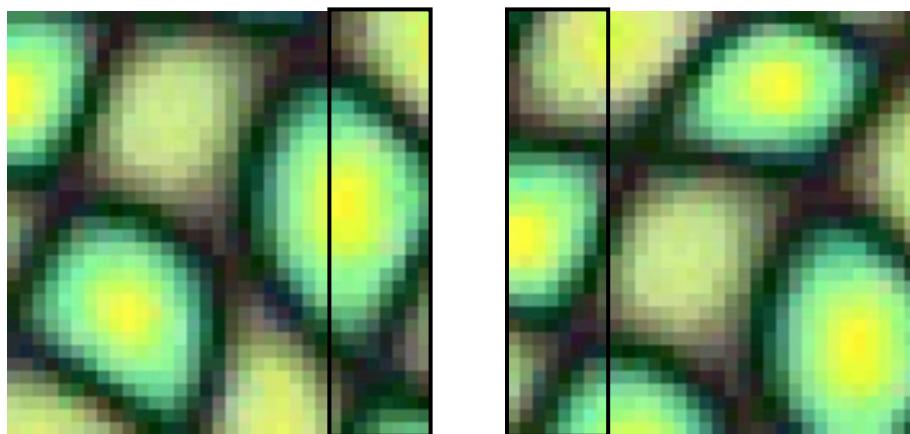


Minimal error
boundary cut

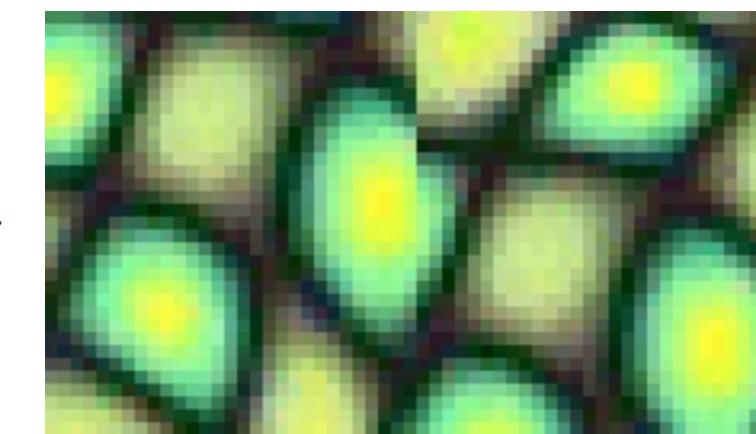


Minimal error boundary

overlapping blocks

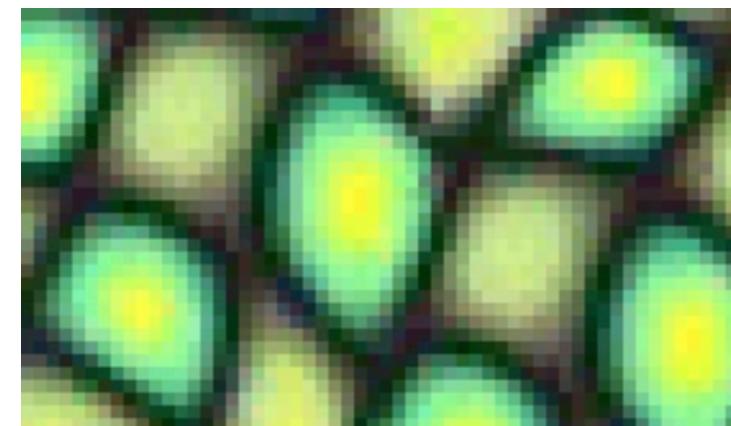


vertical boundary



$$\left(\begin{array}{c} \text{[Heatmap block]} \\ - \\ \text{[Heatmap block]} \end{array} \right)^2 = \text{[Binary mask with red border]}$$

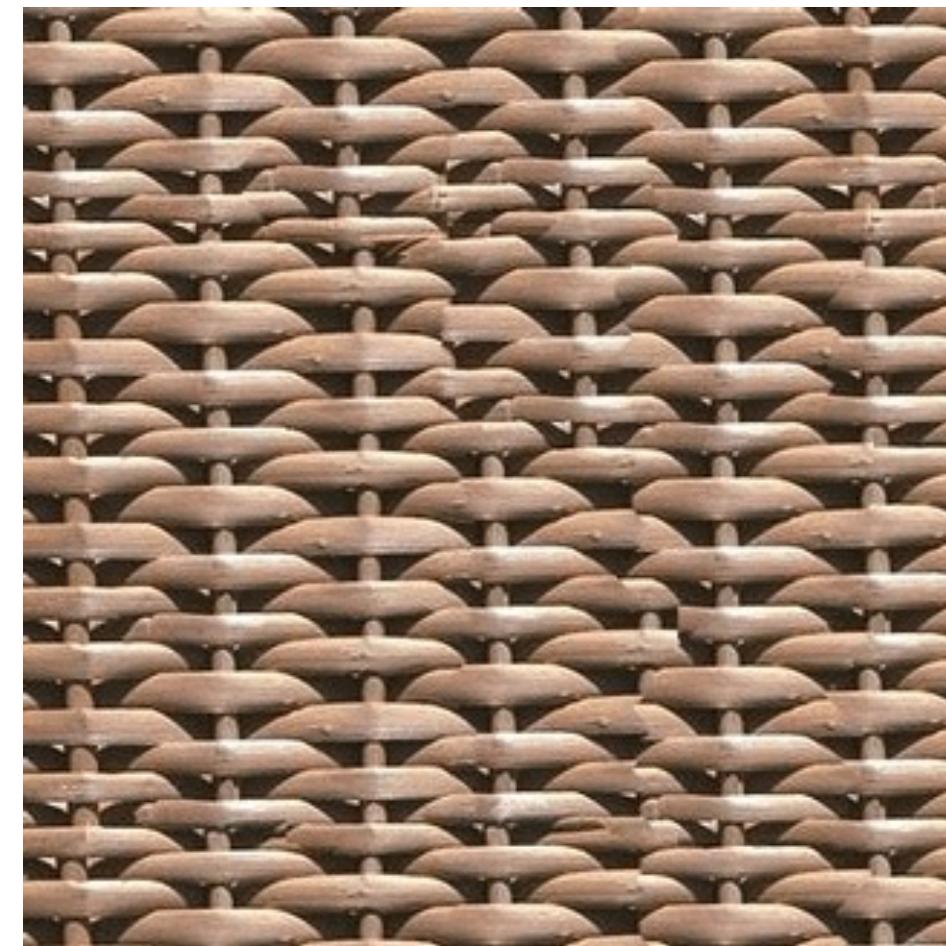
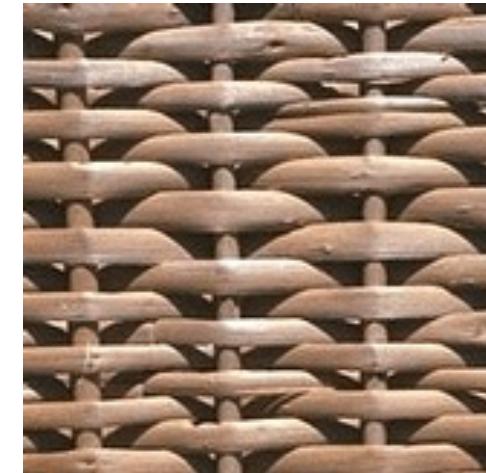
overlap error

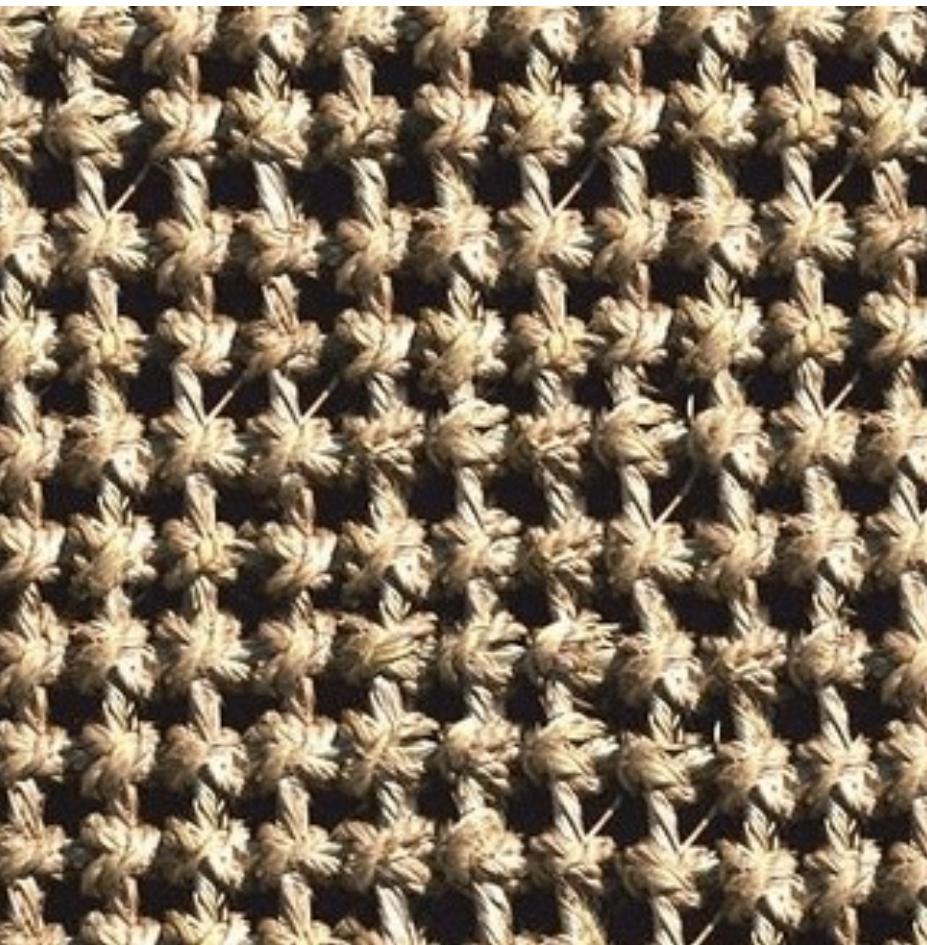


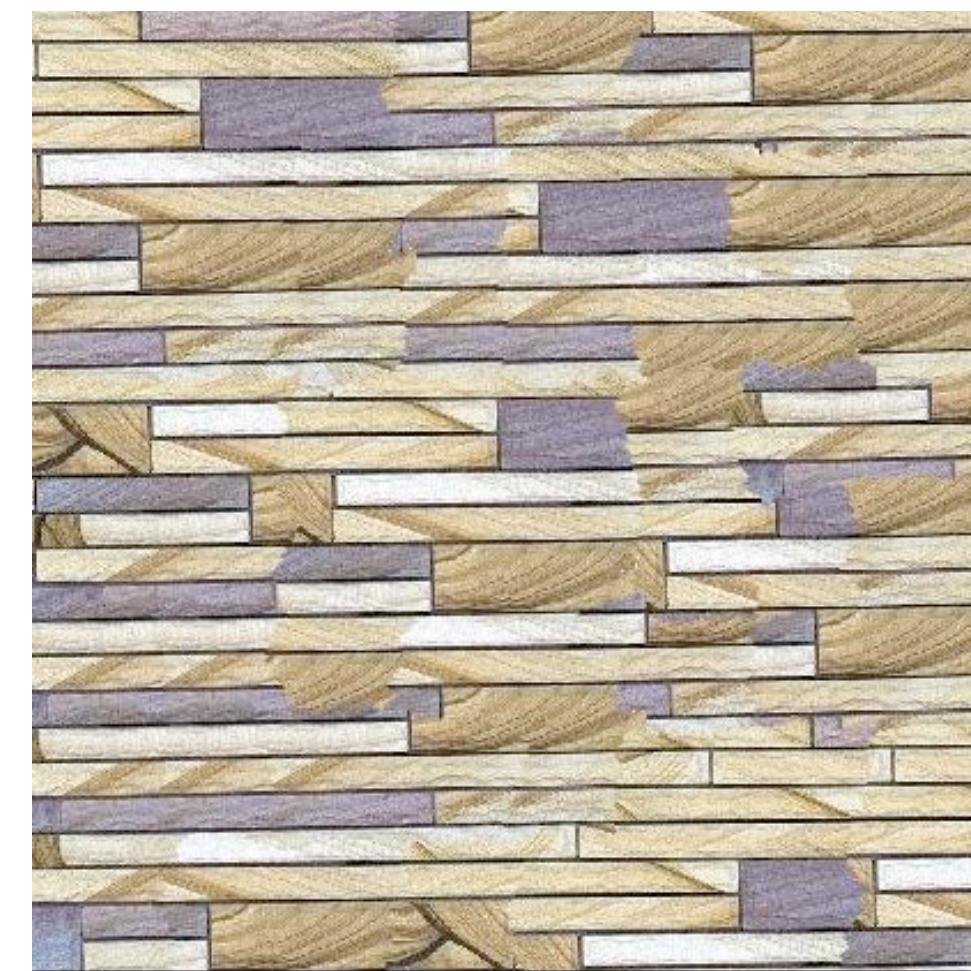
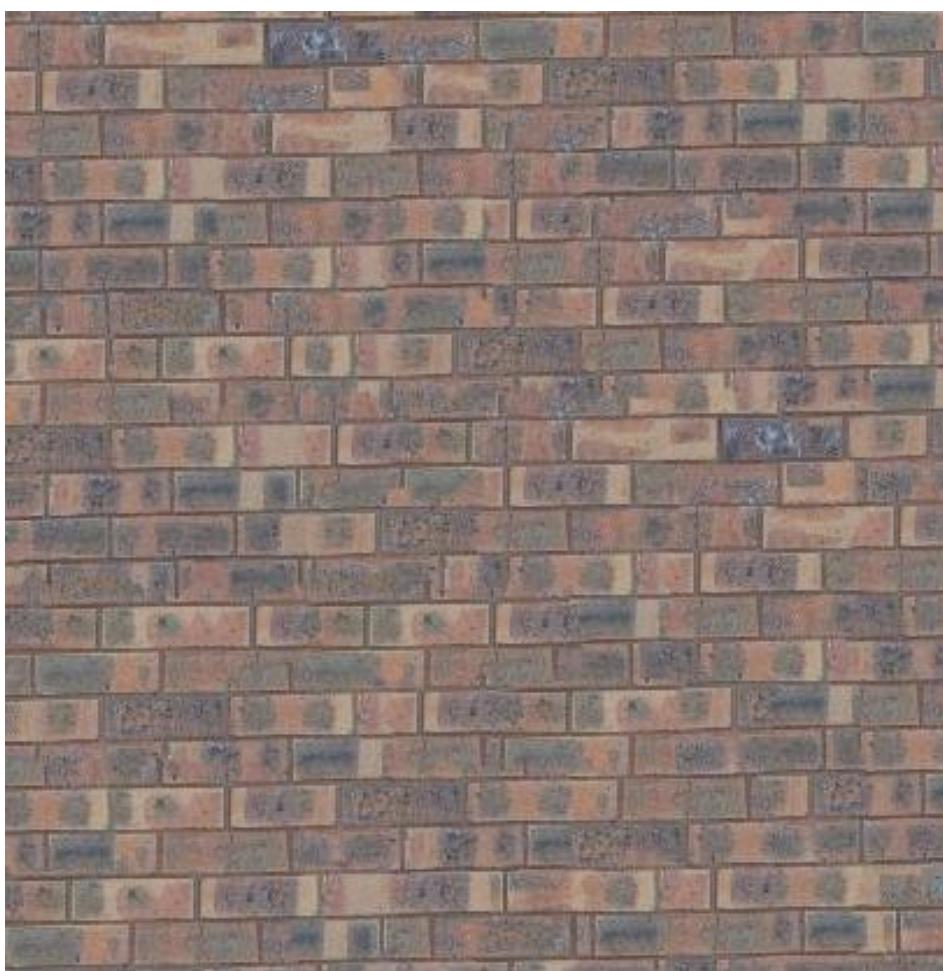
min. error boundary

Our Philosophy

- The “Corrupt Professor’s Algorithm”:
 - Plagiarize as much of the source image as you can
 - Then try to cover up the evidence
- Rationale:
 - Texture blocks are by definition correct samples of texture so problem only connecting them together

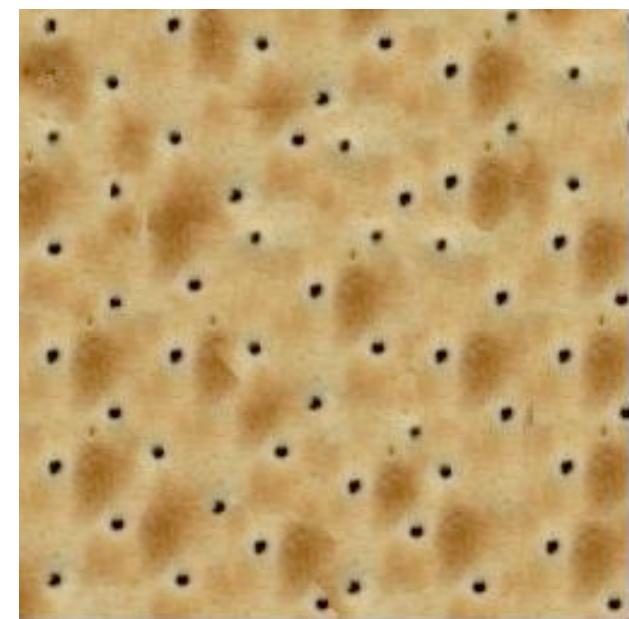
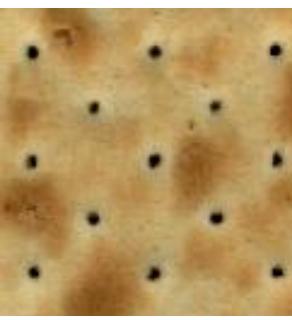






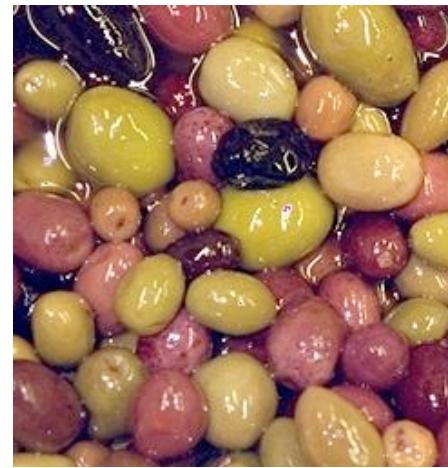


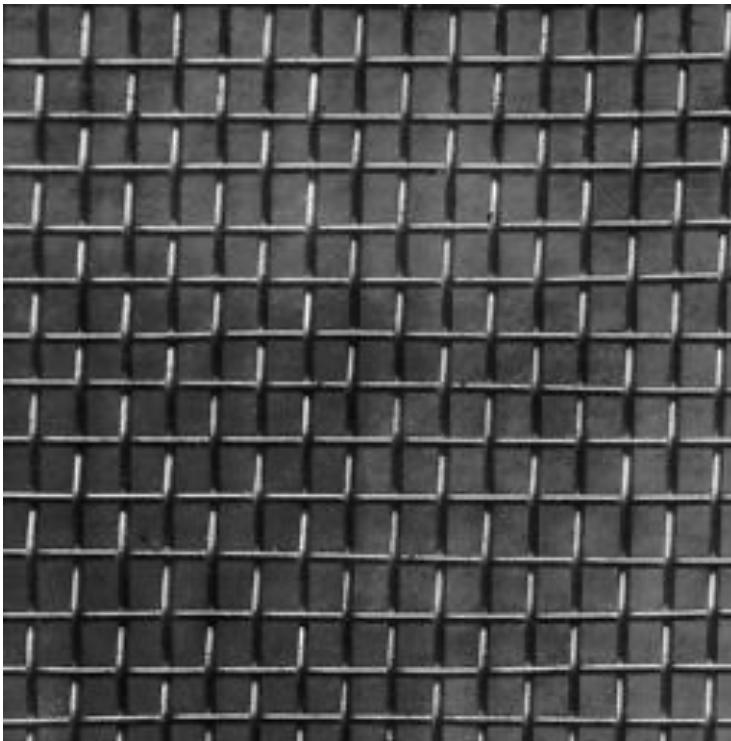




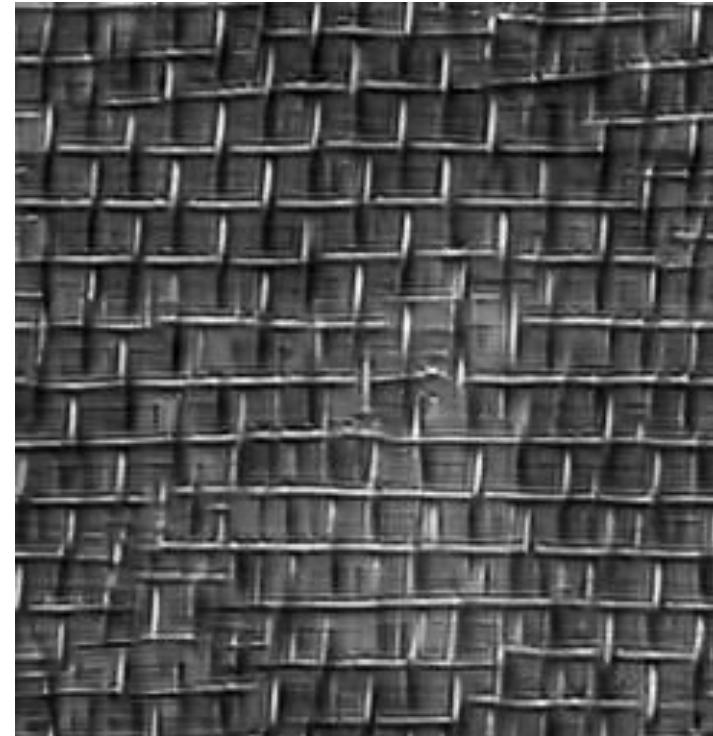


Failures (Chernobyl Harvest)

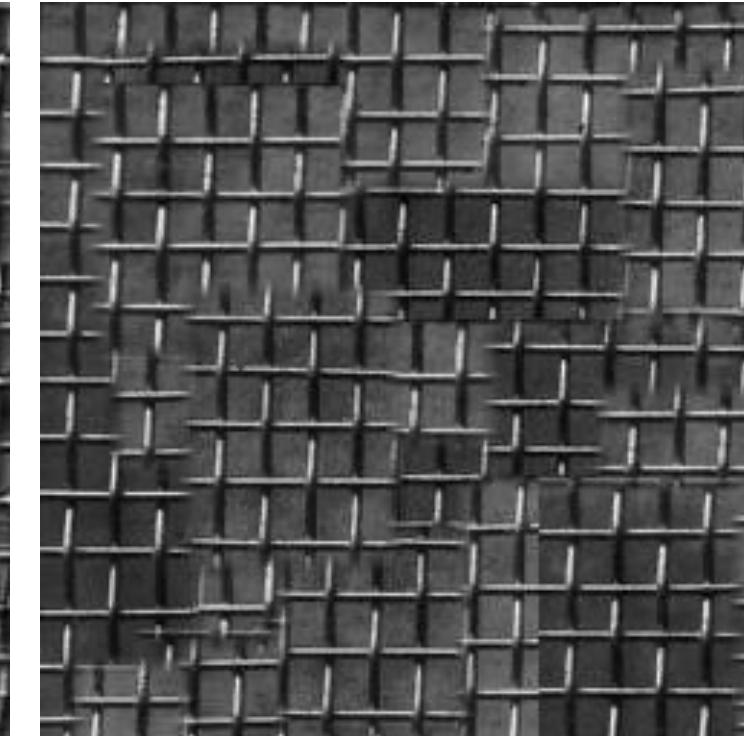




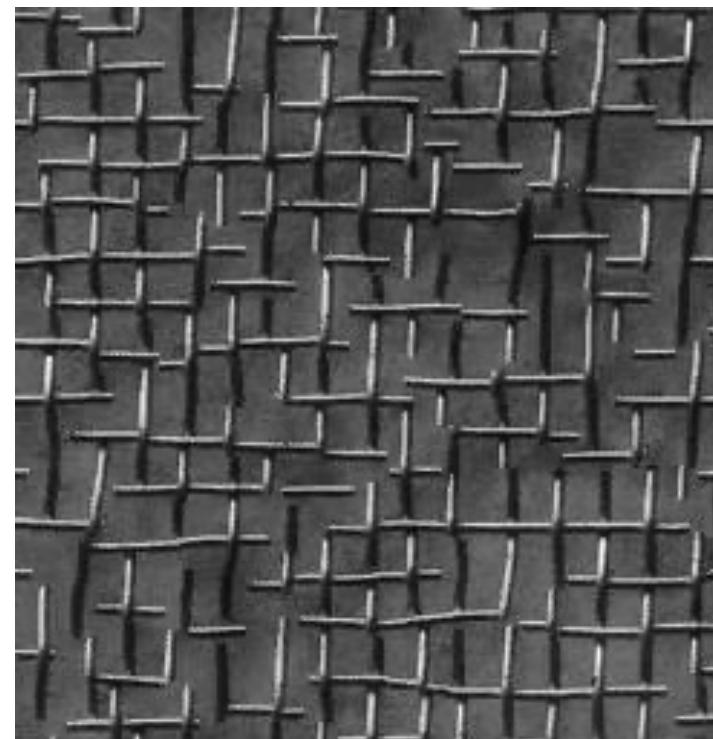
input image



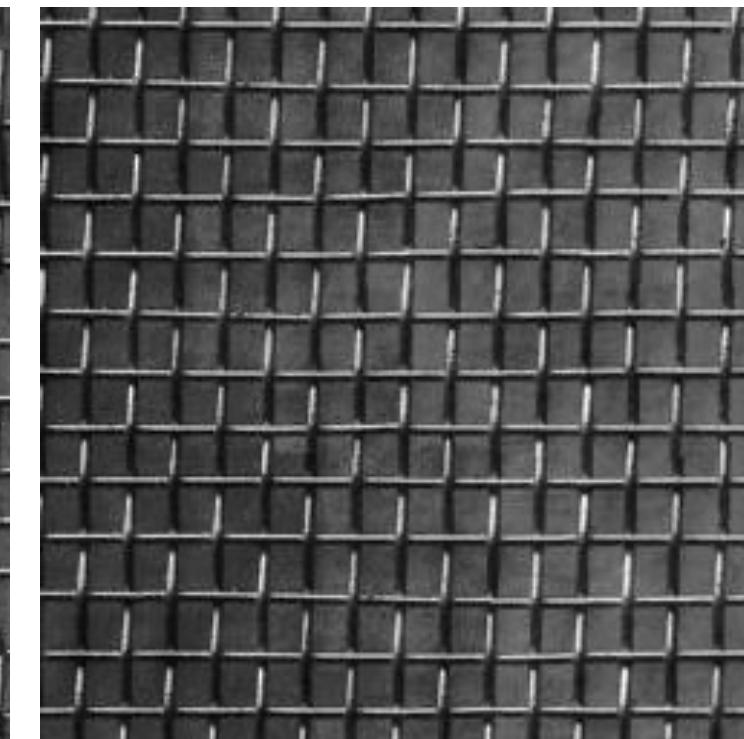
Portilla & Simoncelli



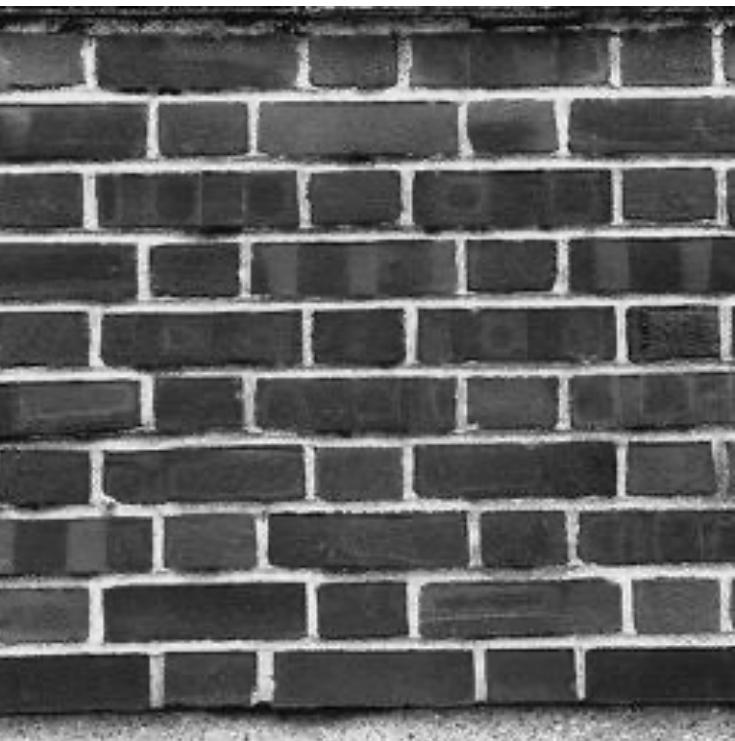
Xu, Guo & Shum



Wei & Levoy



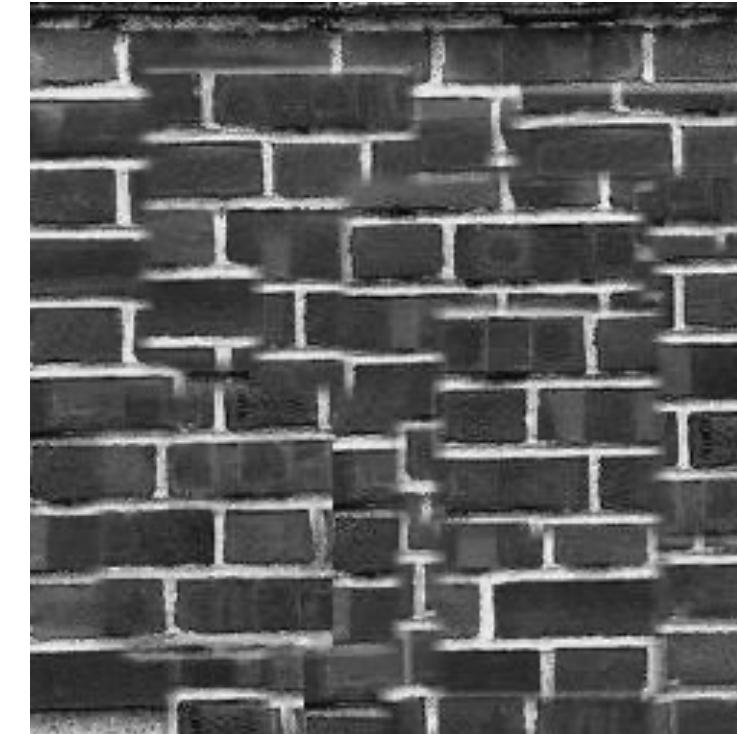
Efros and Freeman



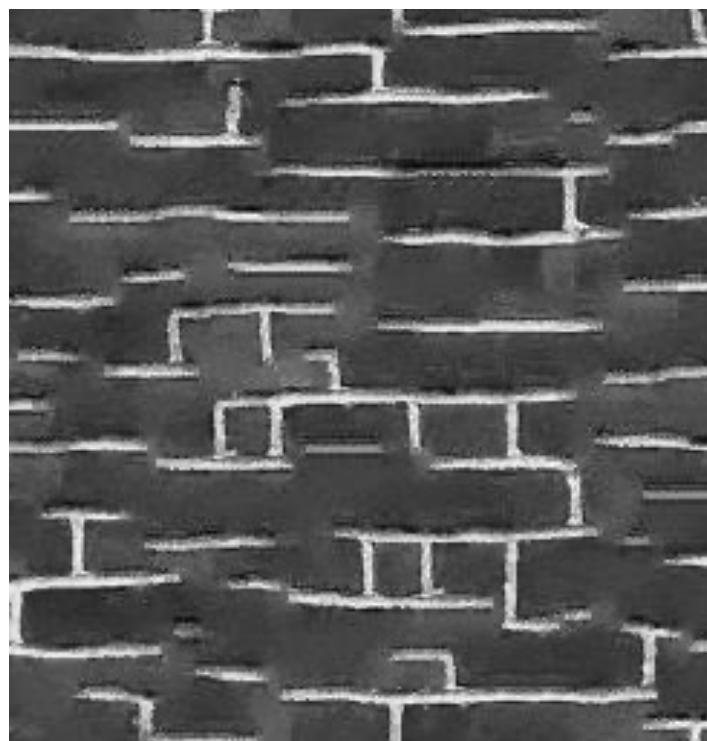
input image



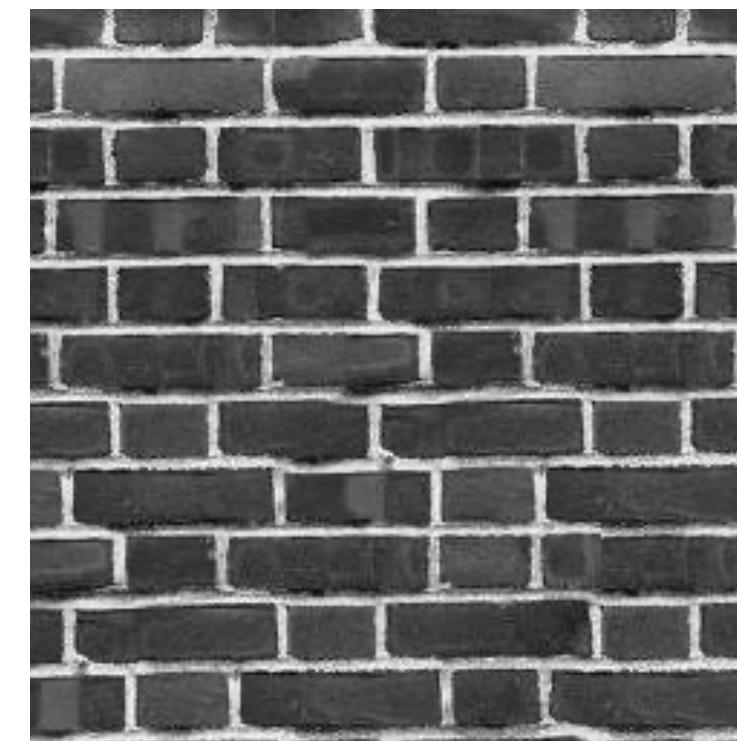
Portilla & Simoncelli



Xu, Guo & Shum



Wei & Levoy



Efros and Freeman

describing the response of that neuron as a function of position—is perhaps the most functional description of that neuron. We seek a single conceptual and mathematical framework to describe the wealth of simple-cell receptive fields neurophysiologically¹⁻³ and inferred especially if such a framework has the virtue of being able to tell us how it helps us to understand the function in a deeper way. Whereas no generic model can account for all simple-cell receptive fields (DOG), difference of offset Gabor, derivative of a Gaussian, higher derivative of a Gaussian, higher derivative of a function, and so on—can be expected to account for all simple-cell receptive fields, we nonetheless have to make some assumptions about the properties of the stimulus.

input image

Portilla & Simoncelli

... sen ois
ols nnuunncce tiarpn, nelolc ewiomsir
car es eince, racy rap:is ss
esoeao so ecxcecd rep ones, so. in.
uogrs e—n-cessarie or is ind hnn
a—ccisrctneseactne dsi one
deientn- eice secimn,ii eisnchrans
n heiatlin.—cicenn,aimnn: cepureoe
se ononass is if emn.
hal dell eieuomn) fitlymr rd thon
cincare trnaciscer tfr:moes fullsiD
in pactuewn coisa-iss runnl re .ll cas
si omnooesl —a nae maeike ne wen
unnting fped ole-can usnsnplm nf
ntu oorenmp —— sreness nnt

Wei & Levoy

and of a visual cortical neuron—the in-
des and so—^{as} Gaussian, ^{that} neuro-
and mathem. ^{the} ^{as}
nt a simple-cell recep^t so ^a ^{as}
uns and inferred ^{tic} ^{bing} ^t ^{function}
sd neurophysiol ^{es} ^{funct} ^{seek} ^a
especially if suc ^{ussians} ^{onal} ^{describ}
t helps us to uirivative ^{ad} ^{single} ^{connec}
eeeper way. ^{er} ^{function}, ^{cent}
tissians (DOG) simple-cell ^{ight} ^{at neu}
eepe ^l ^{offset} ^{Cus}, ^{visam} ^{is perh}
ussia ^{higher} ^{derivative} ^{field} ^{ng} ^{neuro}
ivat ^{scription} ^{of} ^{that} ^{her} ^{an} ^{mathem}
fun alth of simple- ^{conceptual} ^{and} ⁱⁿ ^{seek} ^d ^{cell rec}
implologically¹⁻³ an position—that m

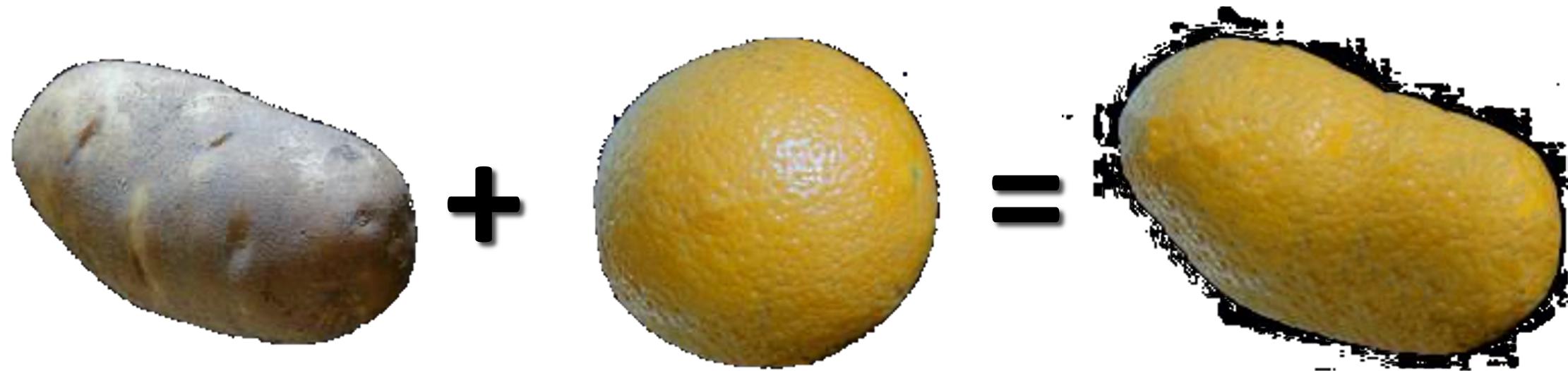
Xu, Guo & Shum

ition—is perk a single conceptual and of that neurube the wealth of simpleal and matheurophysiologically¹⁻³ an simple-cell recially if such a framework¹⁻³ and infer:ps us to understand the mework has perhay. Whereas no ge and the fumeuro:DOG), difference o no generic a single conceptual and n ence of offse the wealth of simple-c higher deriescribing the response of —can be expes a function of position helps us to understand theption of the per way. Whereas no gconceptual an ians (DOG), differencealth of simpl

Efros and Freeman

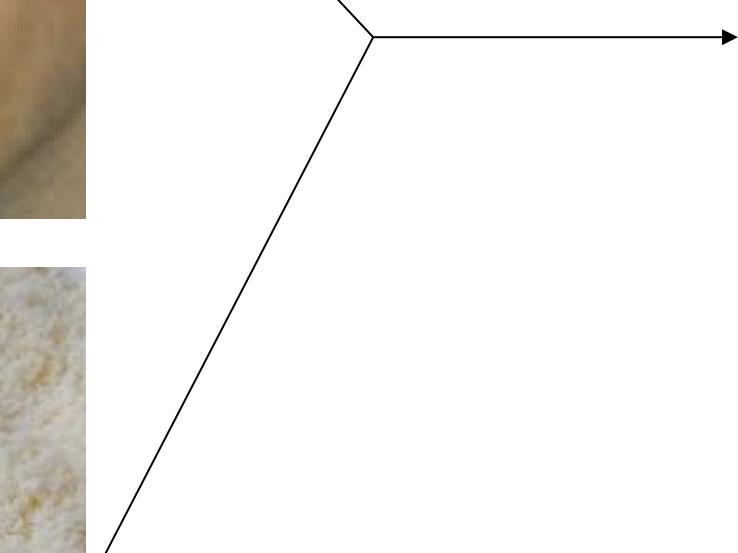
Application: Texture Transfer

- Try to explain one object with bits and pieces of another object:



Texture Transfer

Constraint



Texture sample

Texture Transfer

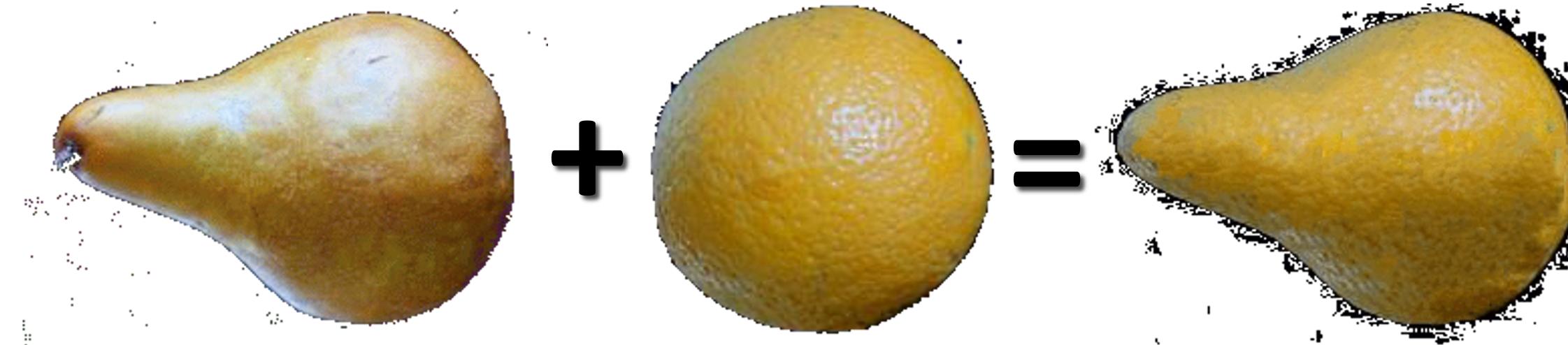
- Take the texture from one image and “paint” it onto another object



Same as texture synthesis, except an additional constraint:

1. Consistency of texture
2. Similarity to the image being “explained”

Texture Transfer



Texture Transfer

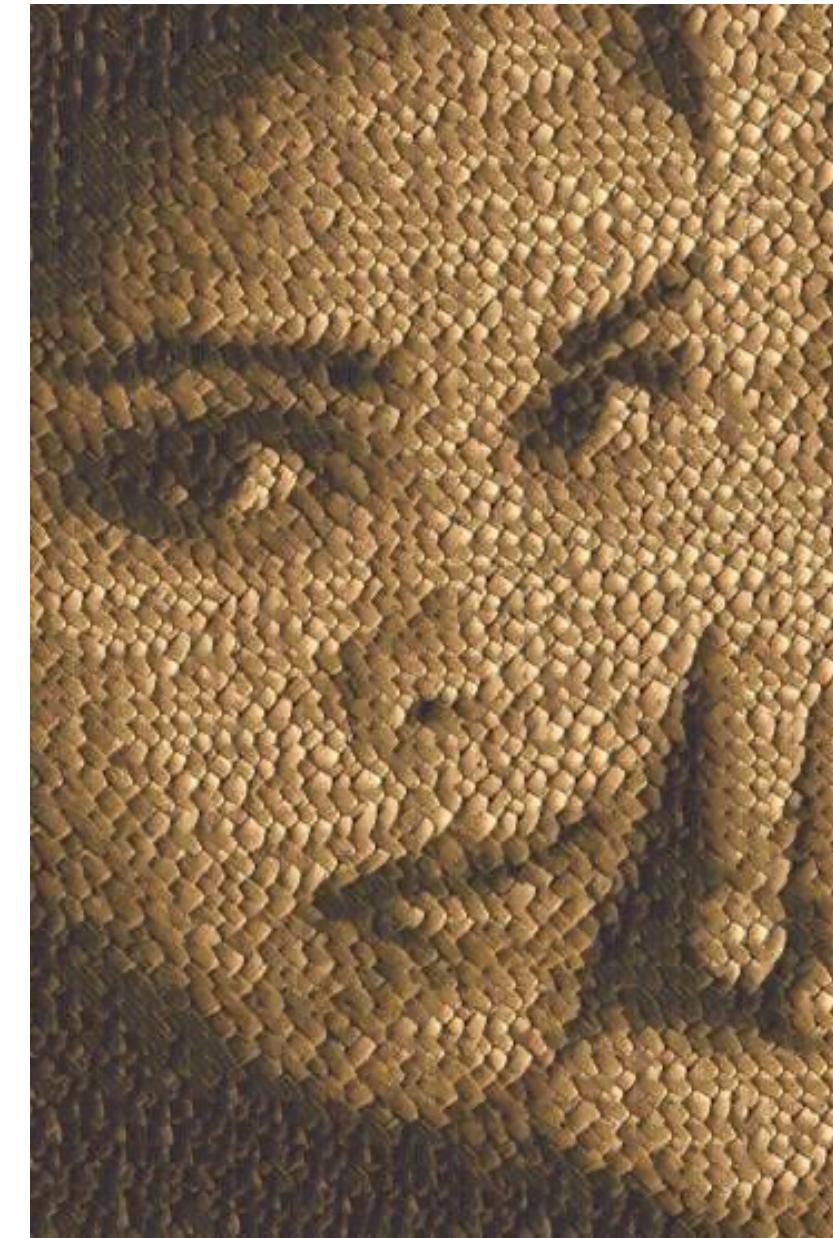
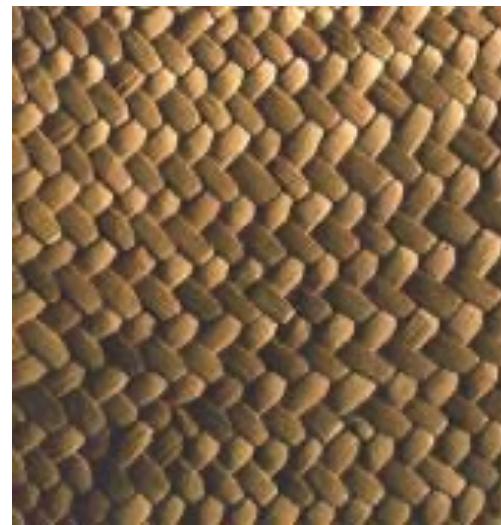


Image Analogies

Aaron Hertzmann^{1,2}

Chuck Jacobs²

Nuria Oliver²

Brian Curless³

David Salesin^{2,3}

¹**New York University**

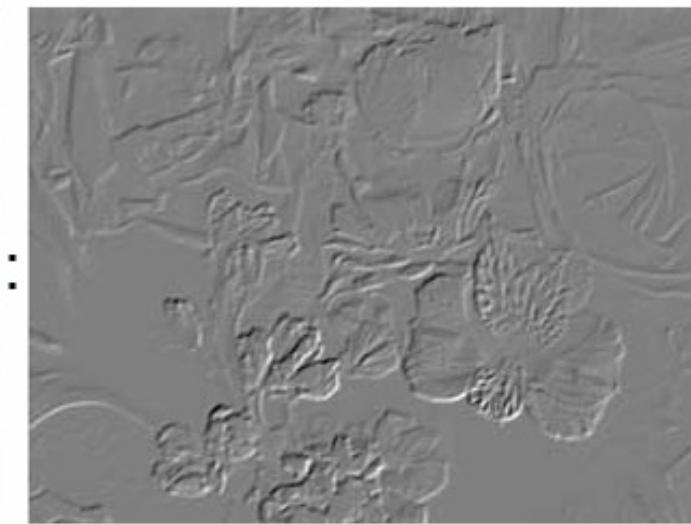
²**Microsoft Research**

³**University of Washington**

Edge Filter



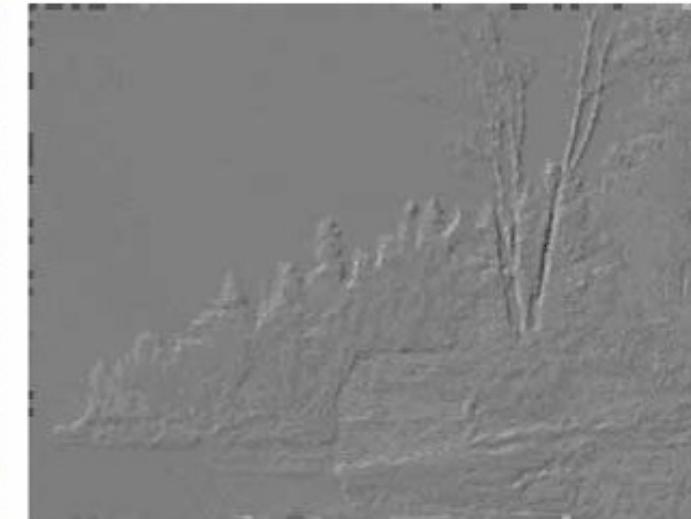
Unfiltered source (A)



Filtered source (A')



Unfiltered target (B)



Filtered target (B')

Artistic Filters



A



A'



B



B'

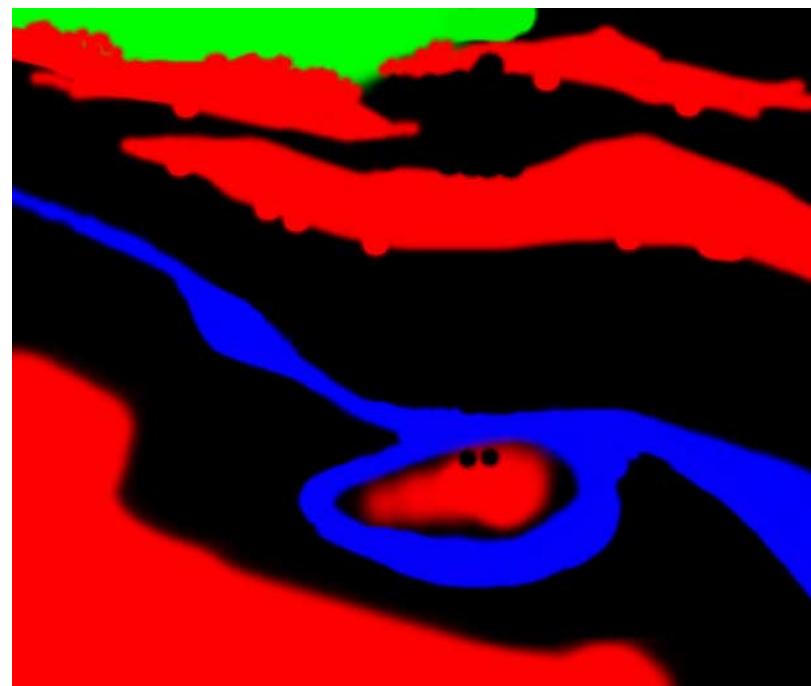
Texture-by-numbers



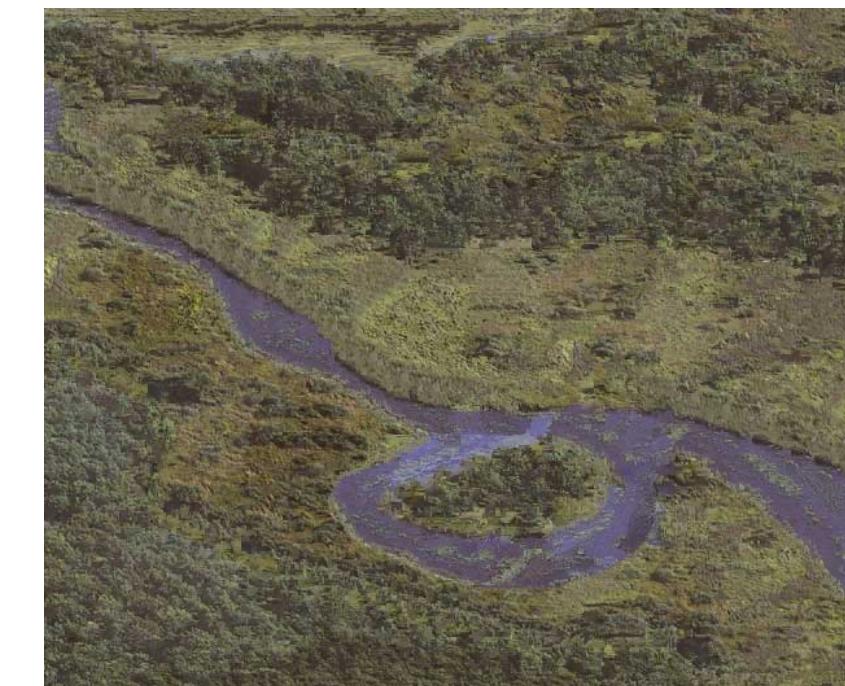
A



A'

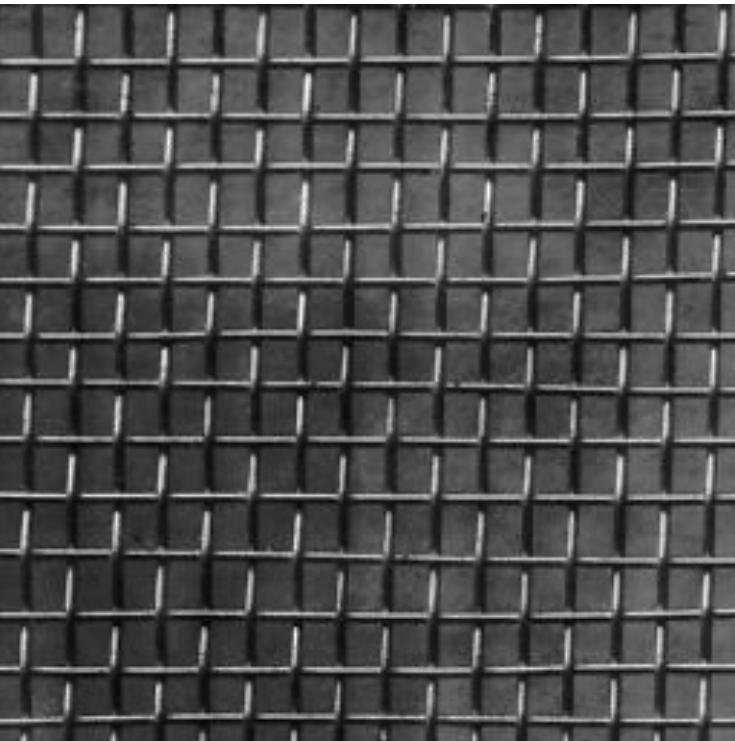


B

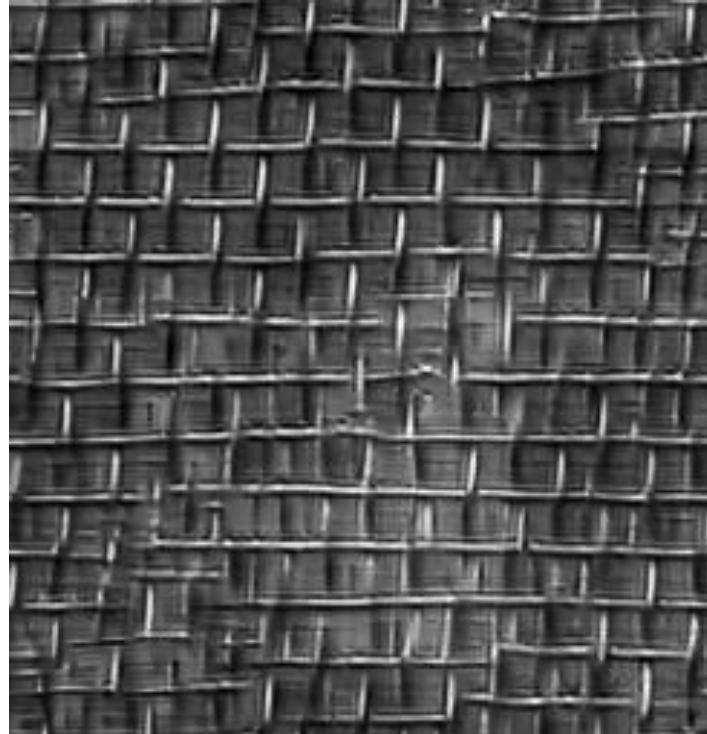


B'

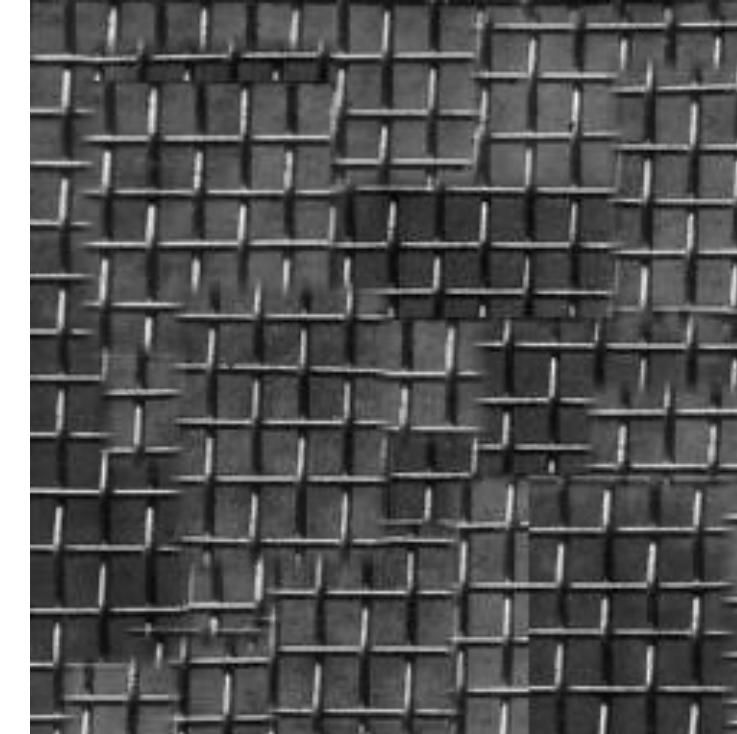
Parametric Texture Synthesis



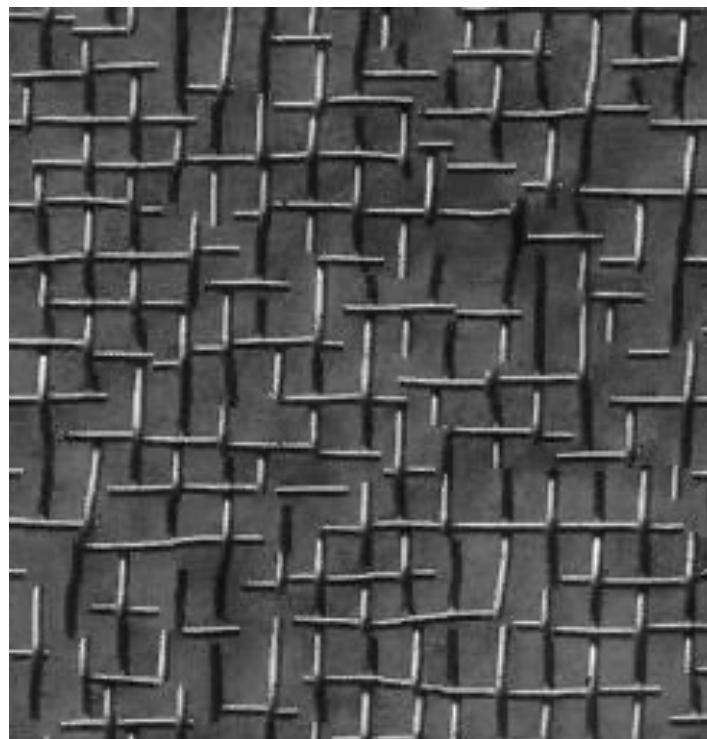
input image



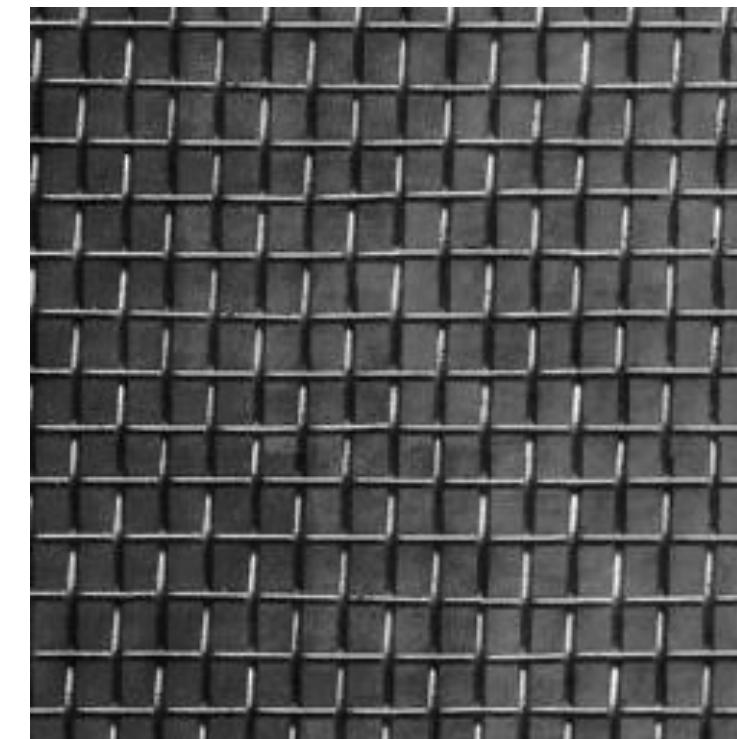
Portilla & Simoncelli



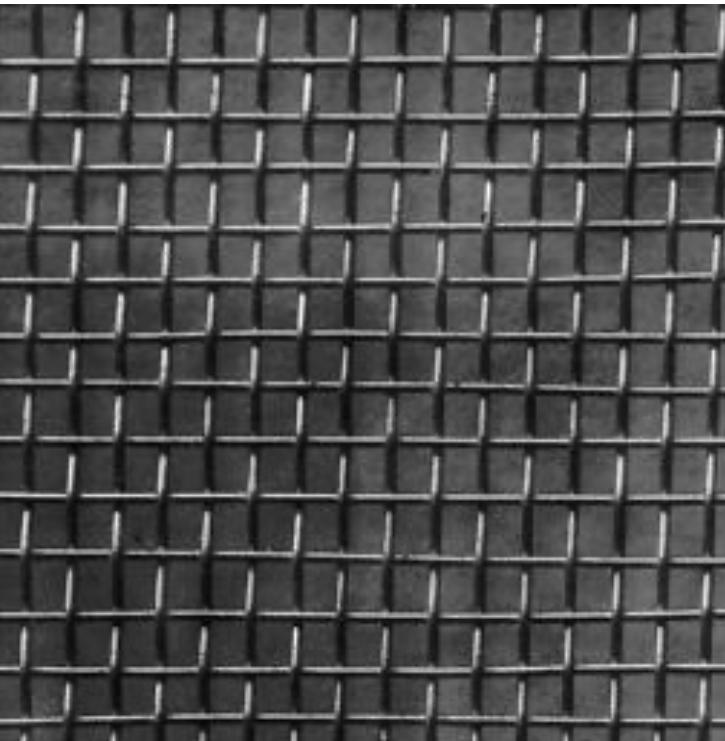
Xu, Guo & Shum



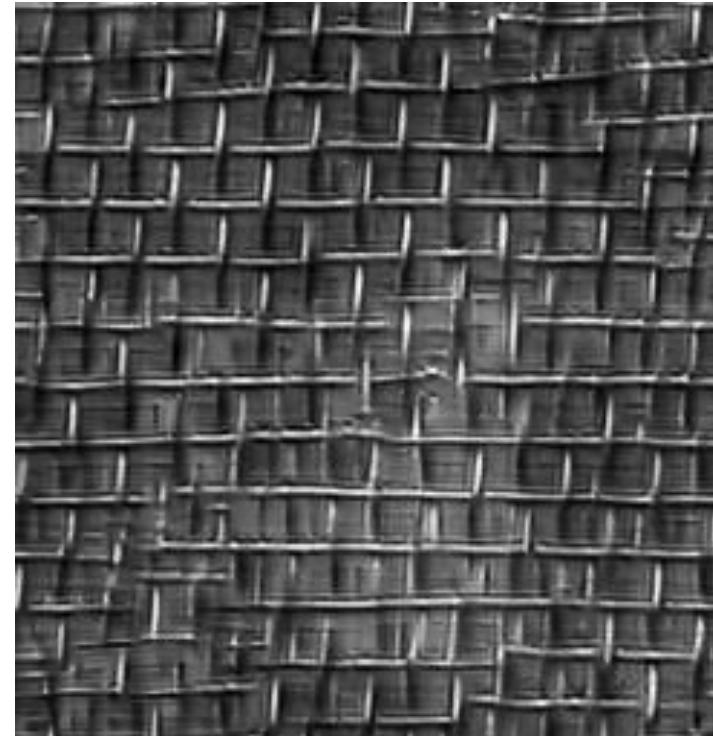
Wei & Levoy



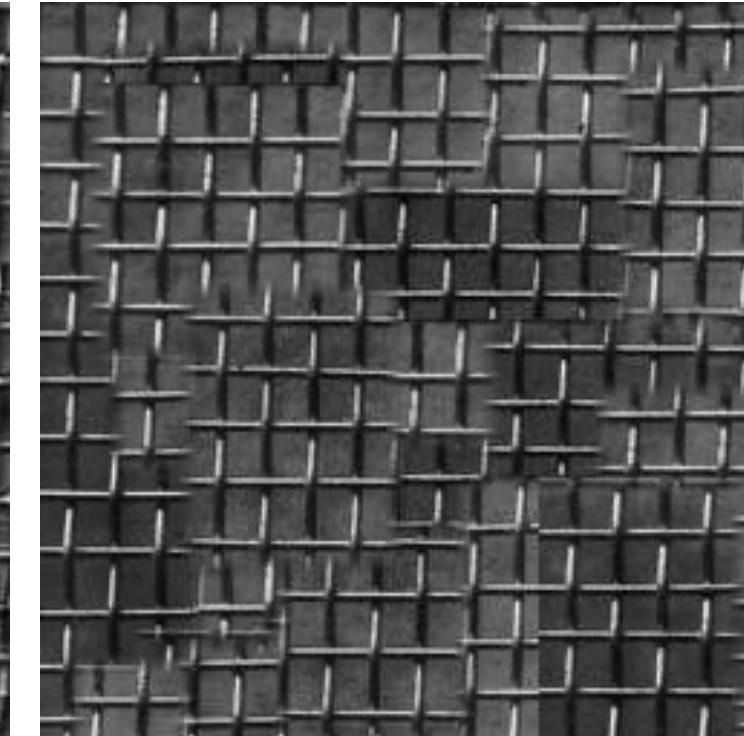
Efros and Freeman



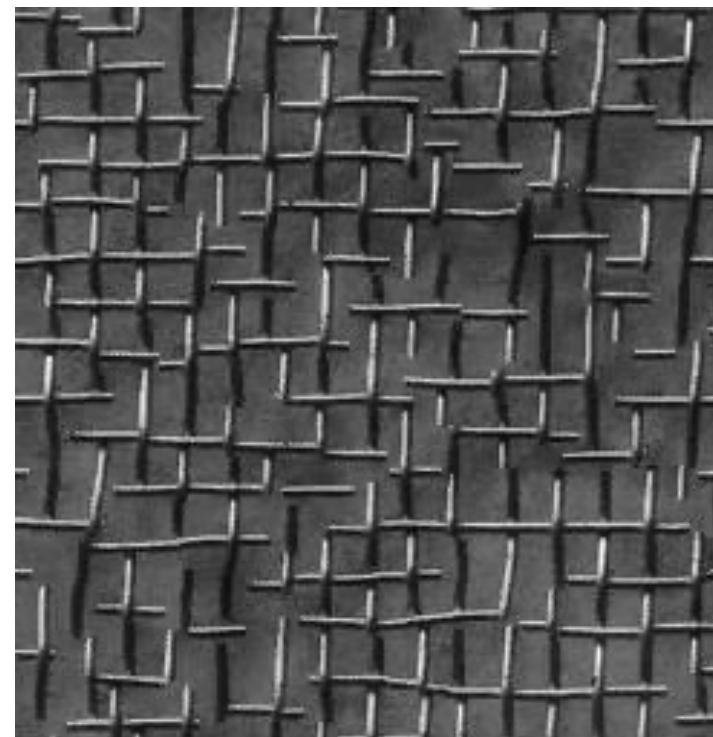
input image



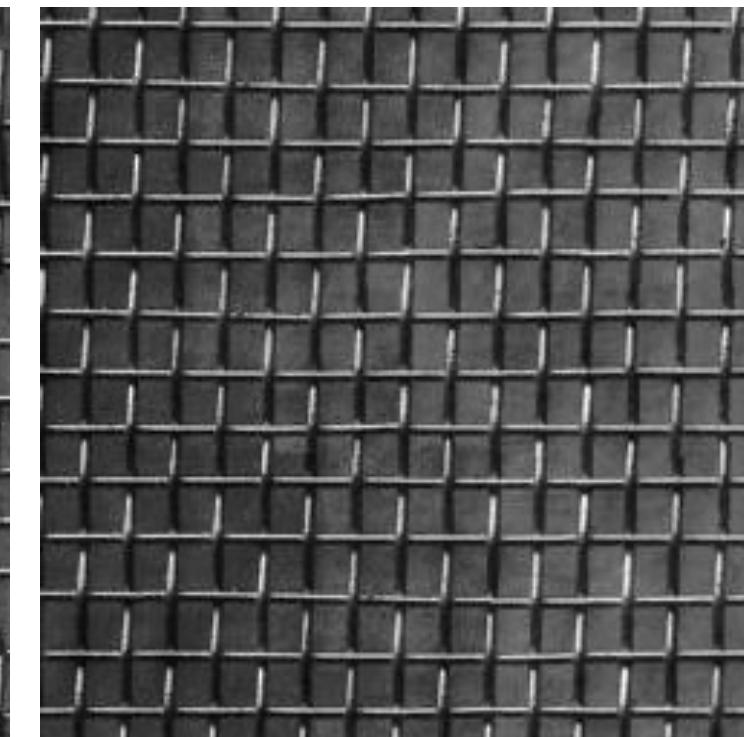
Portilla & Simoncelli



Xu, Guo & Shum

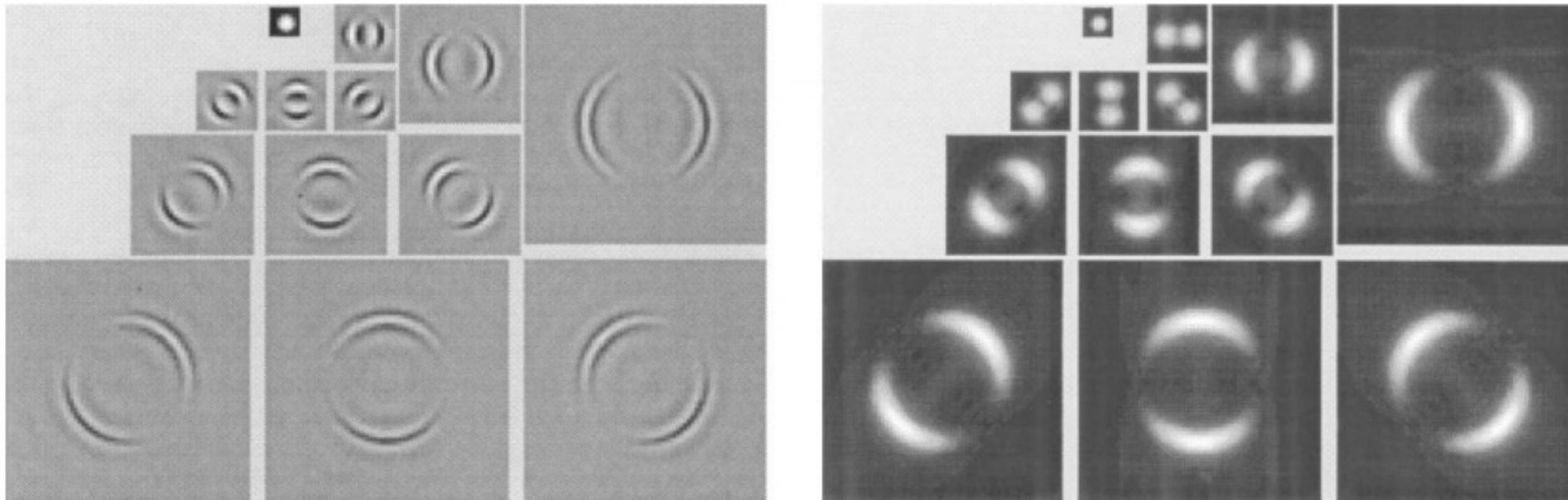


Wei & Levoy



Efros and Freeman

Parametric Texture Synthesis



Histogram and cross-channel correlation using wavelet basis

Statistics $\longrightarrow \mathcal{E}(\phi_j(y)) \approx \mathcal{E}(\phi_j(\hat{y}))$

Wavelet features

A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients
Portilla and Simoncelli, IJCV 1999

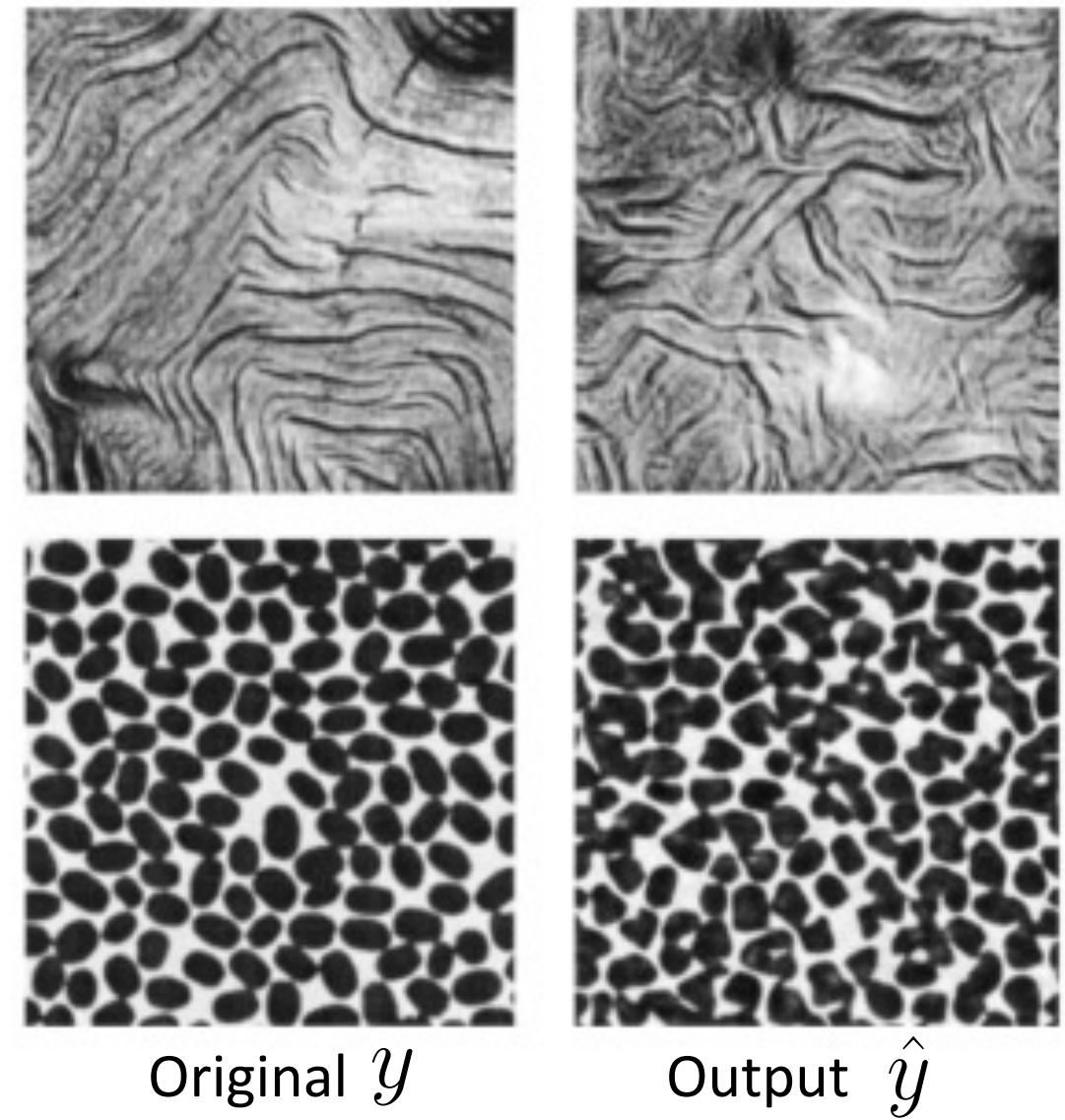
Parametric Texture Synthesis

Objective function

Given input texture y , feature descriptor ϕ ,
and statistics summary function \mathcal{E}

We aim to optimize the output image \hat{y}

$$\hat{y}^* = \arg \min_{\hat{y}} \|\mathcal{E}(\phi_j(\hat{y})) - \mathcal{E}(\phi_j(y))\|$$



Deep Learning Version

Gram matrix:

- Cross Correlation of CNN features
- Invariant to the feature locations

$$V = [v_1, v_2, \dots, v_n]$$

$$G_{ij} = \langle v_i, v_j \rangle \quad G = V^\top V$$

$$\text{Gram}^{(j)}(x) = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(x)_{h,w,c} \phi_j(x)_{h,w,c'}.$$

h, w: pixel locations index

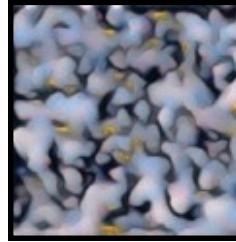
c: channel index

H, W: height and width of feature map

C: the number of total channels

Style Reconstruction (Style Loss)

$$|\text{Gram}(\hat{y}) - \text{Gram}(y)|$$


optimized output 
style image

Gram = Gram Matrix of a deep network's features (e.g., ImageNet classifier)

Style Loss

$$\arg \min_{\hat{y}} \sum_j^M \lambda_j ||\text{Gram}^{(j)}(\hat{y}) - \text{Gram}^{(j)}(y)||^2$$

weight
 \downarrow
 M
 j

(j)-th layer

Portilla & Simoncelli

original



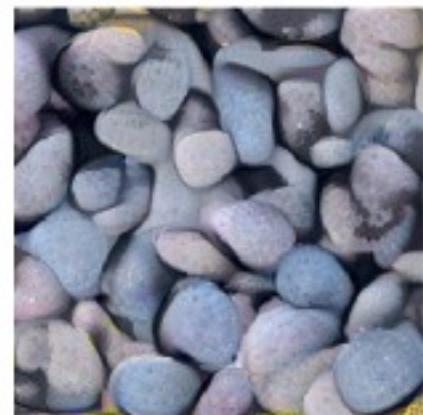
pool4



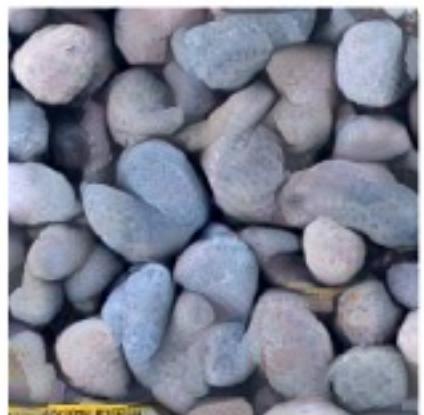
pool3



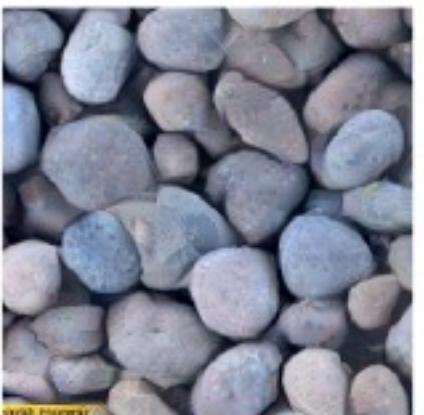
A ~1k parameters



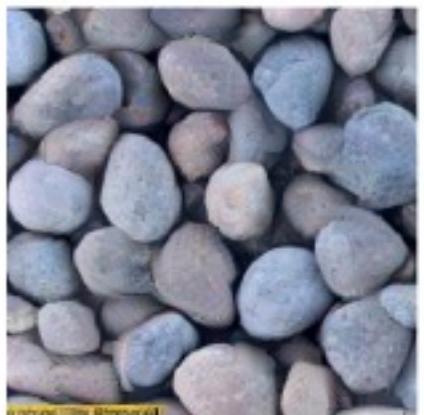
~10k parameters



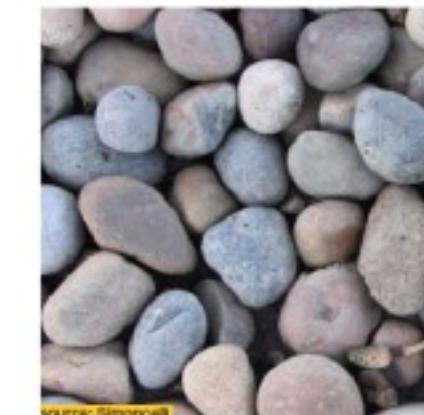
~177k parameters



~852k parameters

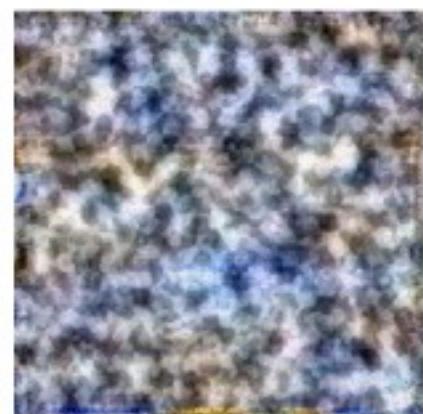


original

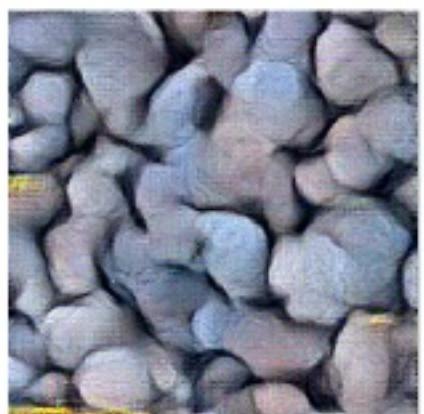


Number of parameters

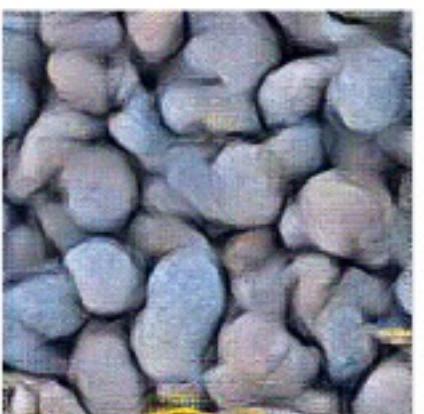
B conv1



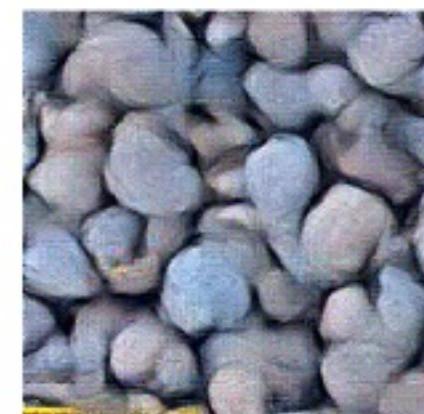
conv2



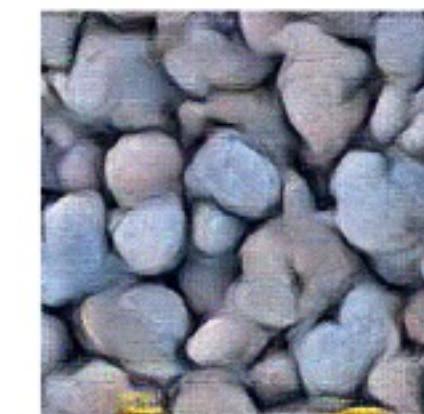
conv3



conv4

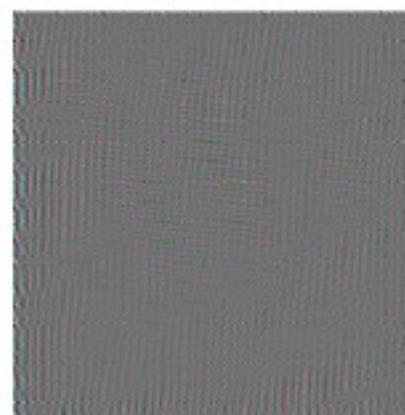


conv5

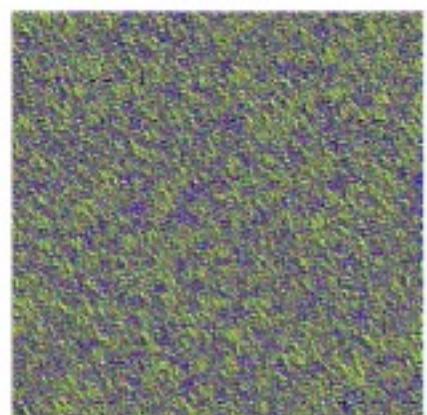


Different layers

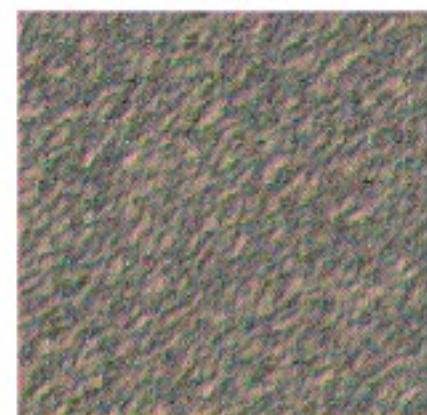
C conv1_1



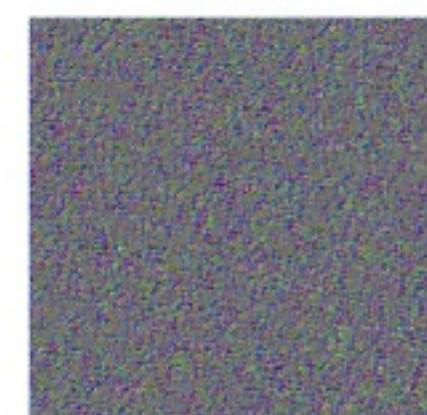
pool1



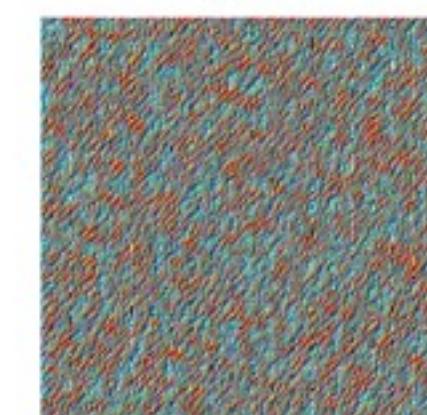
pool2



pool3



pool4



The same network architecture with random weights

Neural Style Transfer



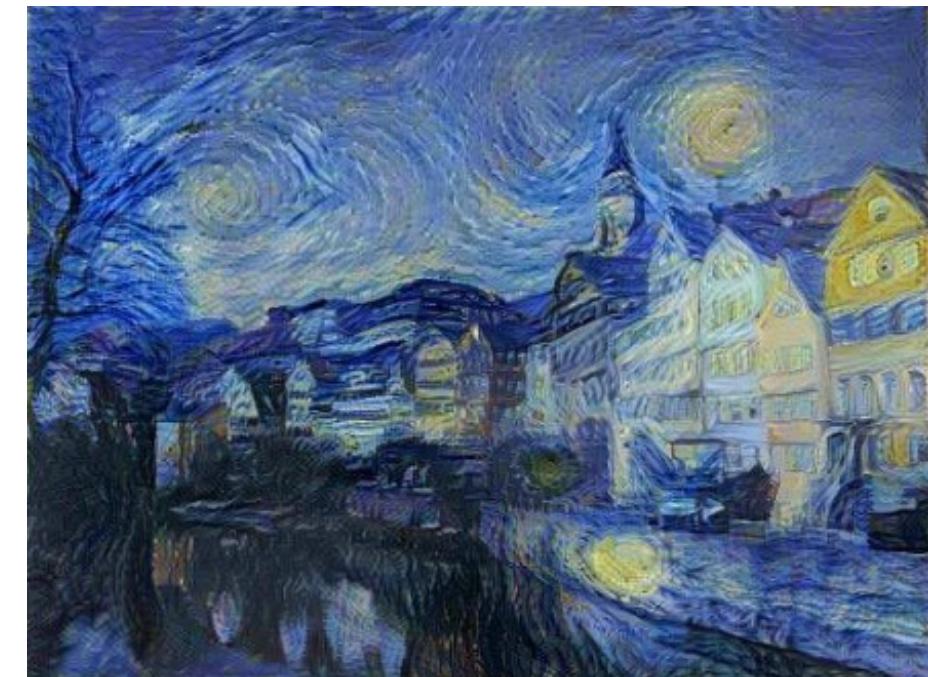
content image

+



style image

=



output result

Content Reconstruction (Perceptual Loss)

$$|\mathcal{F}(\hat{y}) - \mathcal{F}(x)|$$

optimized output content image

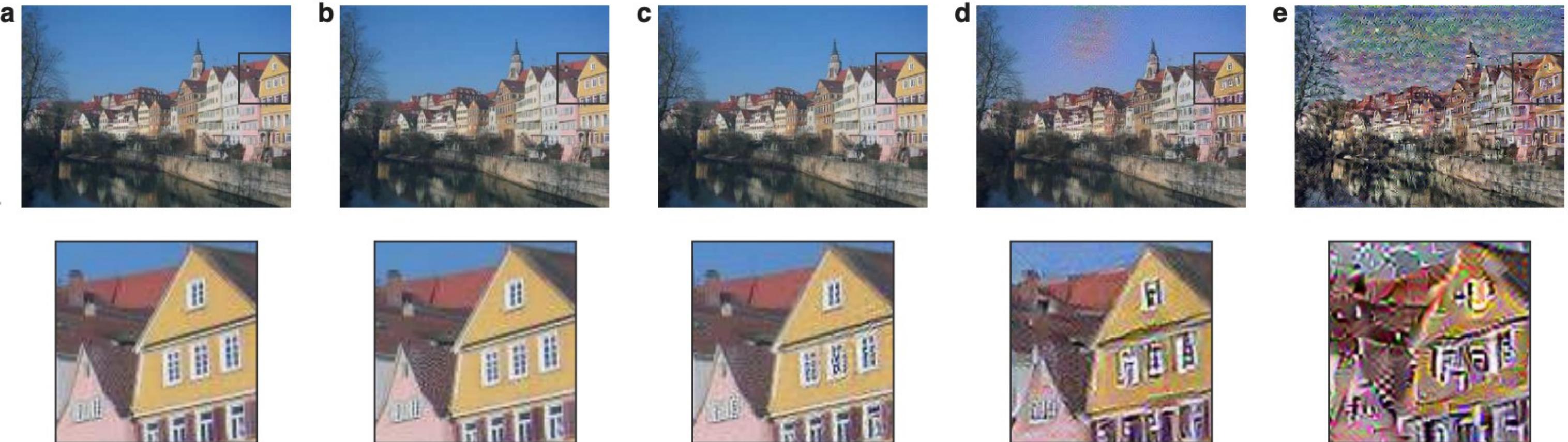
F is a deep network (e.g., ImageNet classifier)

Content Loss

LOSS

$$\arg \min_{\hat{y}} \sum_i \lambda_i ||F^{(i)}(\hat{y}) - F^{(i)}(x)||_1$$

Content Reconstruction (Perceptual Loss)



Conv1_2

Conv2_2

Conv3_2

Conv4_2

Conv5_2

Neural Style Transfer

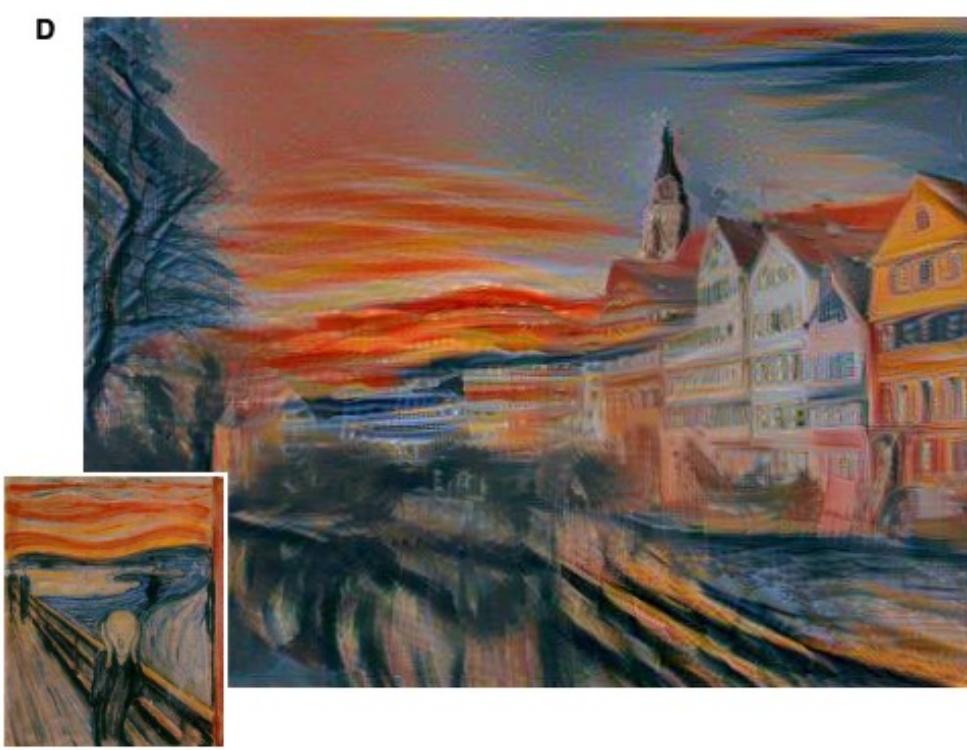
$$|\text{Gram}(\hat{y}) - \text{Gram}(y)|$$

 optimized output  style image

$$+ |\mathbf{F}(\hat{y}) - \mathbf{F}(x)|$$

 optimized output  content image

$$\arg \min_{\hat{y}} \mathcal{L}_{\text{style}}(\hat{y}, y) + \lambda \mathcal{L}_{\text{content}}(\hat{y}, x)$$



Different Initializations

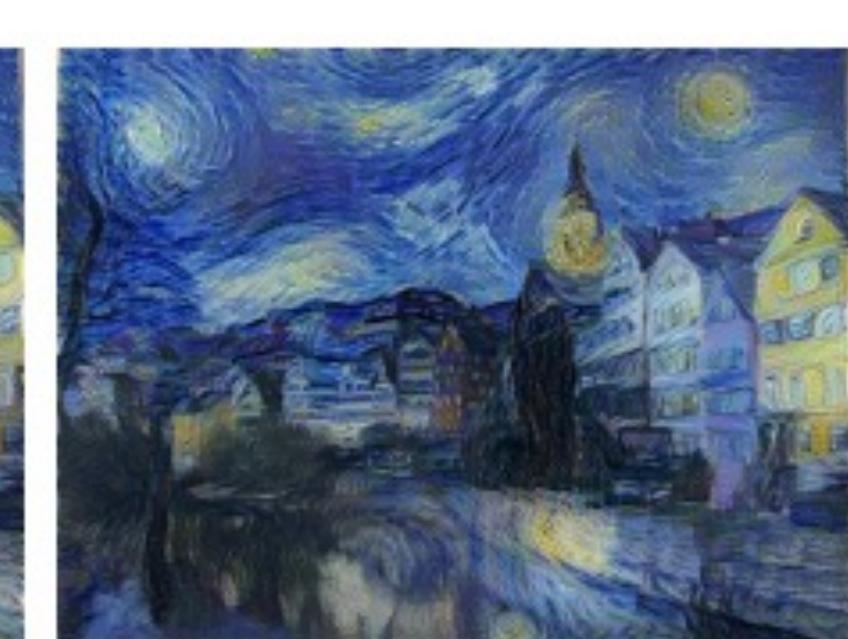
A



B



C



Fast Neural Style Transfer

- Optimization-based method

$$\arg \min_{\hat{y}} \mathcal{L}_{\text{style}}(\hat{y}, y) + \lambda \mathcal{L}_{\text{content}}(\hat{y}, x)$$

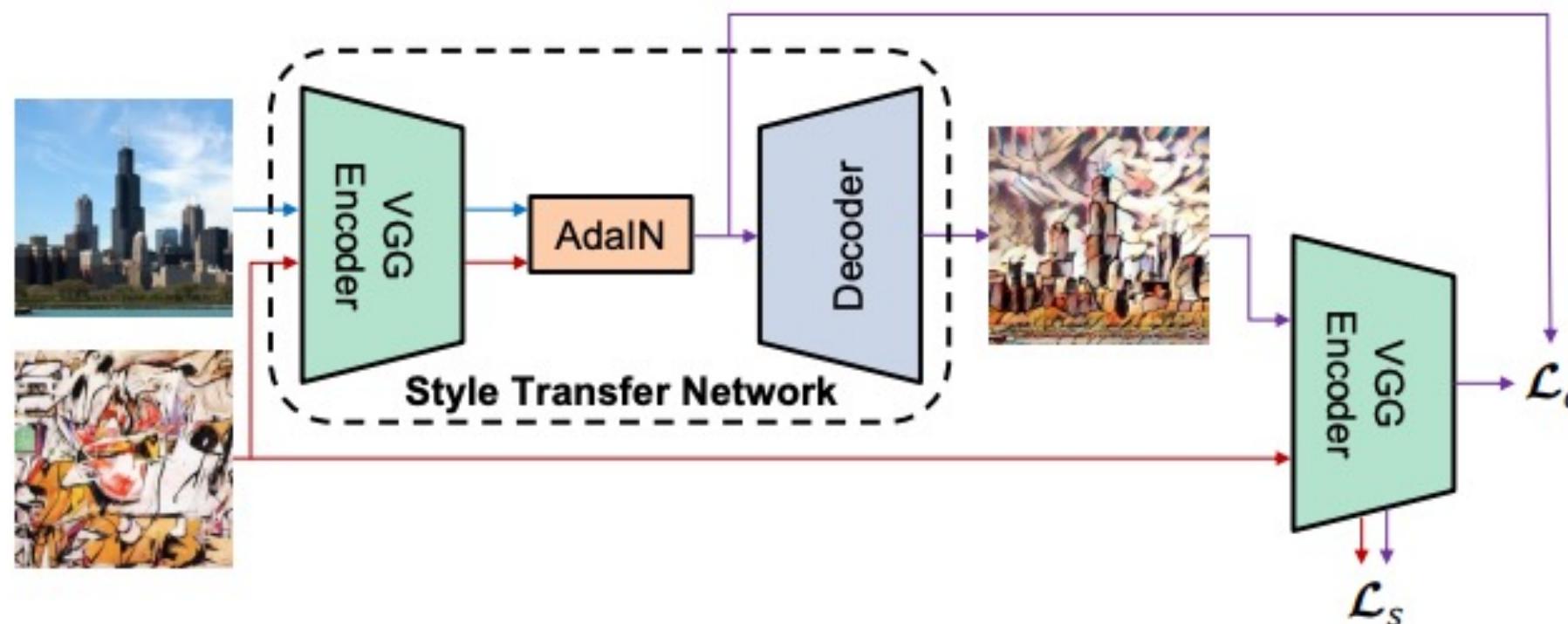
- Feedforward network

$$\arg \min_G \mathbb{E}_x \mathcal{L}_{\text{style}}(G(x), y) + \lambda \mathcal{L}_{\text{content}}(G(x), x)$$

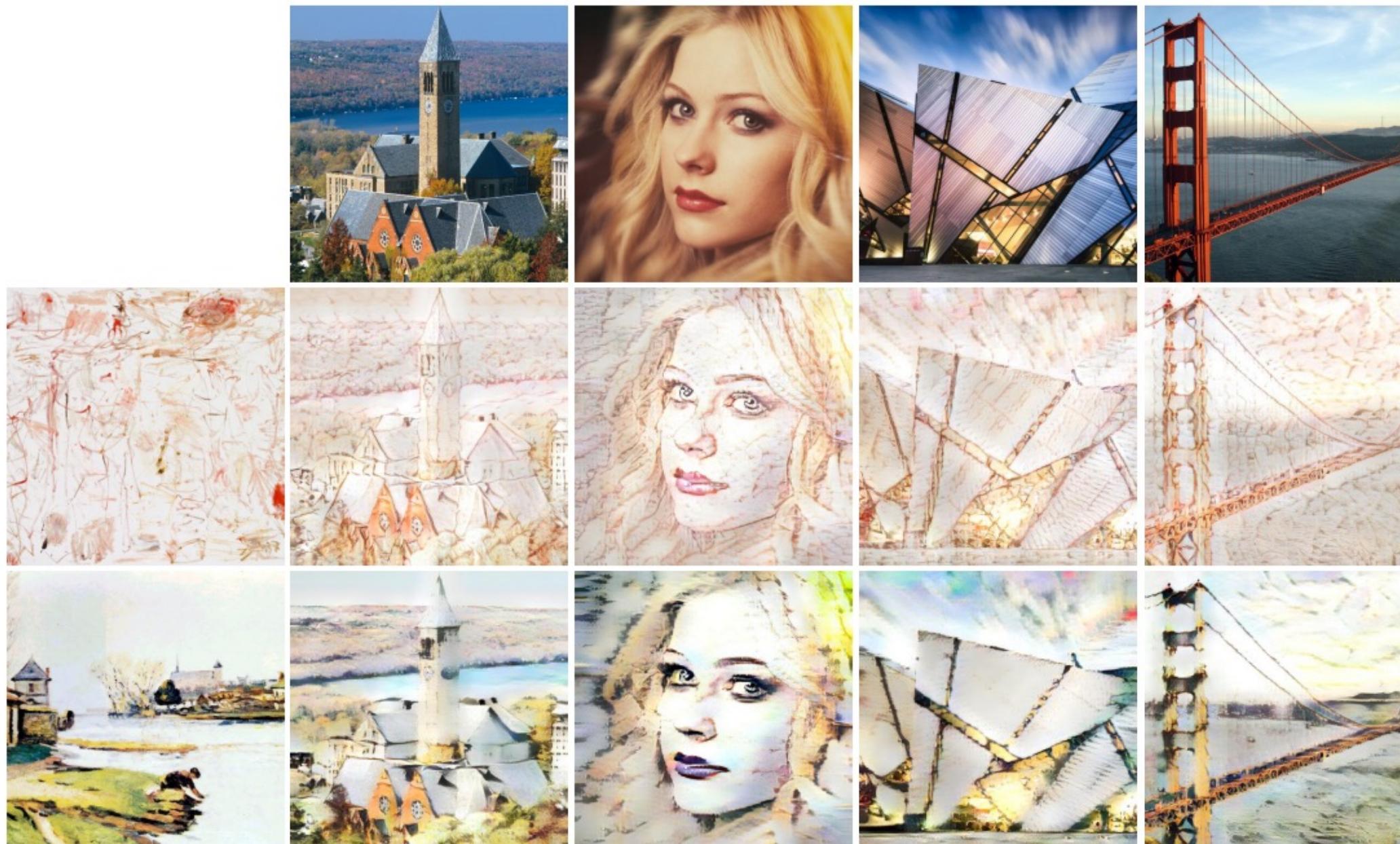
Arbitrary Style Transfer with AdaIN

- Feedforward network with any style

$$\arg \min_G \mathbb{E}_{x,y} \mathcal{L}_{\text{style}}(G(x,y), y) + \lambda \mathcal{L}_{\text{content}}(G(x,y), x)$$



Arbitrary Style Transfer with AdaIN

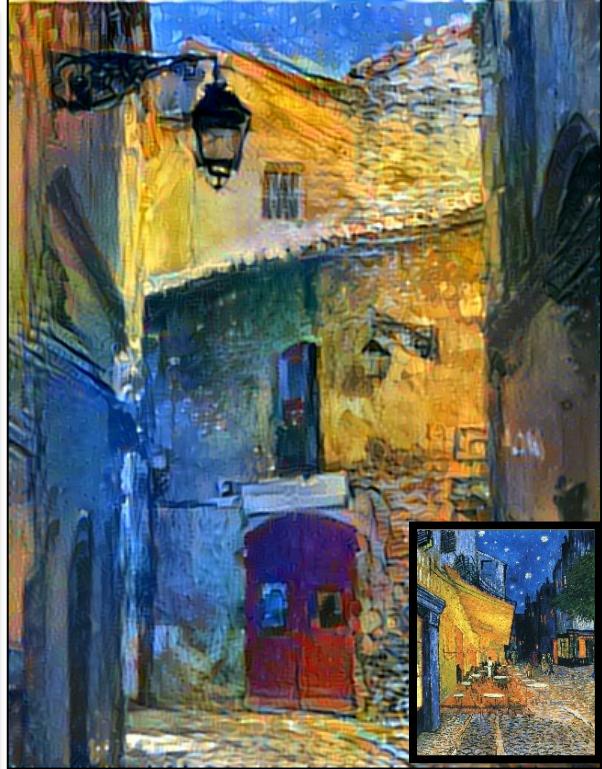


Neural Style Transfer vs. Image-to-Image Translation

Input



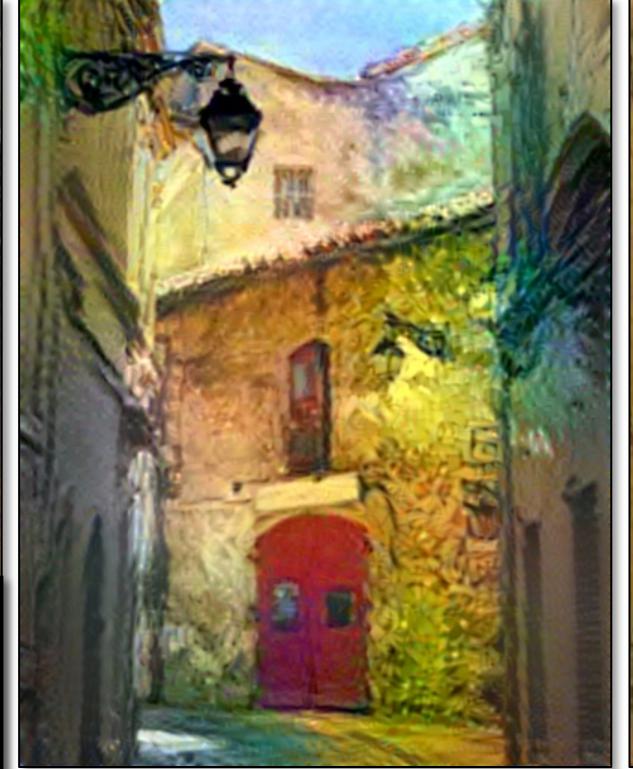
Style Image I



Style image II



Entire collection



CycleGAN

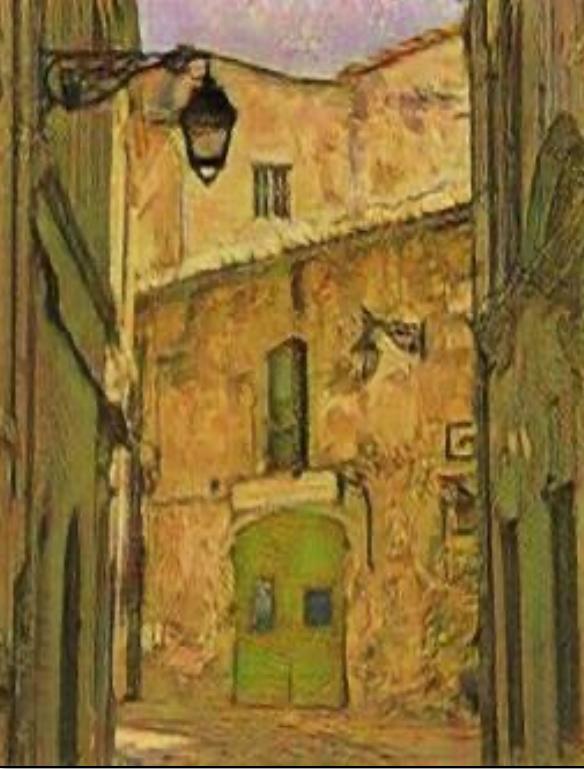


Photo → Van Gogh

Input



Style image I



Style image II



Entire collection



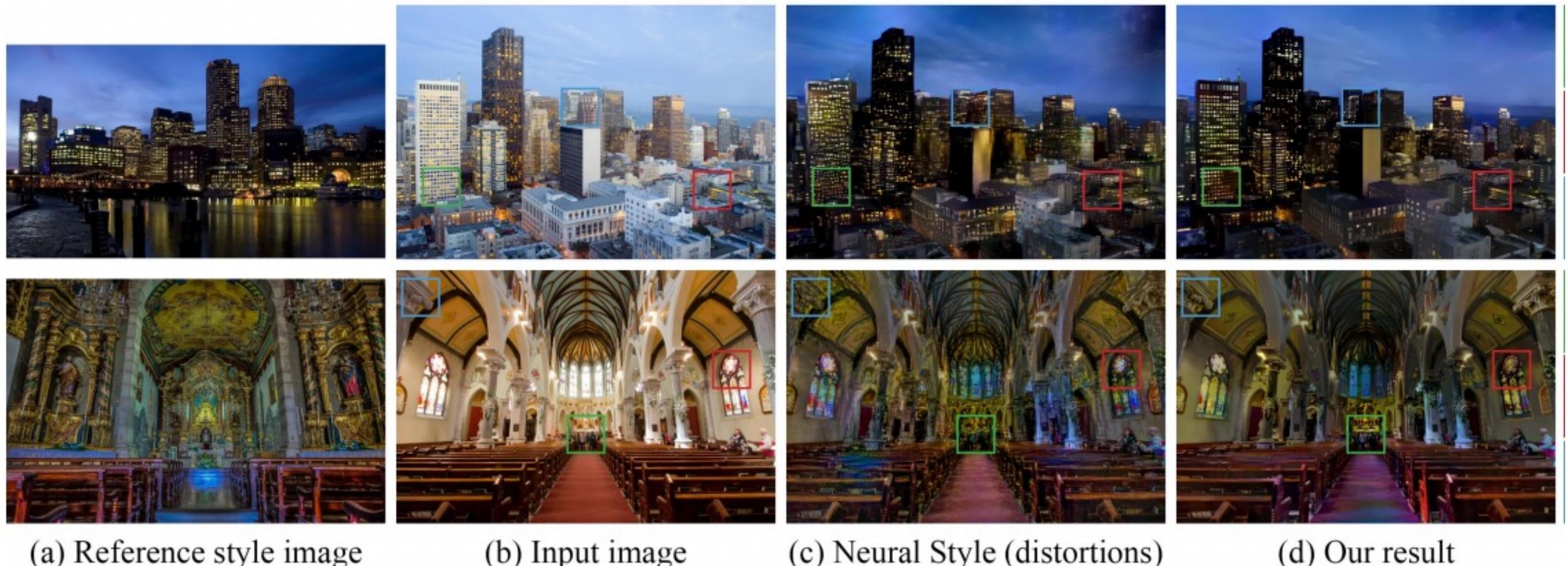
CycleGAN



horse → zebra

Photo Style Transfer

Deep Photo Style Transfer

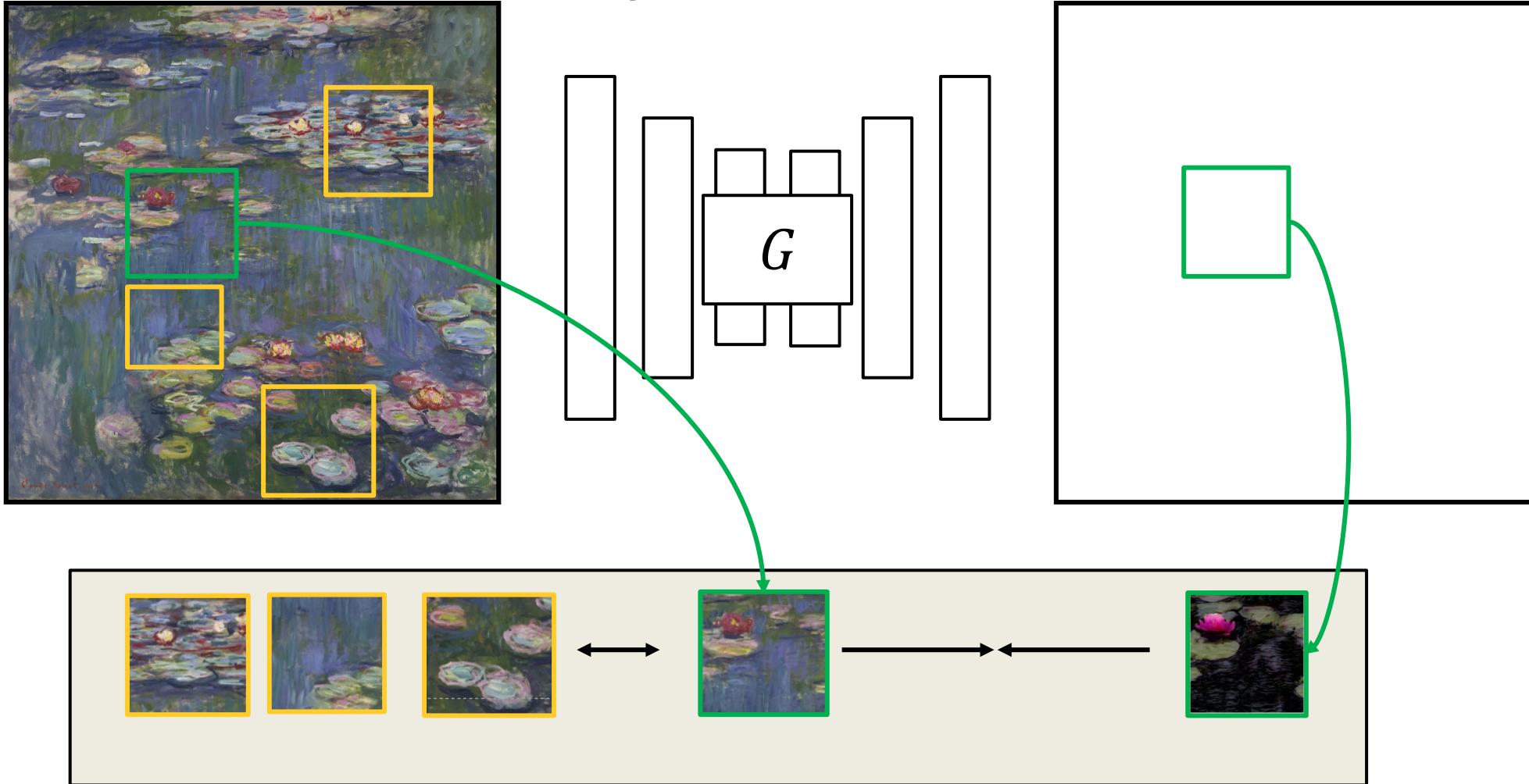


Local color transfer? (hard to transfer texture)

Single Image Translation

Single Image Translation

Claude Monet's painting

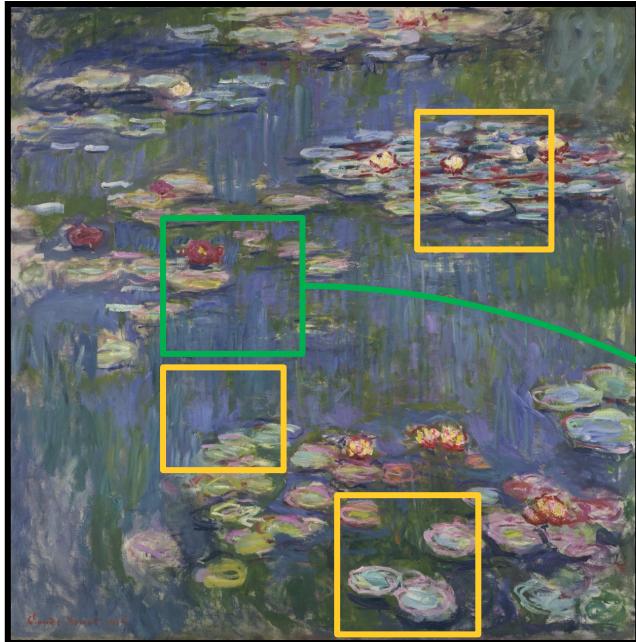


Internal contrastive loss is well-suited for single image translation.

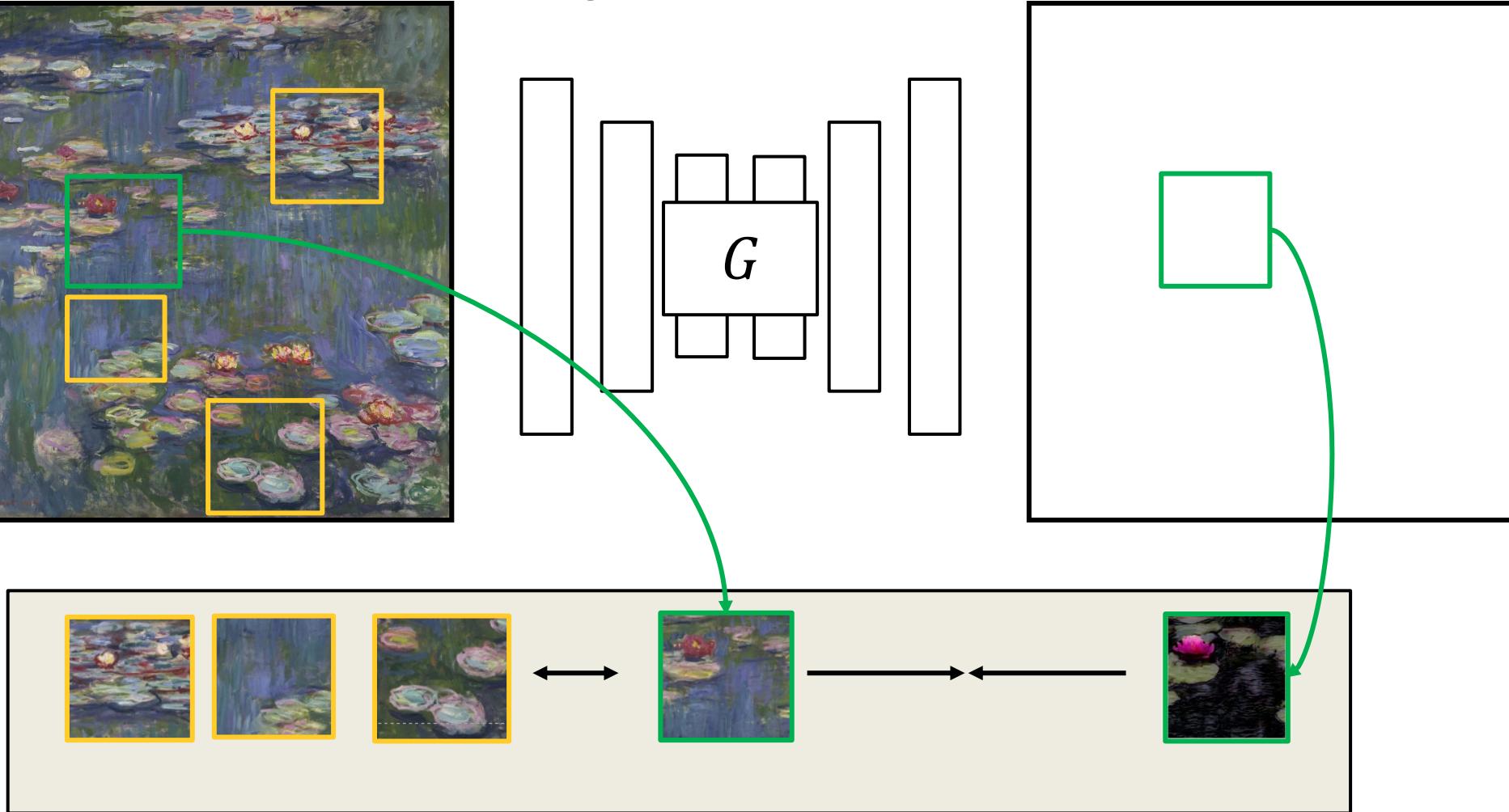
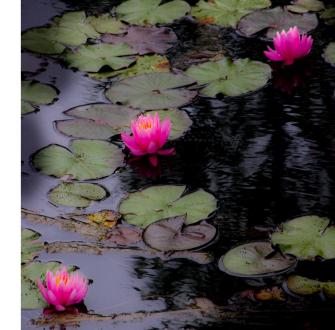
Also see InGAN (Shocher et al., ICCV'19), SinGAN (Shaham et al., ICCV'19)

Single Image Translation

Claude Monet's painting



Reference photo

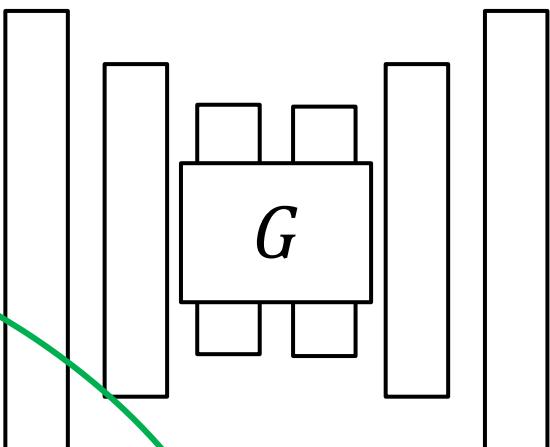
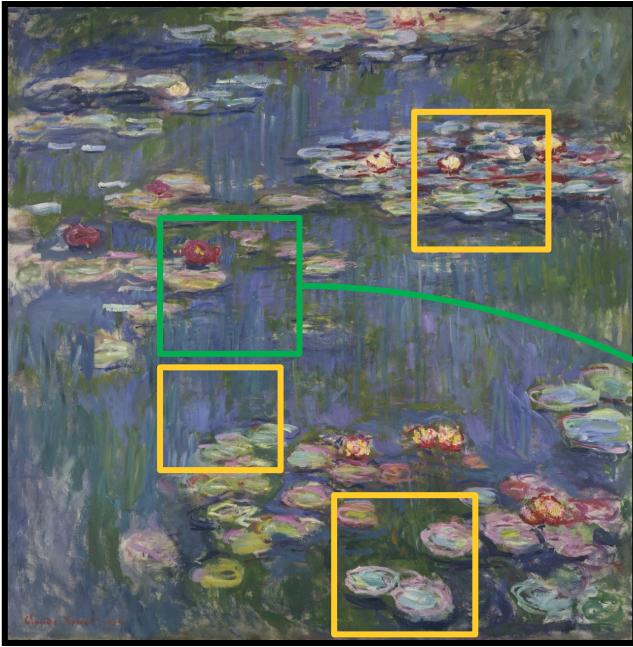


Internal contrastive loss is well-suited for single image translation.

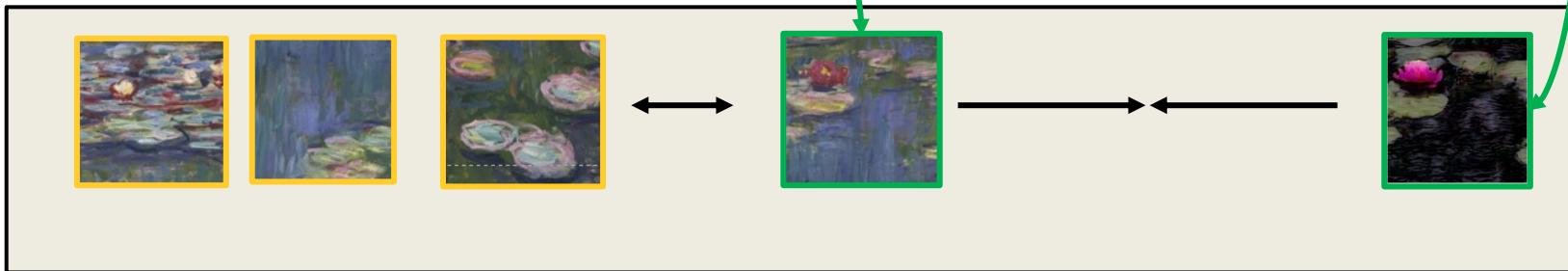
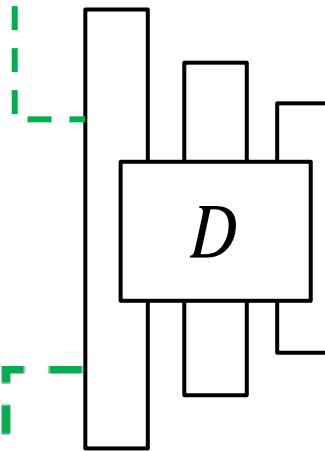
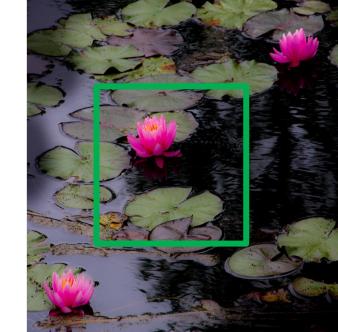
Also see InGAN (Shocher et al., ICCV'19), SinGAN (Shaham et al., ICCV'19)

Single Image Translation

Claude Monet's painting



Reference photo



Internal contrastive loss is well-suited for single image translation.

Also see InGAN (Shocher et al., ICCV'19), SinGAN (Shaham et al., ICCV'19)



Painting

Reference



Painting

Reference



Painting



Gatys et al. CVPR'16

Reference



Painting

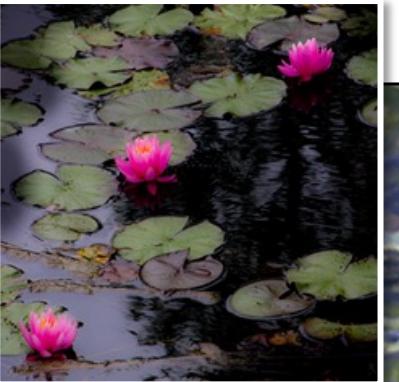


Single Image translation (CUT)

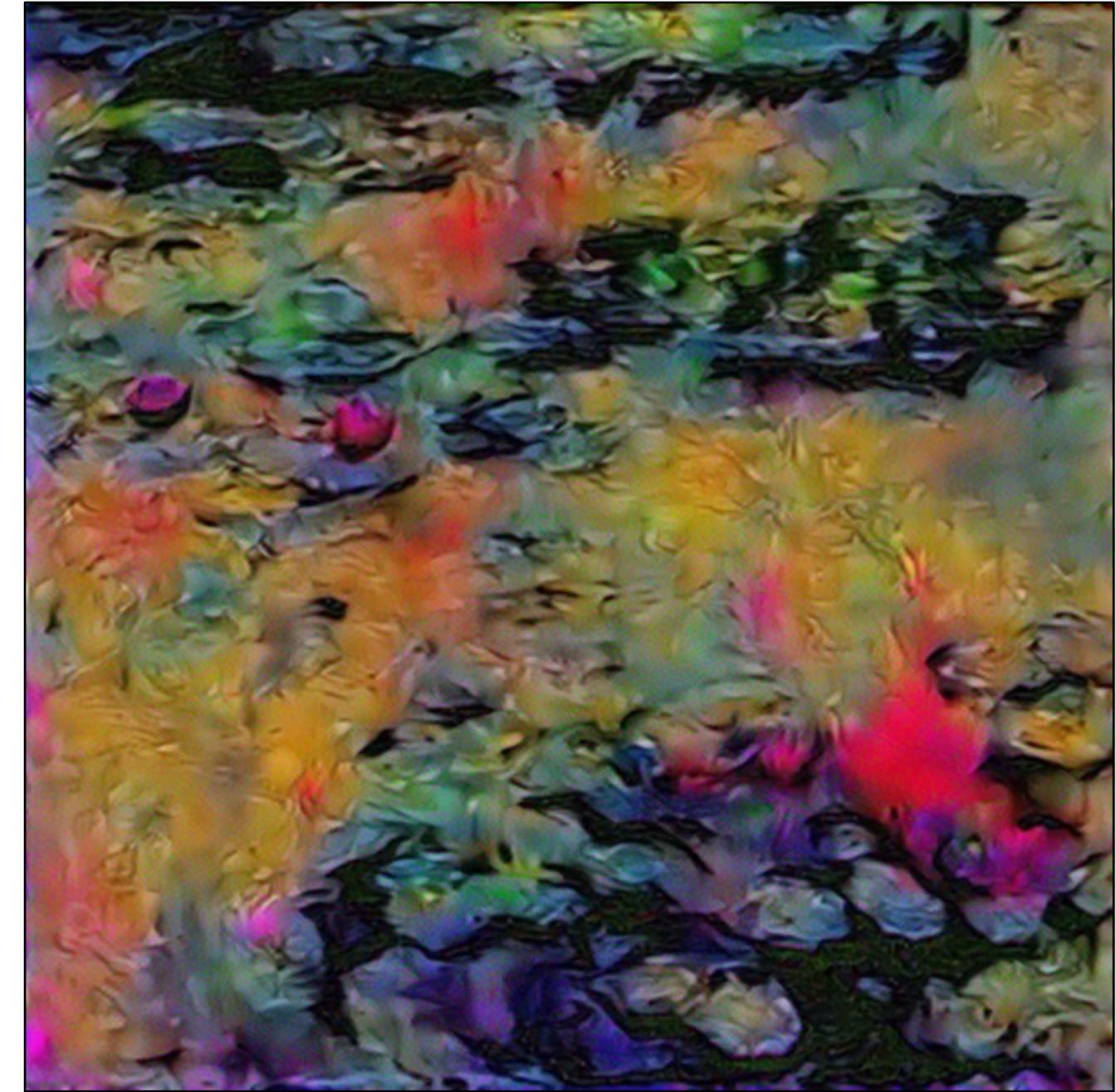


Painting

Reference

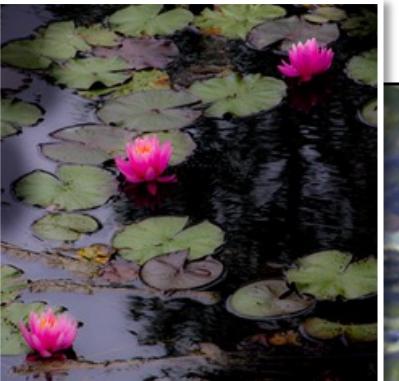


Painting

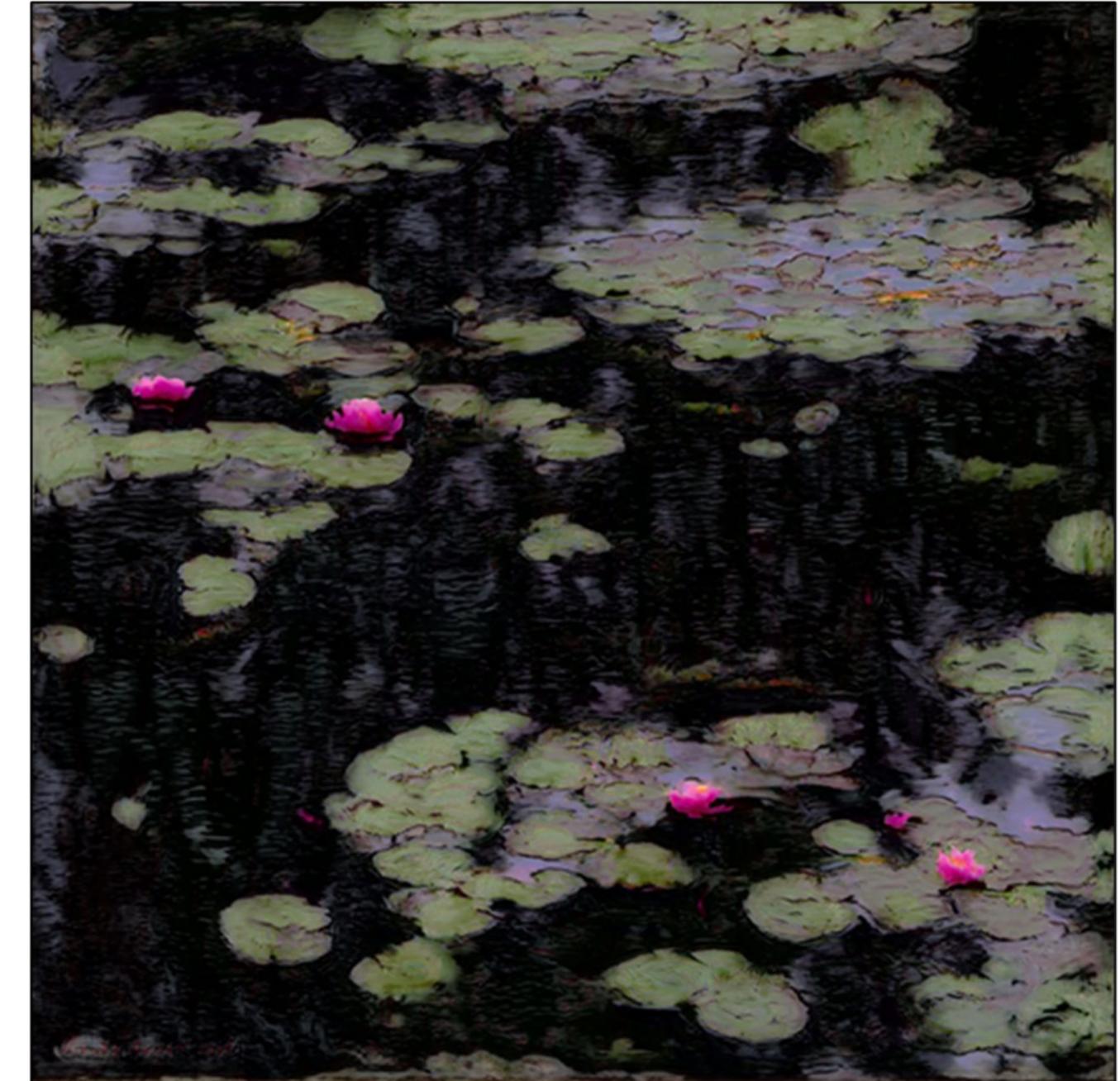


Gatys et al. CVPR'16

Reference



Painting



Single Image translation (CUT)

Style Transfer vs. Image-to-Image Translation

- Data (how to define Style)
 - A single image? A collection of images
- Applications
 - Photo -> Painting (Neural Style Transfer, Image-to-Image Translation)
 - Photo -> Photo (Image-to-Image Translation, Photo Style Transfer (Color))
 - Painting -> Photo (Image-to-Image Translation, Deep Image Analogy)
- Algorithms:
 - Patch-based method (or dense correspondence)
 - Optimization-based method
 - Feed-forward network
- Loss functions
 - Style Loss: GAN loss, Gram matrix loss
 - Content Loss: Perceptual Loss, Cycle-consistency loss, Contrastive Loss (InfoNCE)