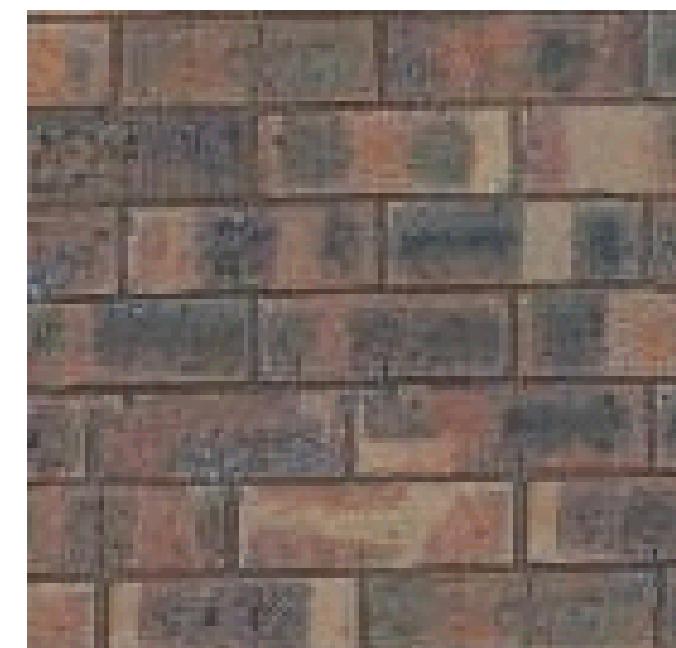
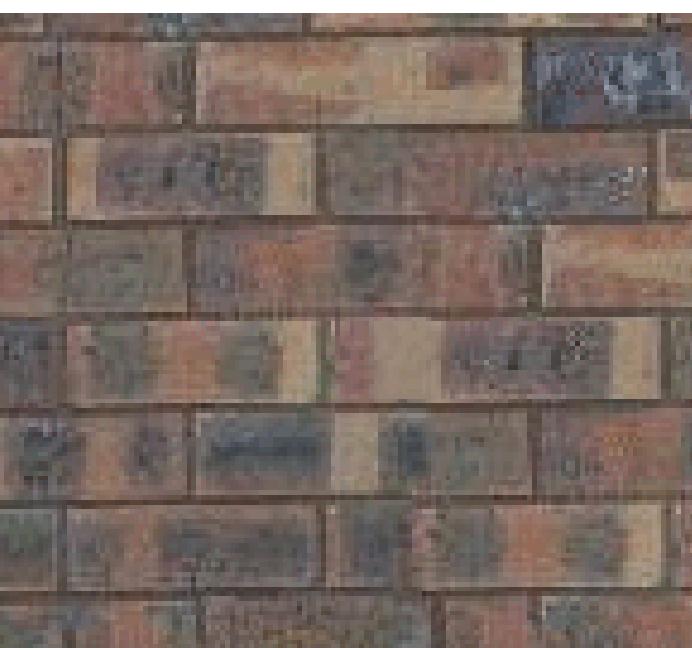
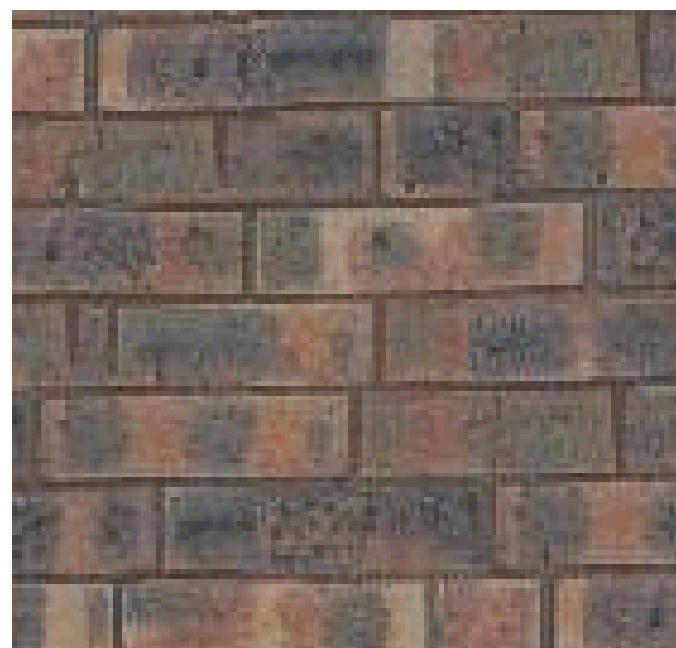


Julesz Revisited: Generating Images from Noise



Alexei Efros, CS280, UC Berkeley, Spring 2023

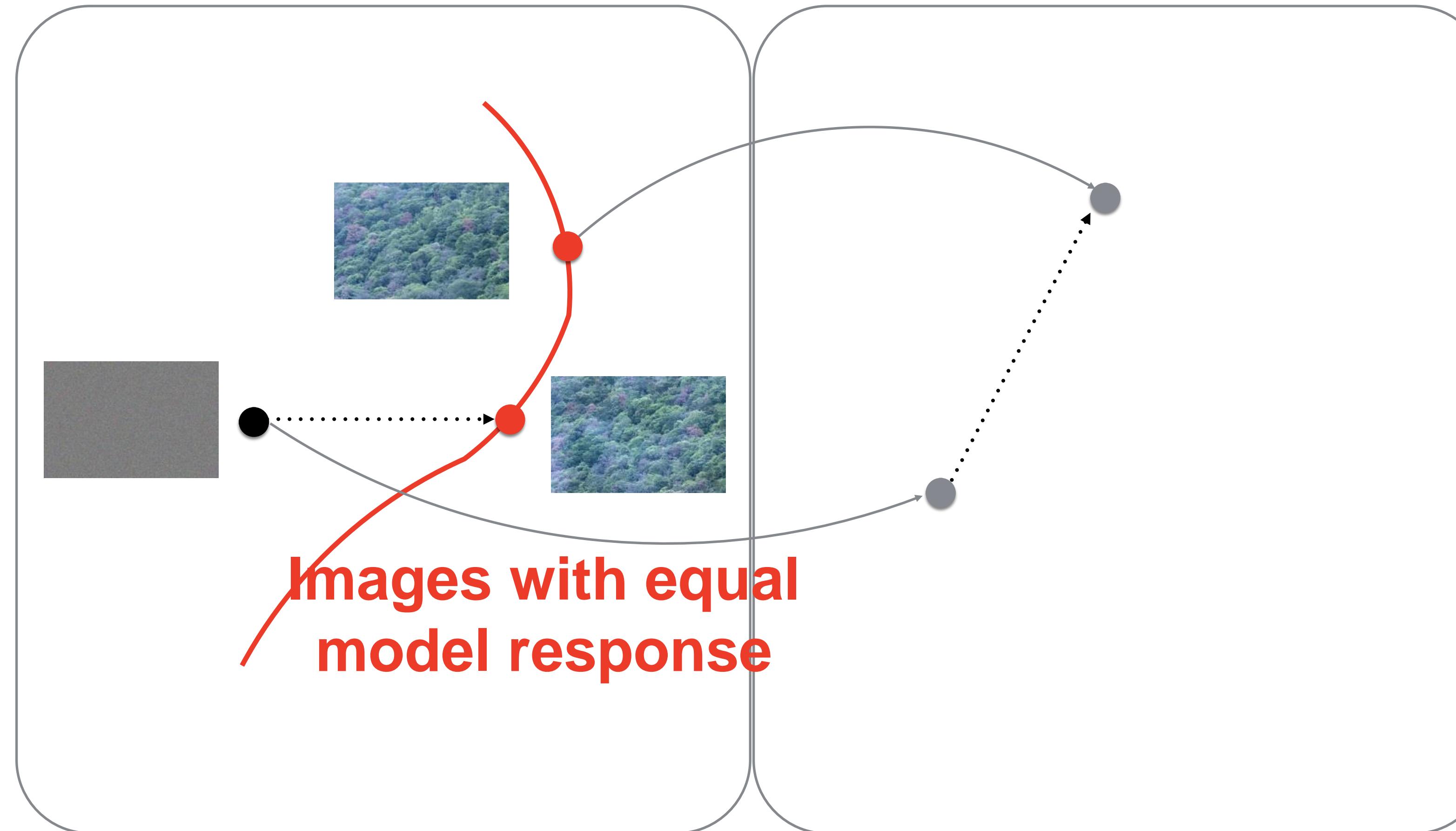
Texture as samples from distribution



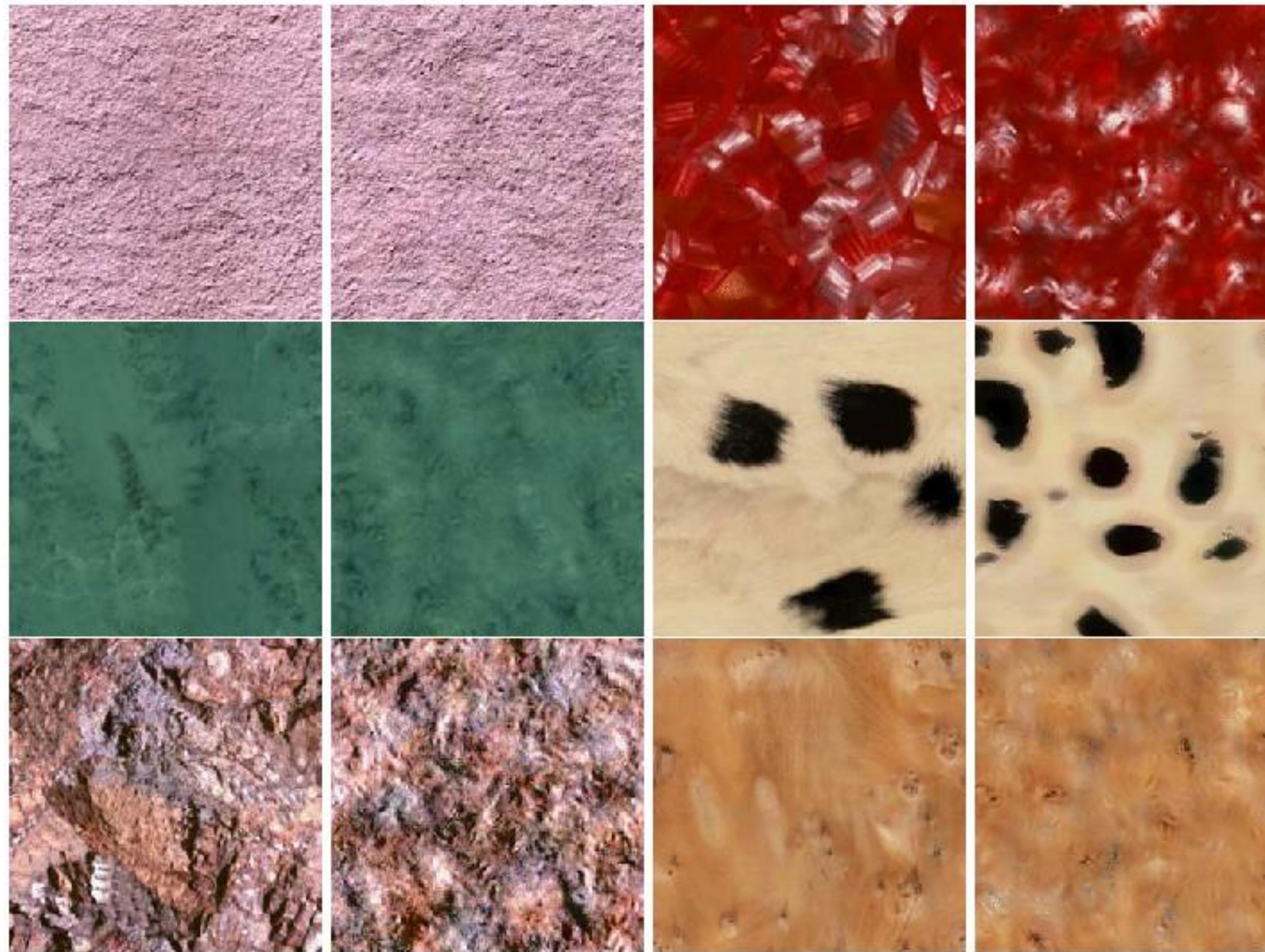
Texture Synthesis

Image Space

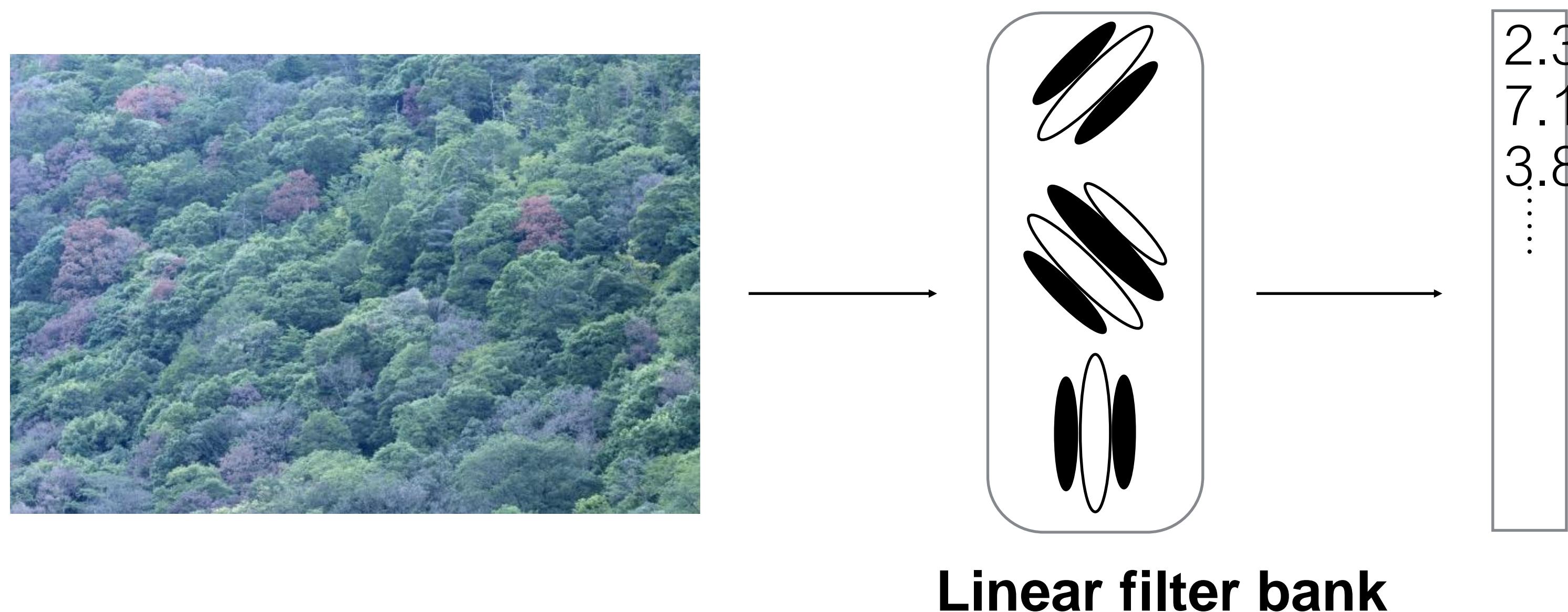
Model Space



Heeger & Bergen, SIGGRAPH'95

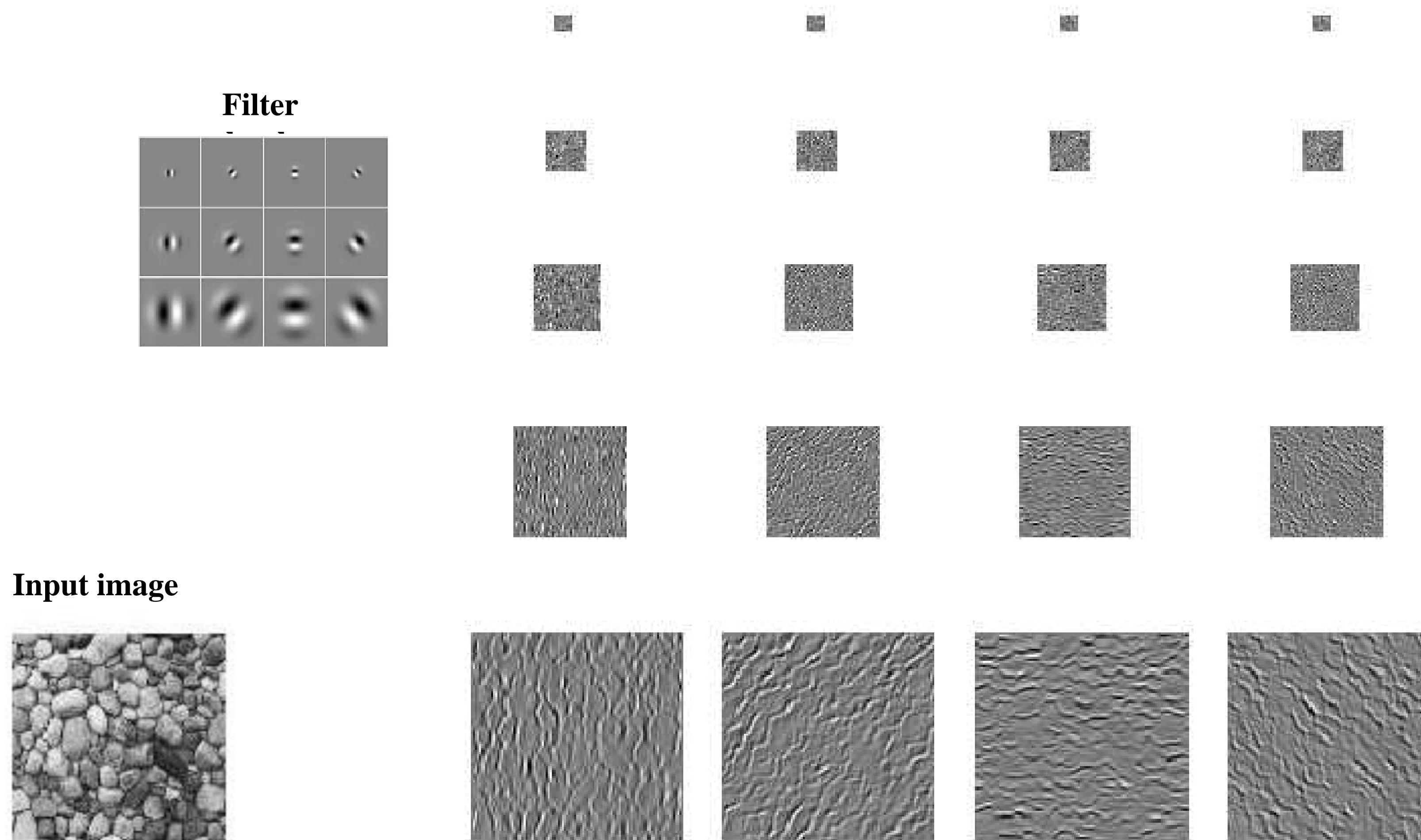


Early Vision Texture Models

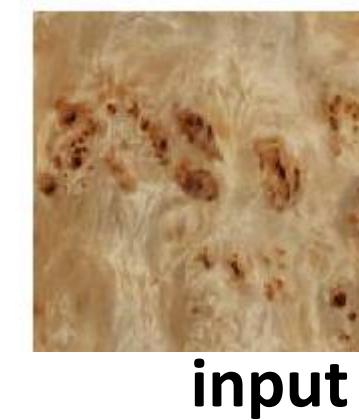


Heeger & Bergen (1995)
Portilla & Simoncelli (2000)

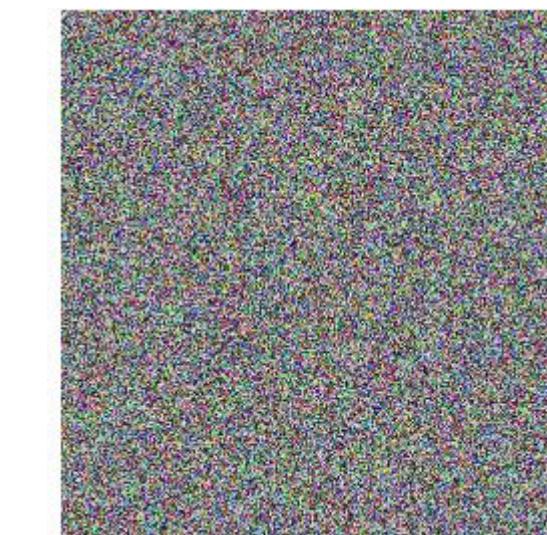
Multi-scale filter decomposition (steerable pyramid)



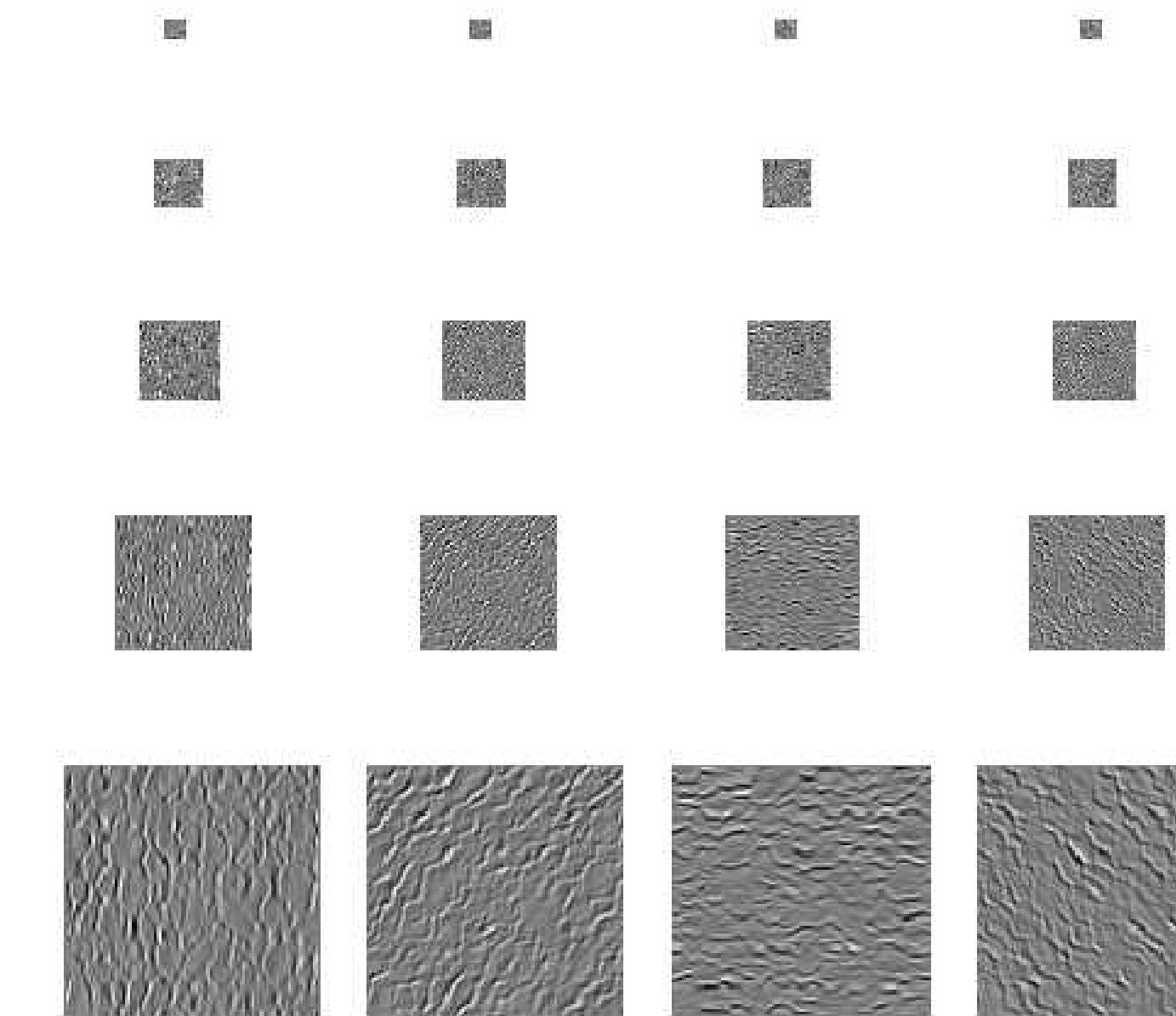
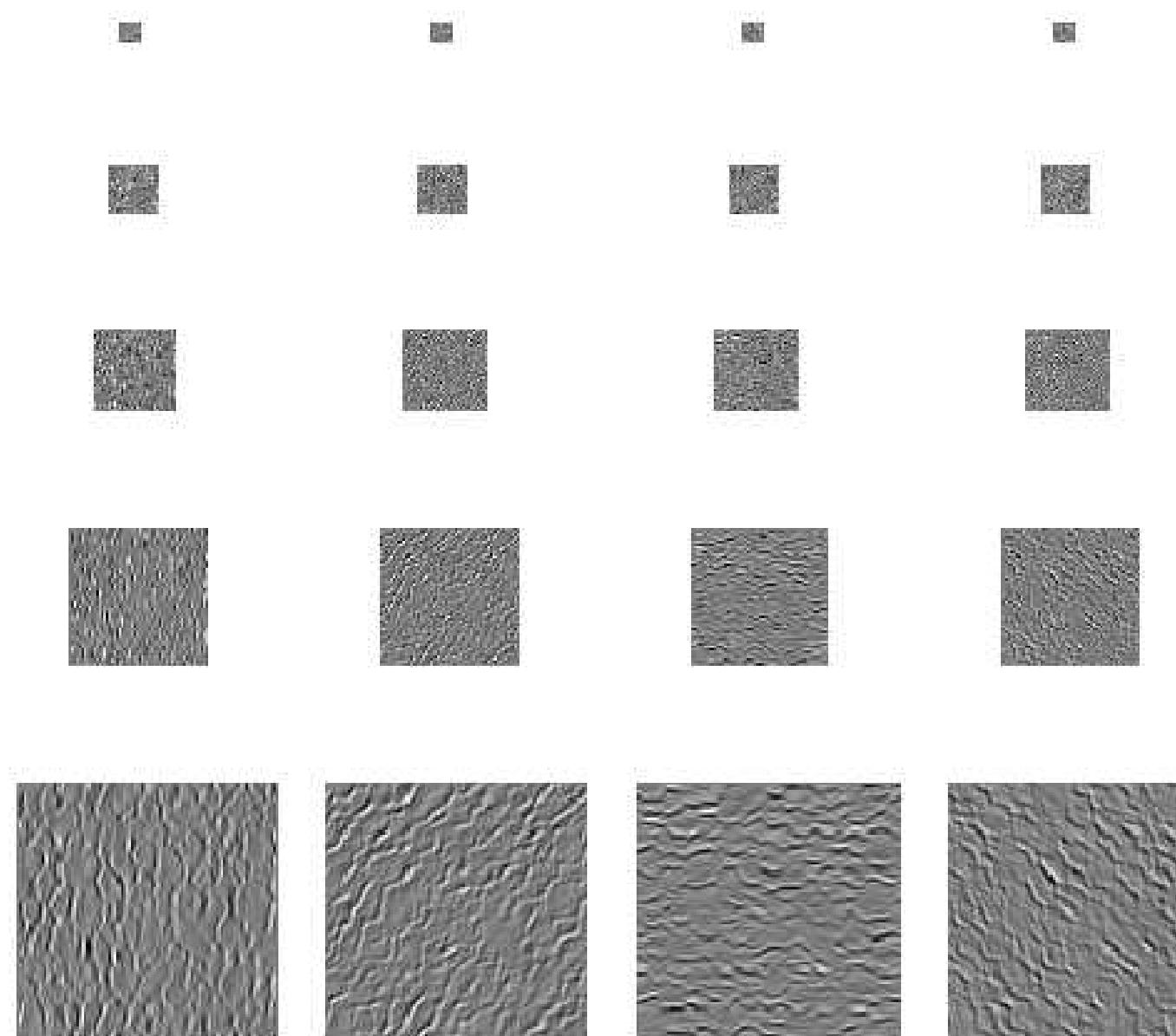
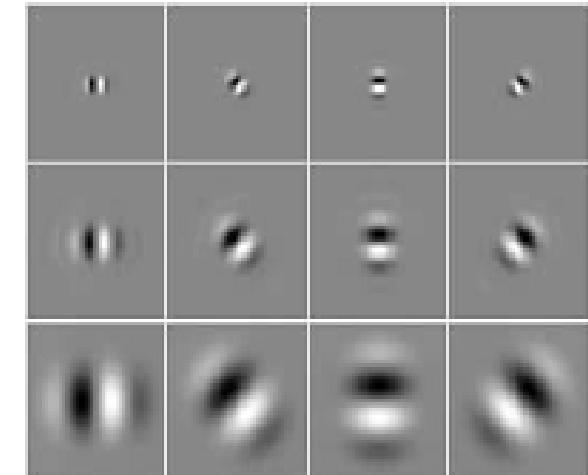
Step 1: Convolve with filterbank



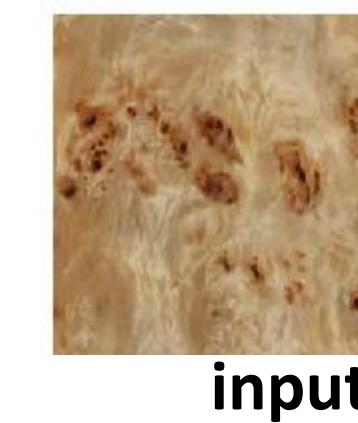
input



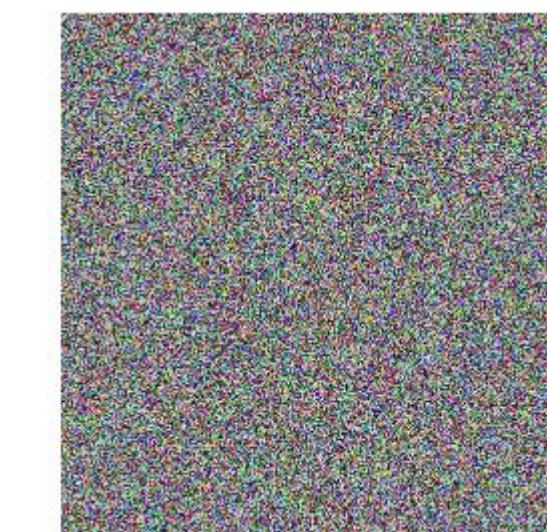
Noise image



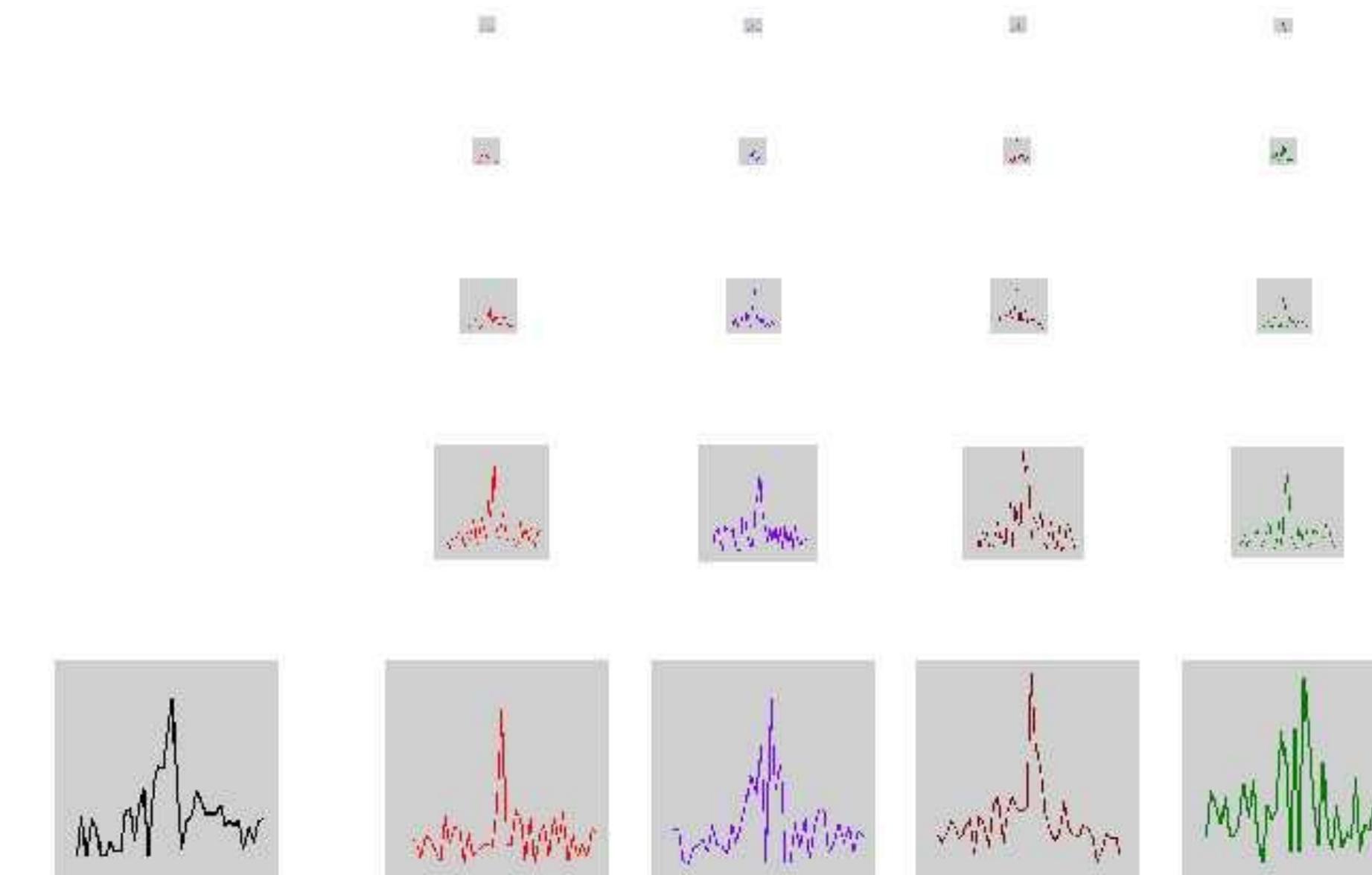
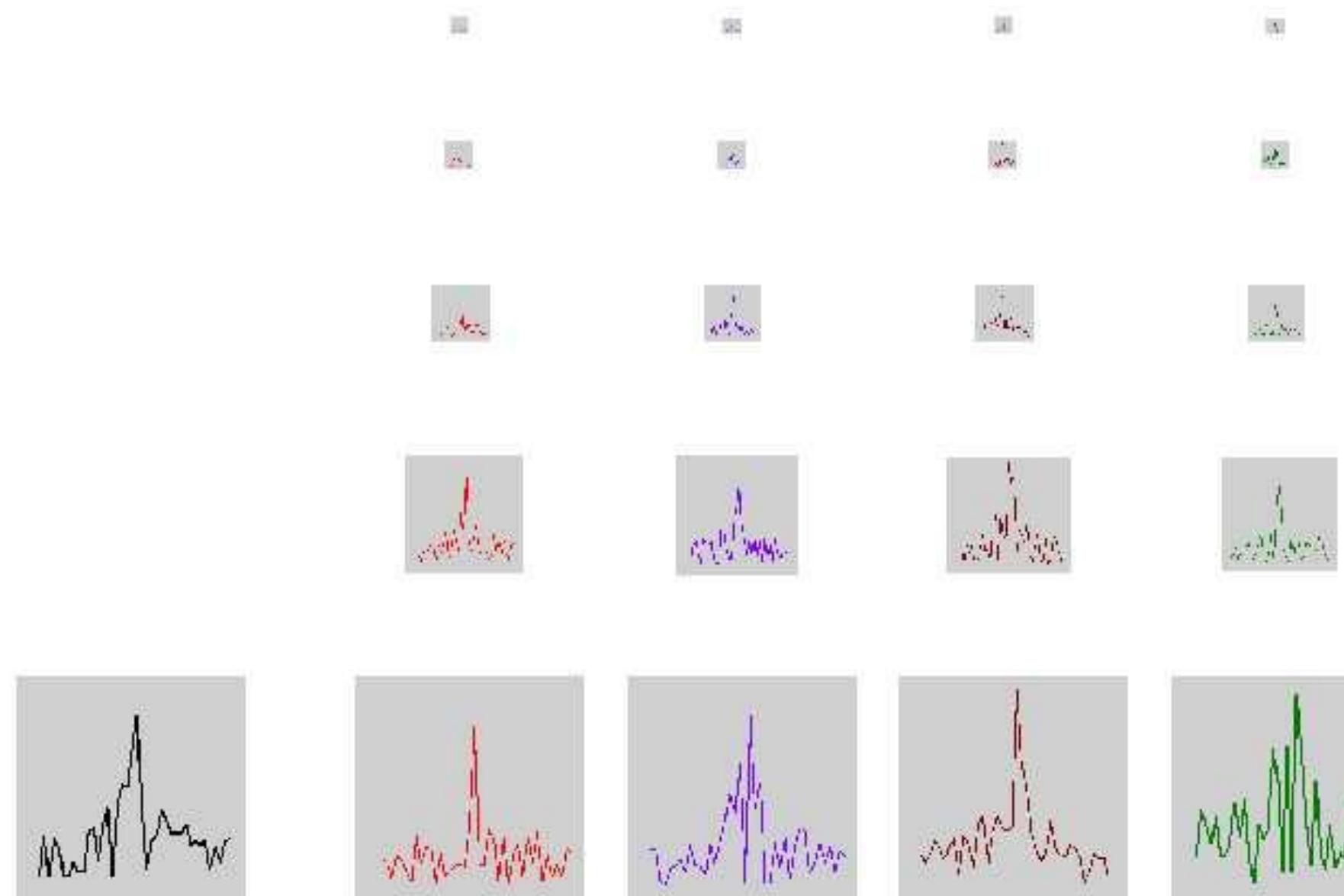
Step 2: match per-channel histograms



input



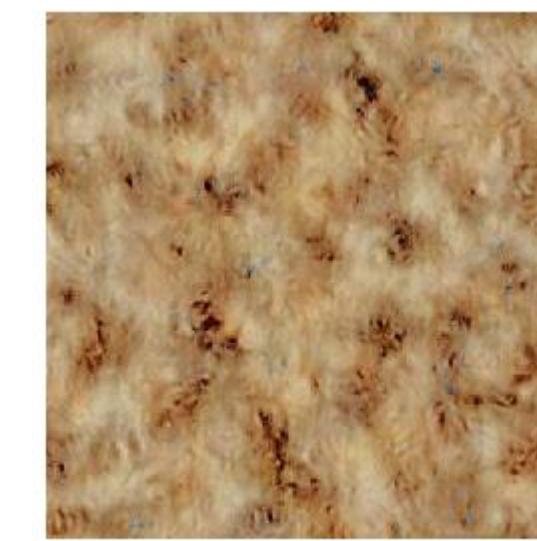
Noise image



Step 3: collapse pyramid and repeat!

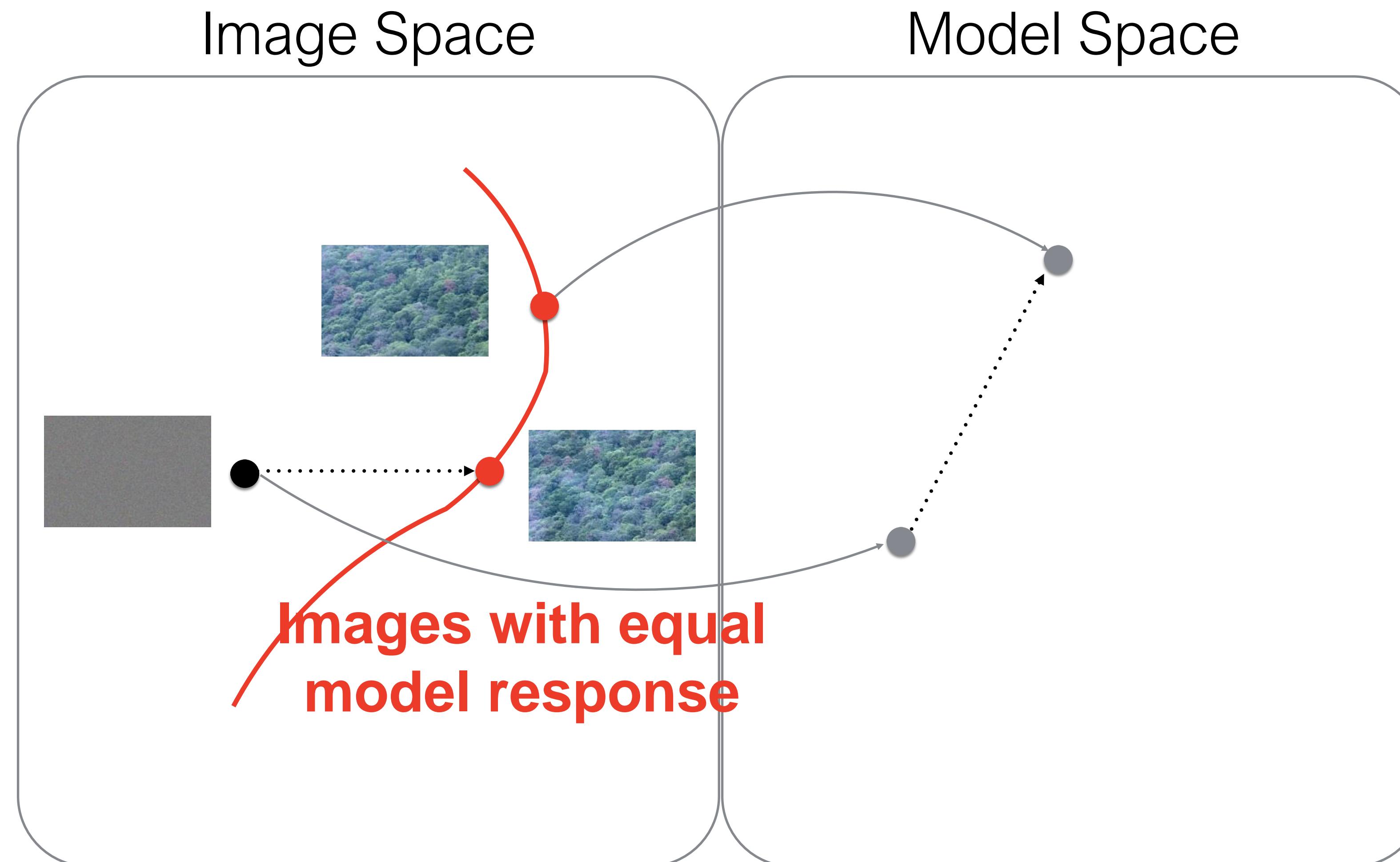


input

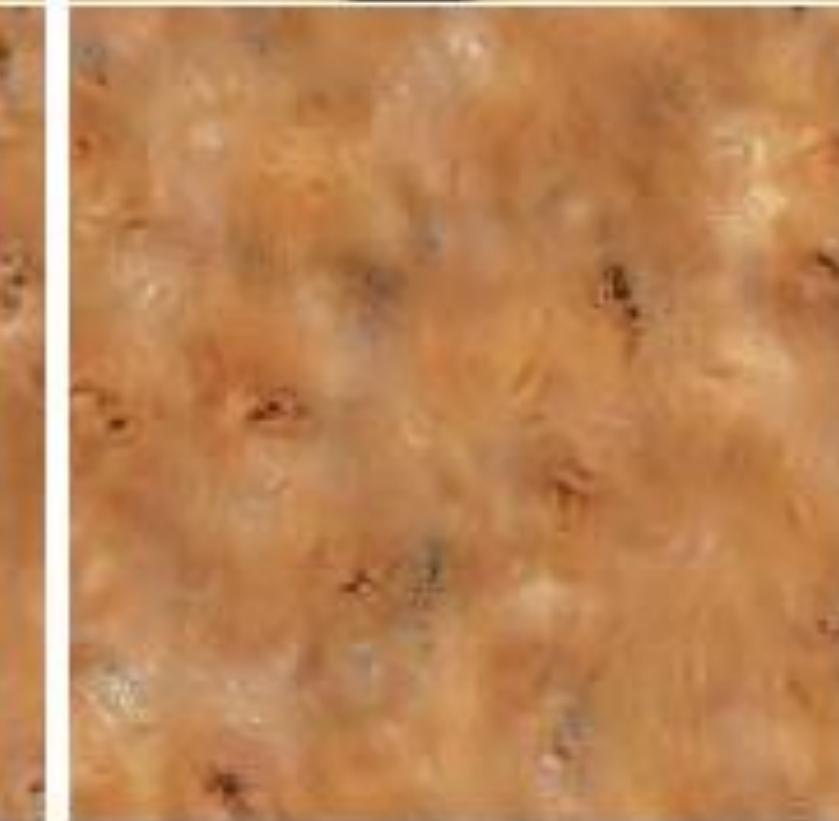
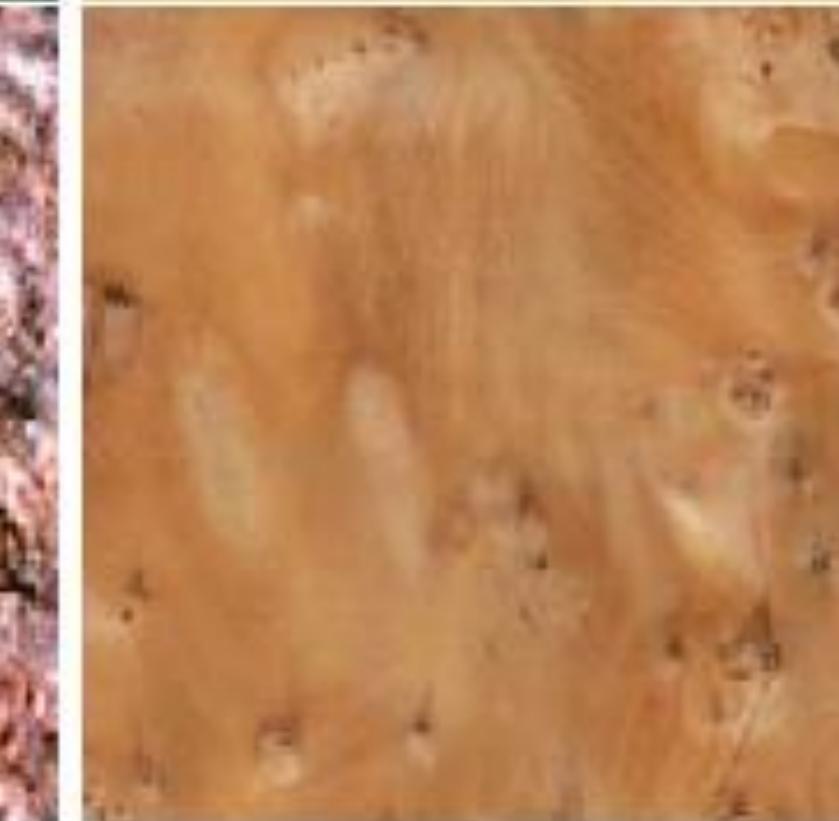
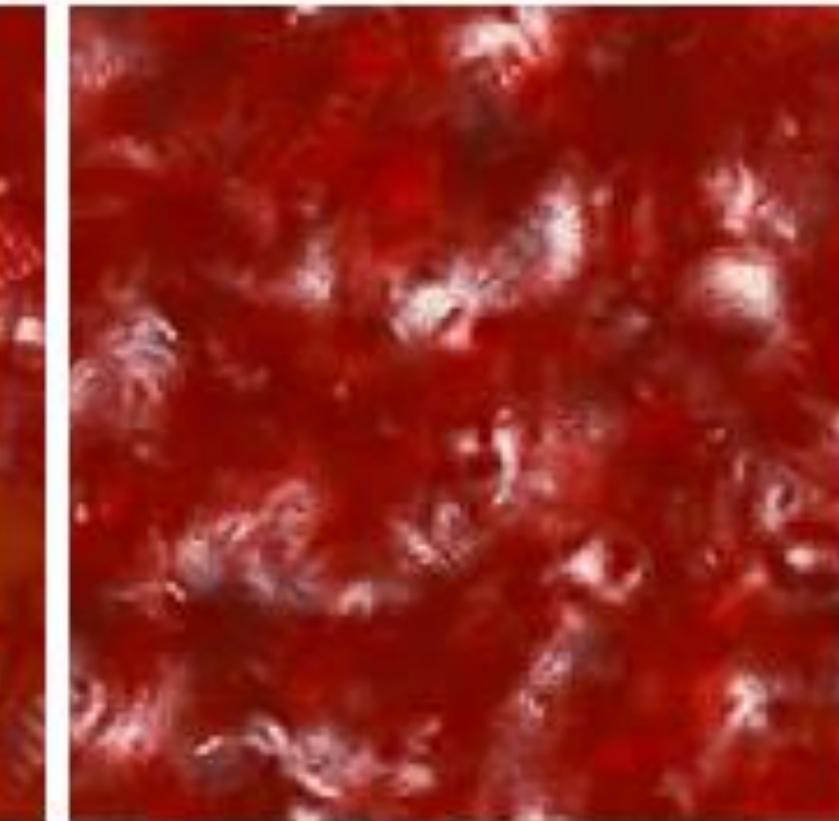


Noise image

Texture Synthesis



Portilla & Simoncelli (2000)



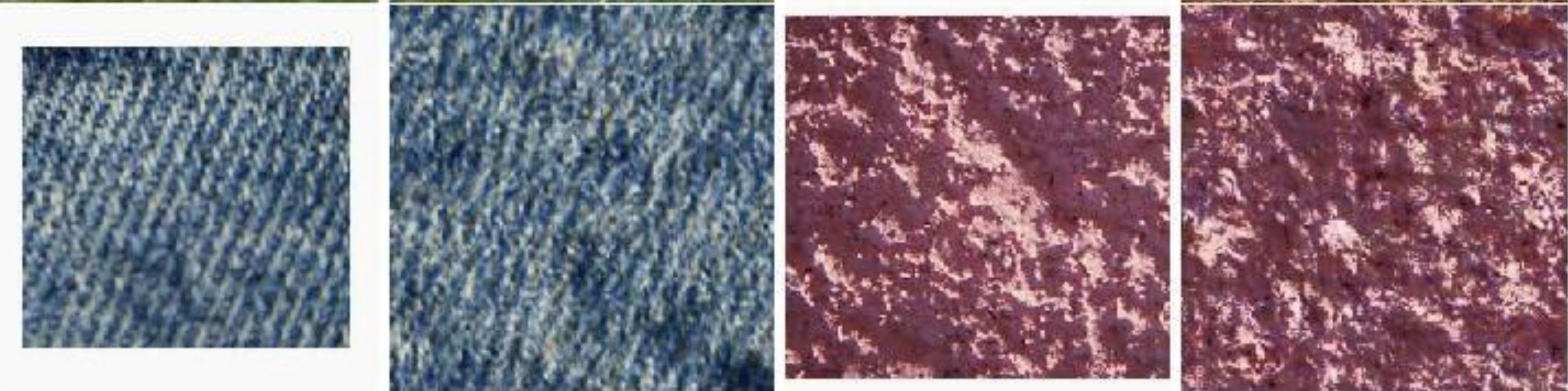




Figure 7: (Left pair) Inhomogeneous input texture produces blotchy synthetic texture. (Right pair) Homogenous input.

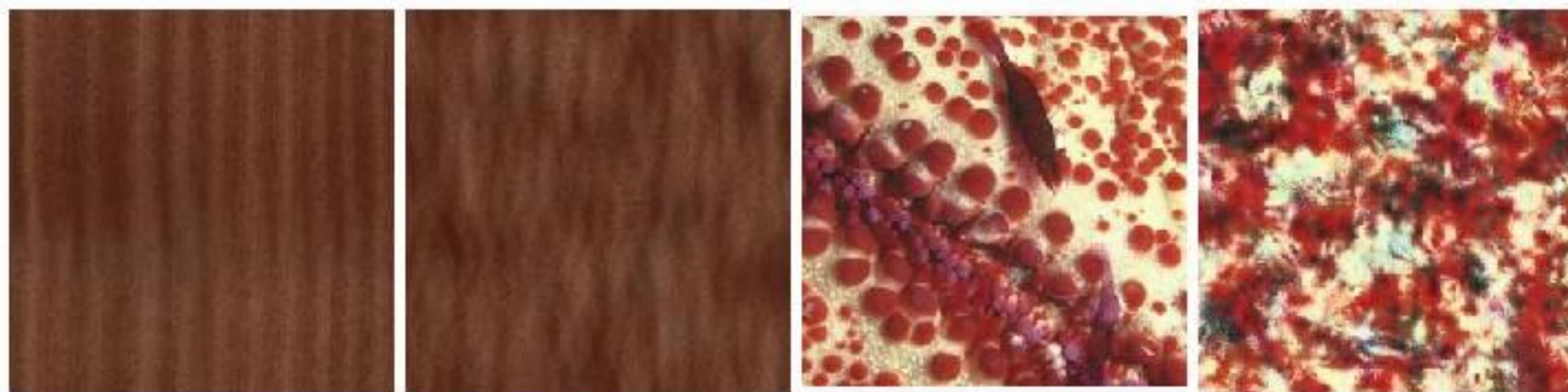


Figure 8: Examples of failures: wood grain and red coral.



Figure 9: More failures: hay and marble.

Simoncelli & Portilla '98+

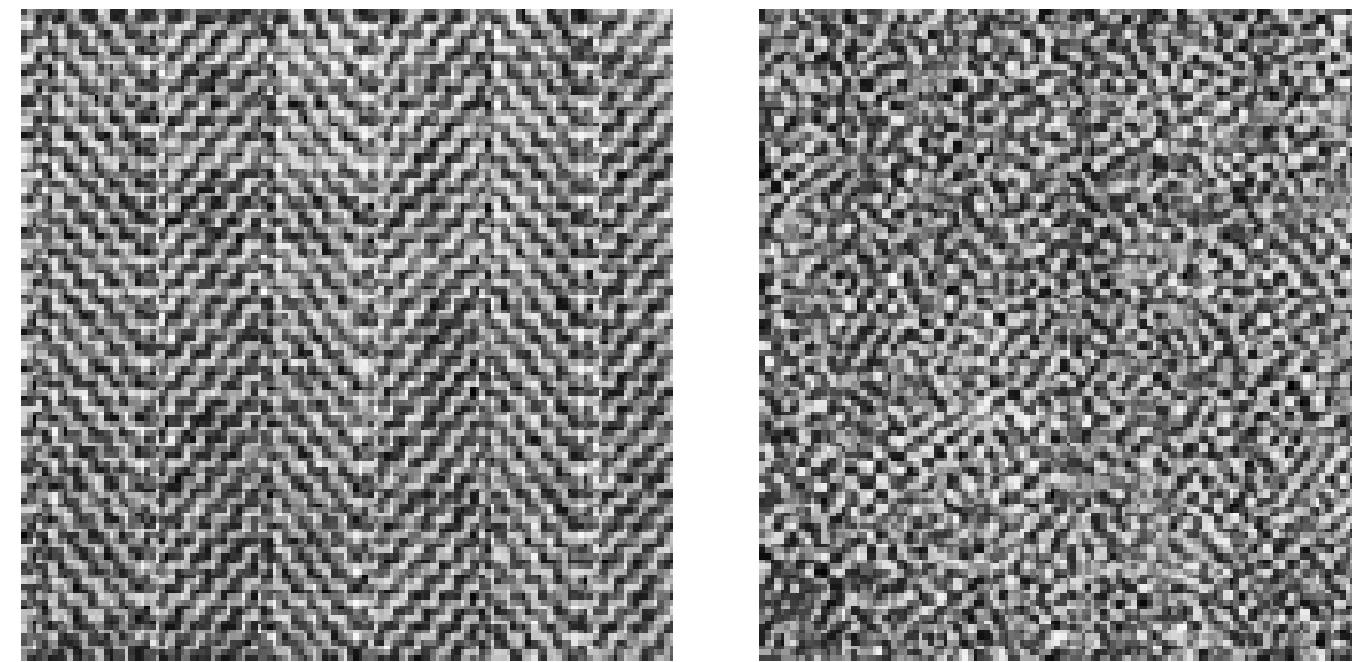
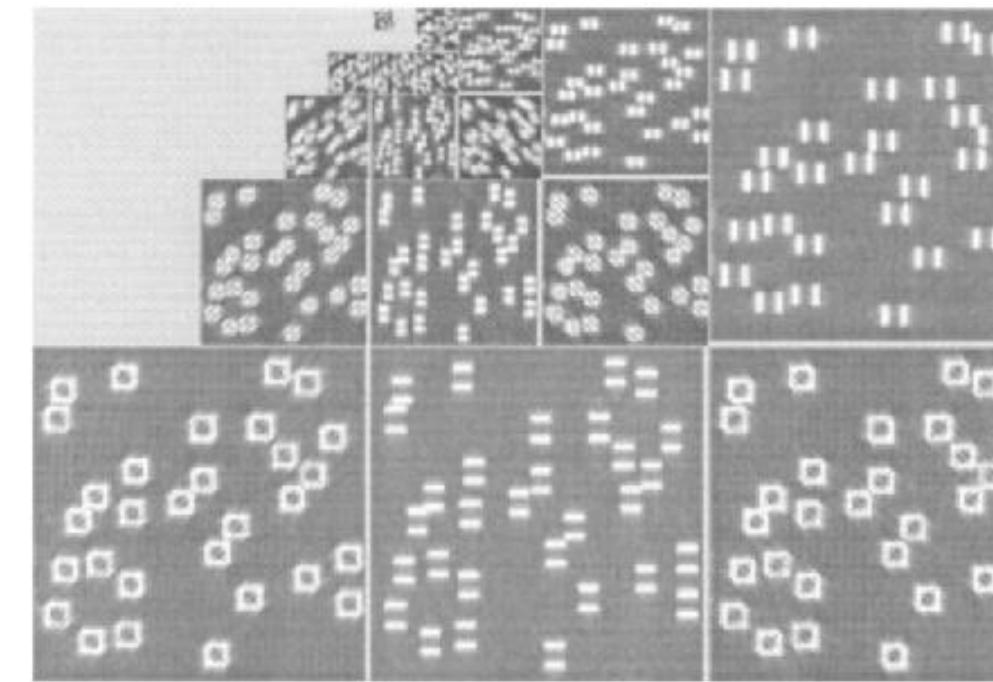
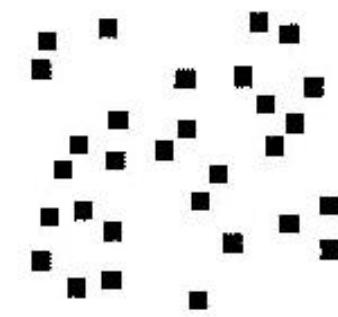


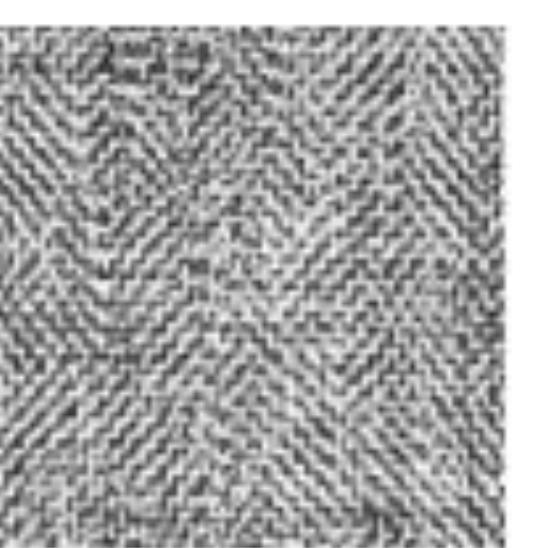
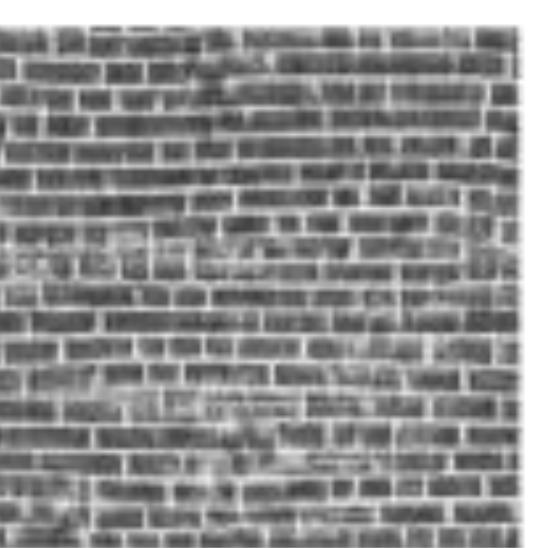
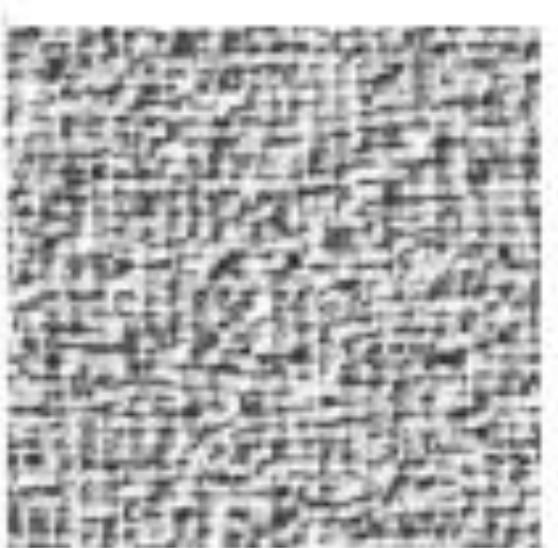
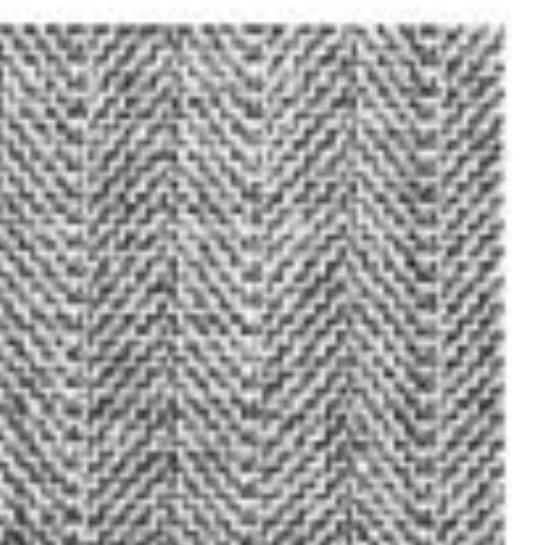
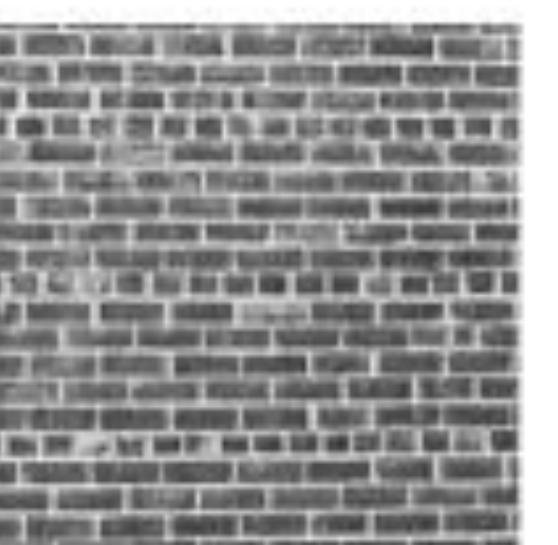
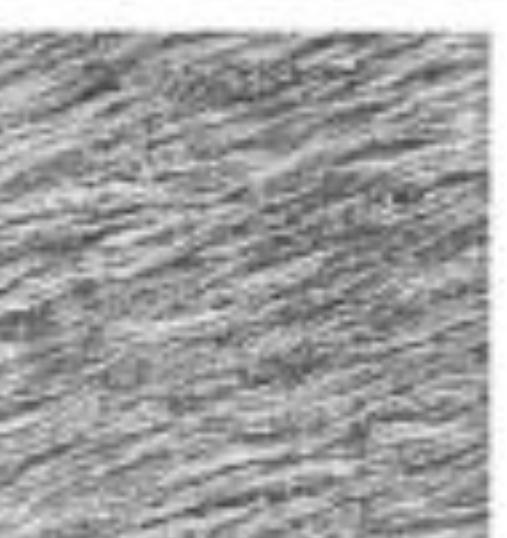
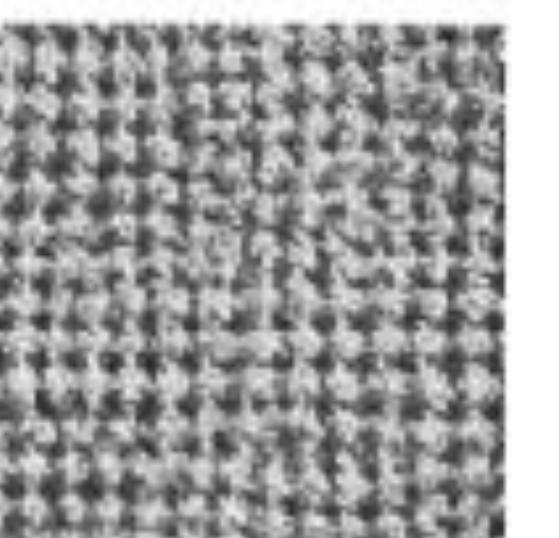
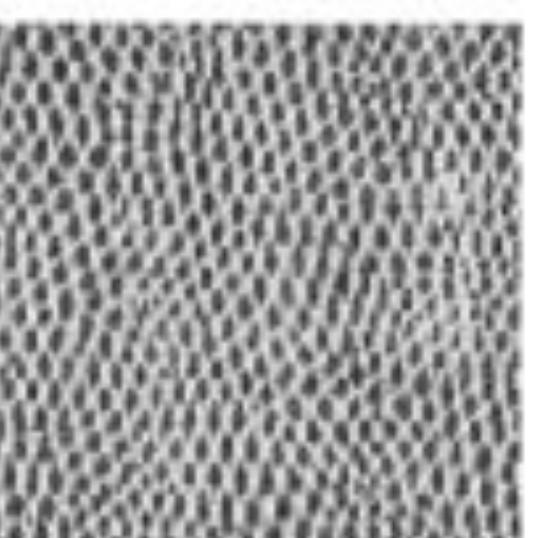
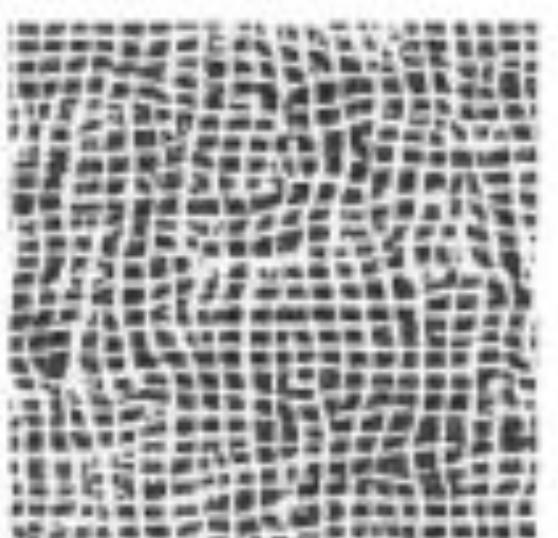
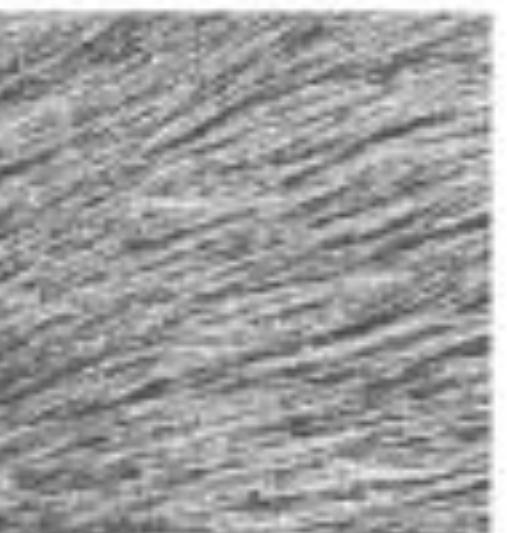
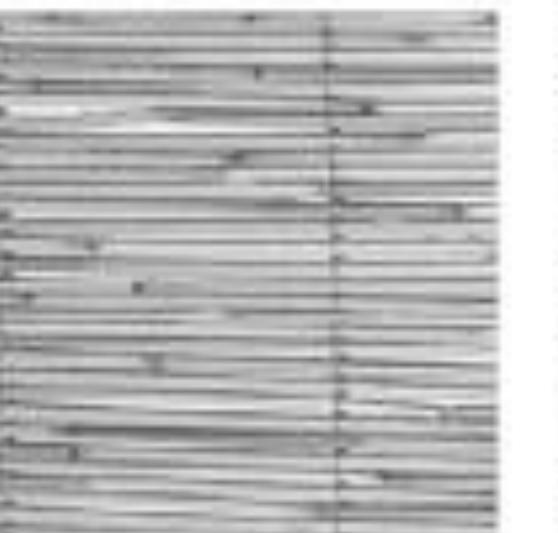
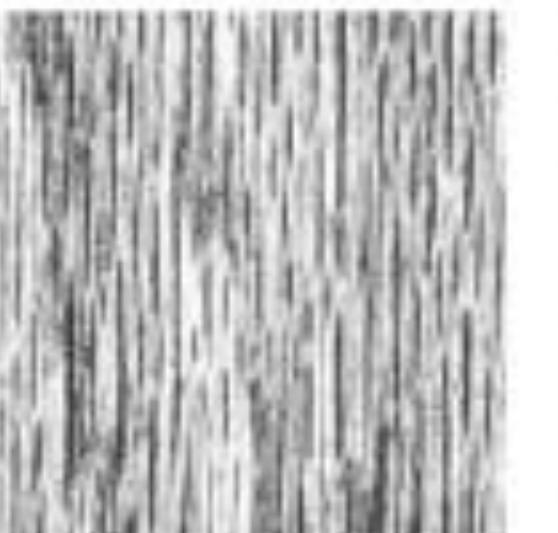
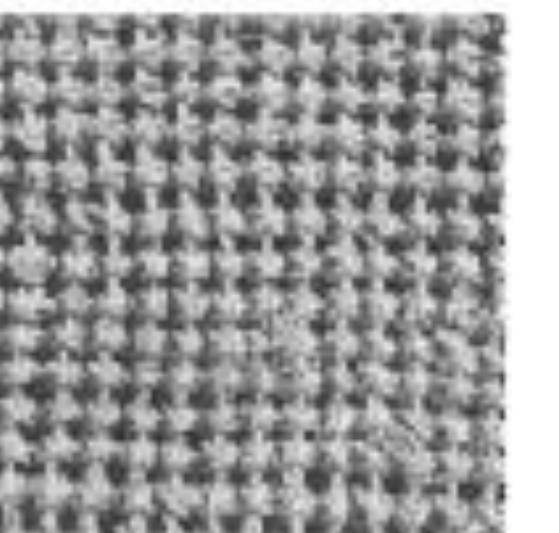
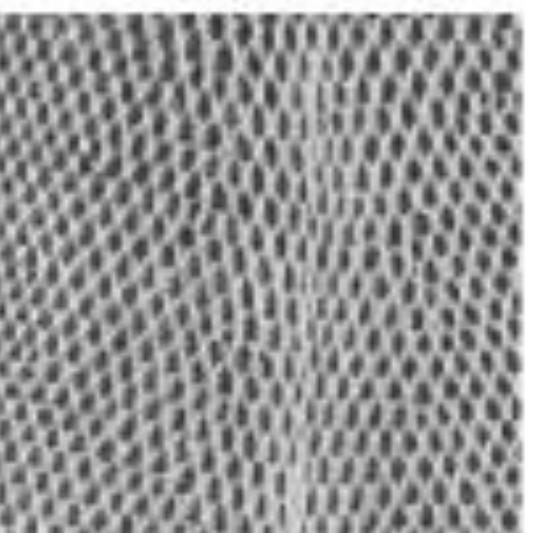
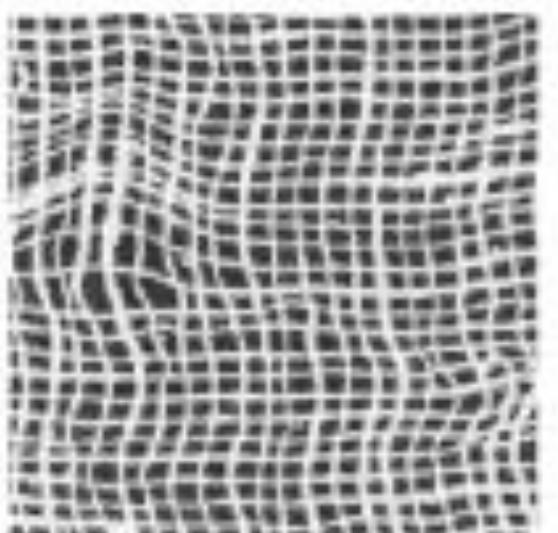
Figure 1. Textures with matching marginal statistics.

- Marginal statistics are not enough
- Neighboring filter responses are highly correlated
 - an edge at low-res will cause an edge at high-res
- Let's match 2nd order statistics too!
- J Portilla and E P Simoncelli. *A Parametric Texture Model based on Joint Statistics of Complex Wavelet Coefficients*. Int'l Journal of Computer Vision. 40(1):49-71, October, 2000.

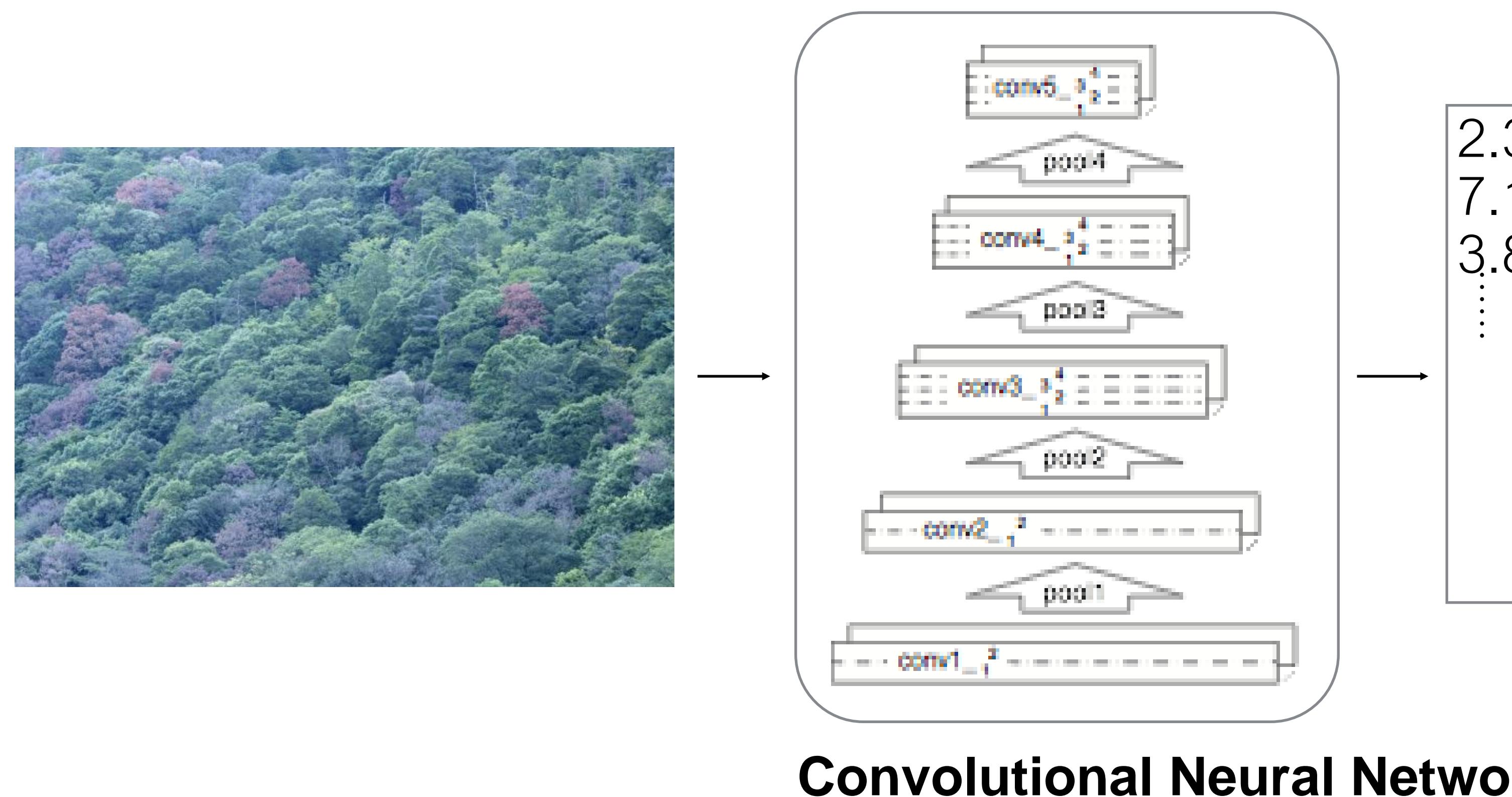
Simoncelli & Portilla '98+



- Match joint histograms of pairs of filter responses at adjacent spatial locations, orientations, and scales.
- Optimize using repeated projections onto statistical constraint surfaces



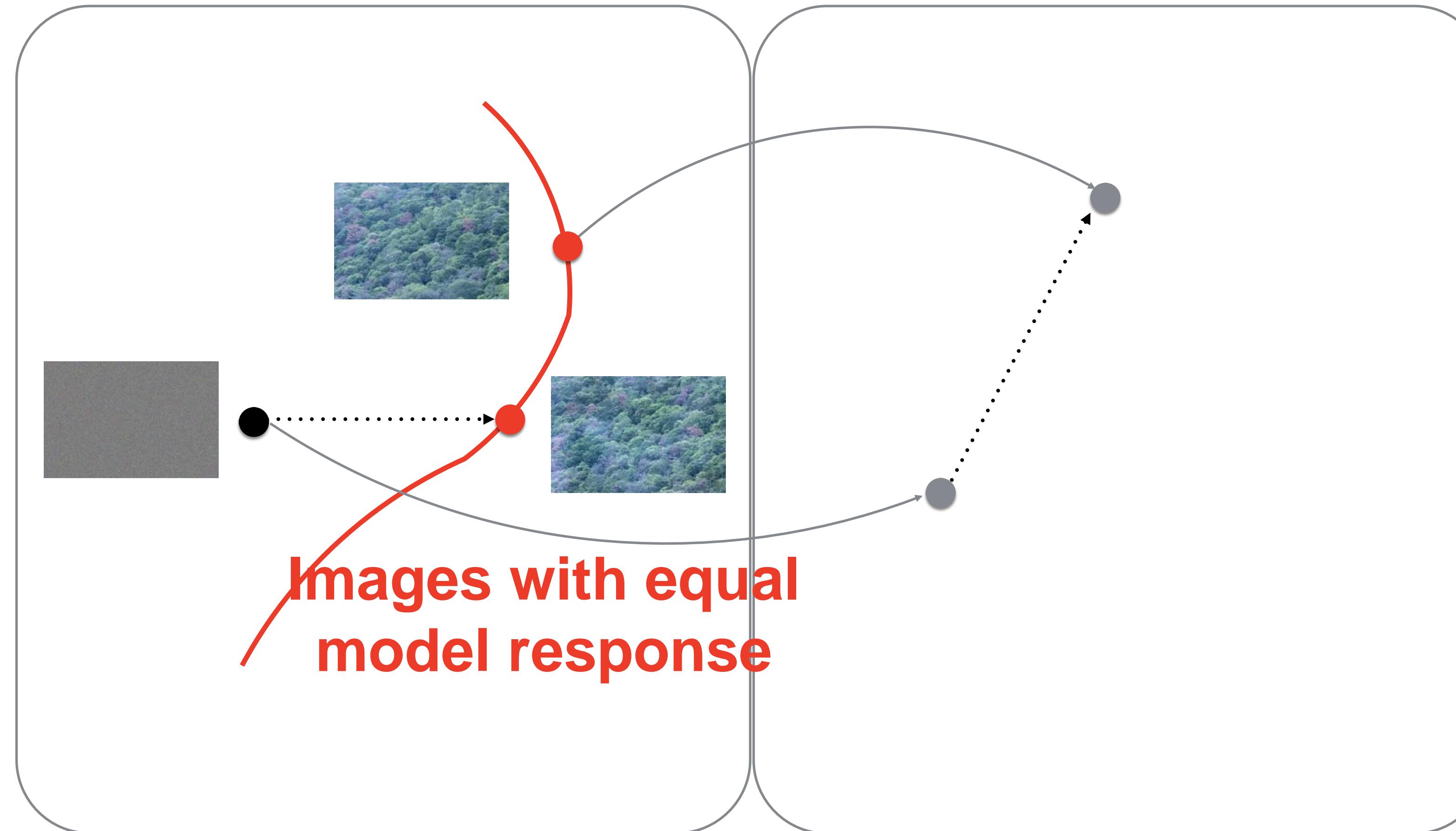
Convolutional Neural Network Texture Model



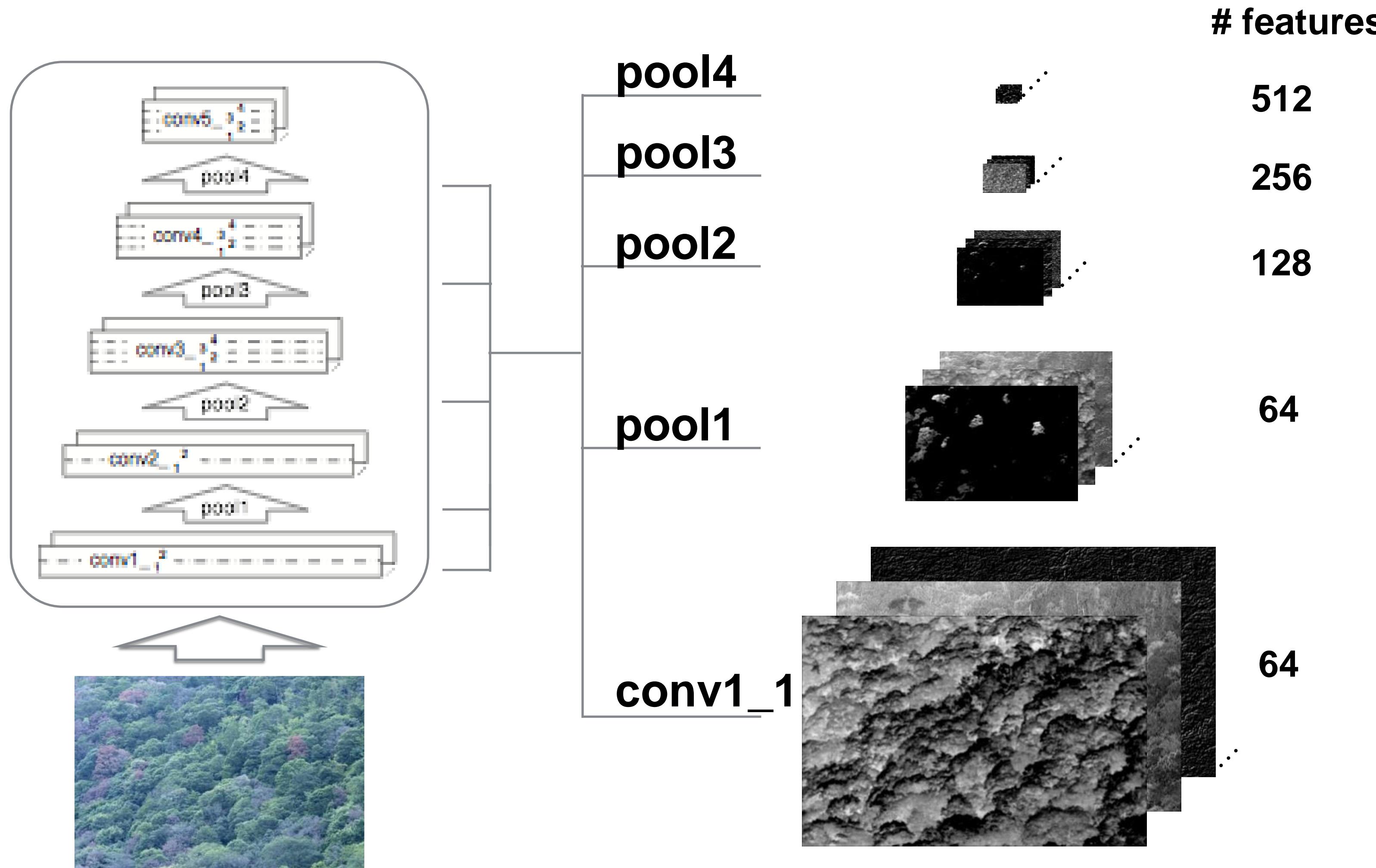
Texture Synthesis

Image Space

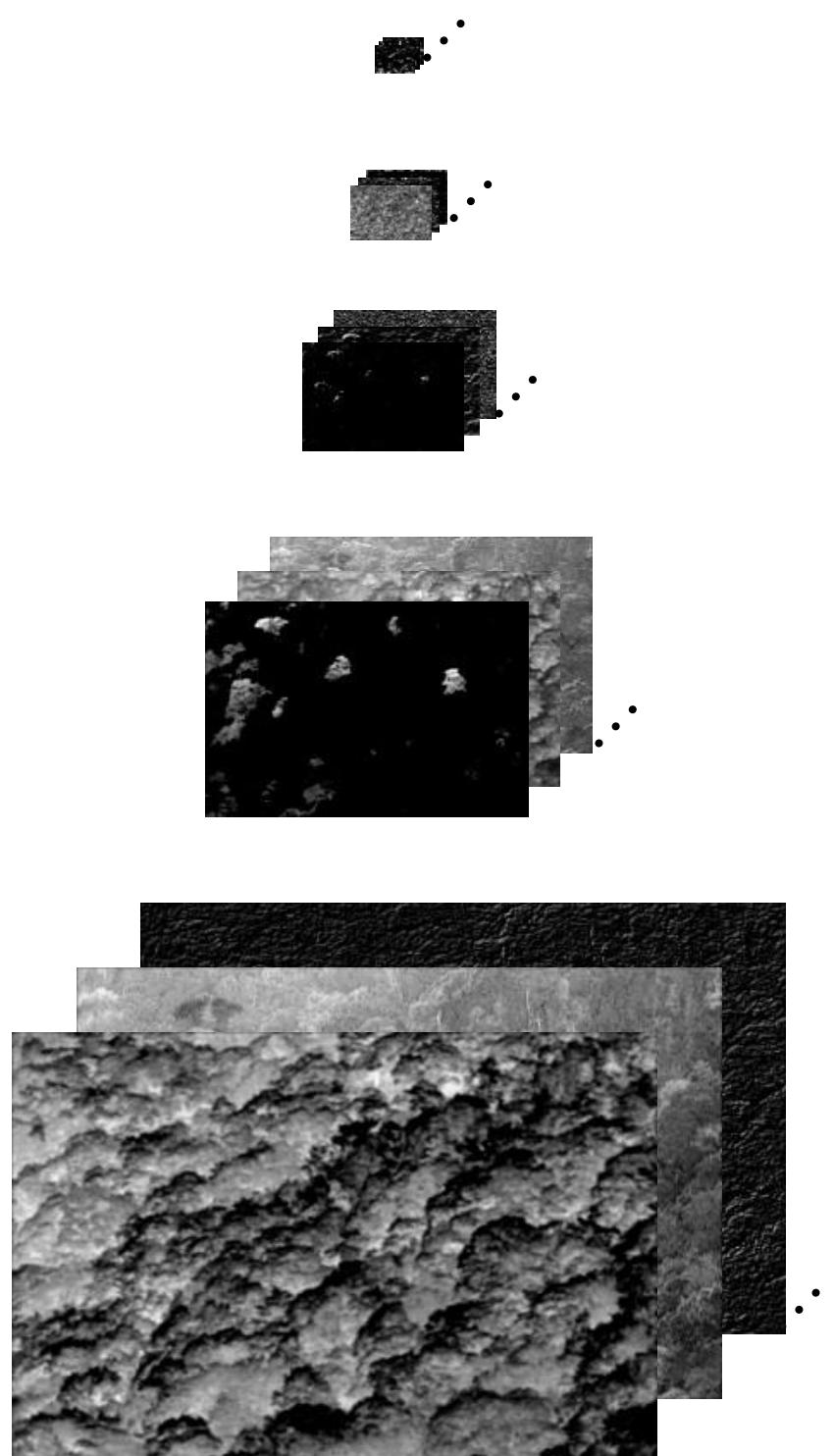
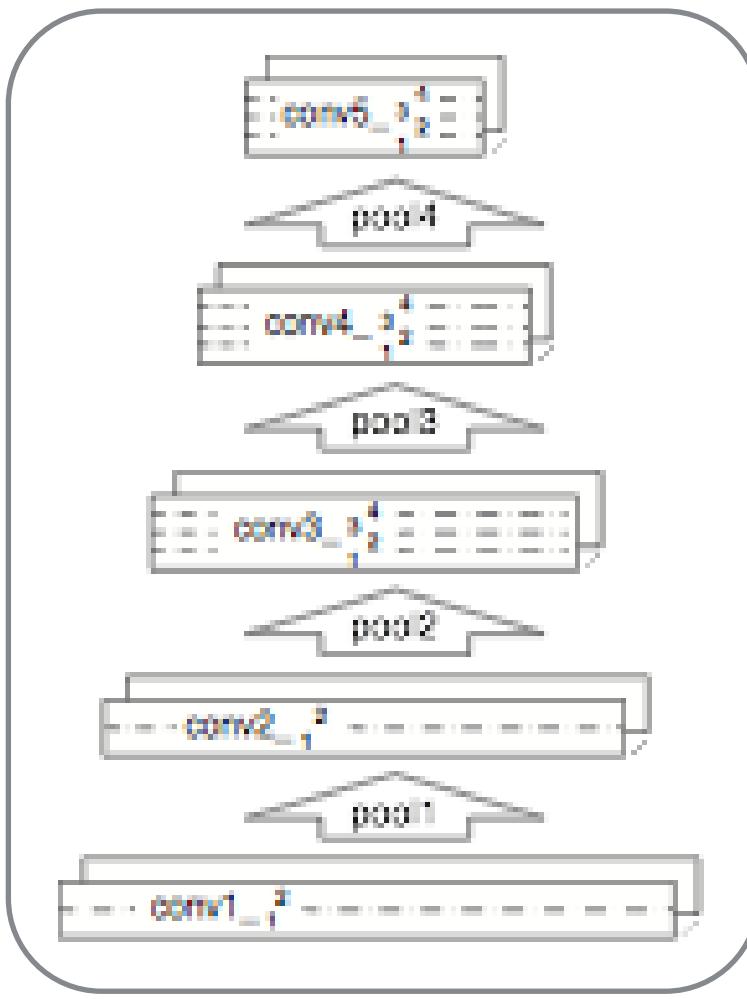
Model Space



CNN - Multiscale Filter Bank



CNN - Texture Features



$$F = [\bar{f}_1, \bar{f}_2, \bar{f}_3, \dots, \bar{f}_N]^T$$

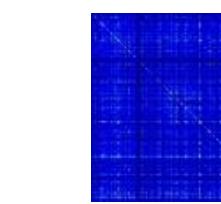
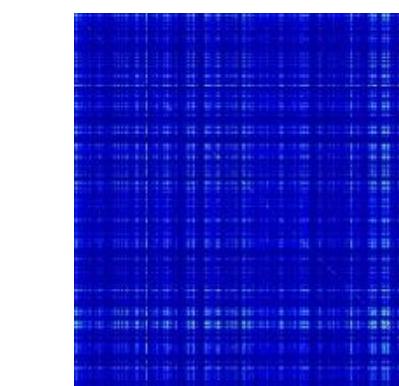
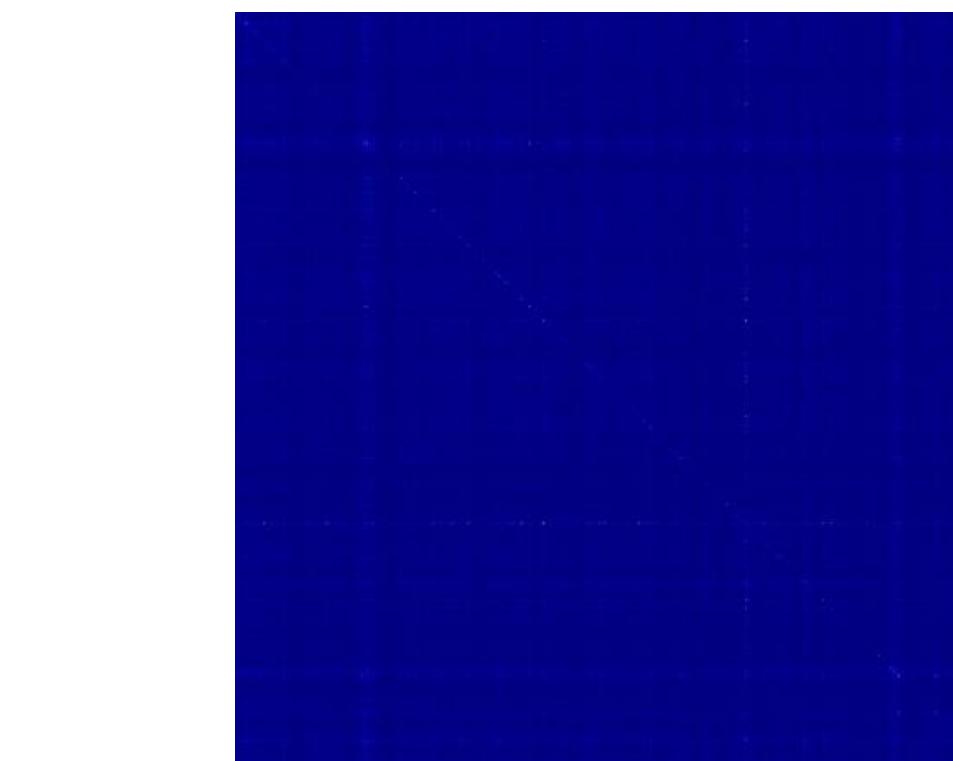
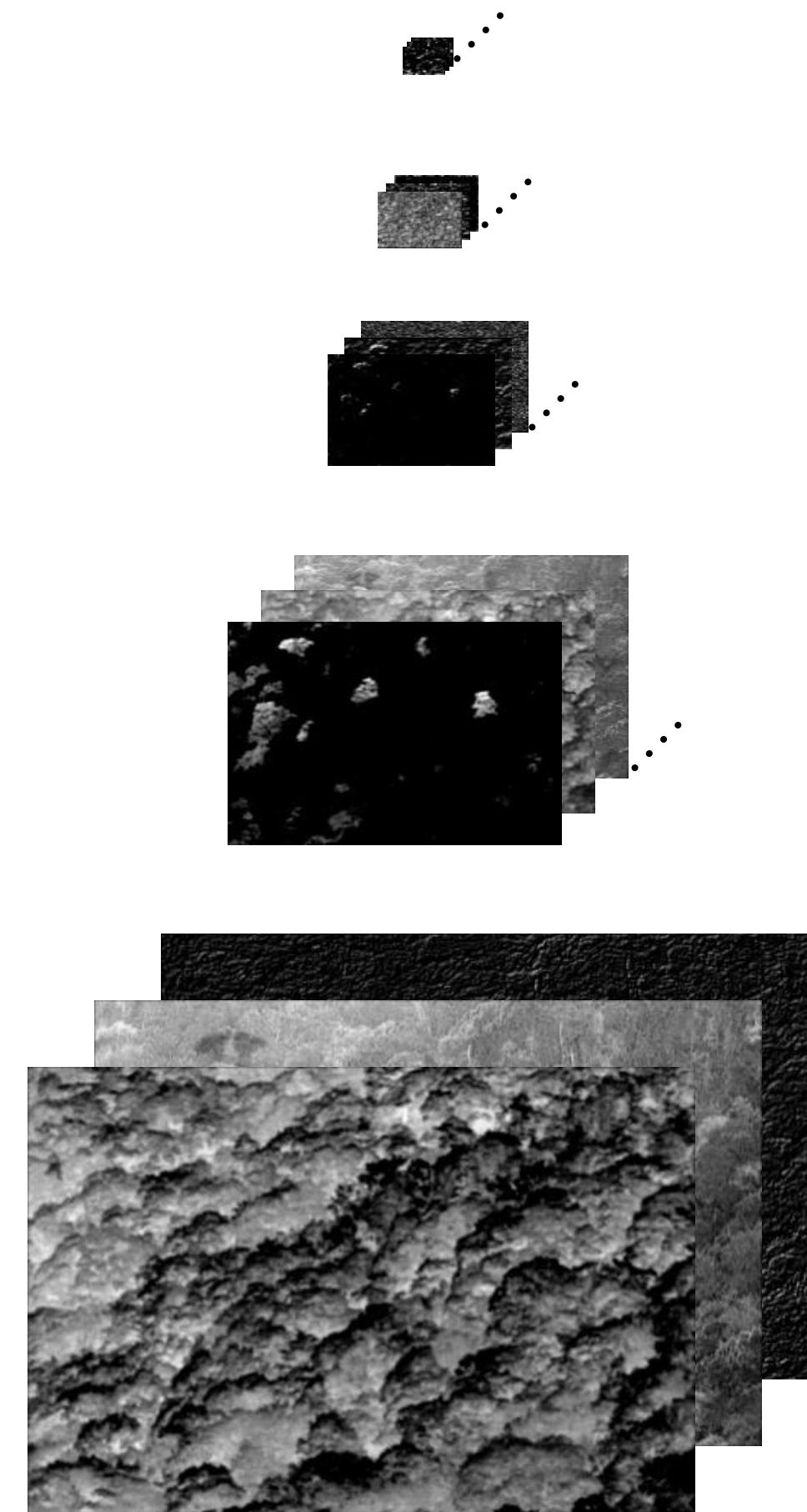
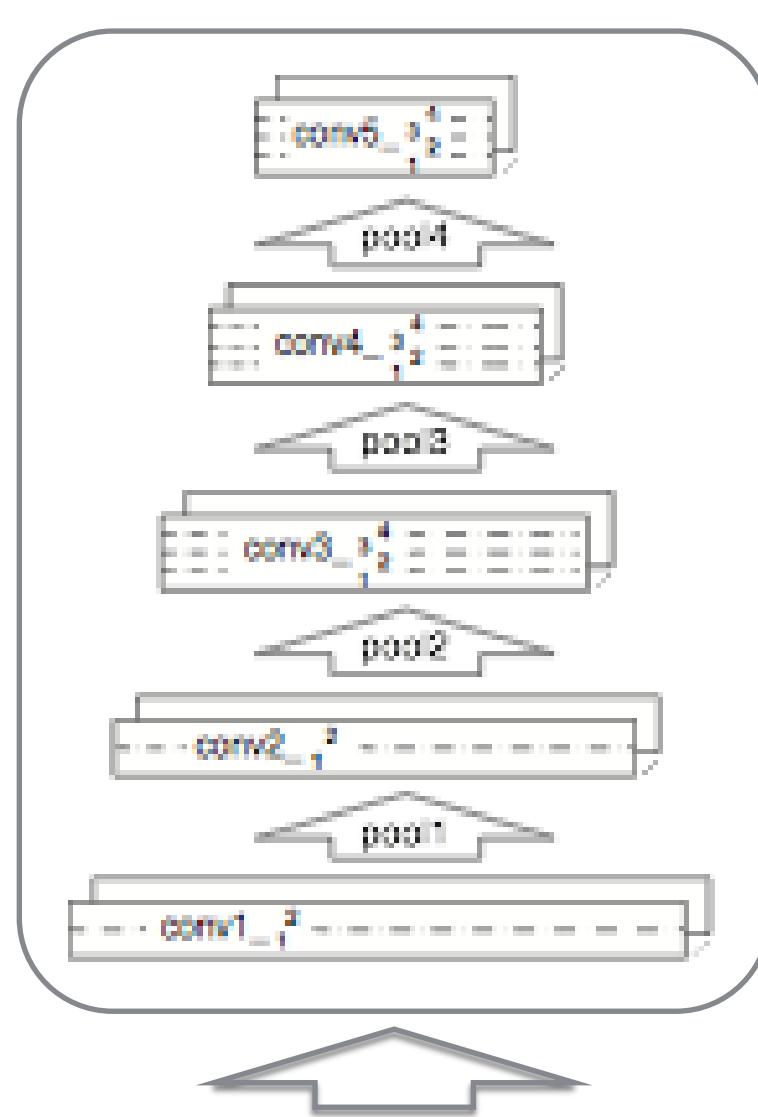
$$G = FF^T$$

$$= \begin{pmatrix} \langle \bar{f}_1, \bar{f}_1 \rangle & \cdots & \langle \bar{f}_1, \bar{f}_N \rangle \\ \langle \bar{f}_2, \bar{f}_1 \rangle & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \langle \bar{f}_N, \bar{f}_1 \rangle & \cdots & \langle \bar{f}_N, \bar{f}_N \rangle \end{pmatrix}$$

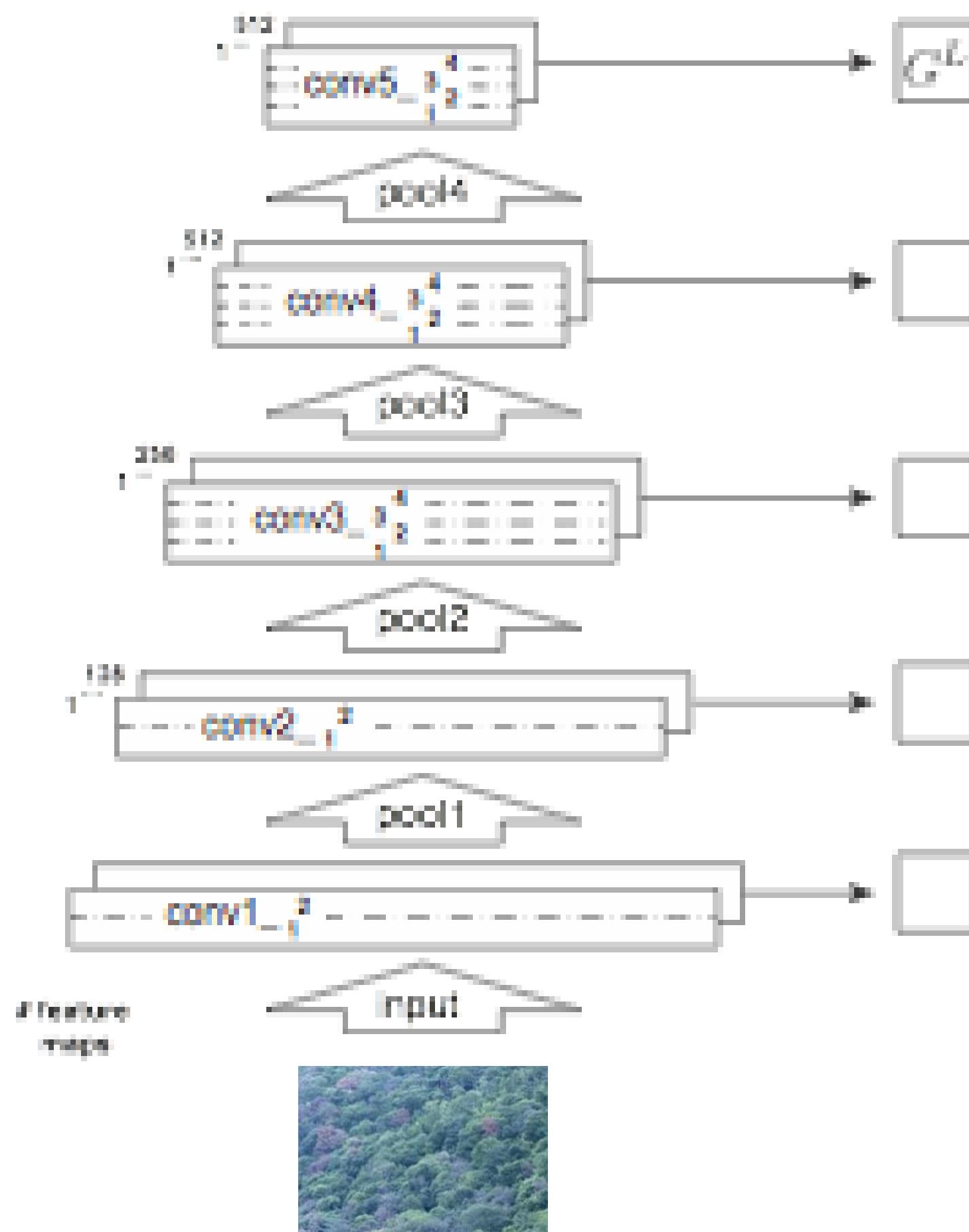
$$\langle \bar{f}_i, \bar{f}_j \rangle = \sum_k F_{ik} F_{jk}$$

CNN-Texture Features

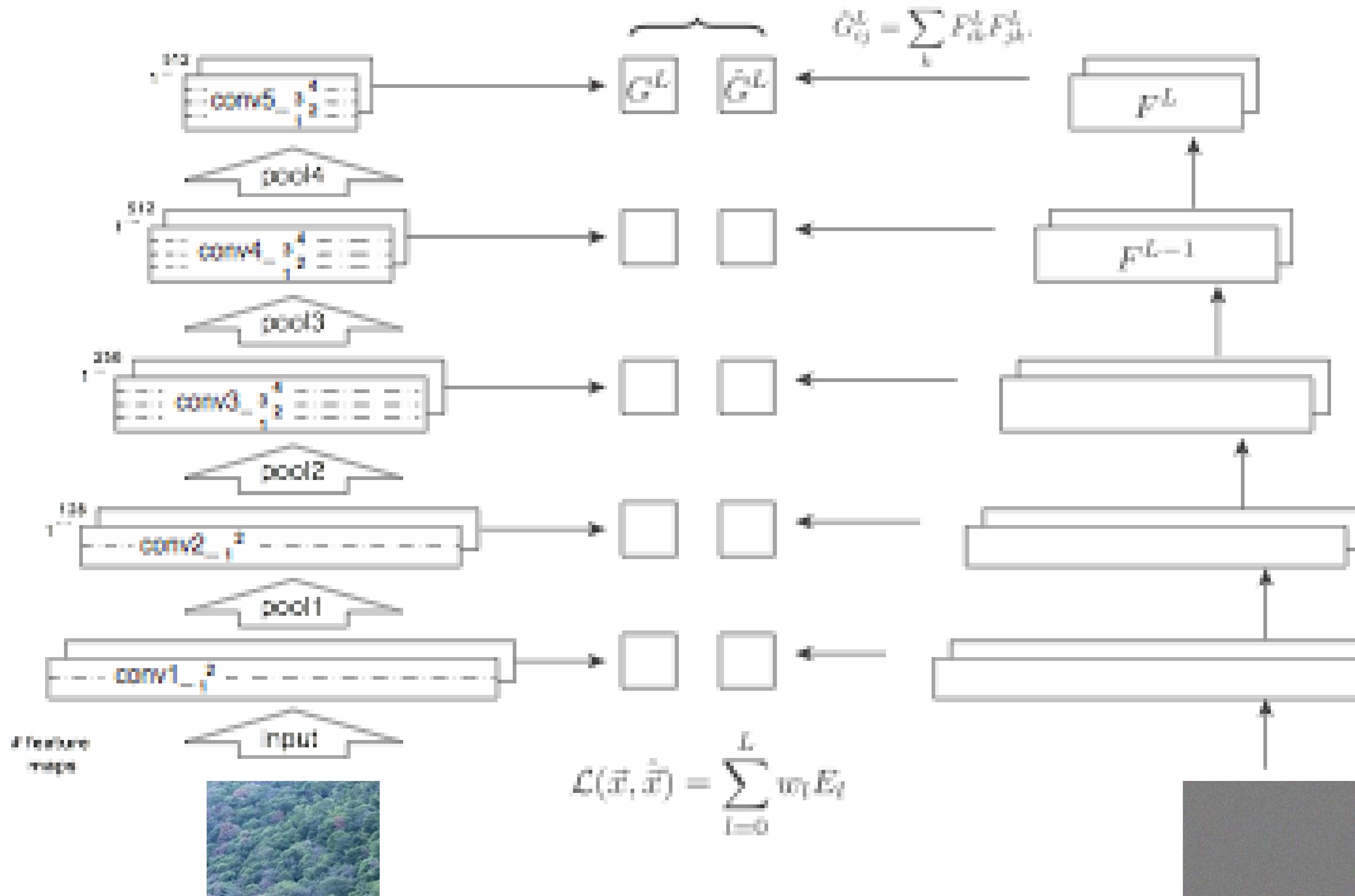
Gram Matrices



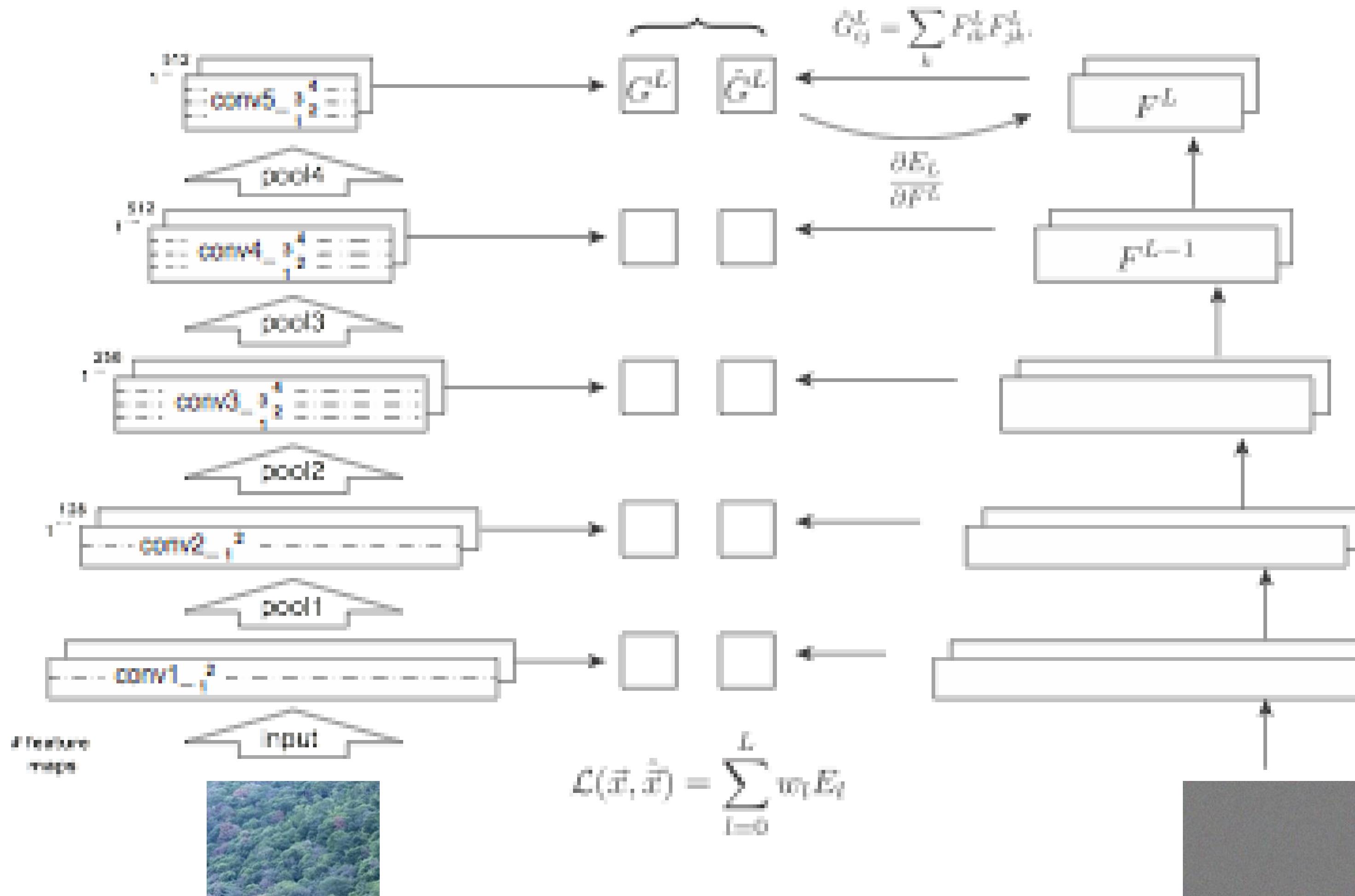
Texture Synthesis



Texture Synthesis



Texture Synthesis



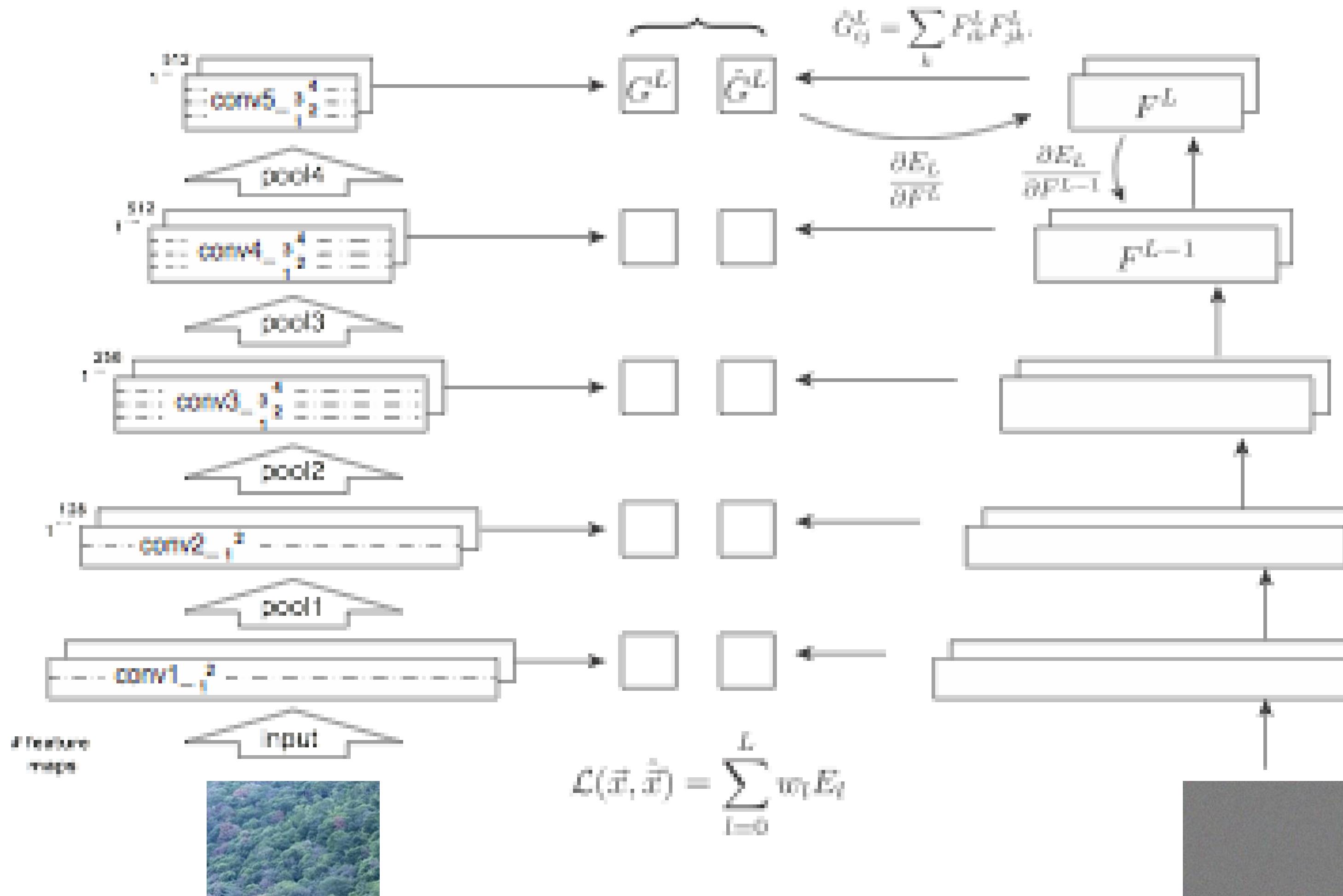
$$E_l = \sum_i (G^L_i - \tilde{G}^L_i)^2$$

$$G^L_i = \sum_k F^L_{ik} F^L_{jk}$$

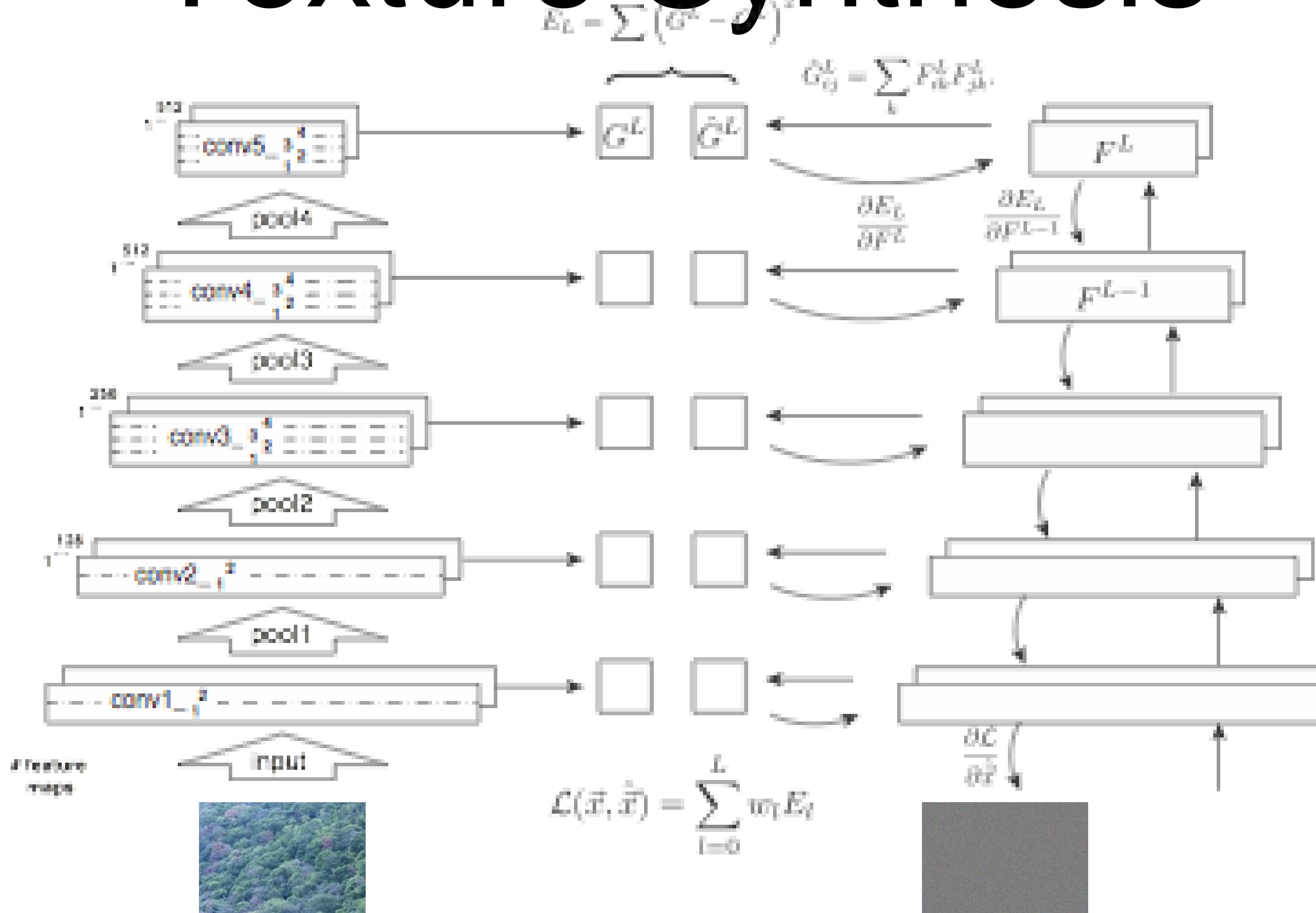
$$\frac{\partial E_l}{\partial F^L}$$

$$\mathcal{L}(\vec{x}, \hat{\vec{x}}) = \sum_{l=0}^L w_l E_l$$

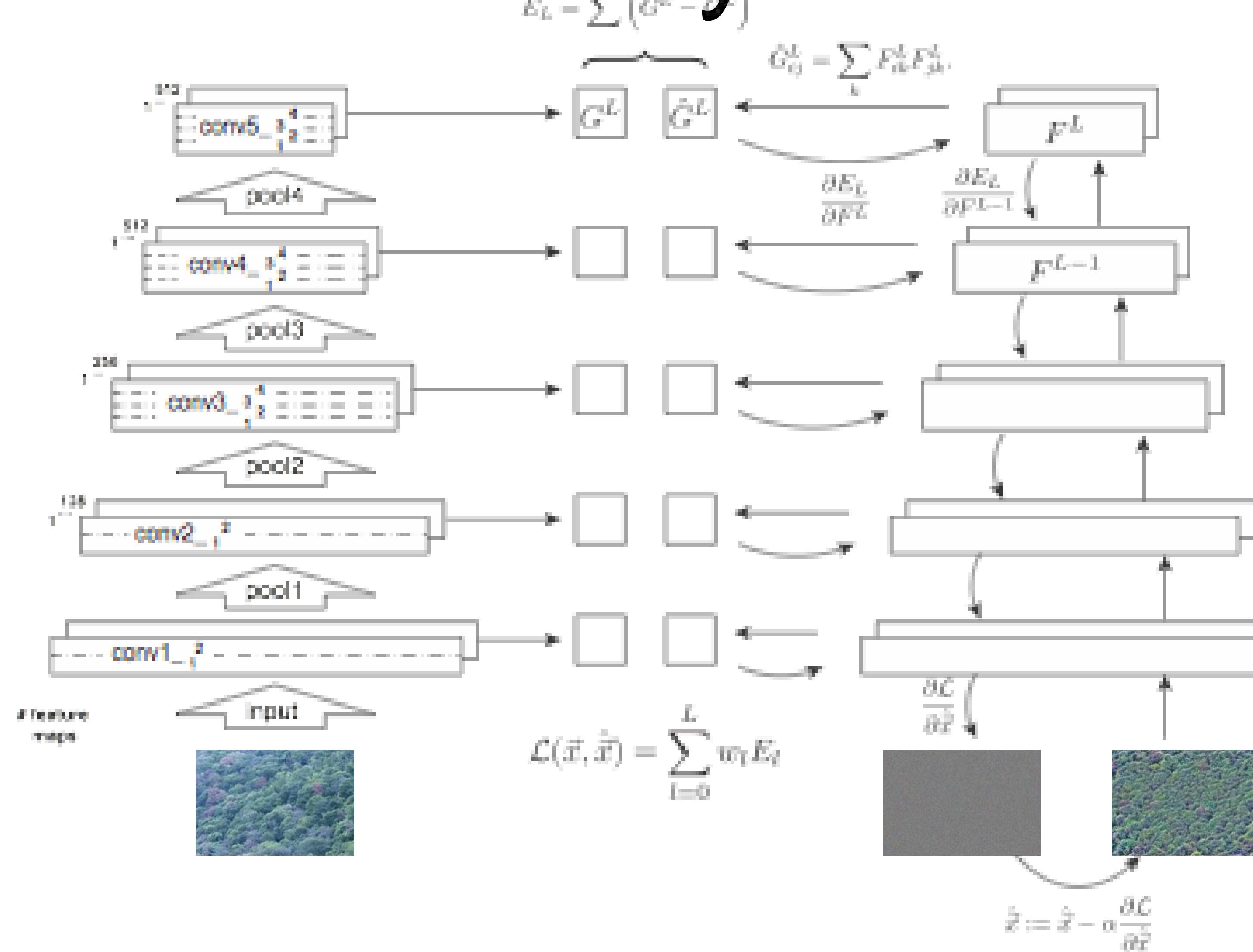
Texture Synthesis



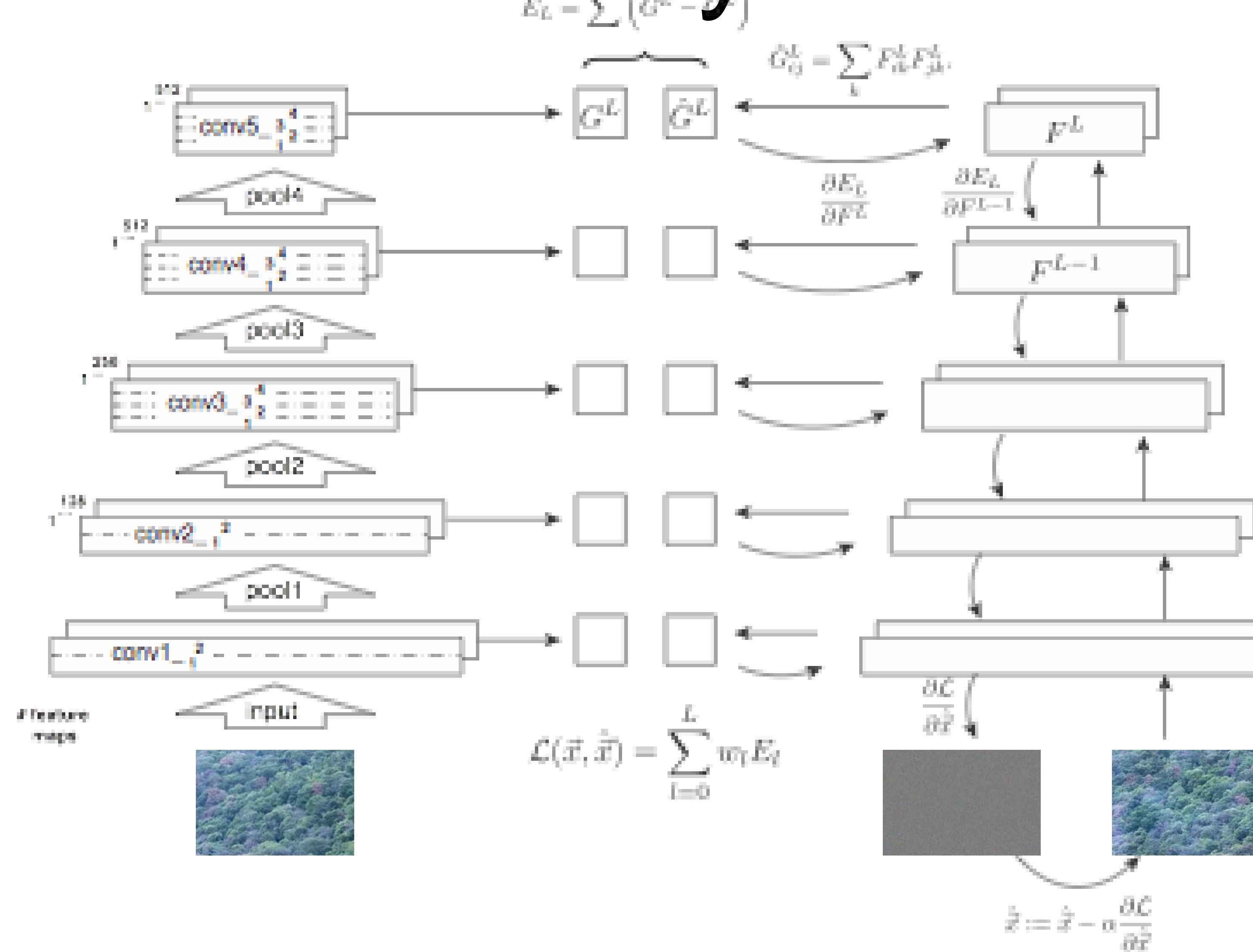
Texture Synthesis



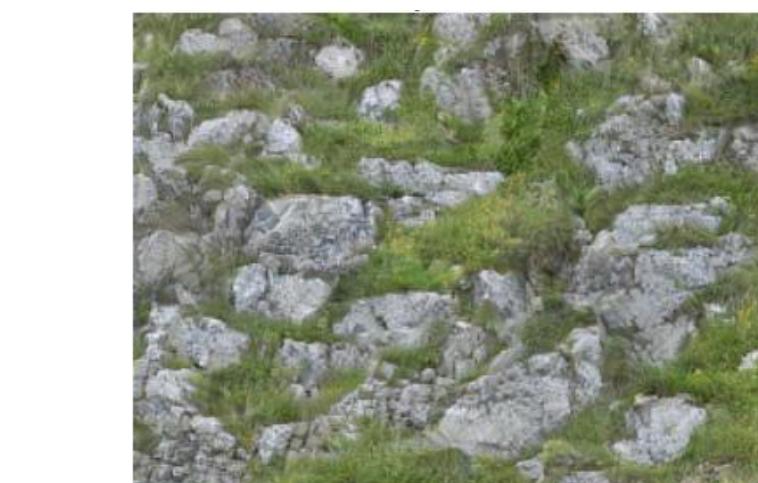
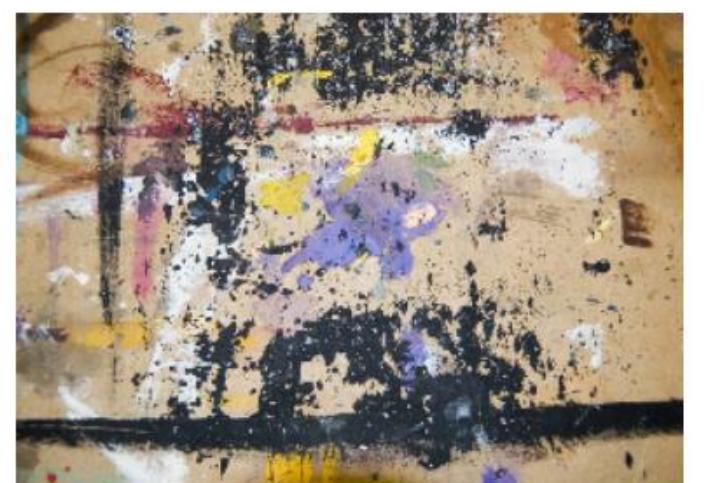
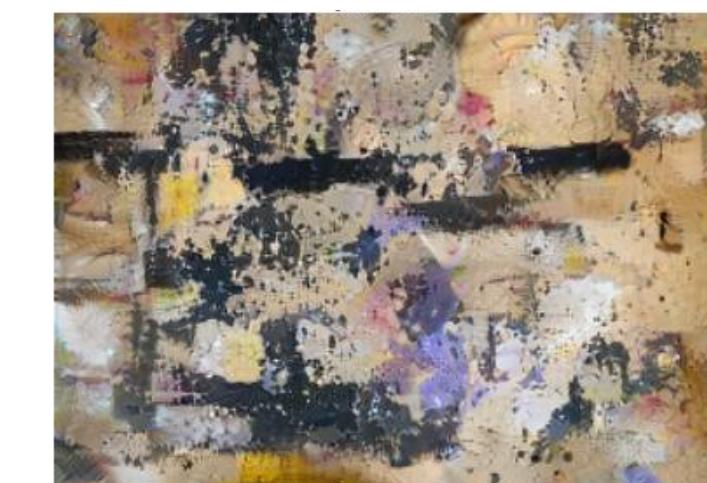
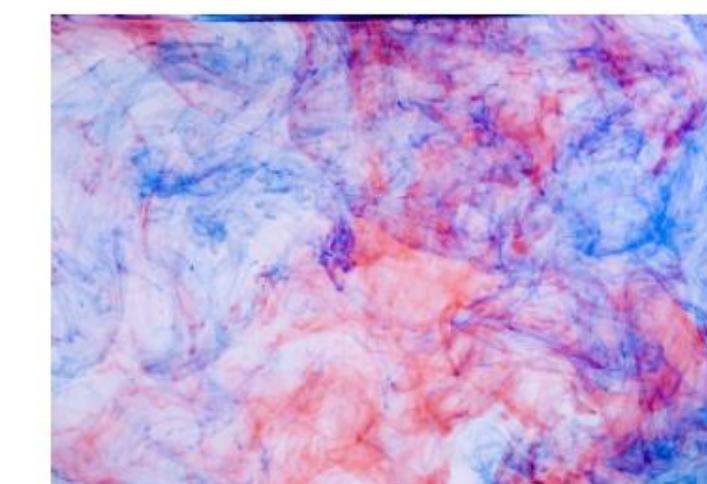
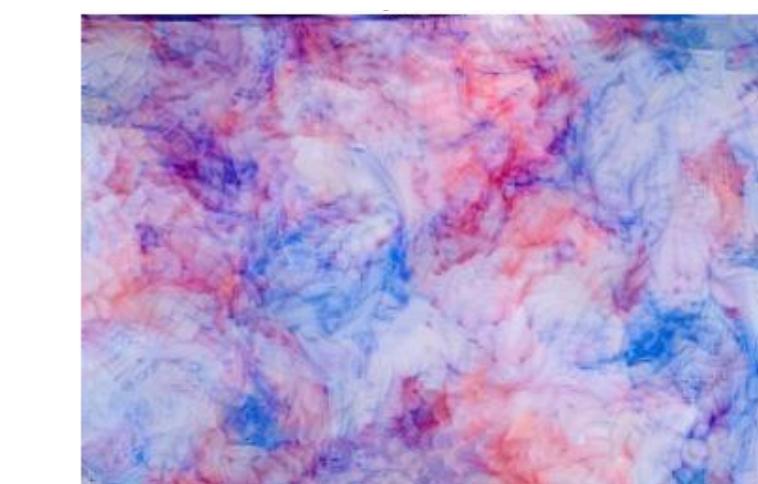
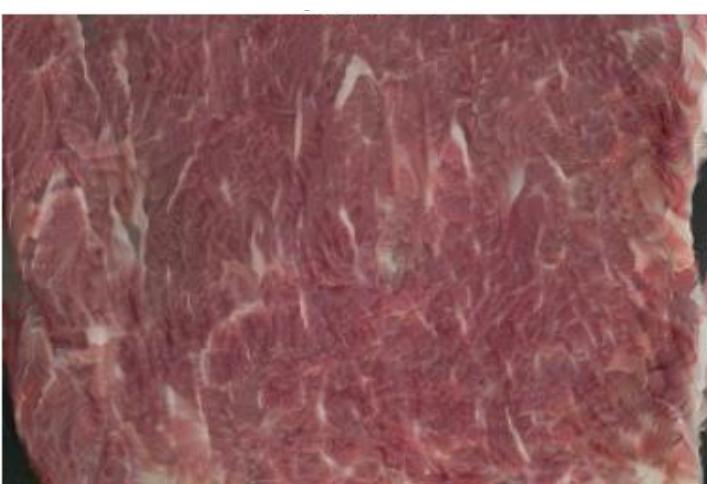
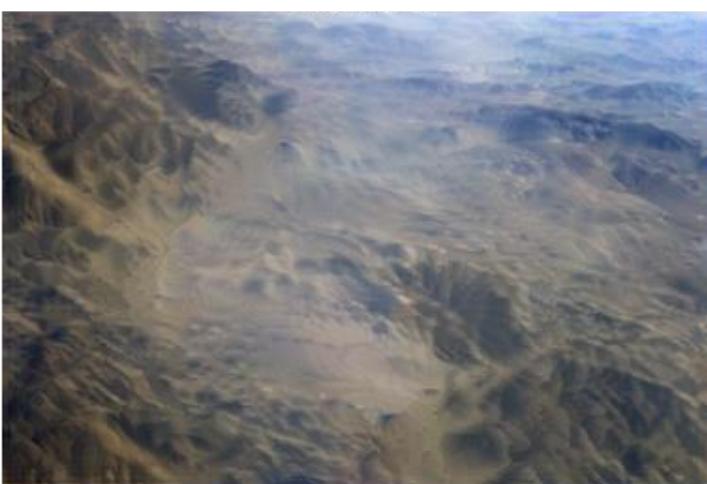
Texture Synthesis



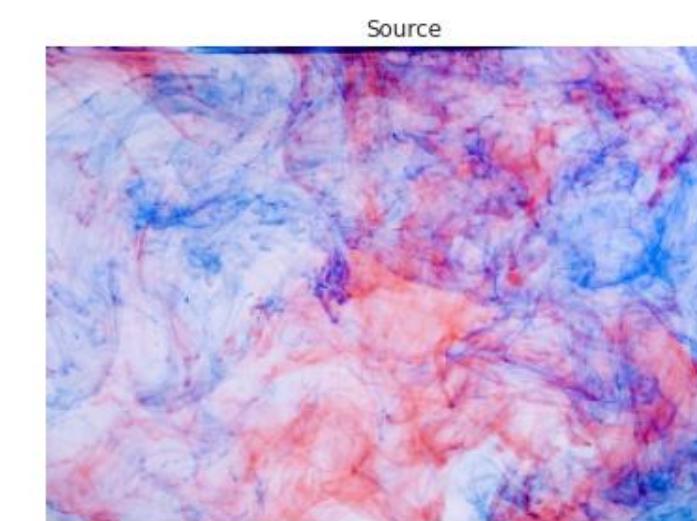
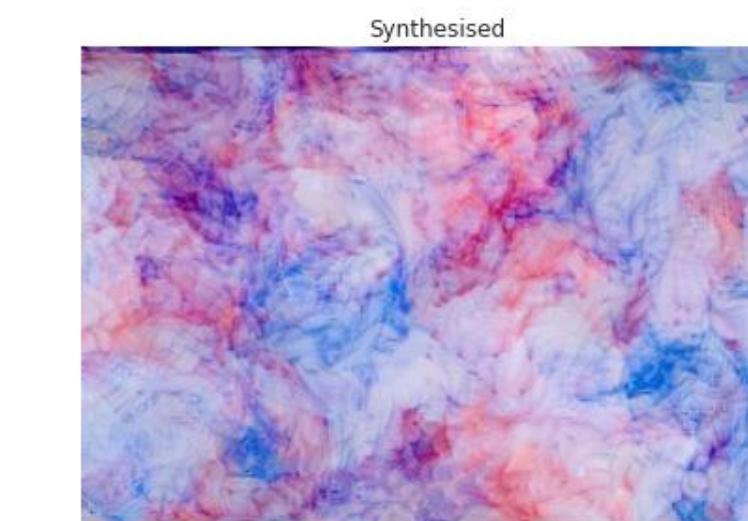
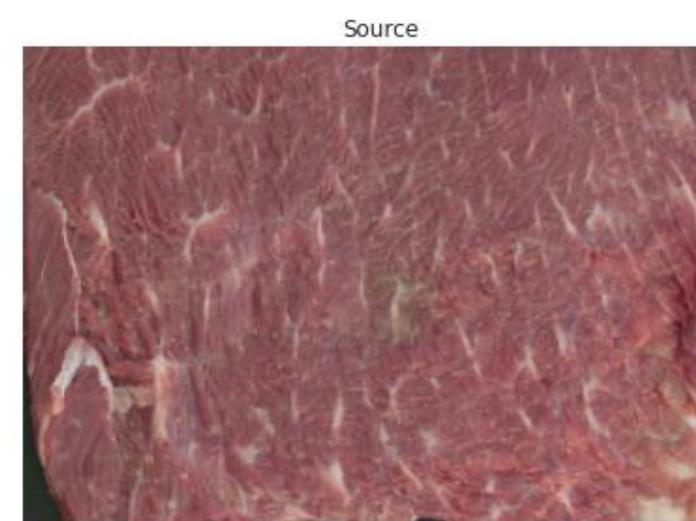
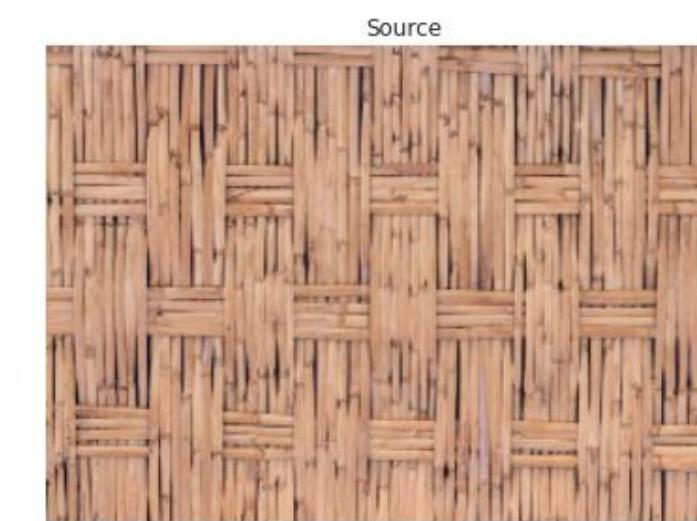
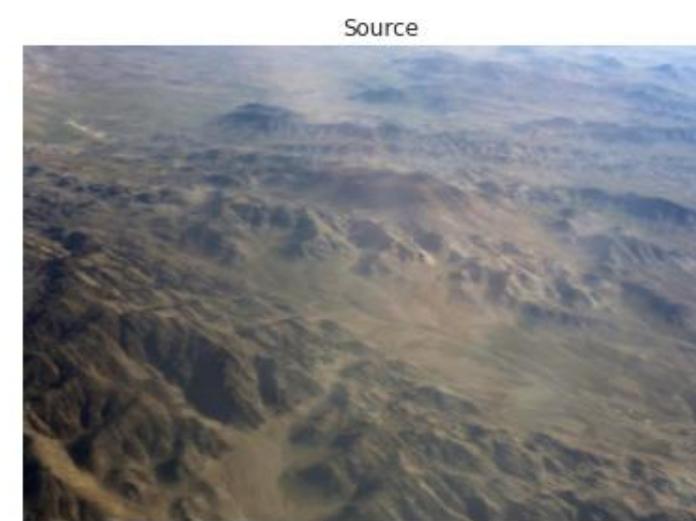
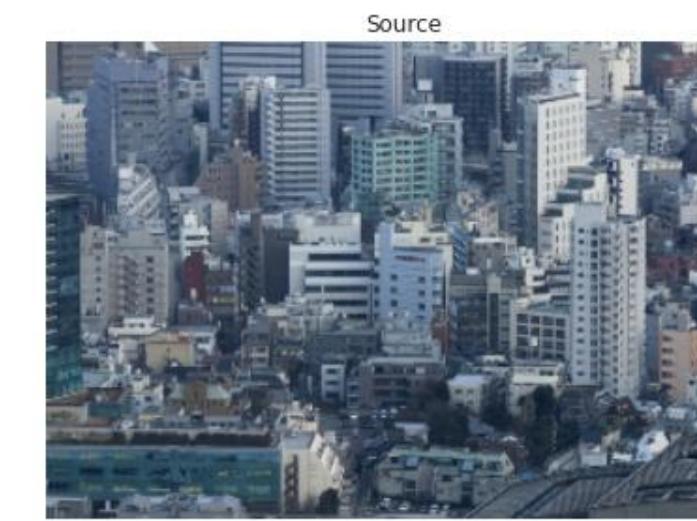
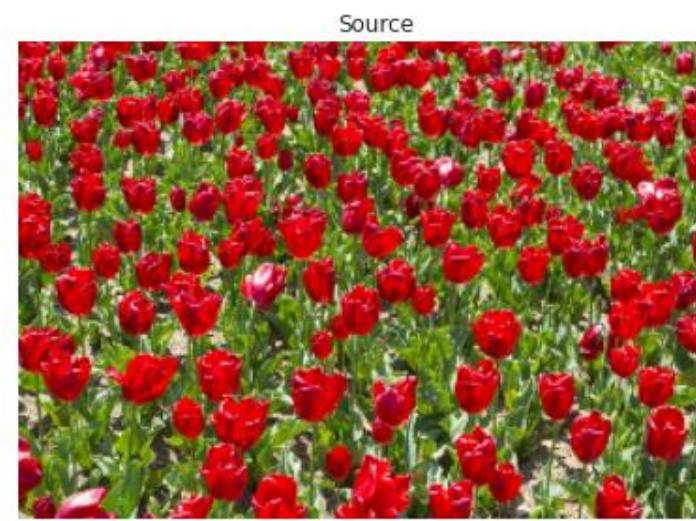
Texture Synthesis



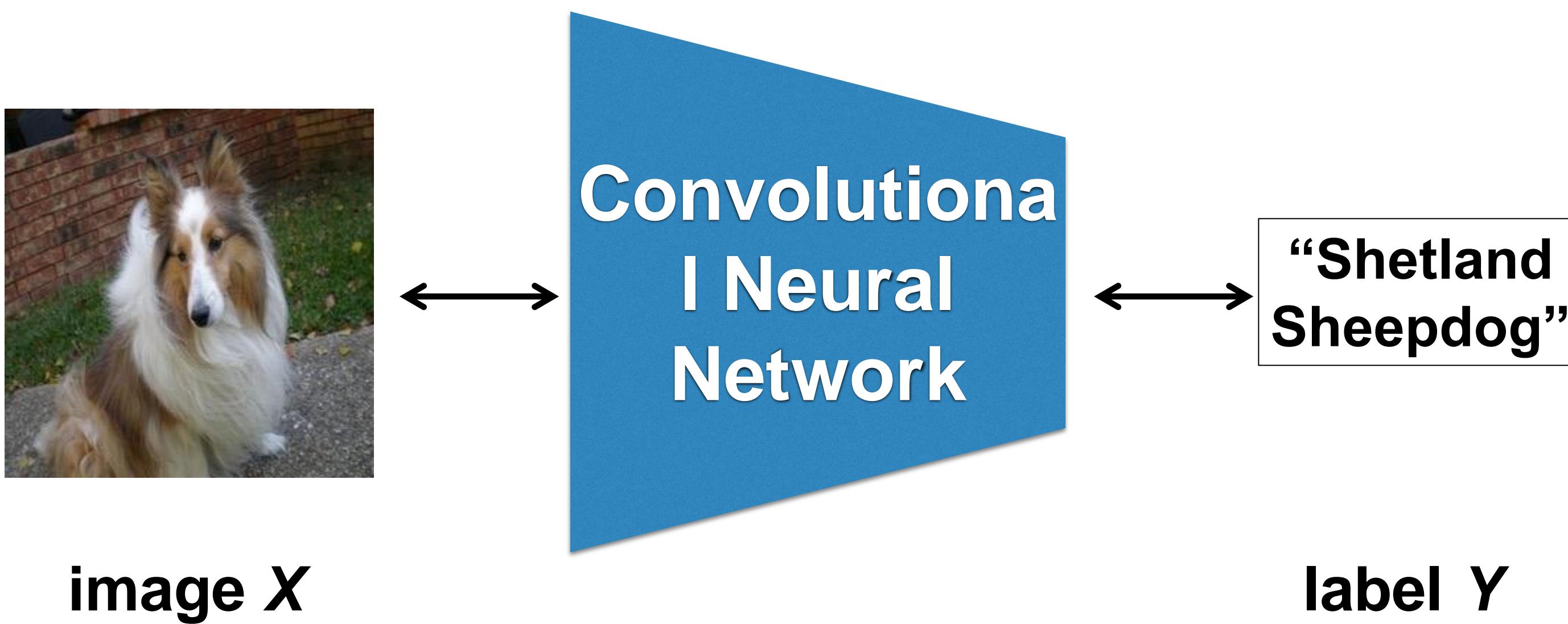
Test Julesz' Conjecture



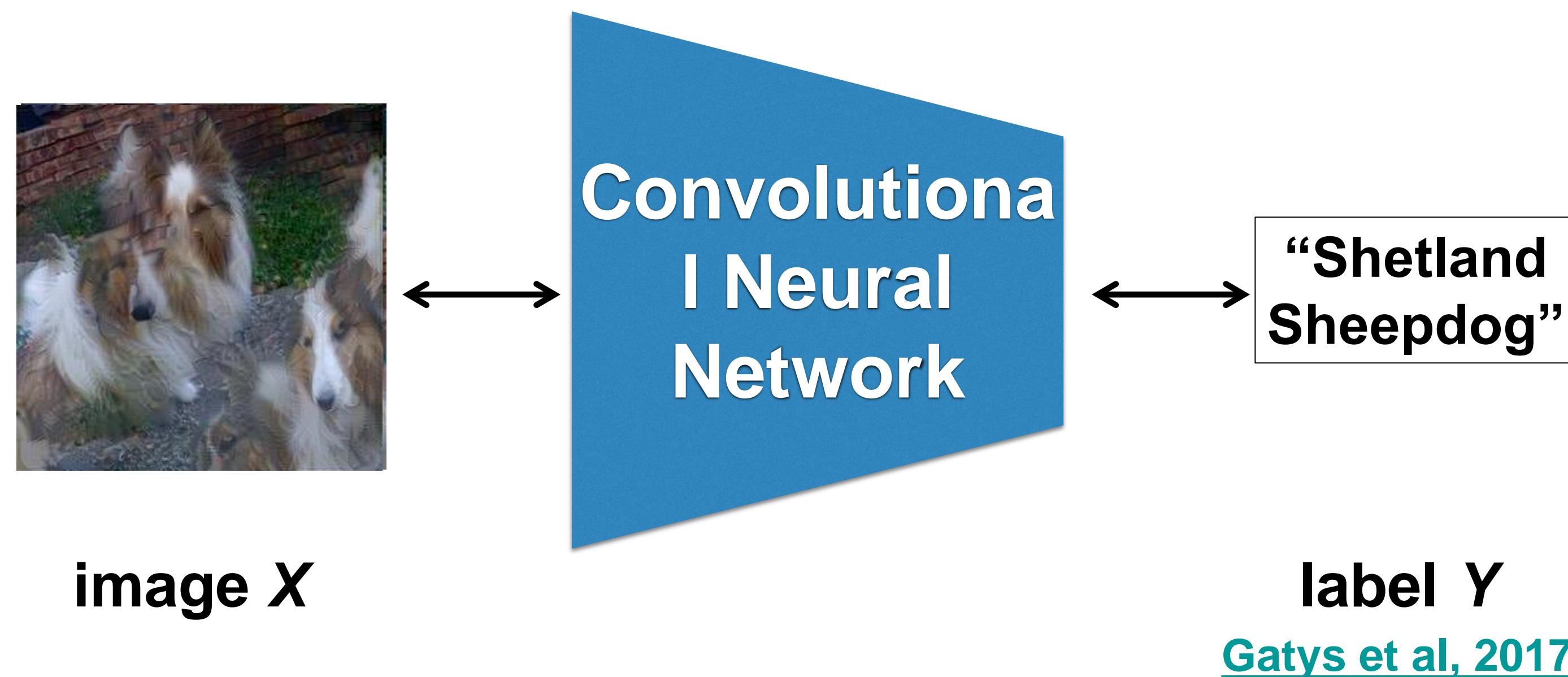
Test Julesz' Conjecture



ImageNet Recognition is just Texture Recognition



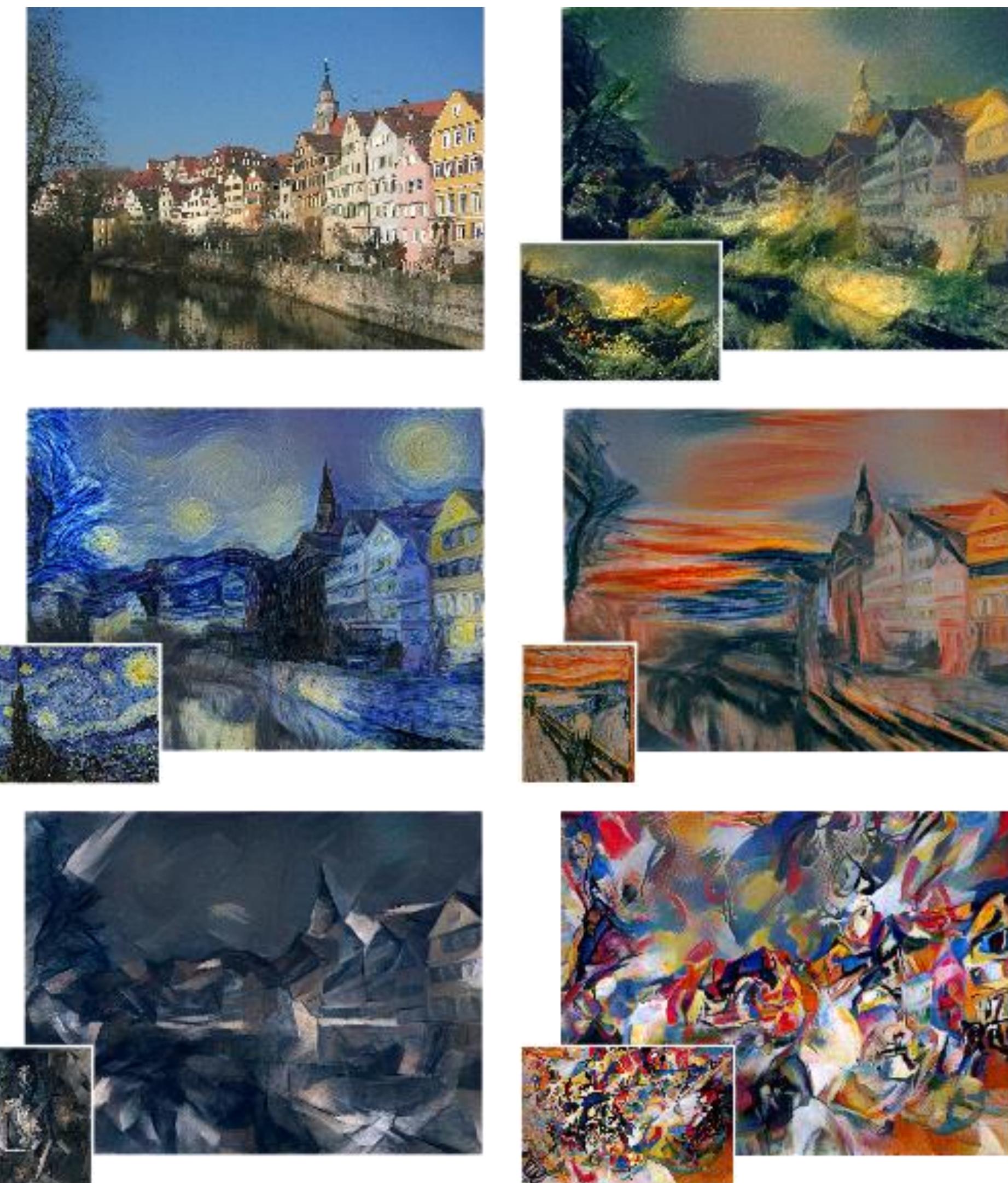
ImageNet Recognition is just Texture Recognition



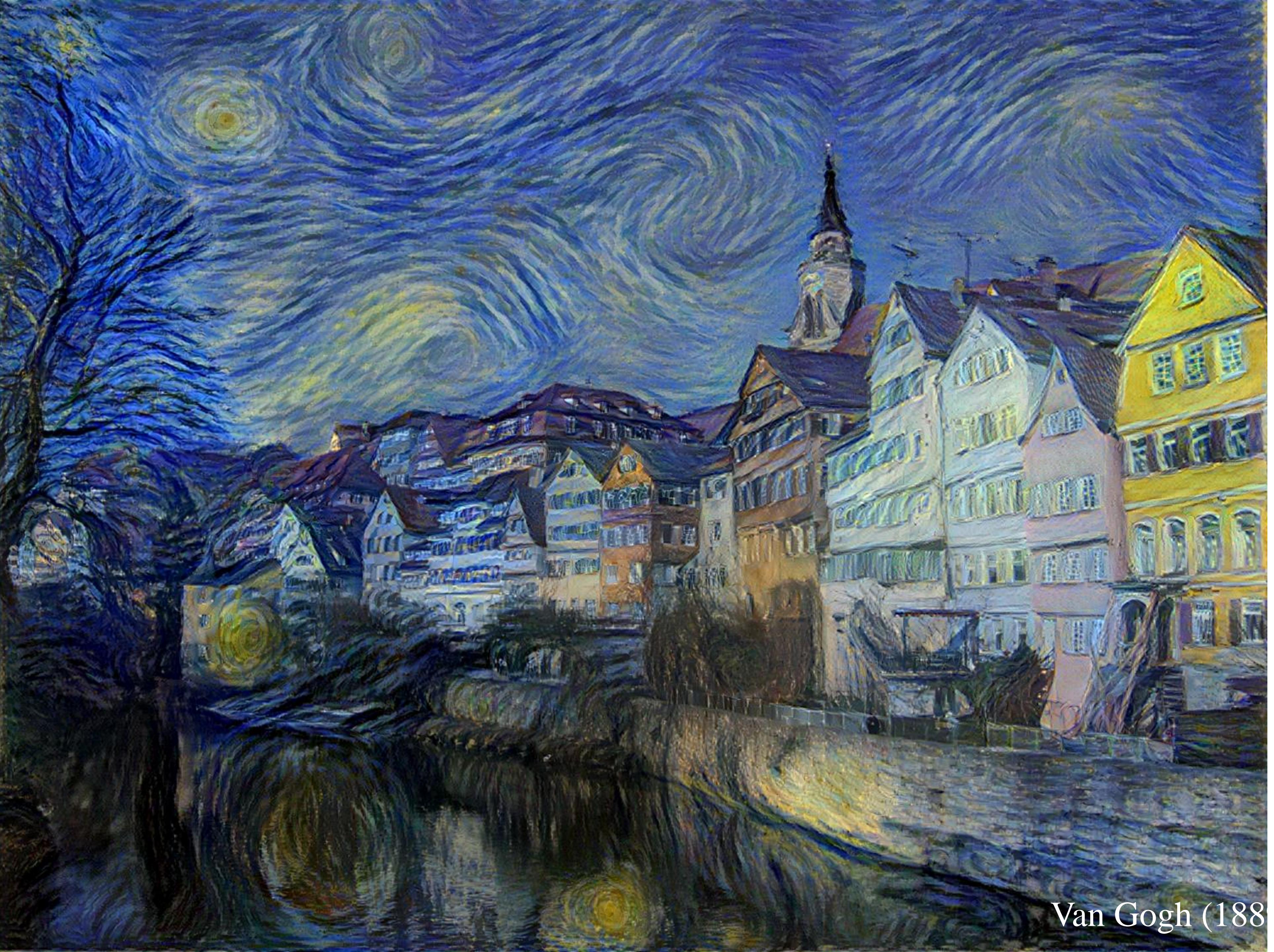
A Neural Algorithm of Artistic Style



Gatys, Ecker, Bethge (arXiv 2015)







Van Gogh (188



Picasso (1910)



Munch (1893)

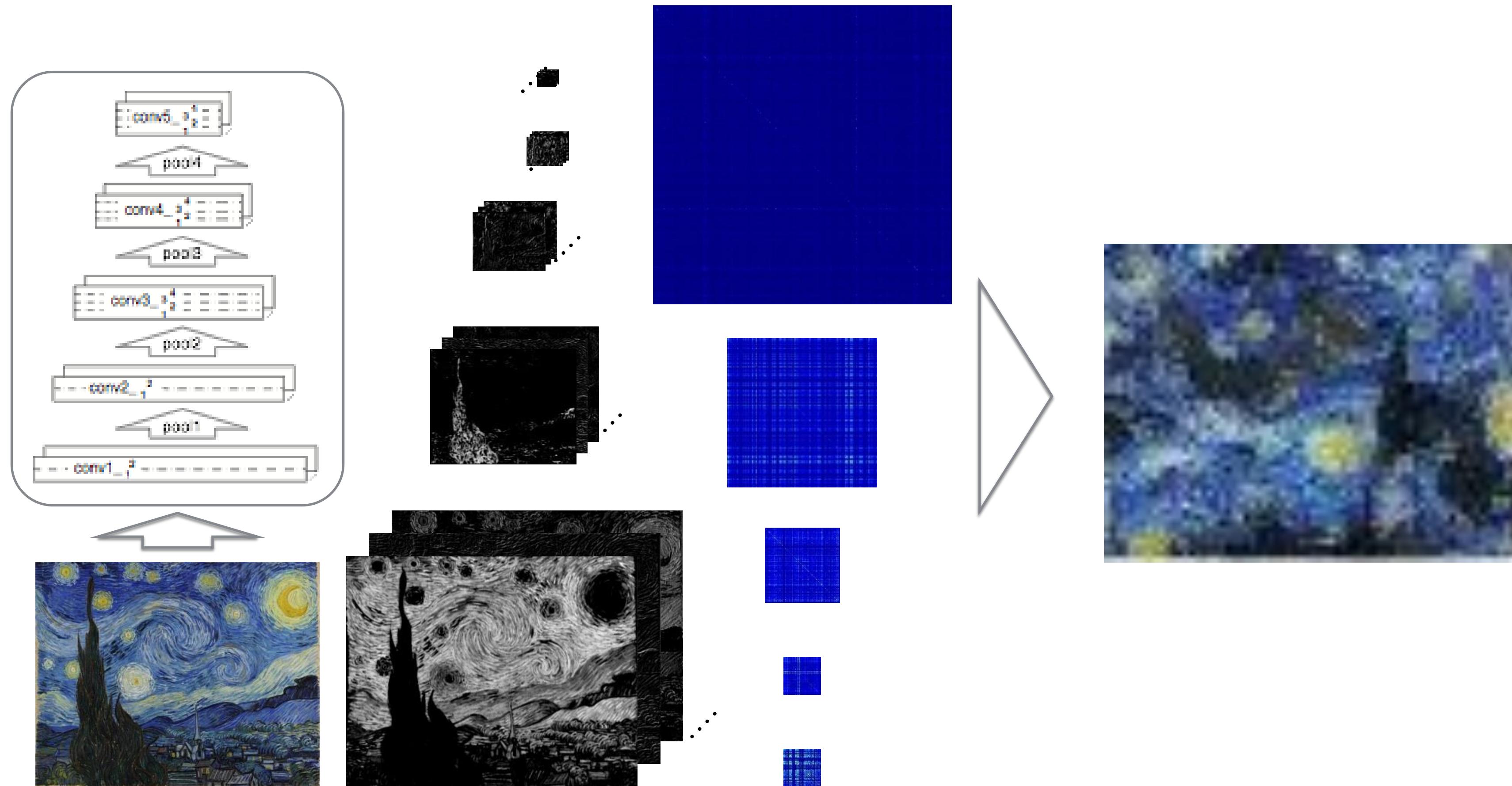


Turner (1805)



Kandinsky (191)

CNN - Texture Synthesis

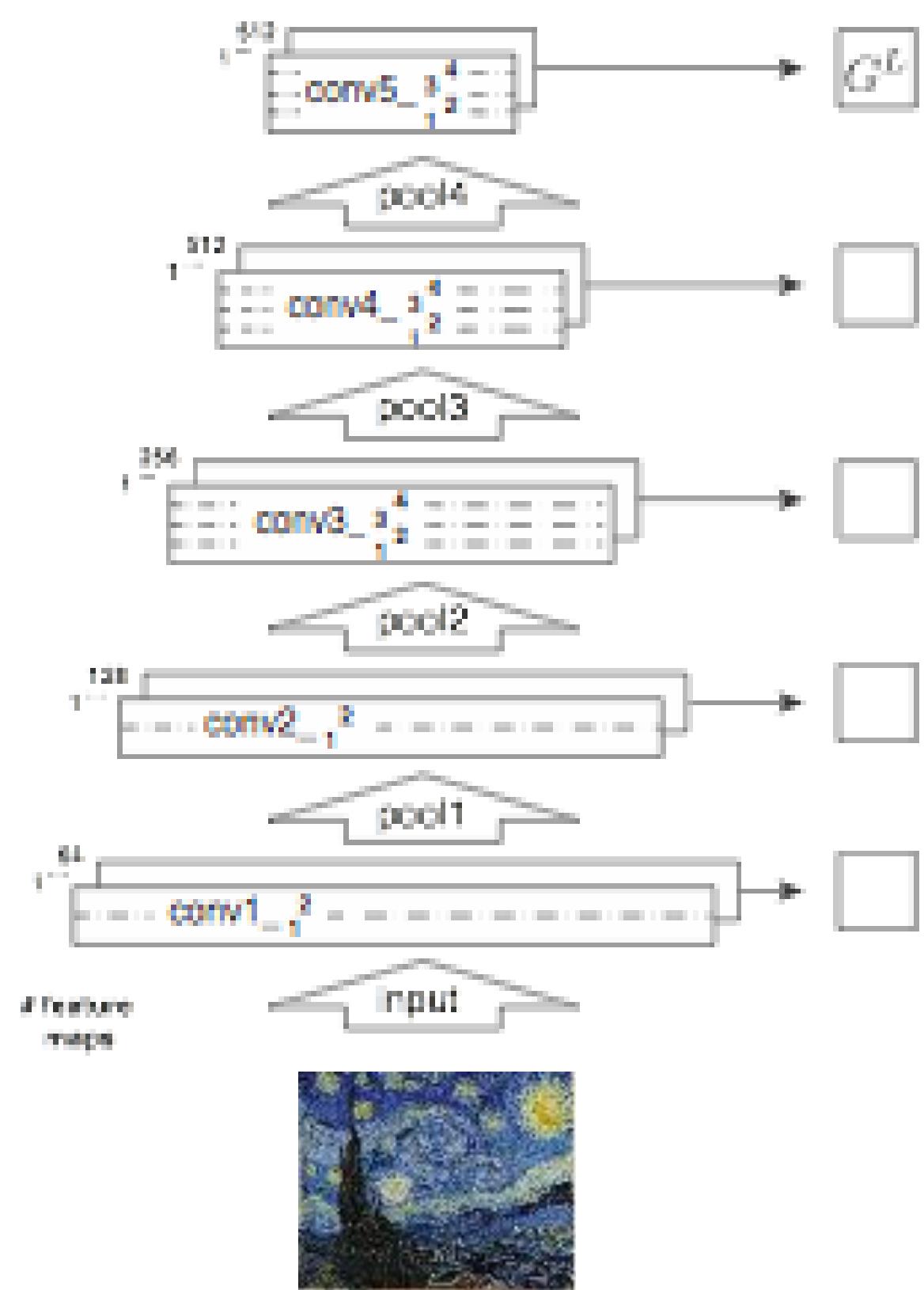


Gatys et al. (NIPS 2015)

Artistic Style Transfer



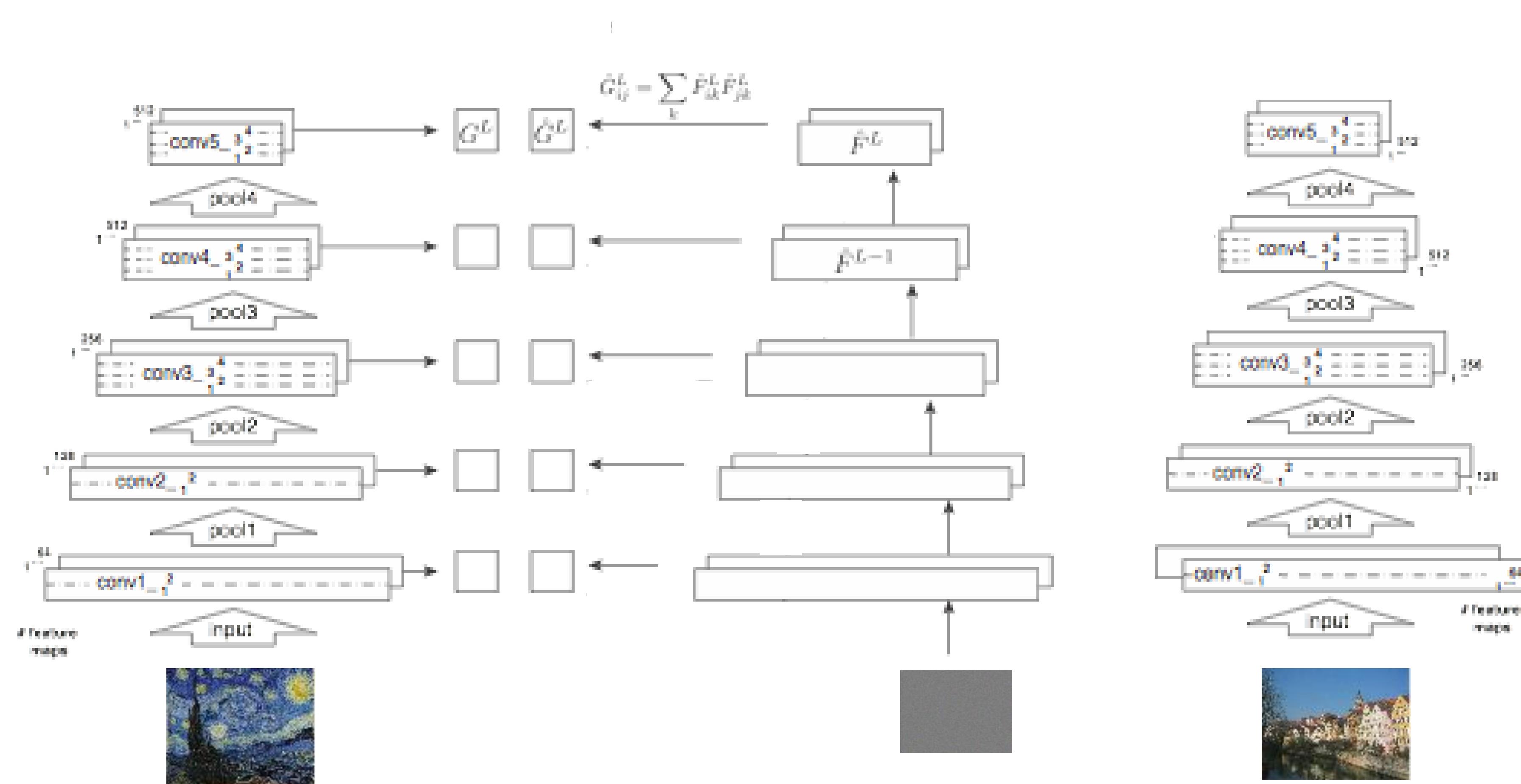
Artistic Style Transfer



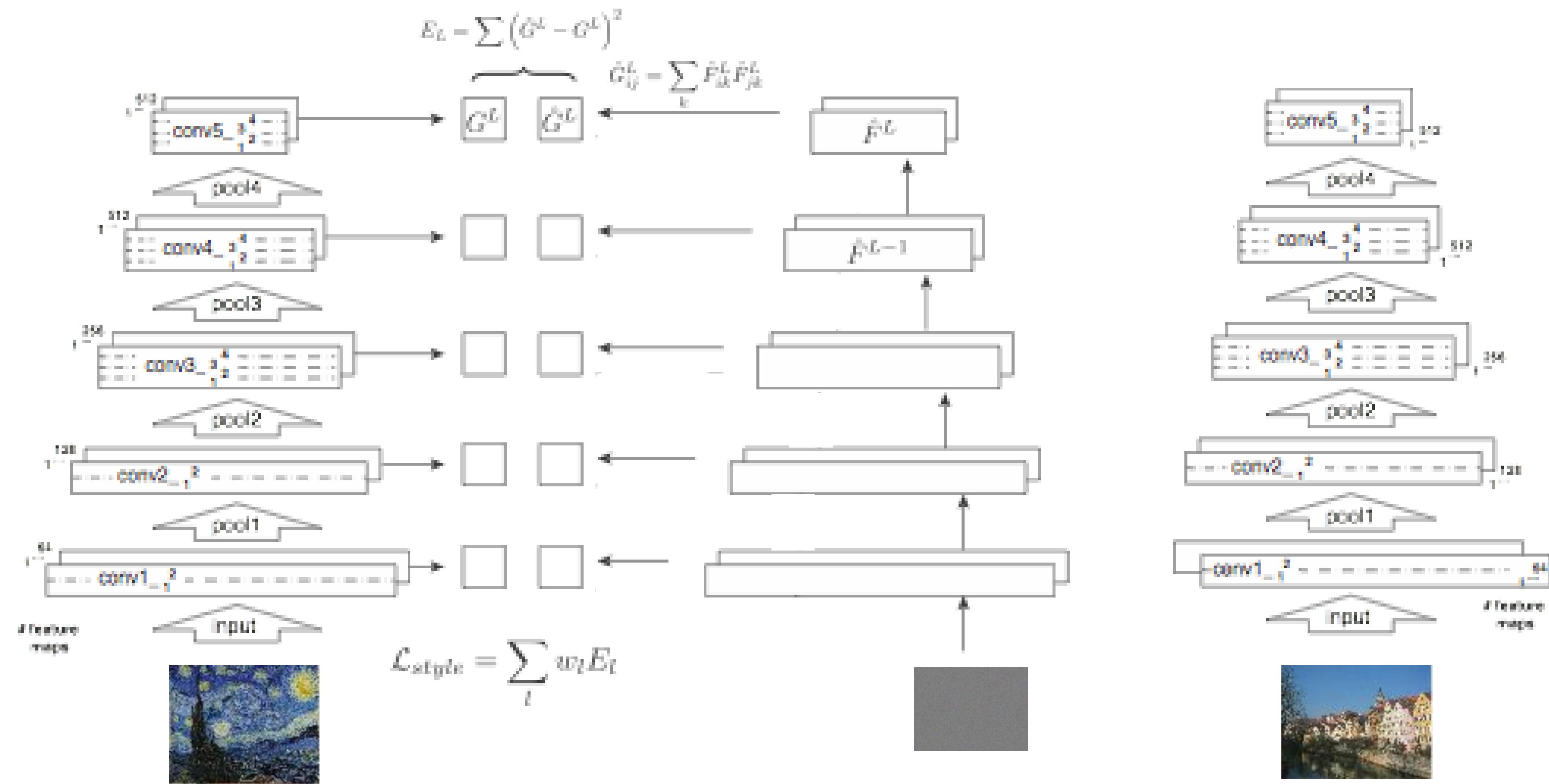
Artistic Style Transfer



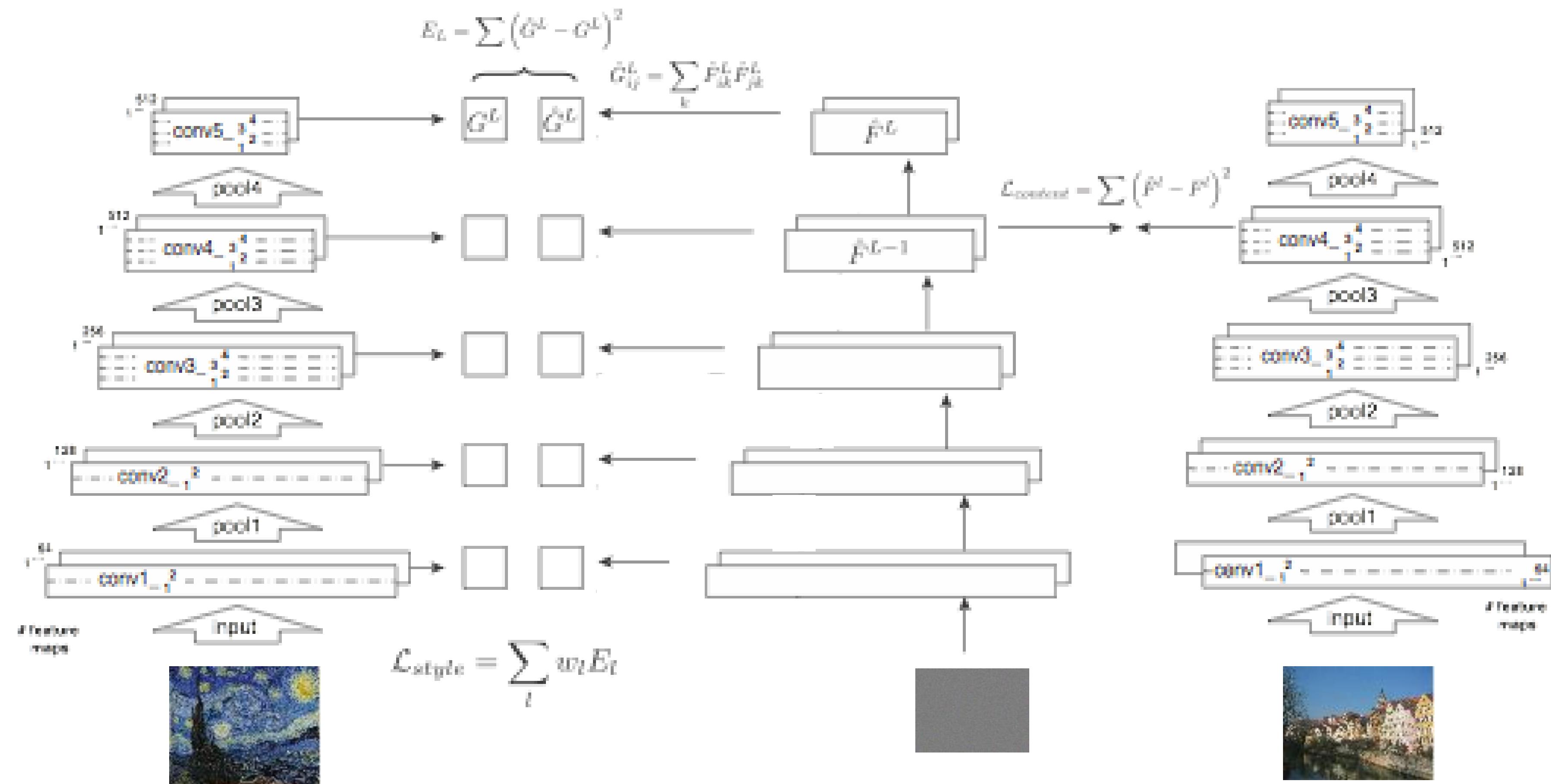
Artistic Style Transfer



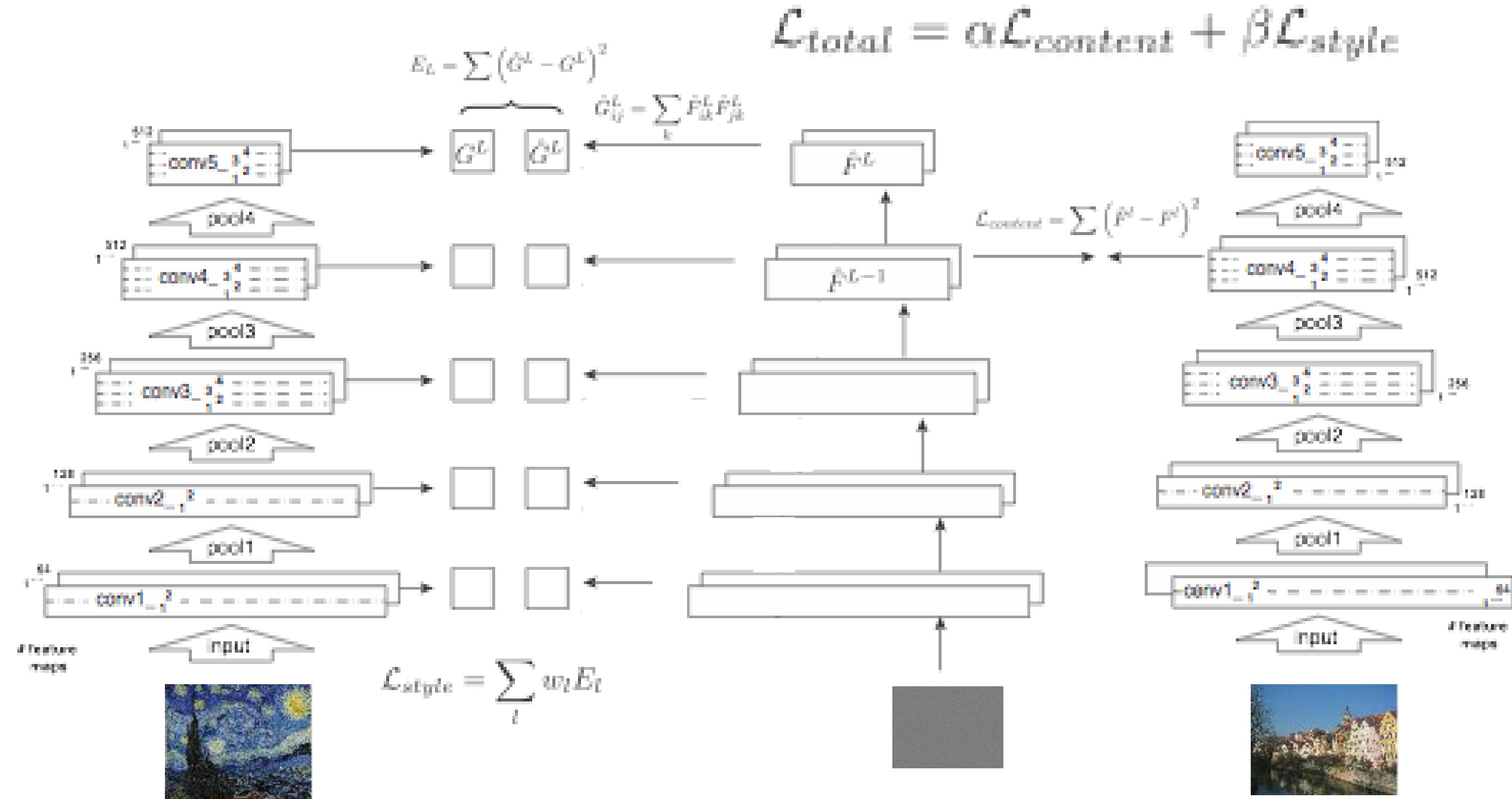
Artistic Style Transfer



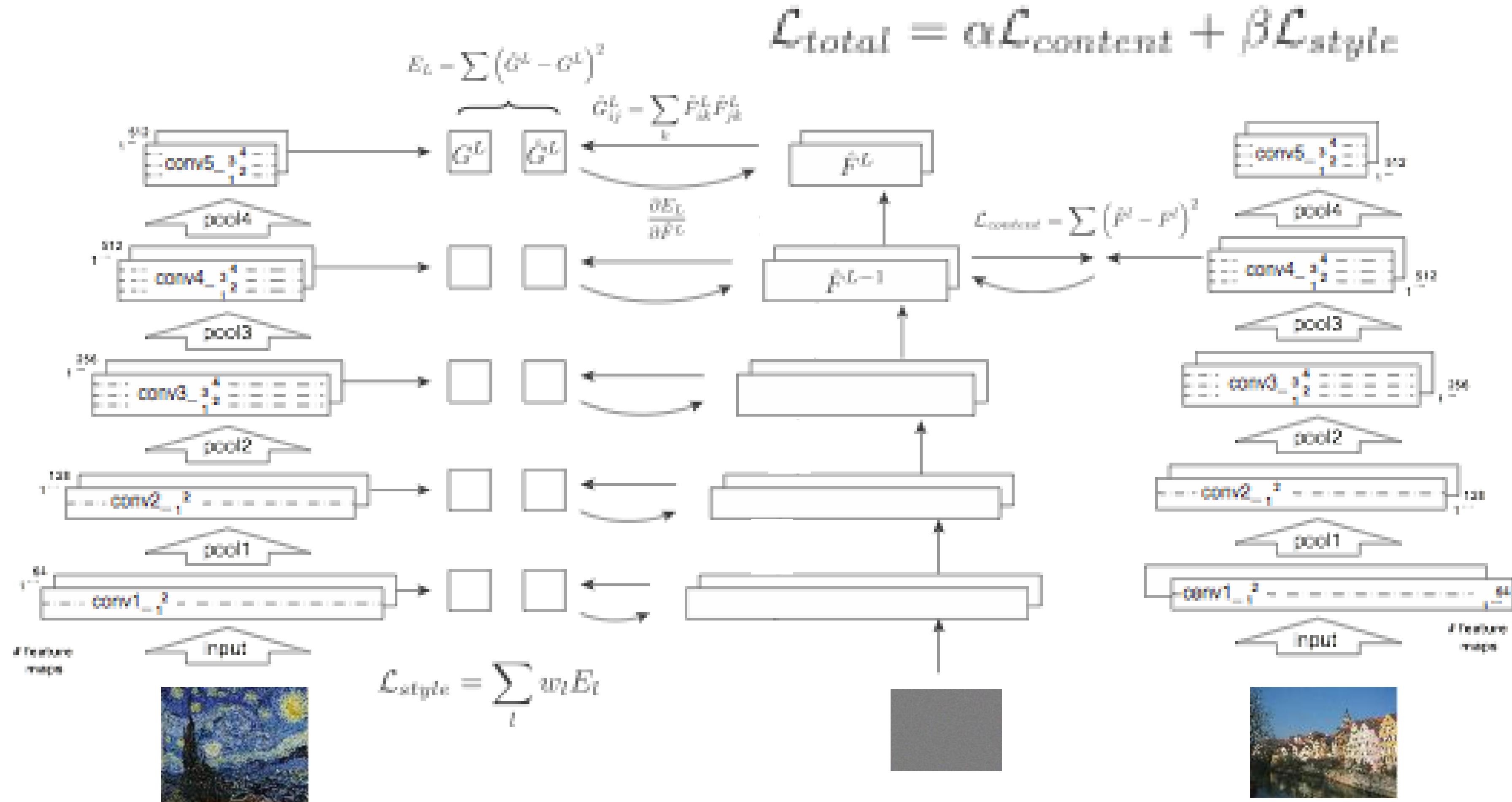
Artistic Style Transfer



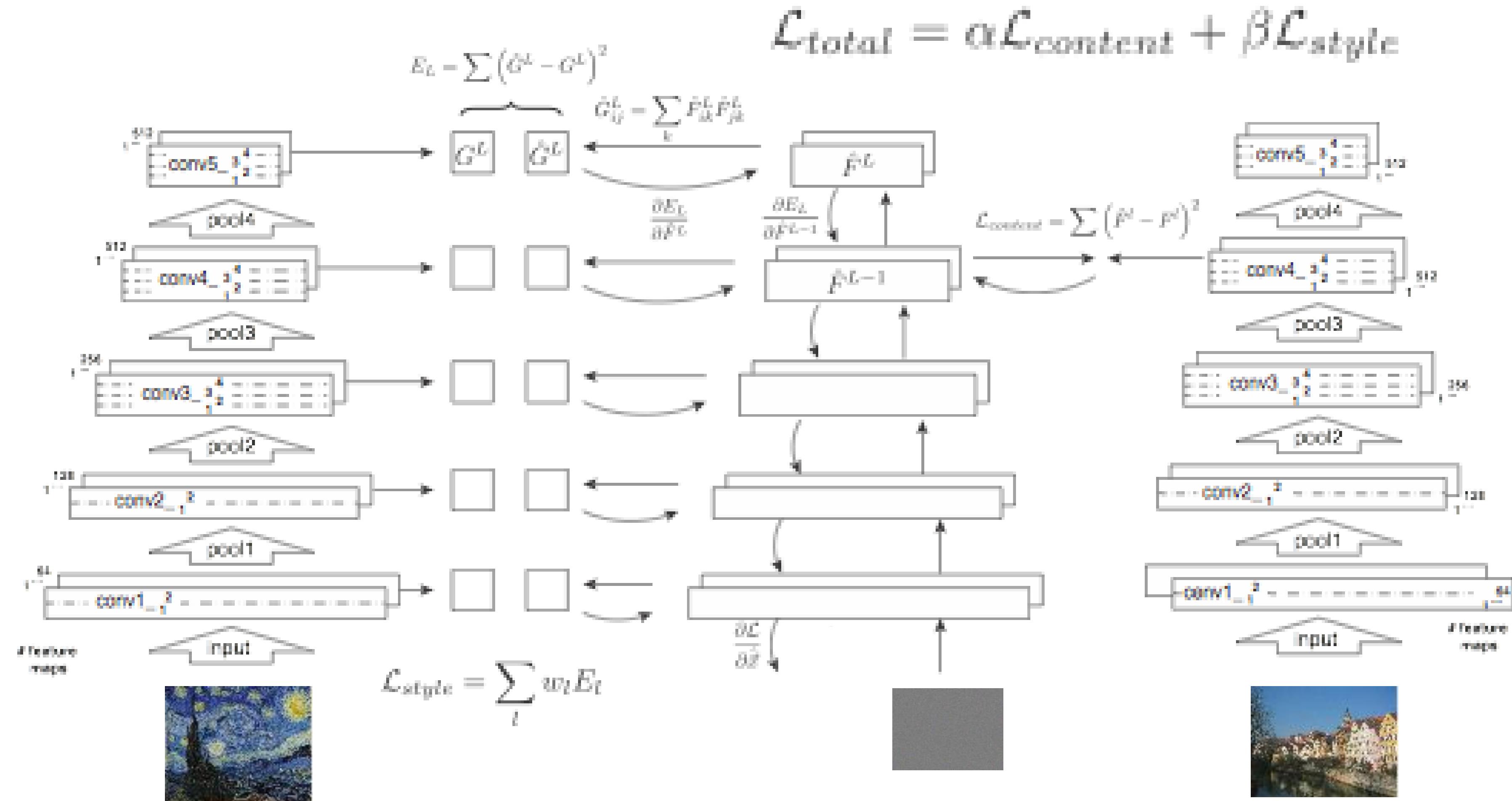
Artistic Style Transfer



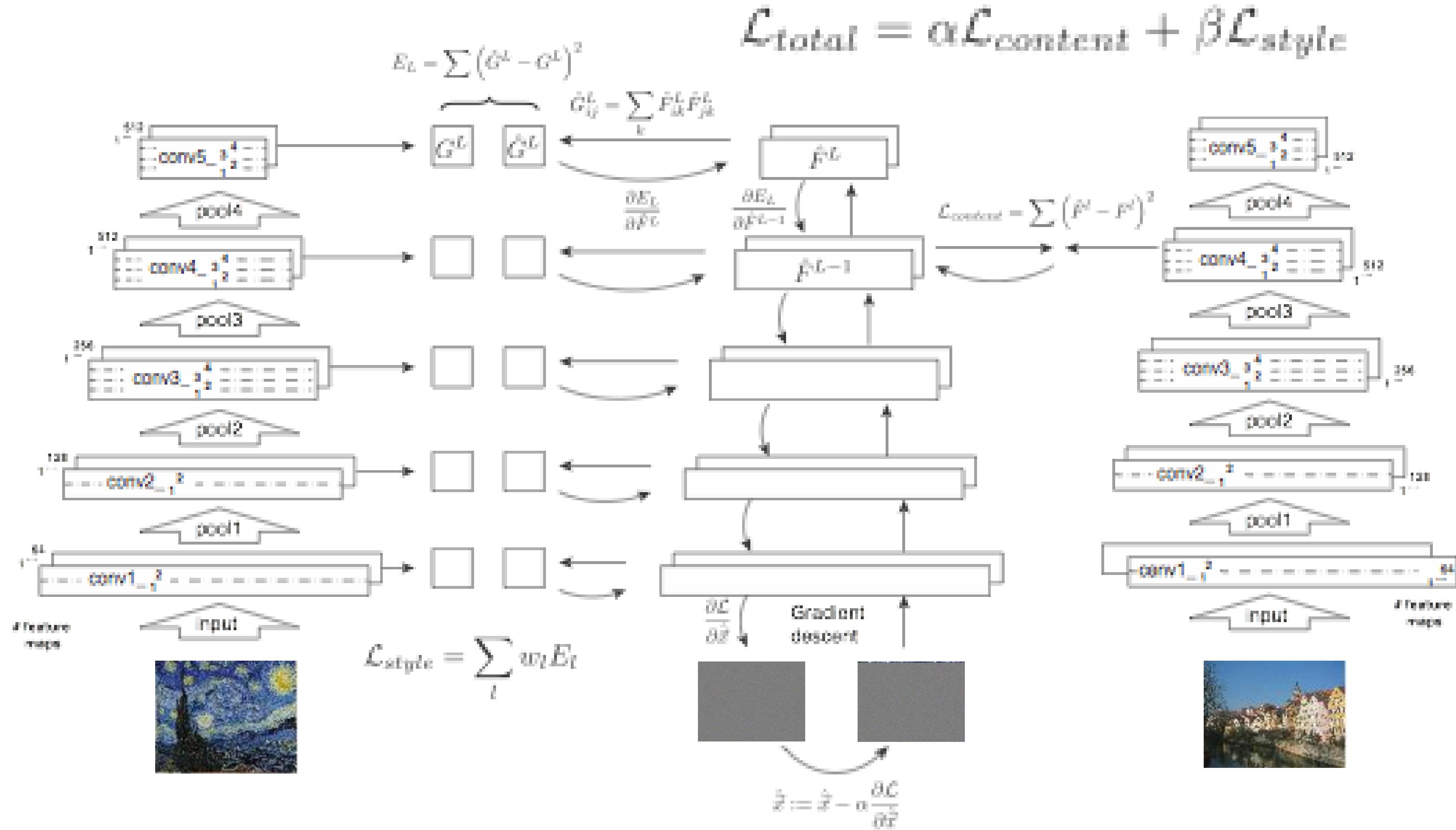
Artistic Style Transfer



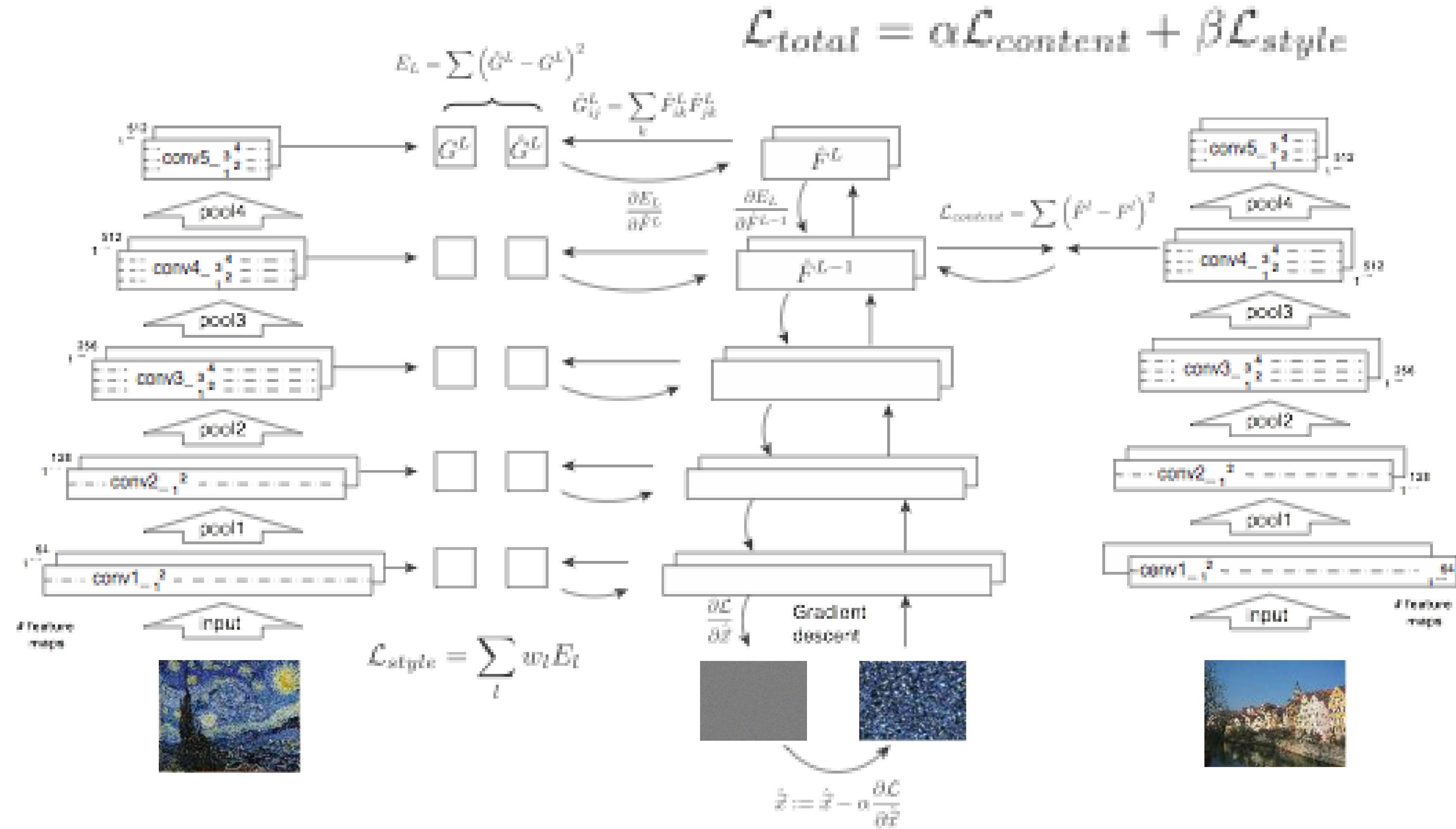
Artistic Style Transfer



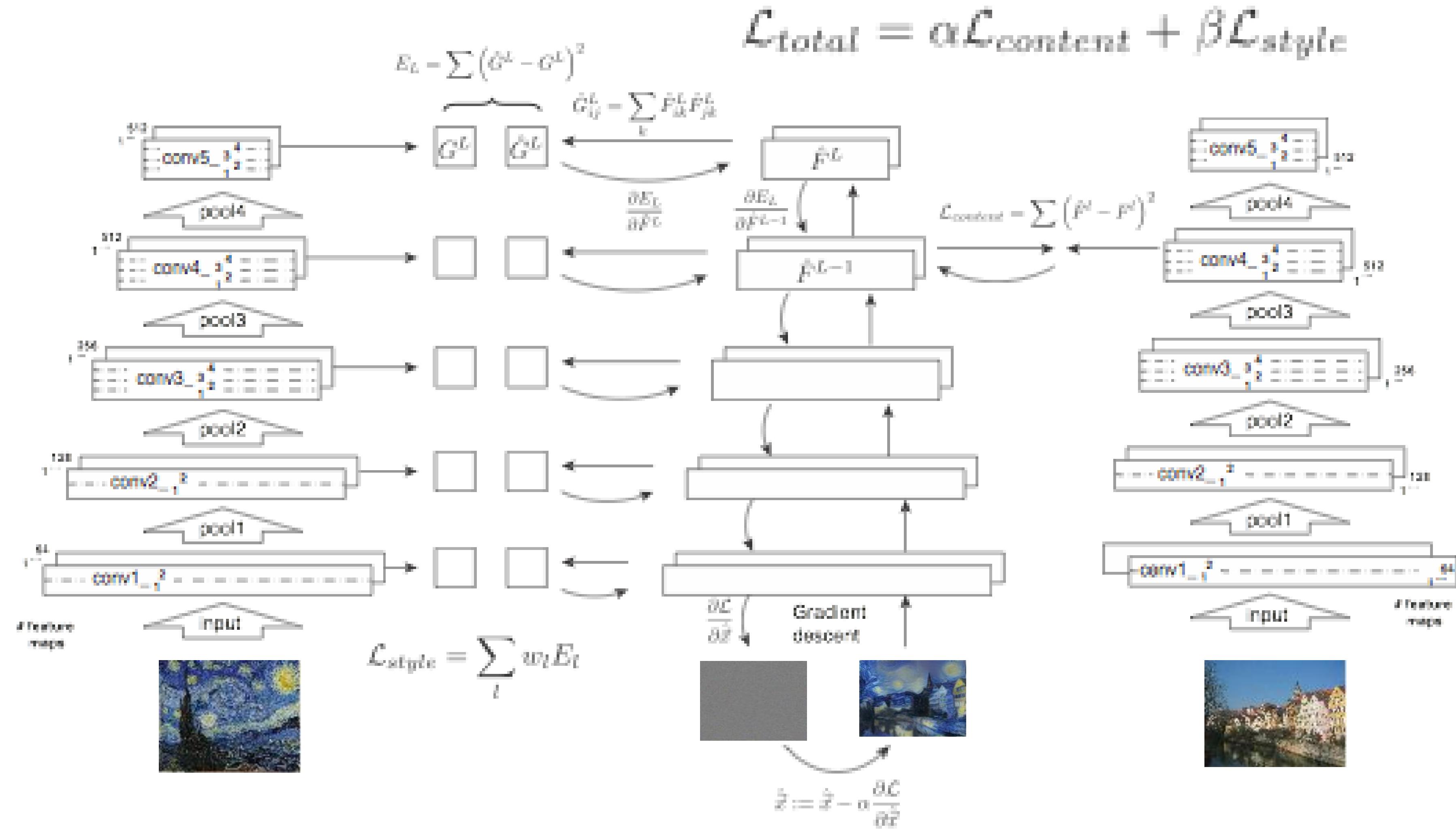
Artistic Style Transfer



Artistic Style Transfer



Artistic Style Transfer



Relative Weighting of Content and Style

1e-4



1e-3



1e-2



1e-1



Different Reconstruction Layers

Conv2_2

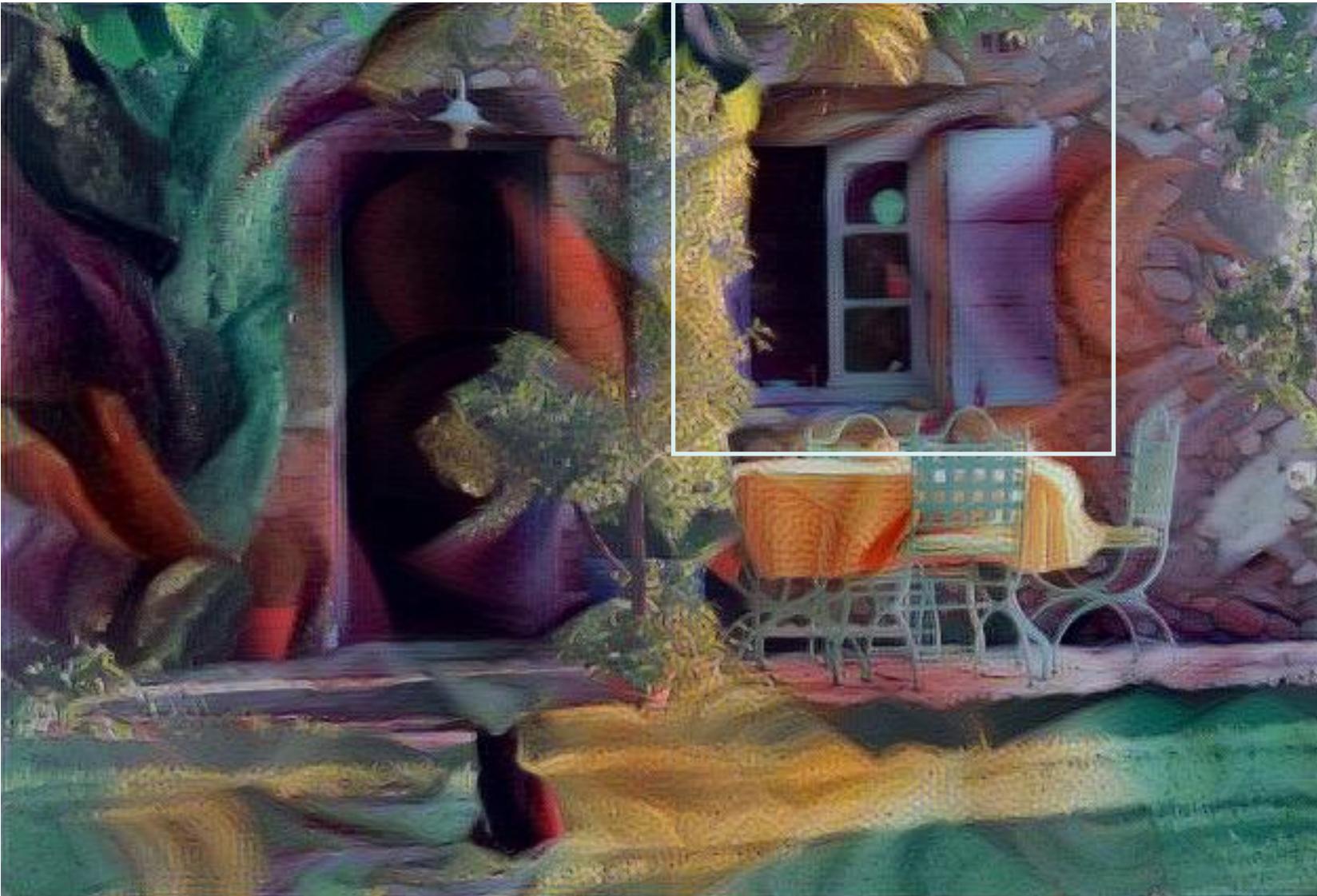


Conv4_2

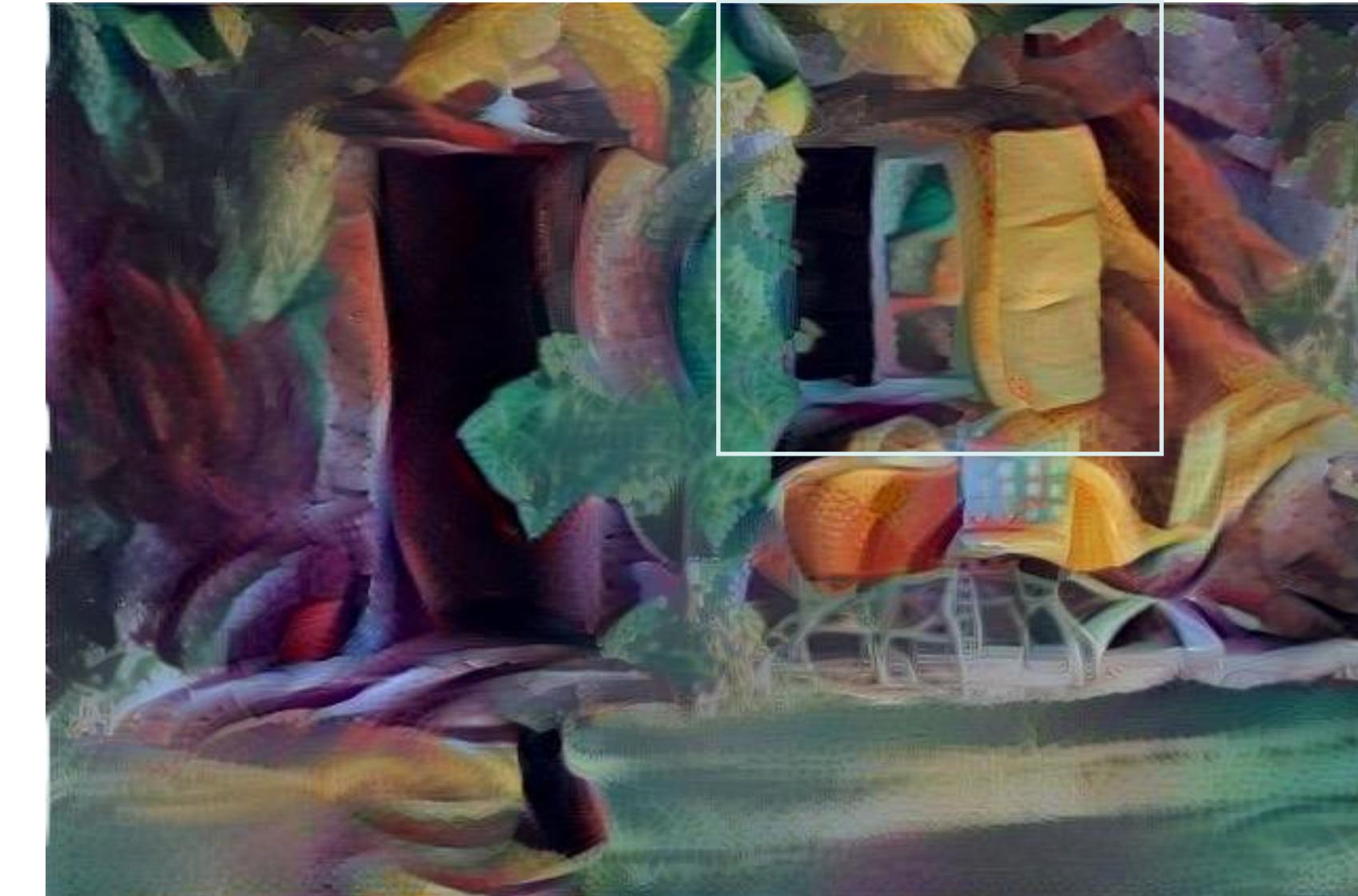


Different Reconstruction Layers

Conv2_2



Conv4_2



Different Reconstruction Layers

Original



Conv2_2



Conv4_2



General Style Transfer

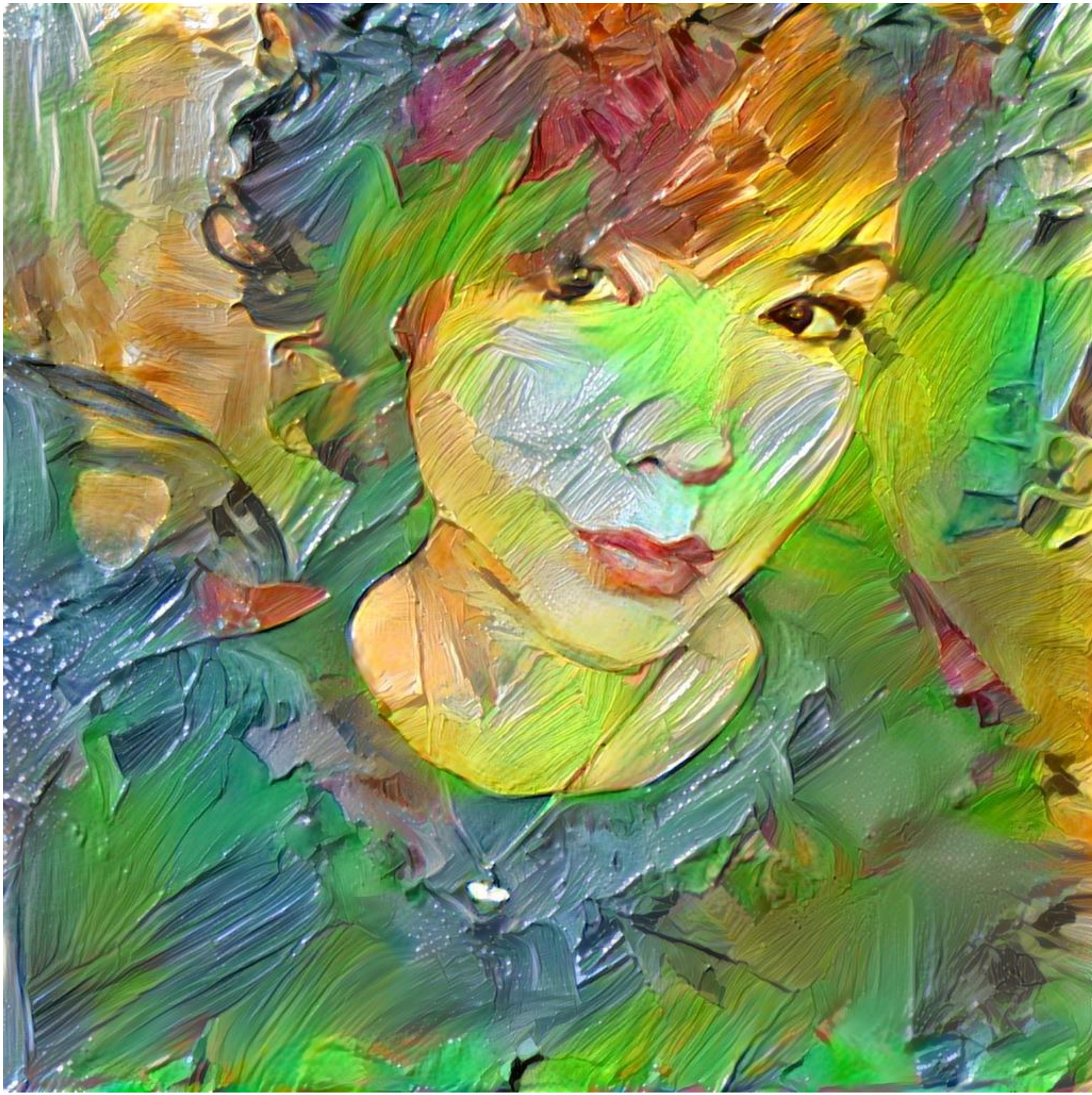


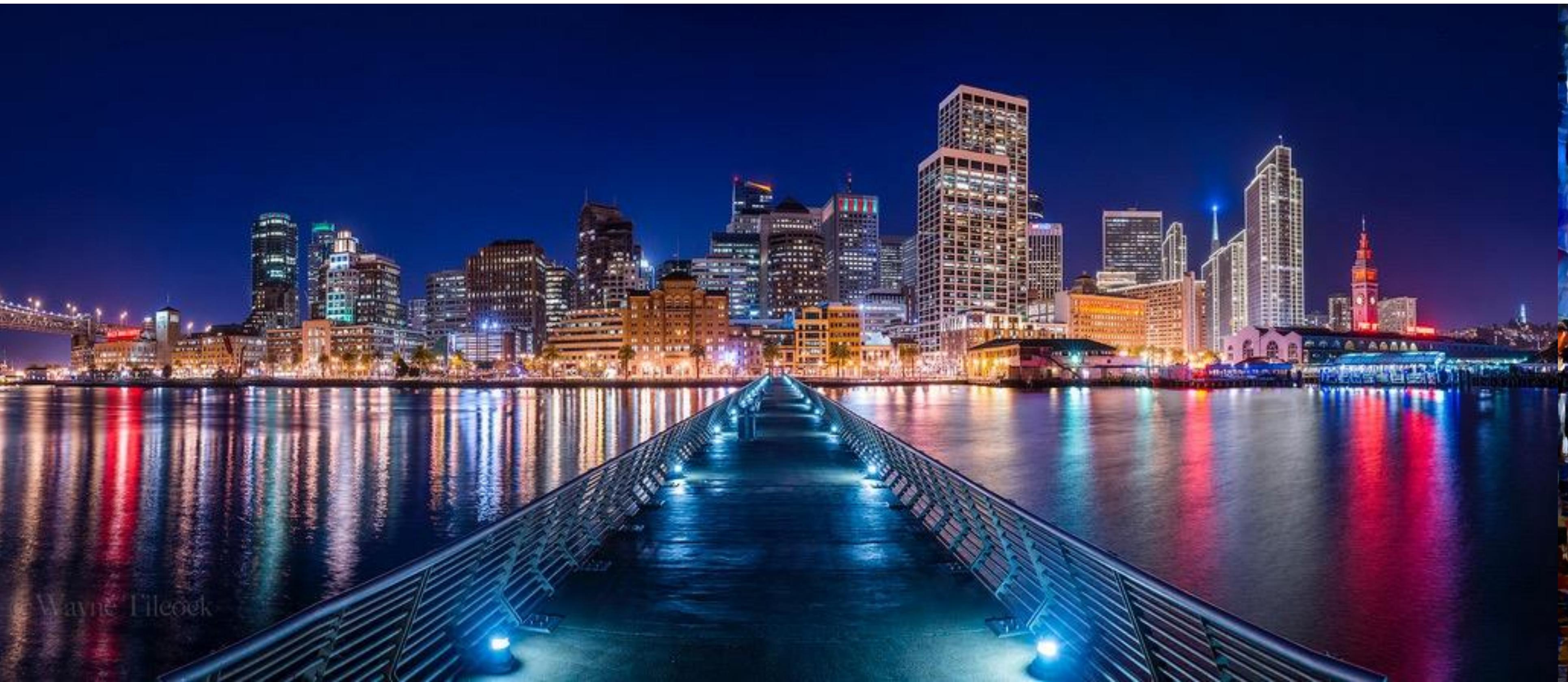
General Style Transfer







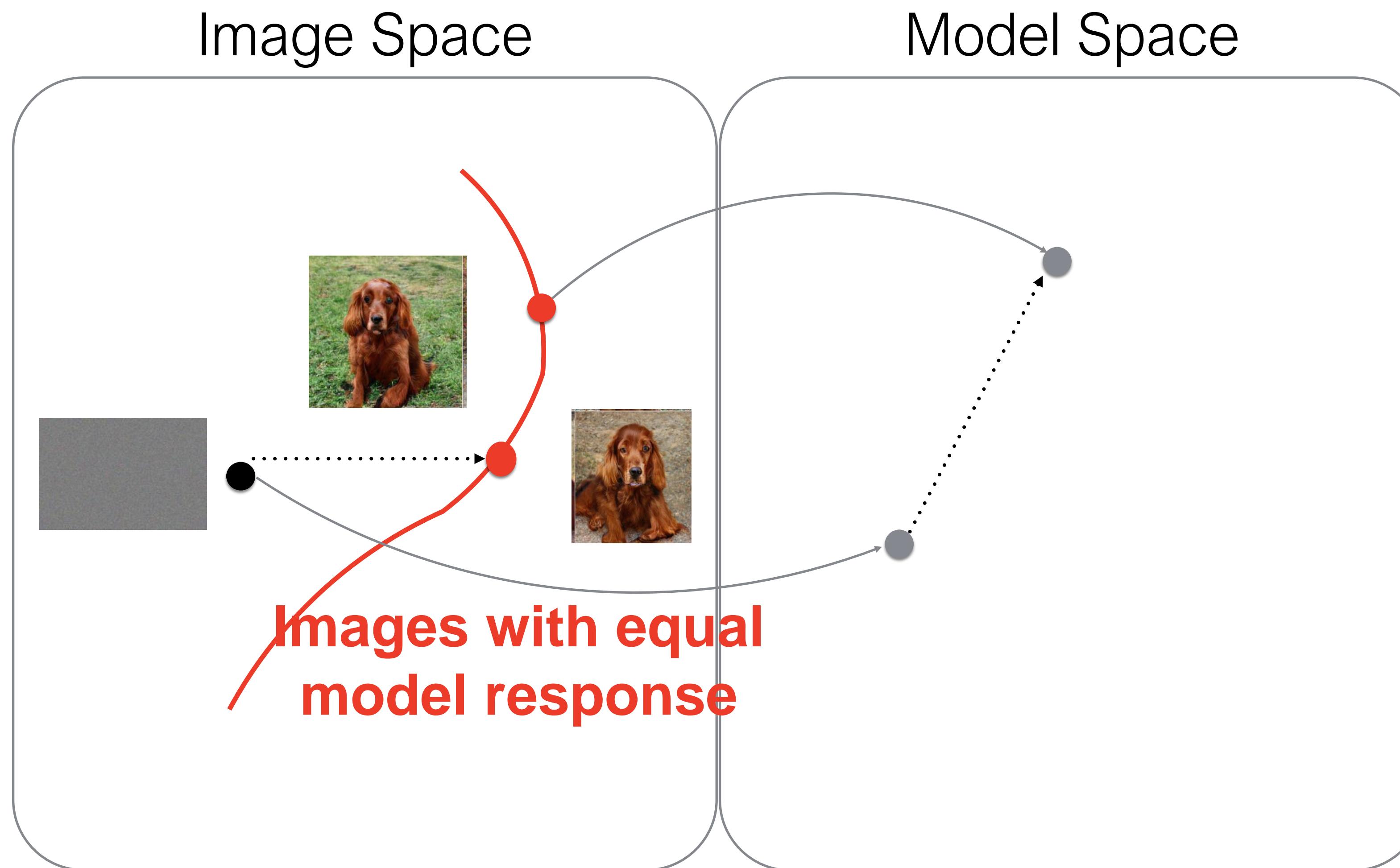




© Wayne Tilcock



“Texture” could be interpreted broadly



Images as Texture

A probabilistic image jigsaw puzzle solver

Taeg Sang Cho[†], Shai Avidan[‡], William T. Freeman[†]

[†] Massachusetts Institute of Technology

[‡] Tel-Aviv University

taegsang@mit.edu, shaiavidan@gmail.com, billf@mit.edu

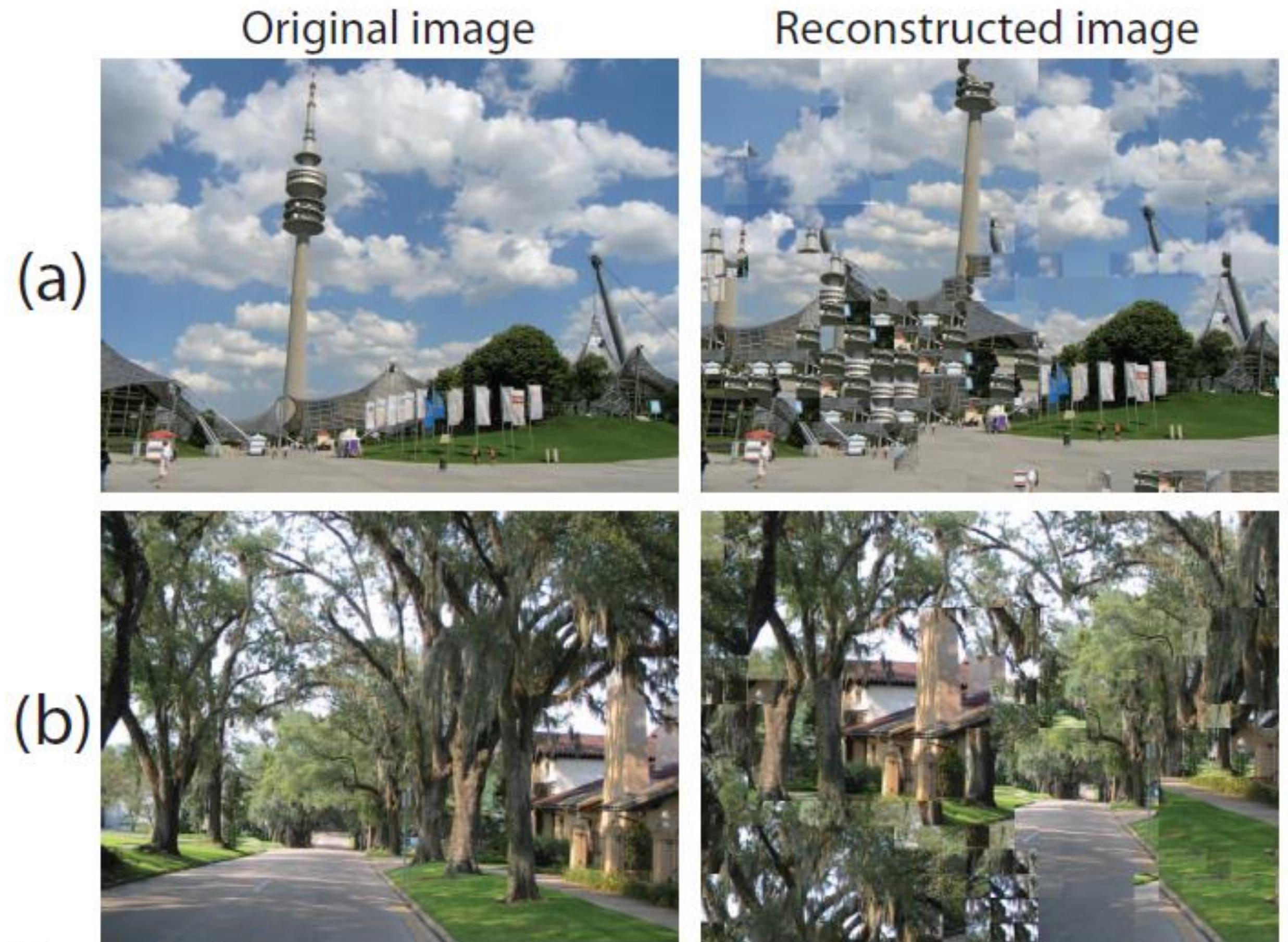
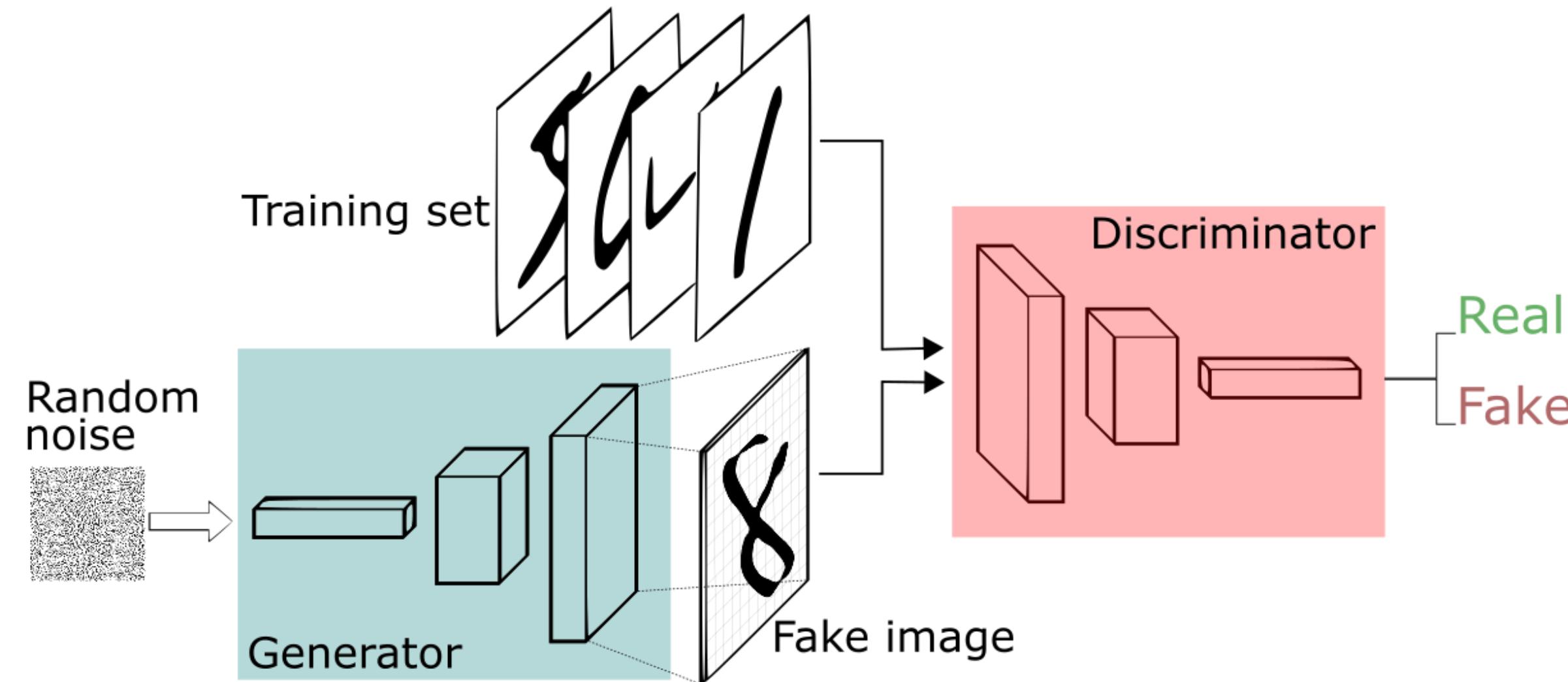


Figure 5. Two examples ((a) image 8, (b) image 20) of reconstructed images using the estimated local evidence.

GANs as Texture Synthesis



- **Conjecture:** GANs might be learning the “right” features to match for natural images





Imagen

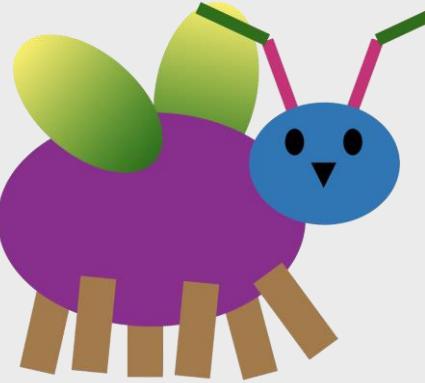
Imagen



DALL-E 2

diffusion

By Steve Seitz



**Language
Generator**

(+ pixels)

Parti

**Image
Generator**

(+ language)

Imagen

Generate 100 images



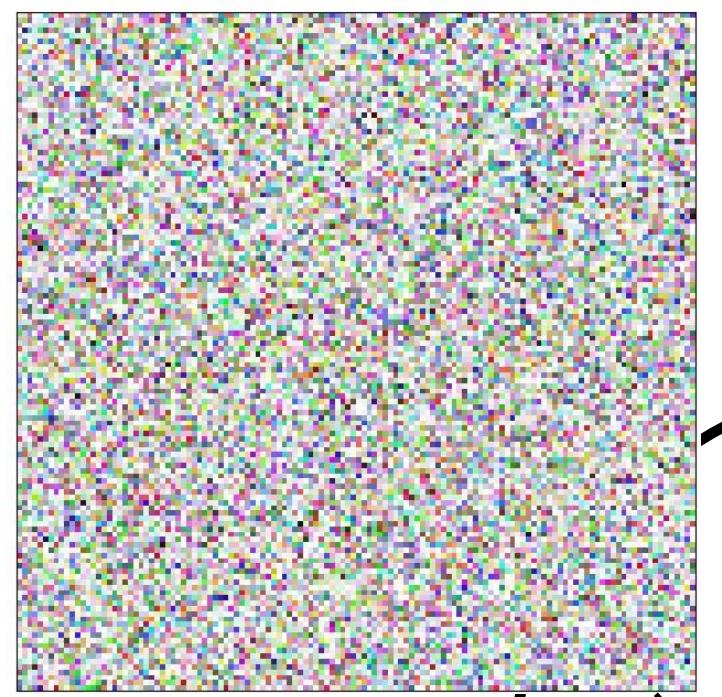
Generate 100 images

slide from Steve Seitz's [video](#)

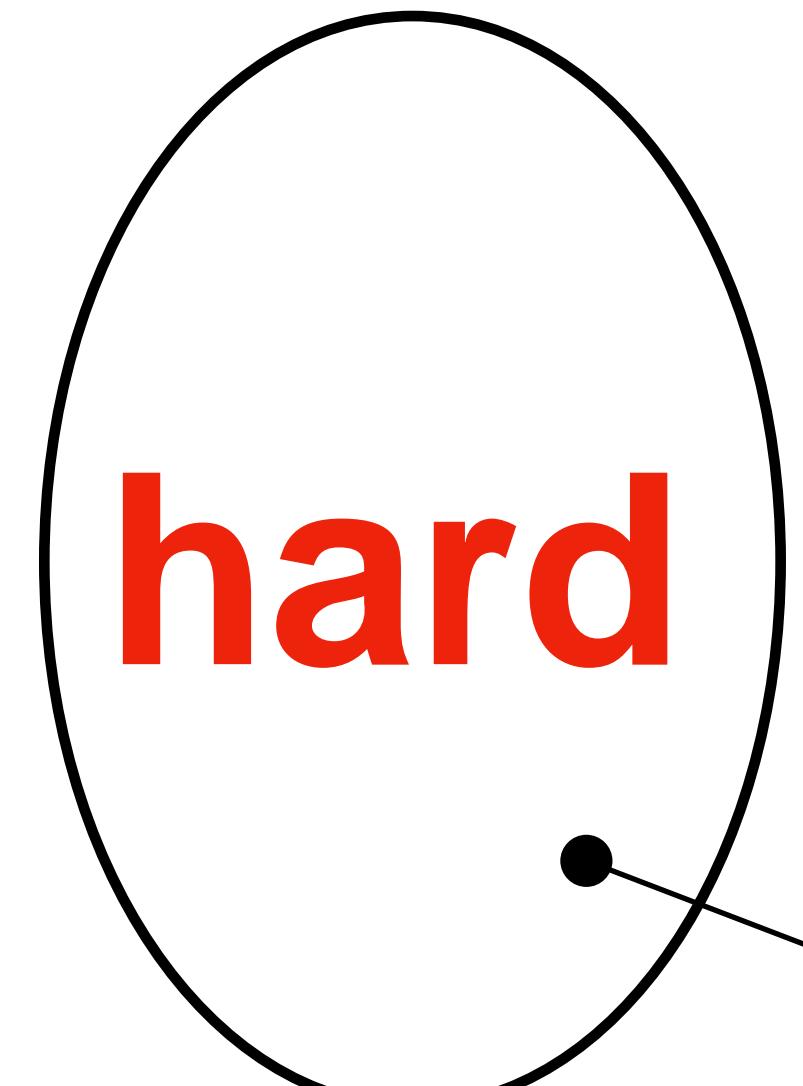


slide from Steve Seitz's [video](#)

Generate 100 images of raspberries

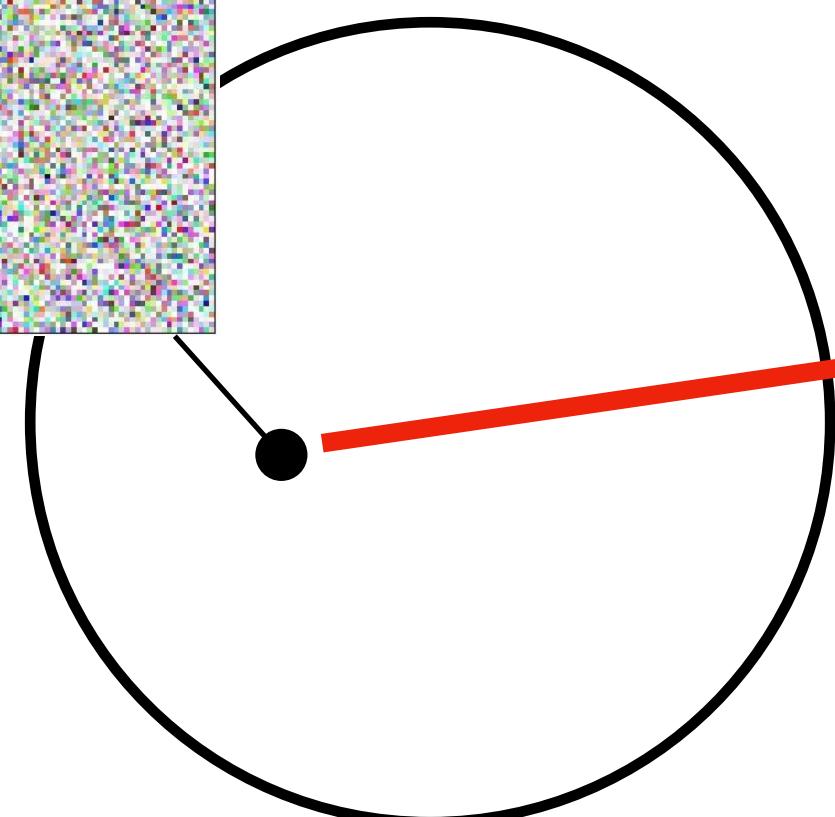
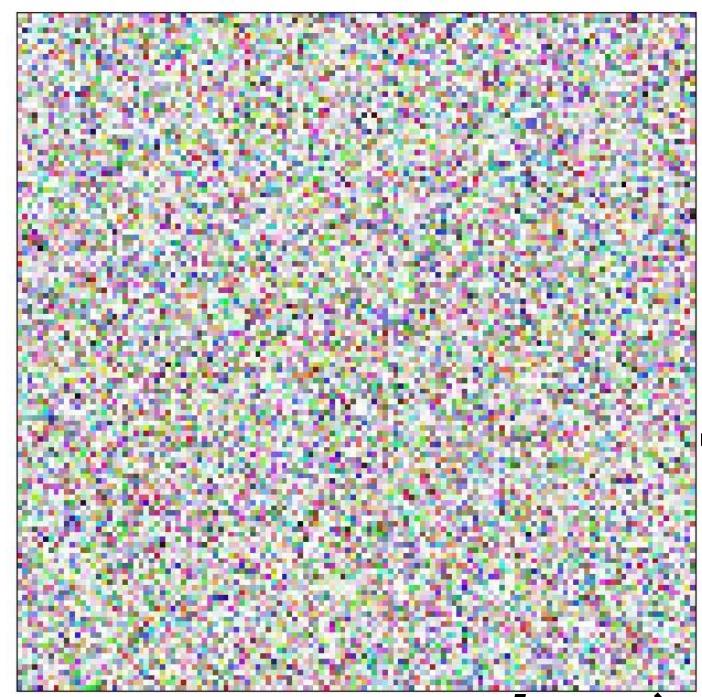


random
images



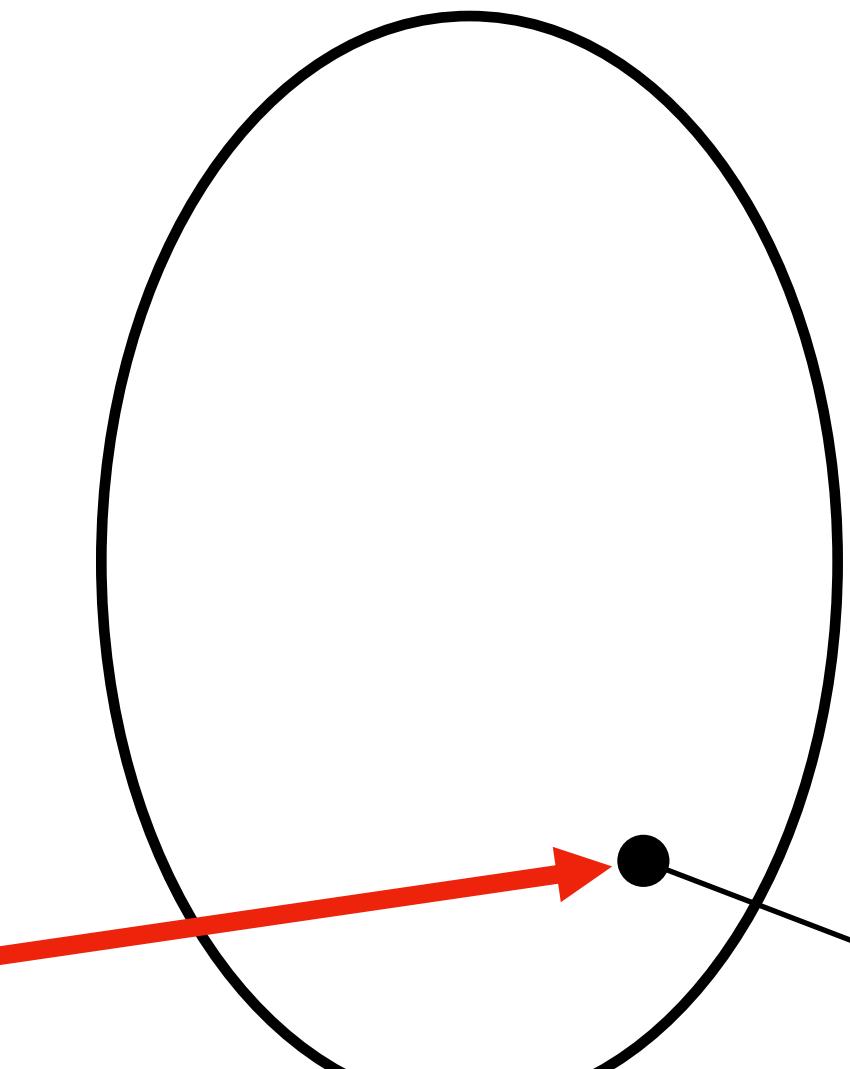
raspberry
images



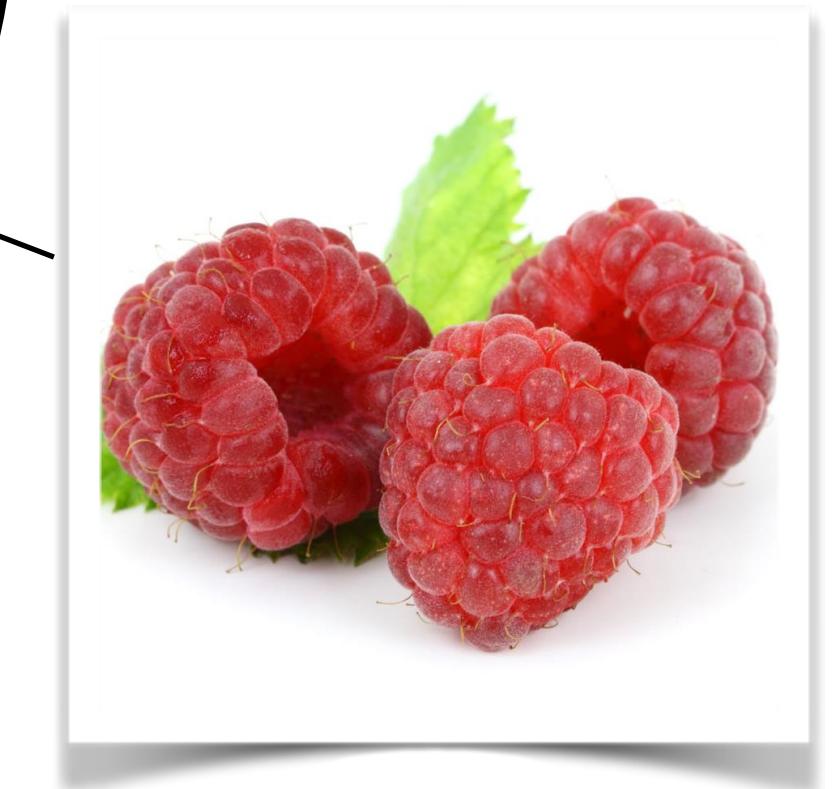


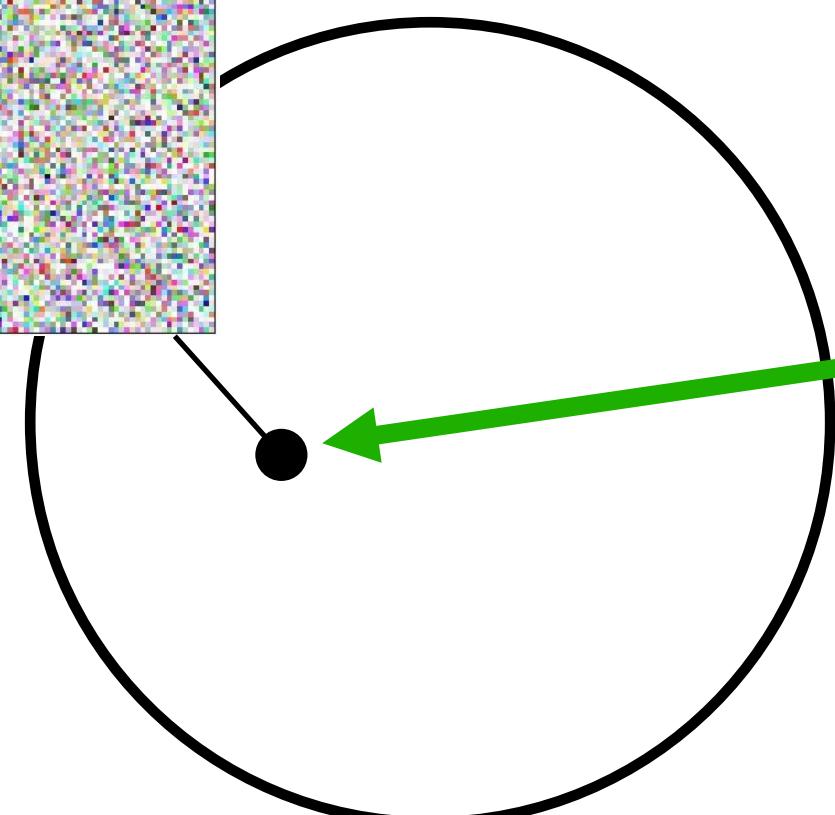
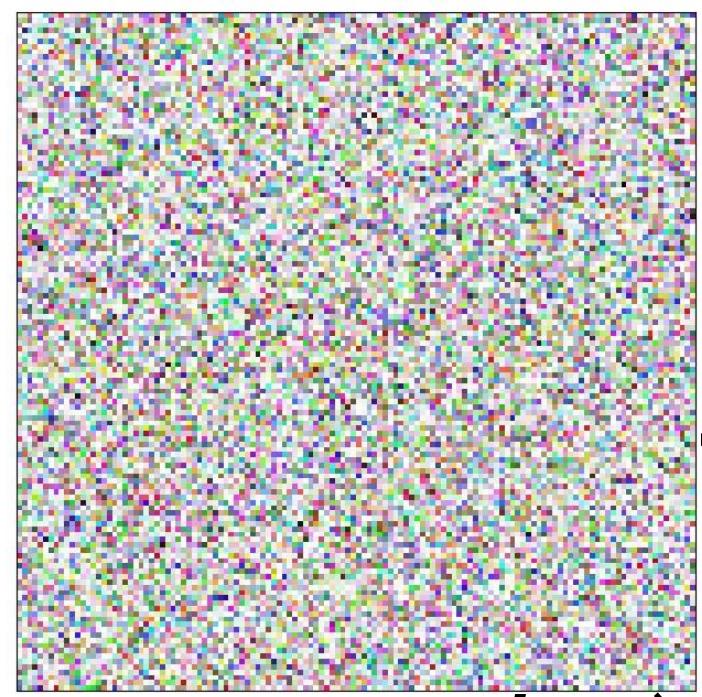
random
images

hard



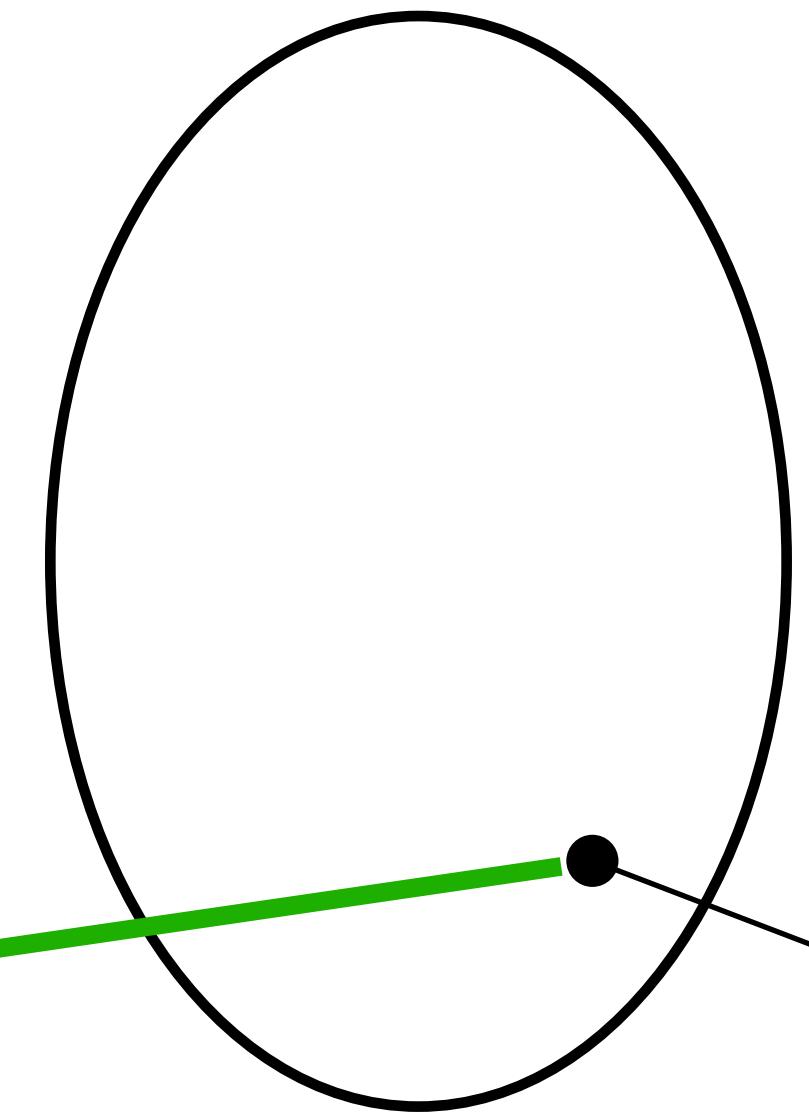
raspberry
images



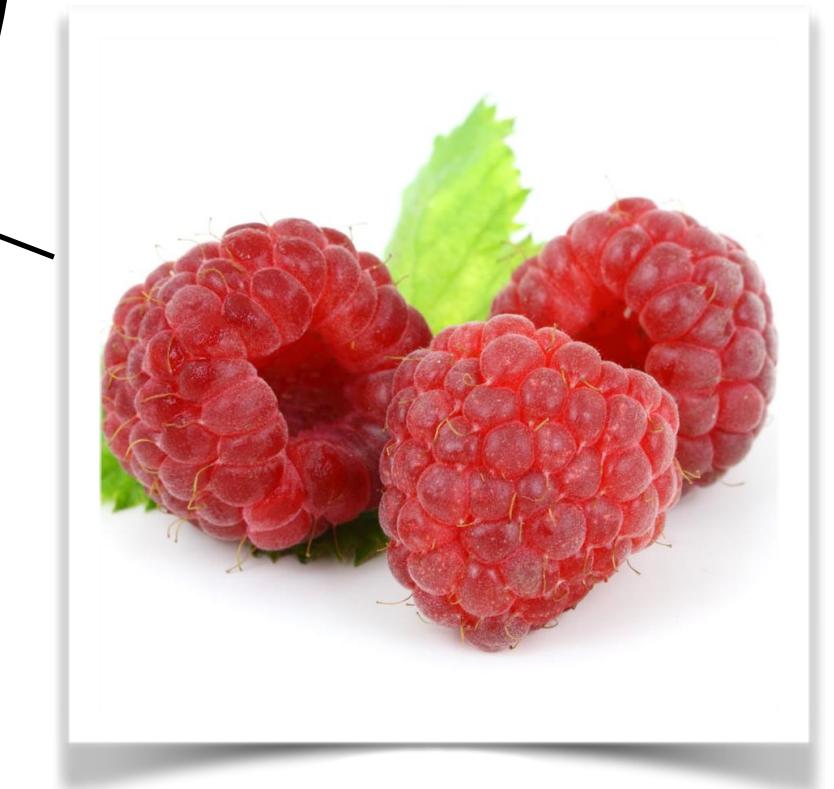


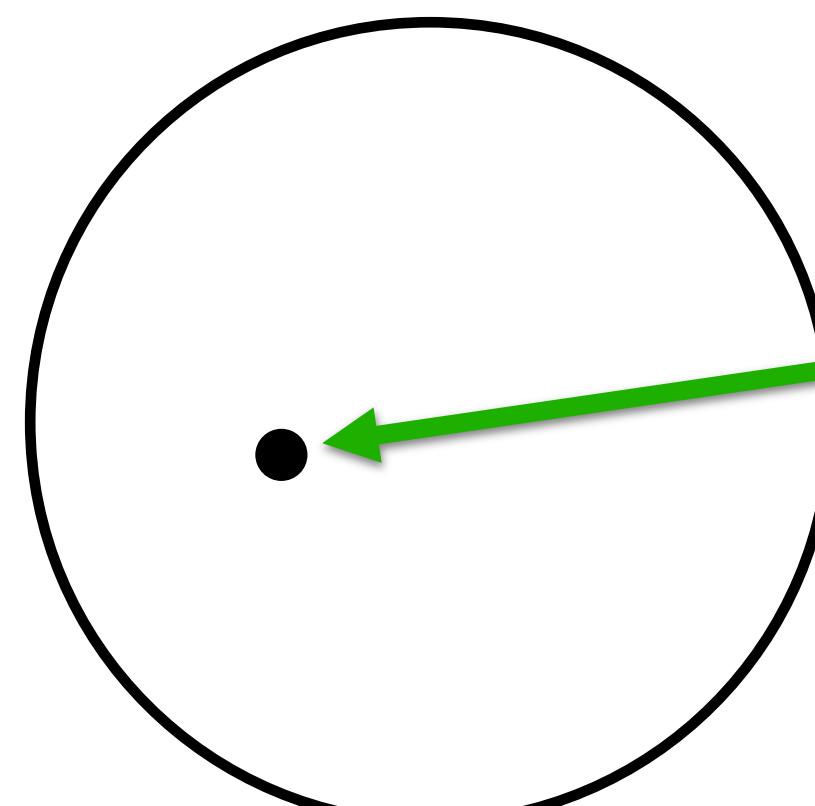
random
images

easy

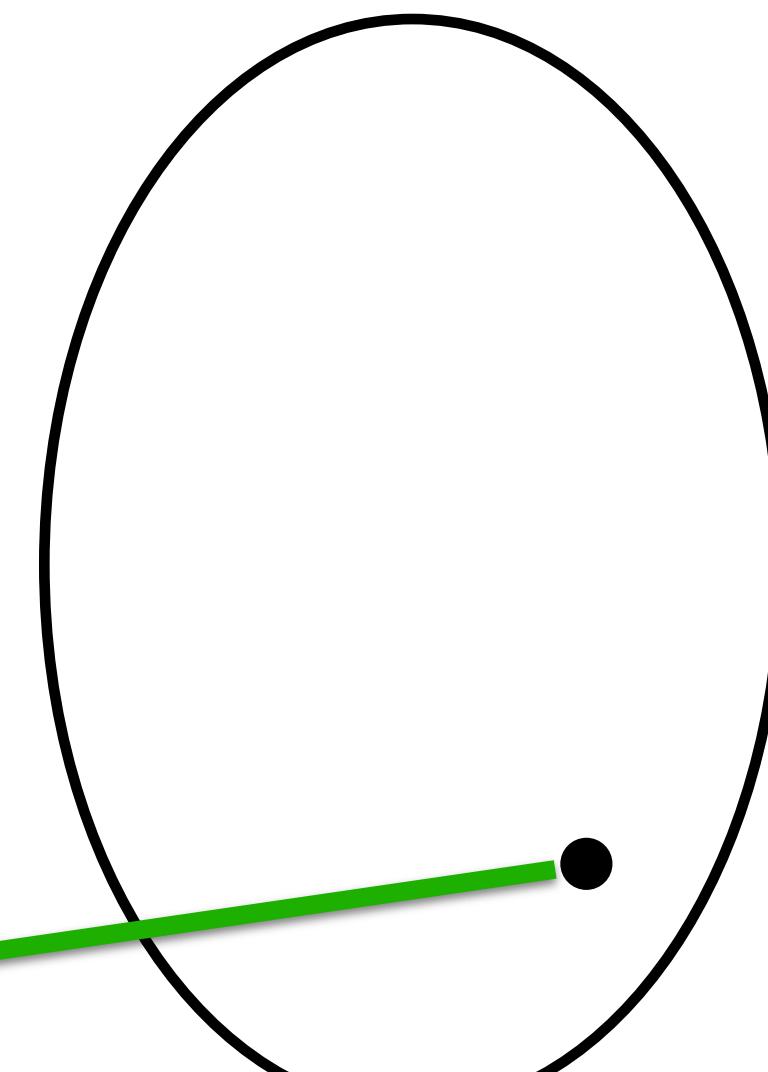
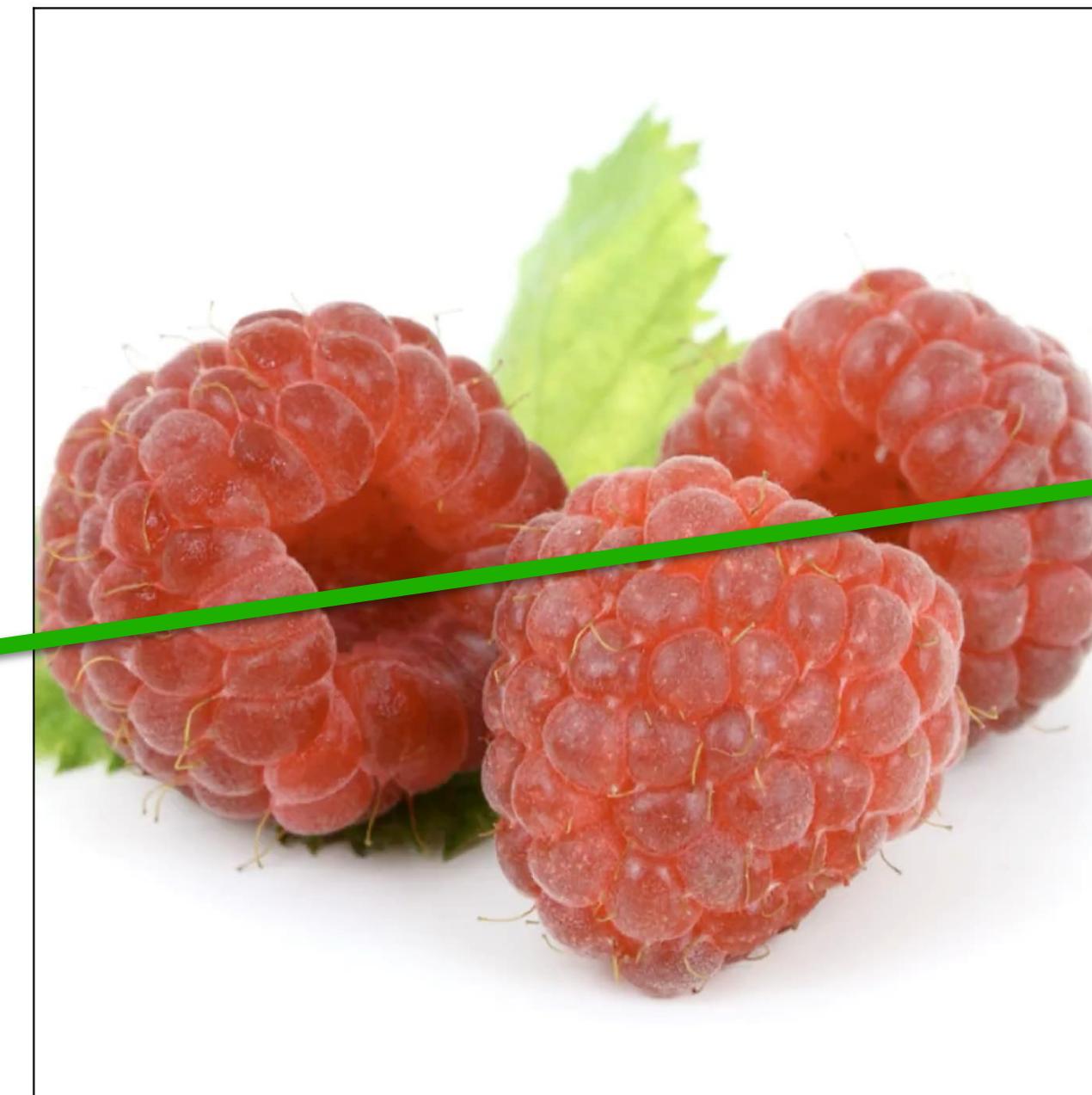


raspberry
images

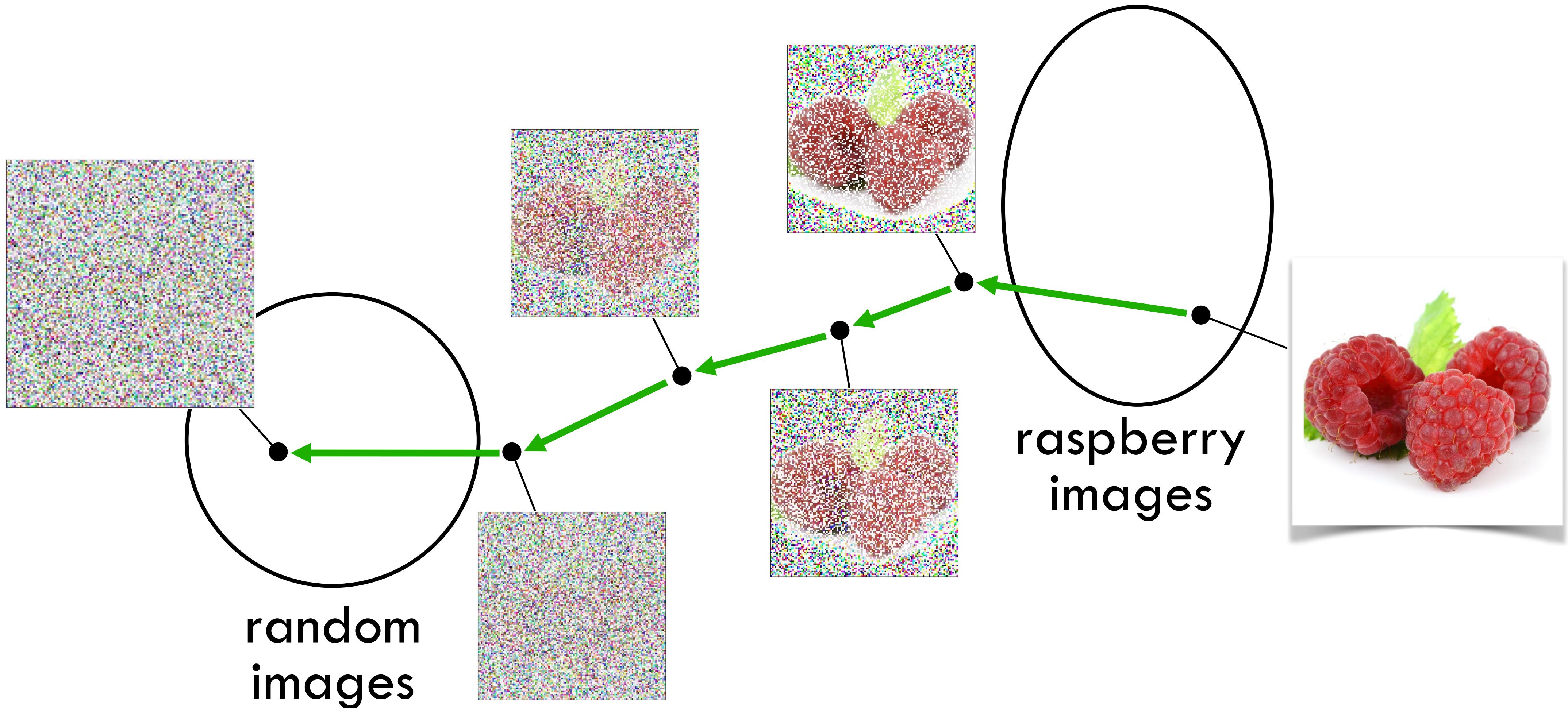


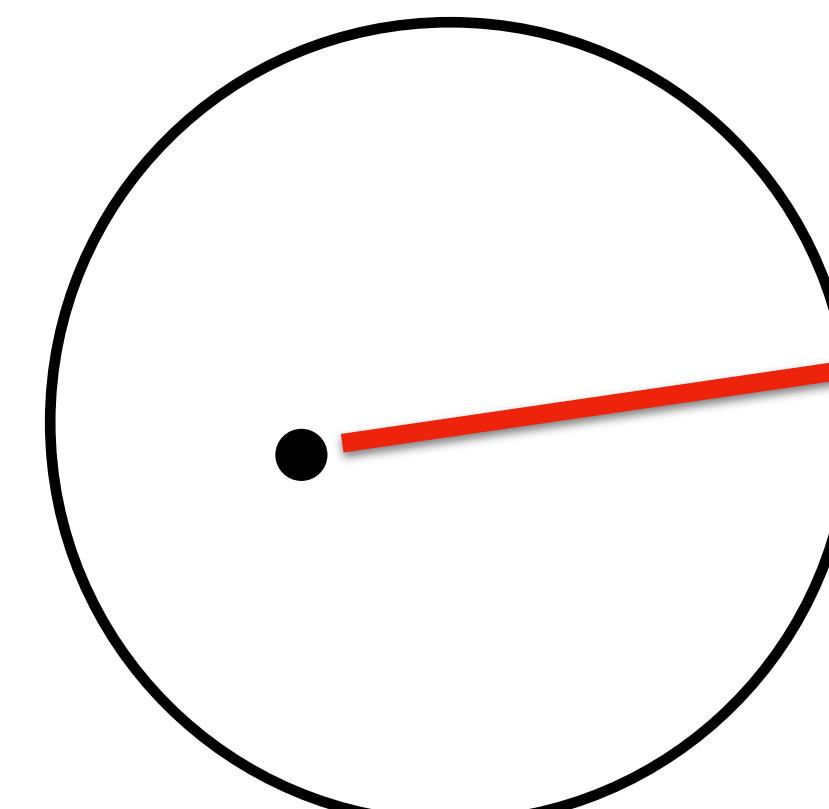


random
images

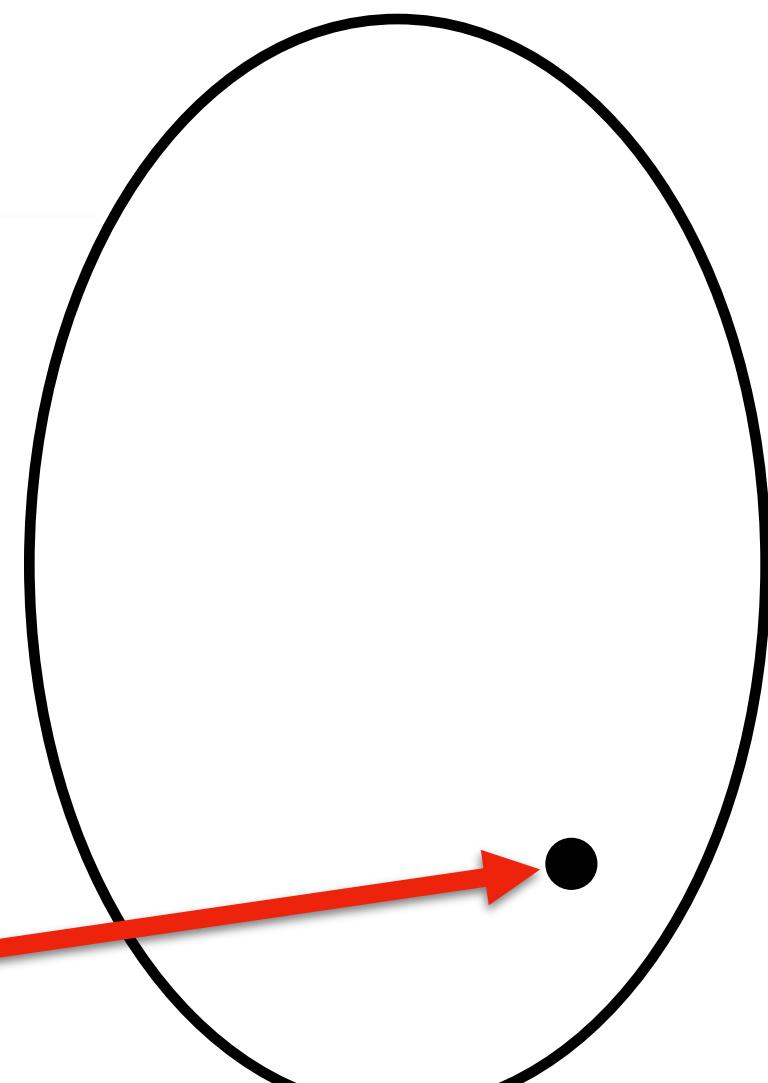
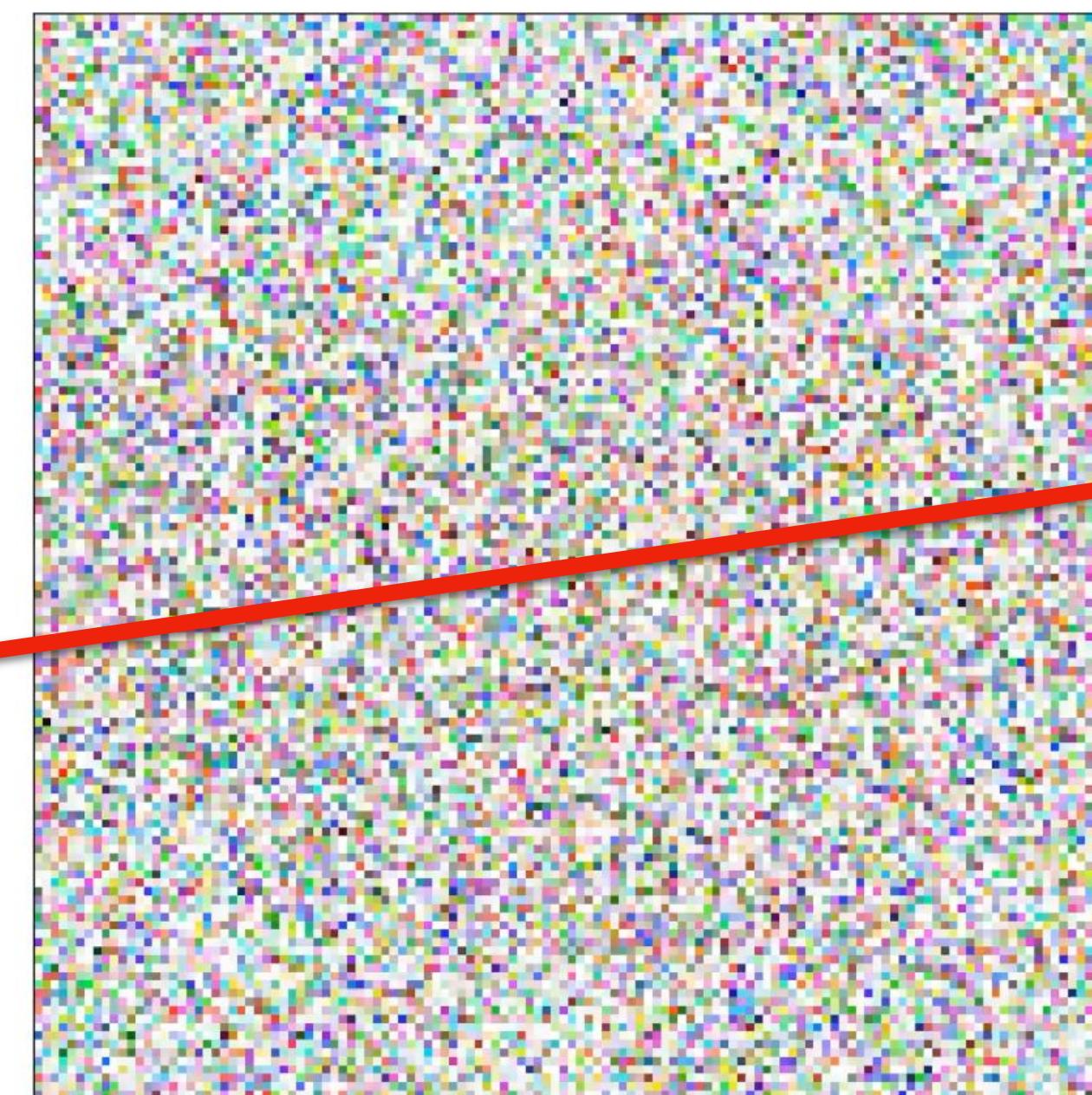


raspberry
images

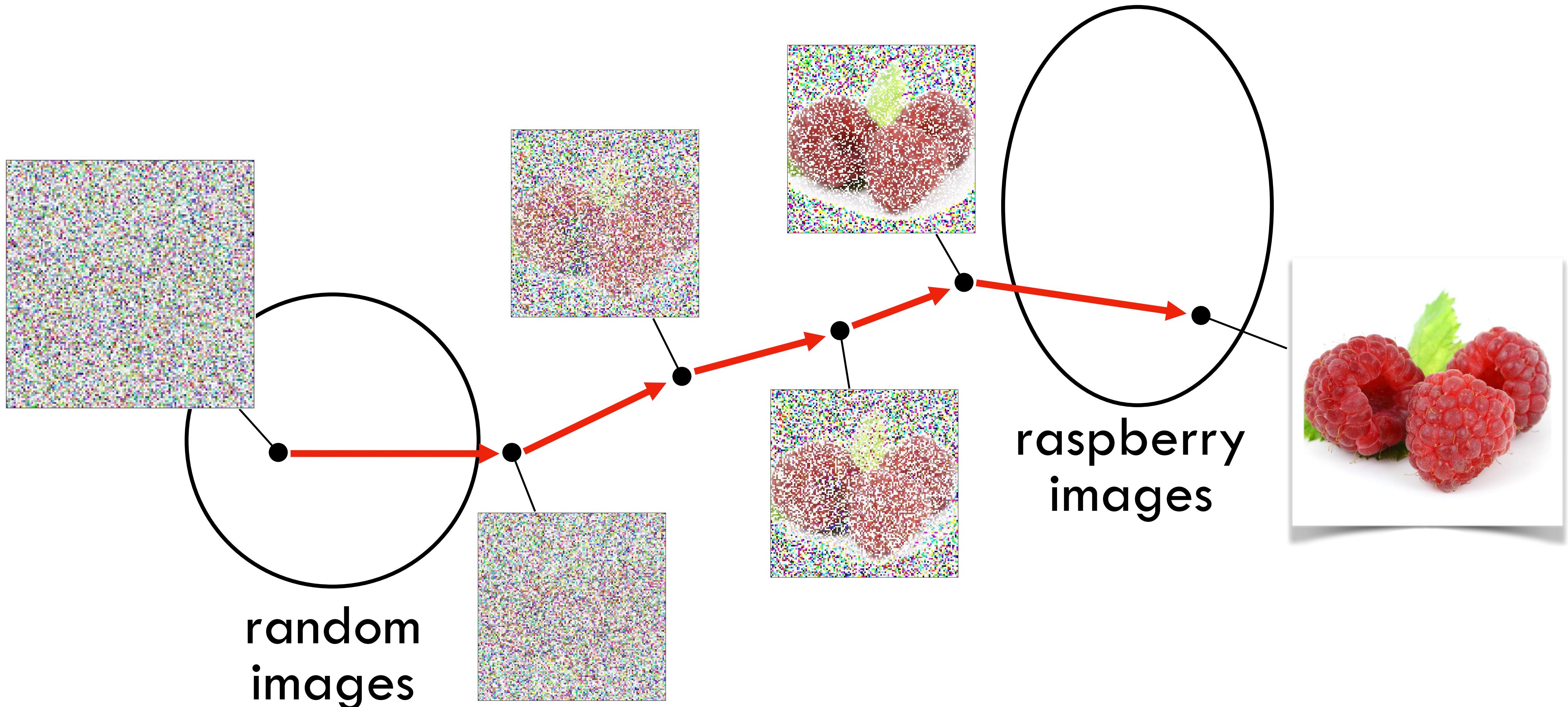


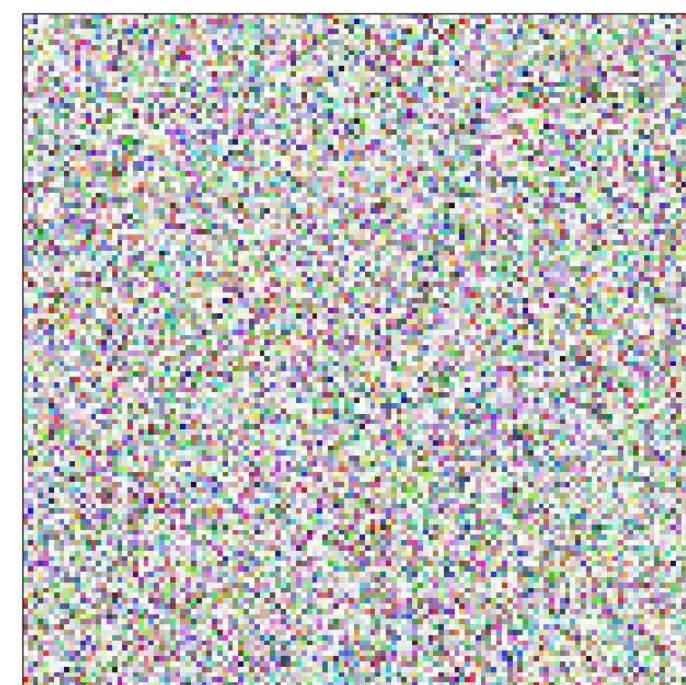


random
images

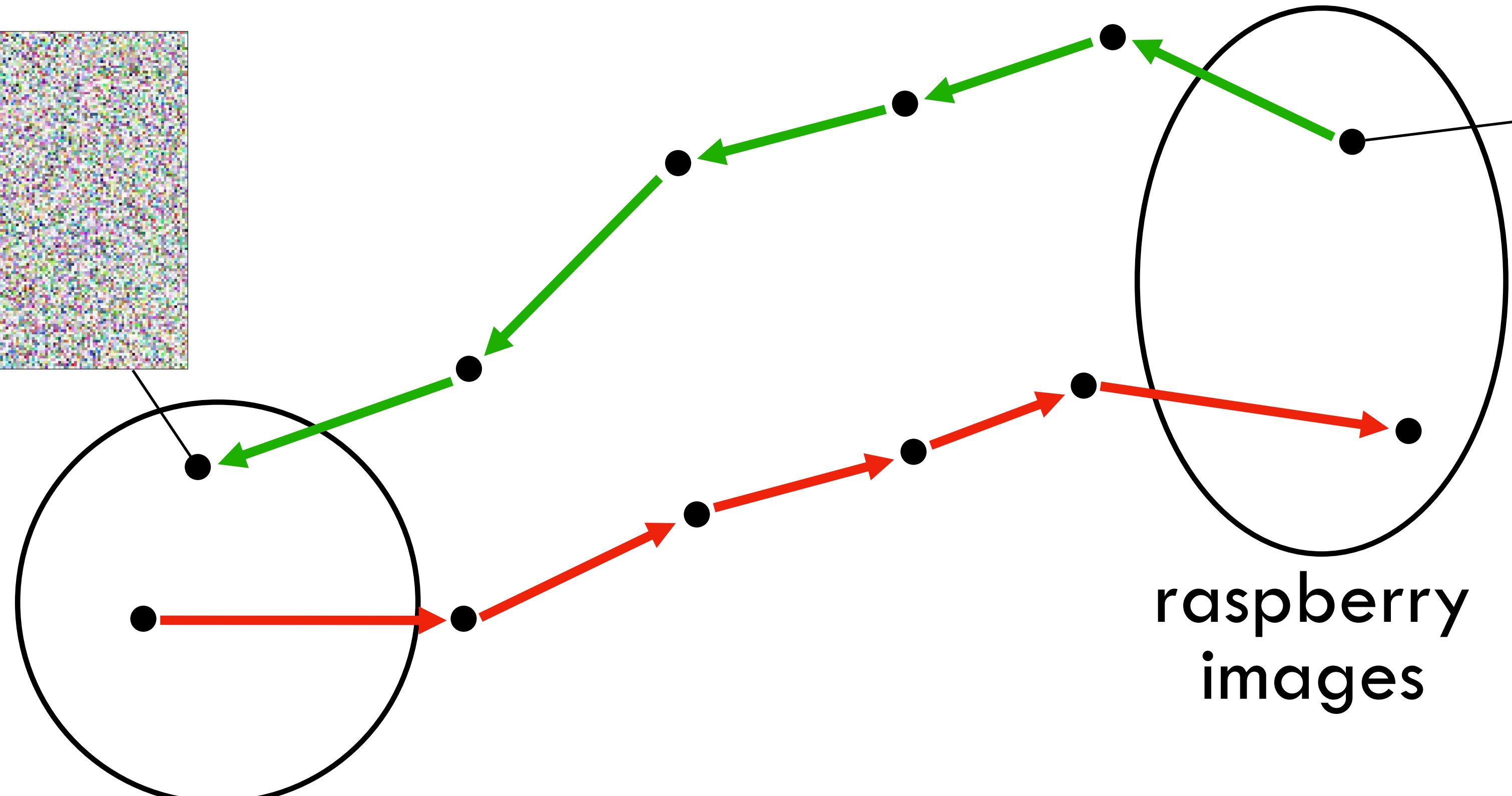


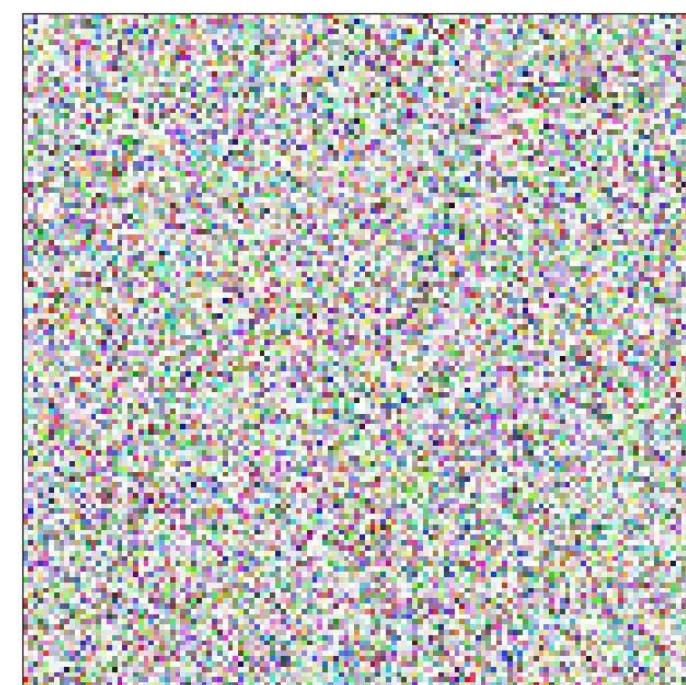
raspberry
images



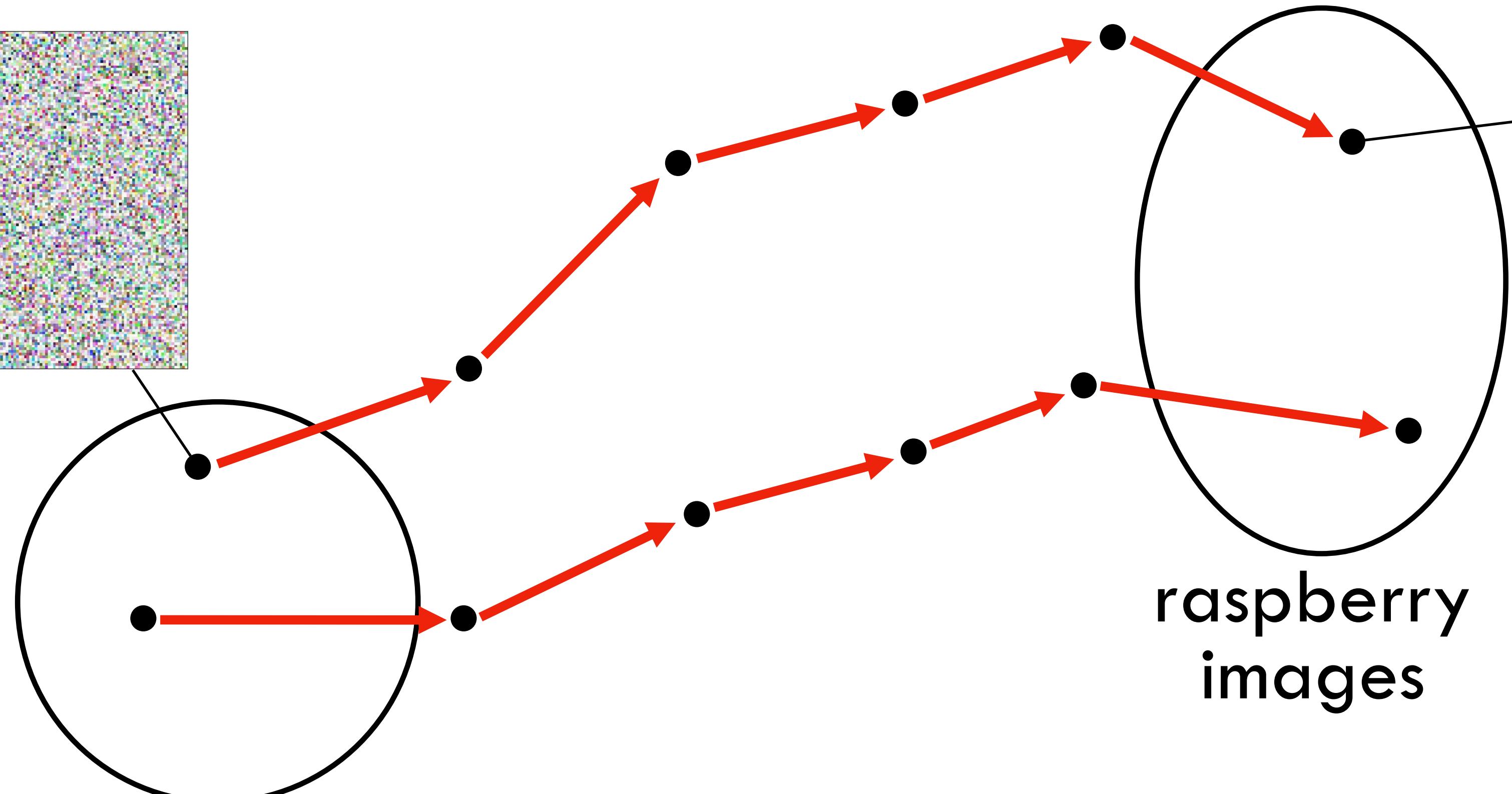


random
images

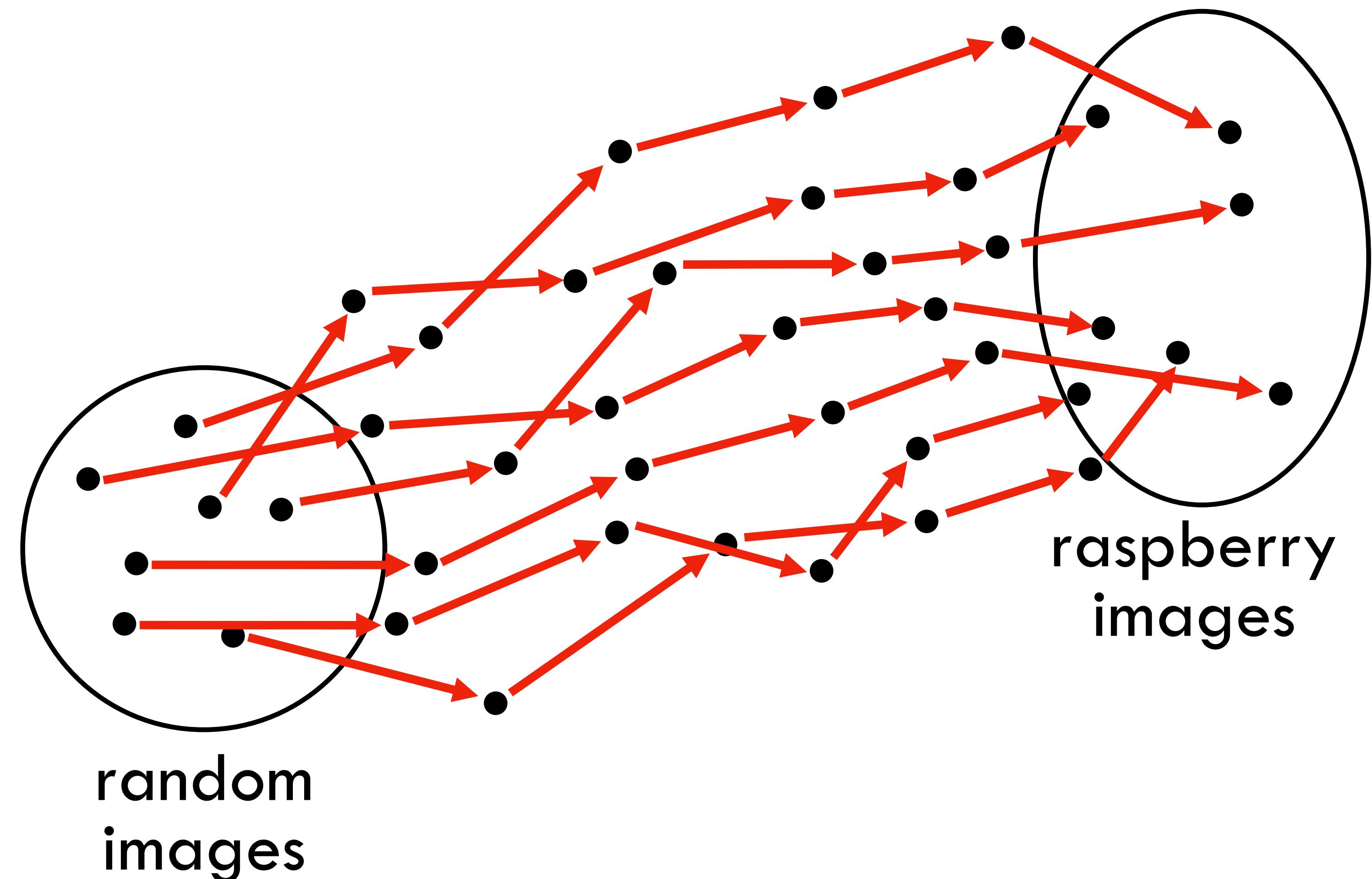


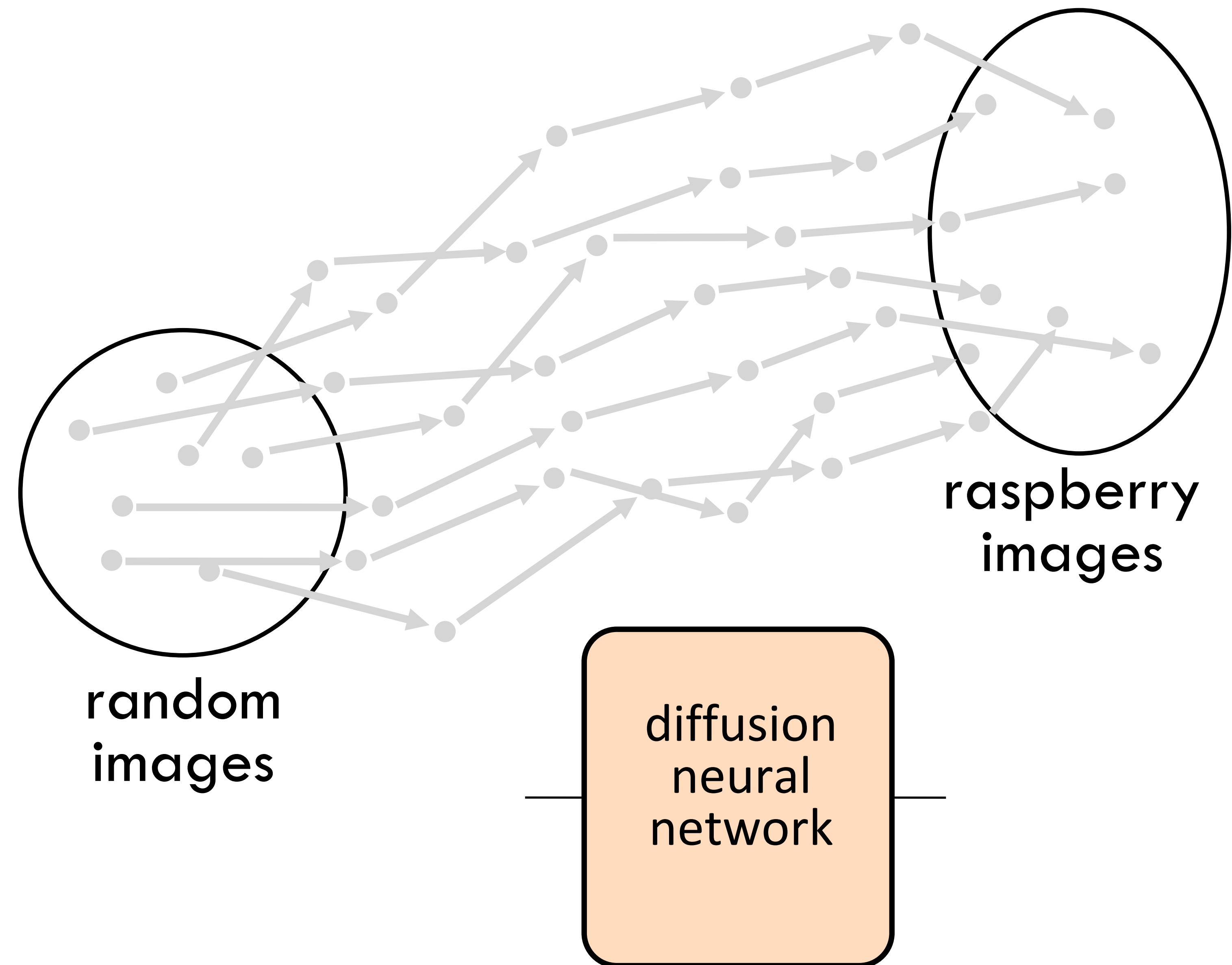


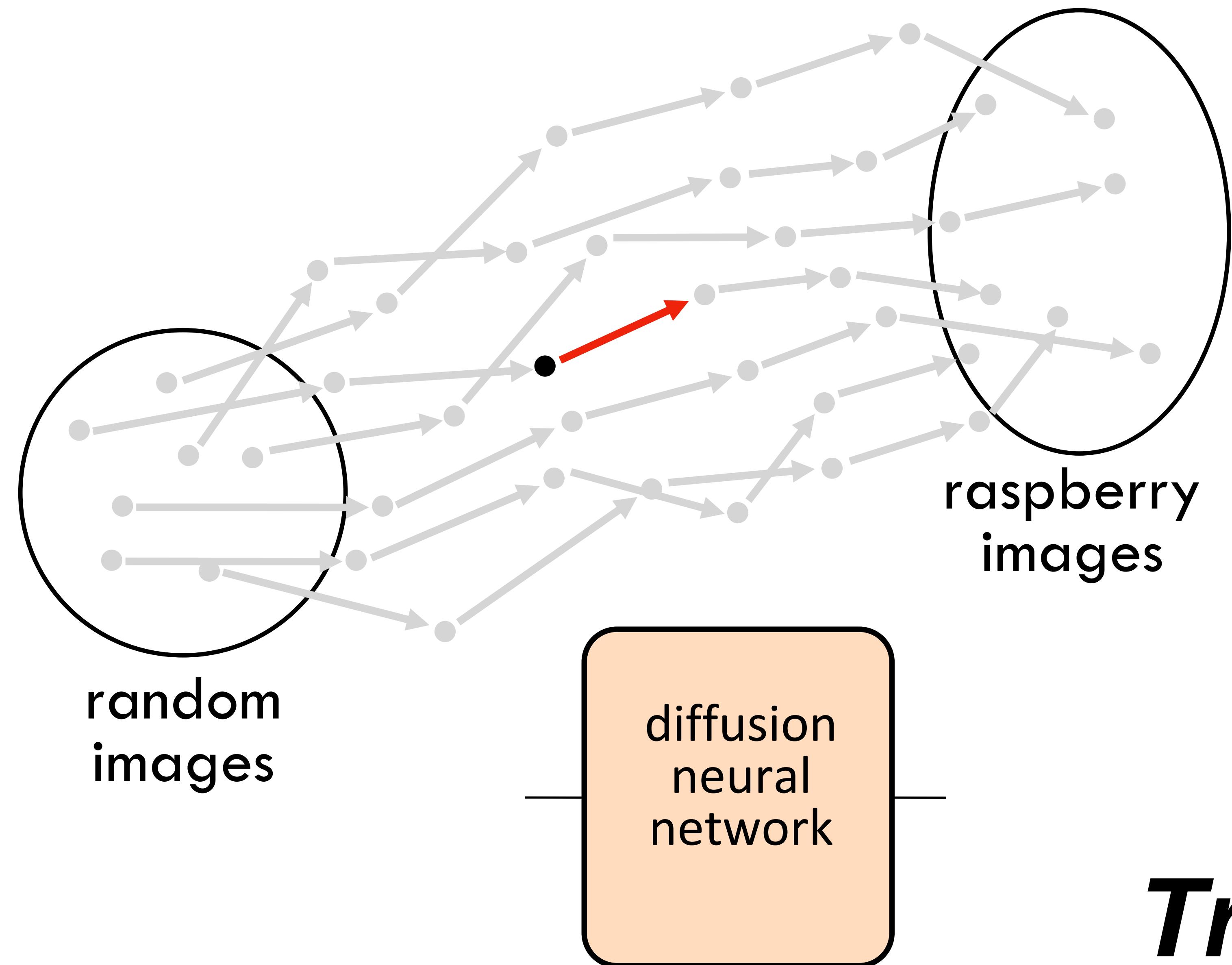
random
images



raspberry
images

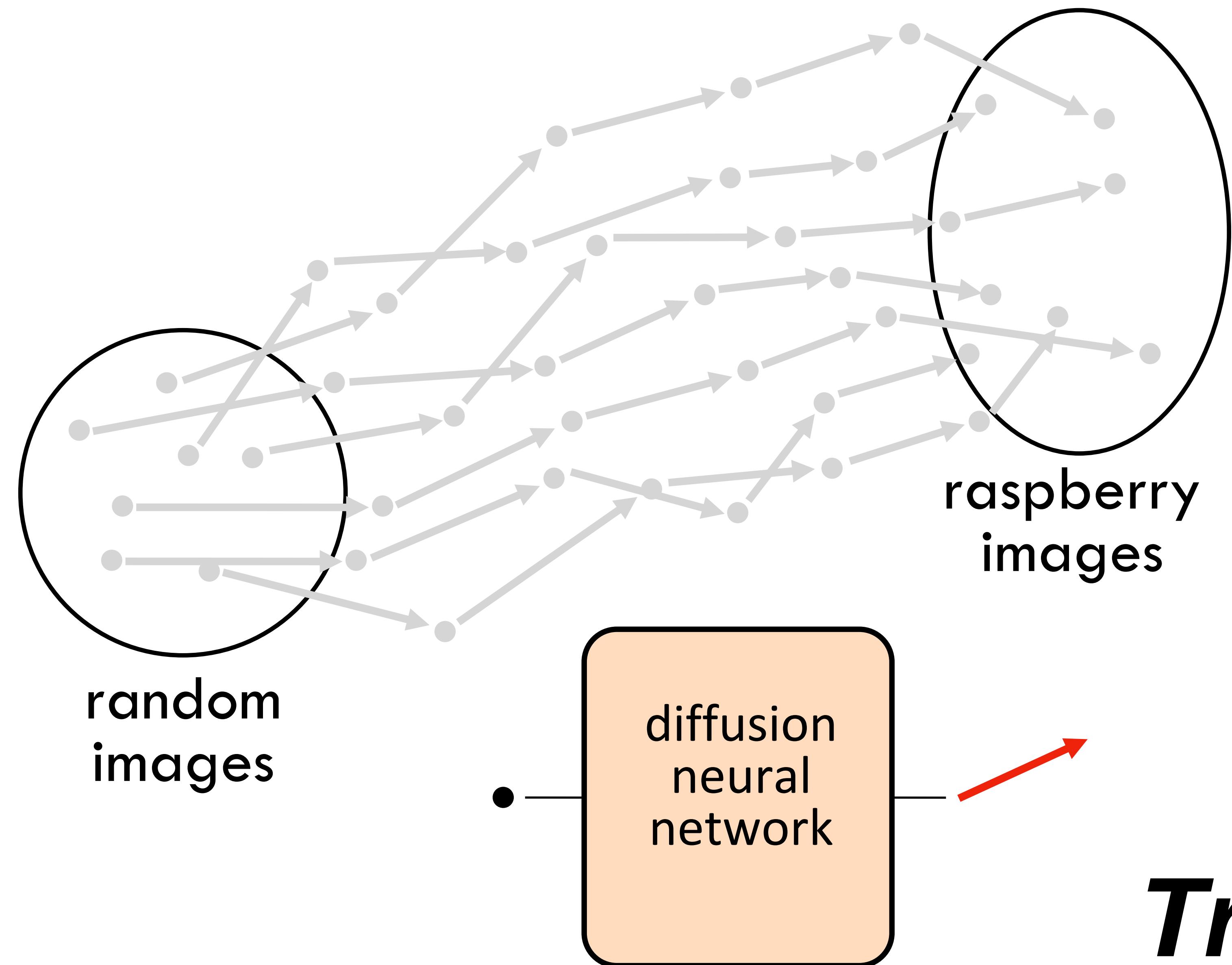


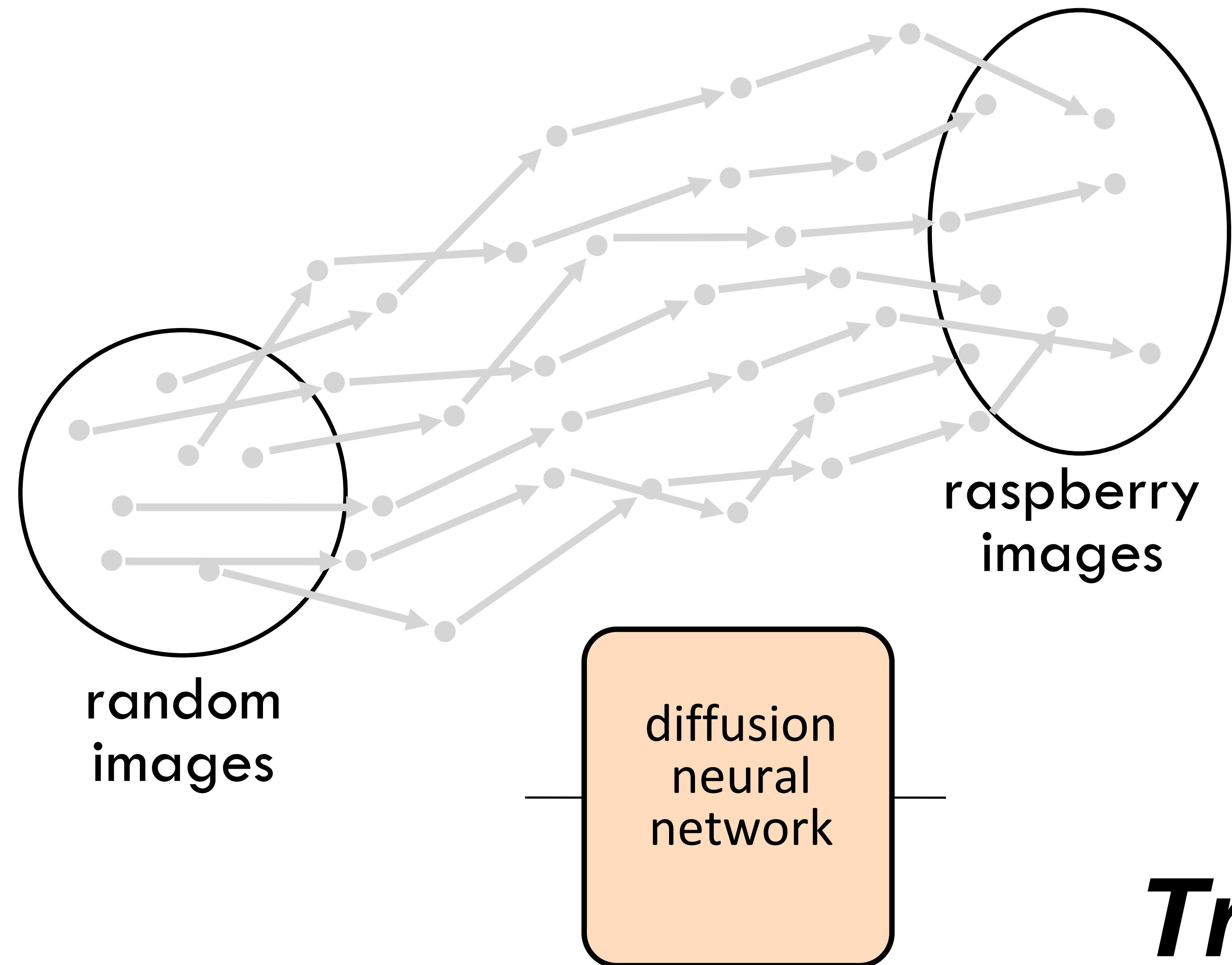




Training

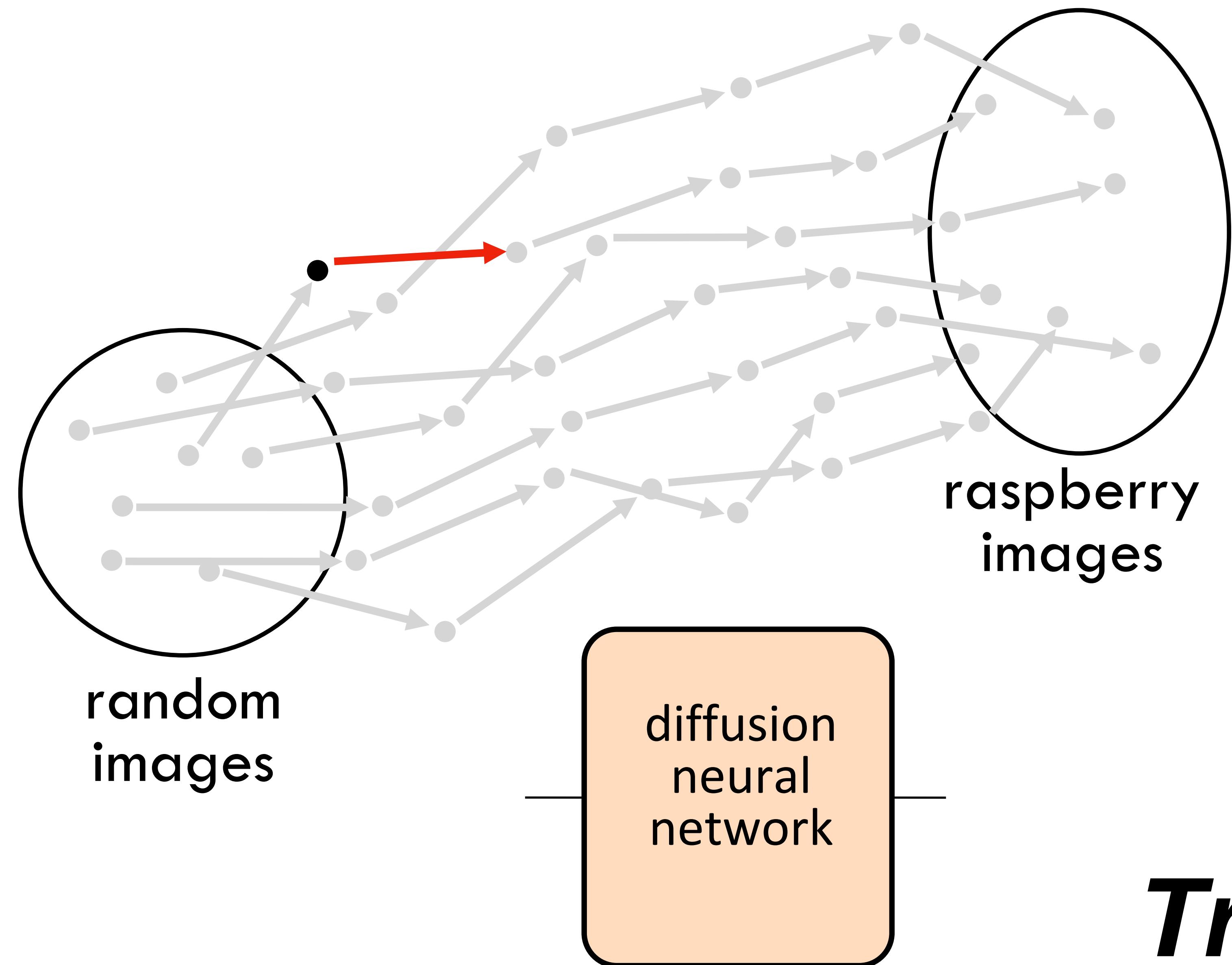
slide from Steve Seitz's [video](#)





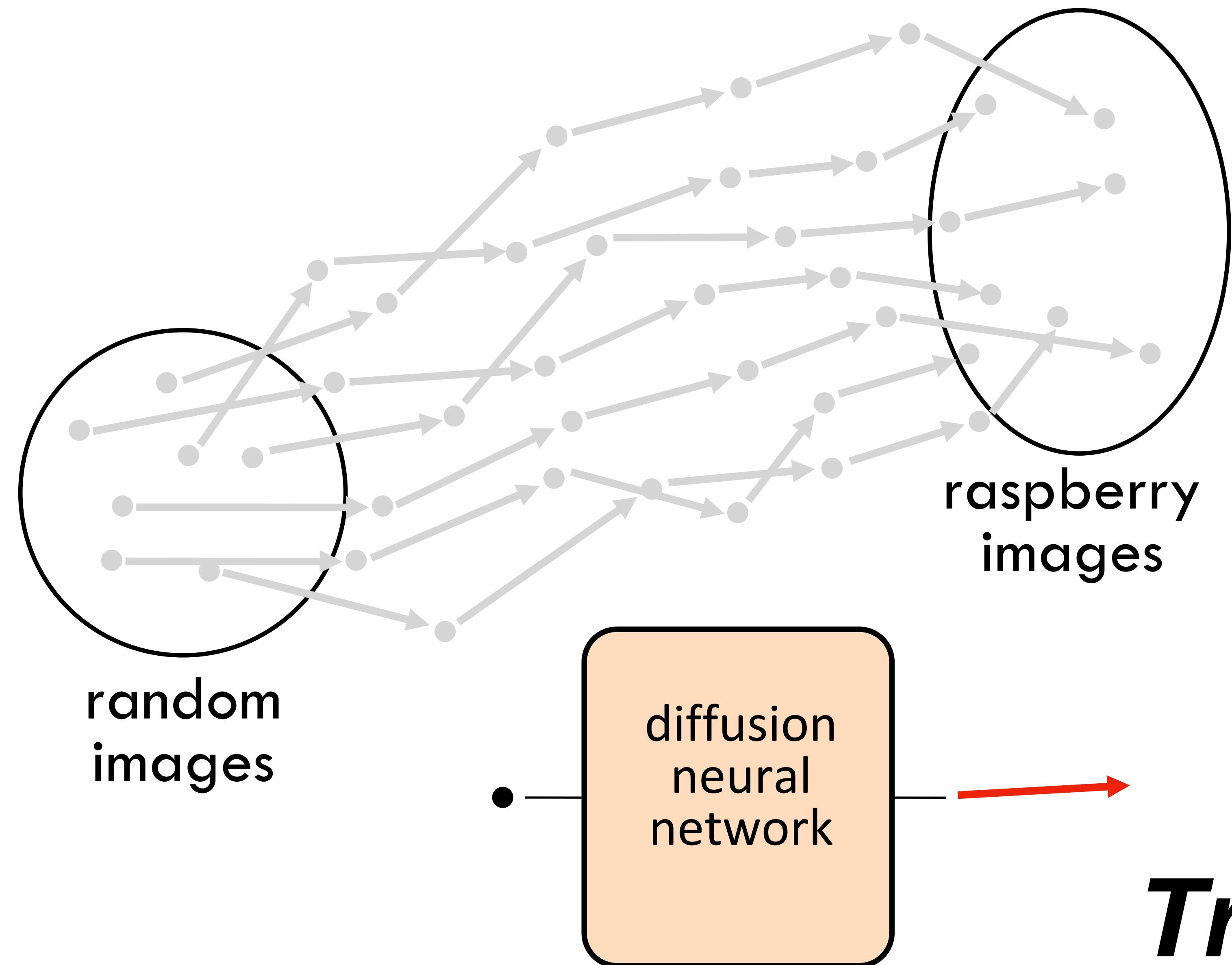
Training

slide from Steve Seitz's [video](#)



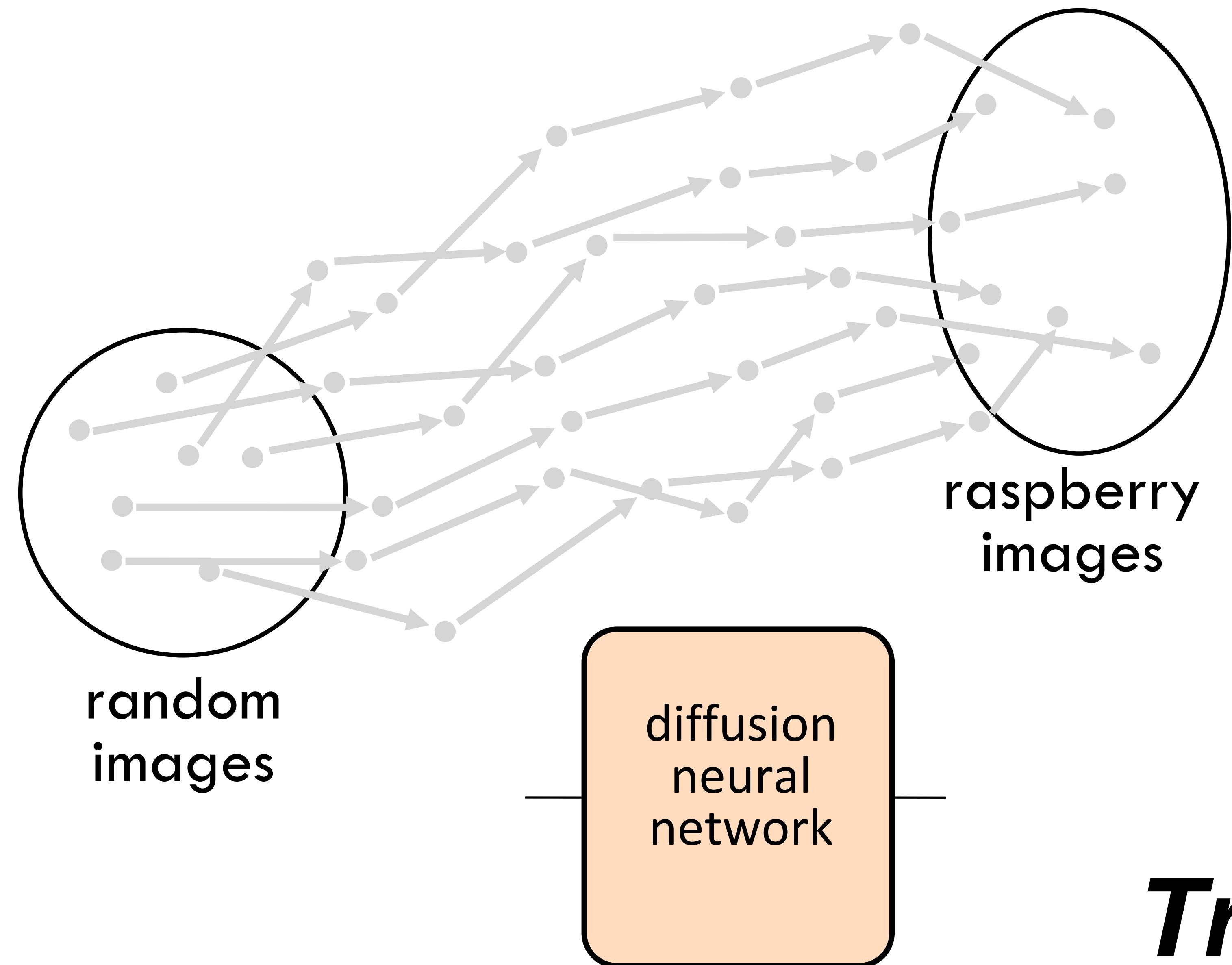
Training

slide from Steve Seitz's [video](#)



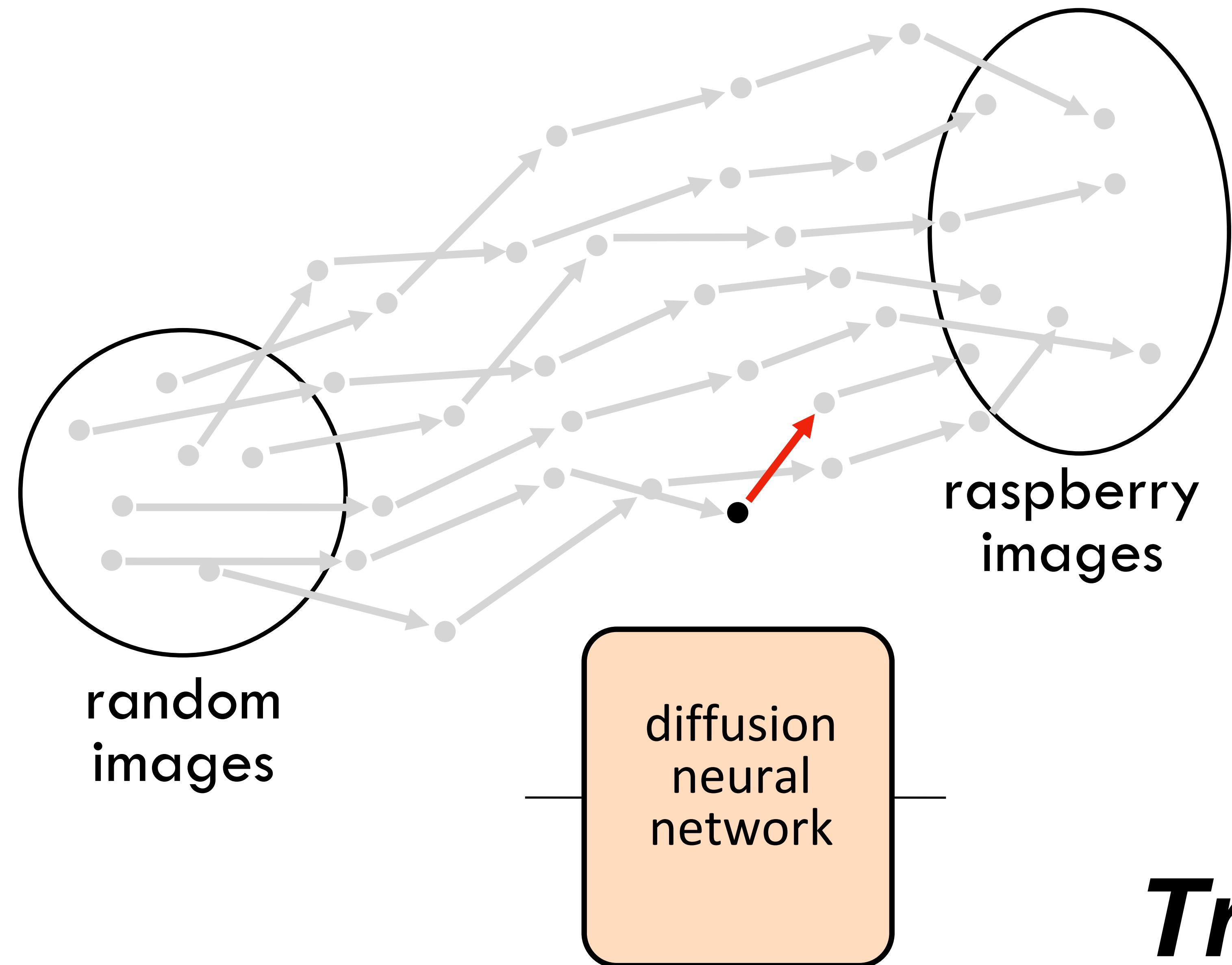
Training

slide from Steve Seitz's [video](#)



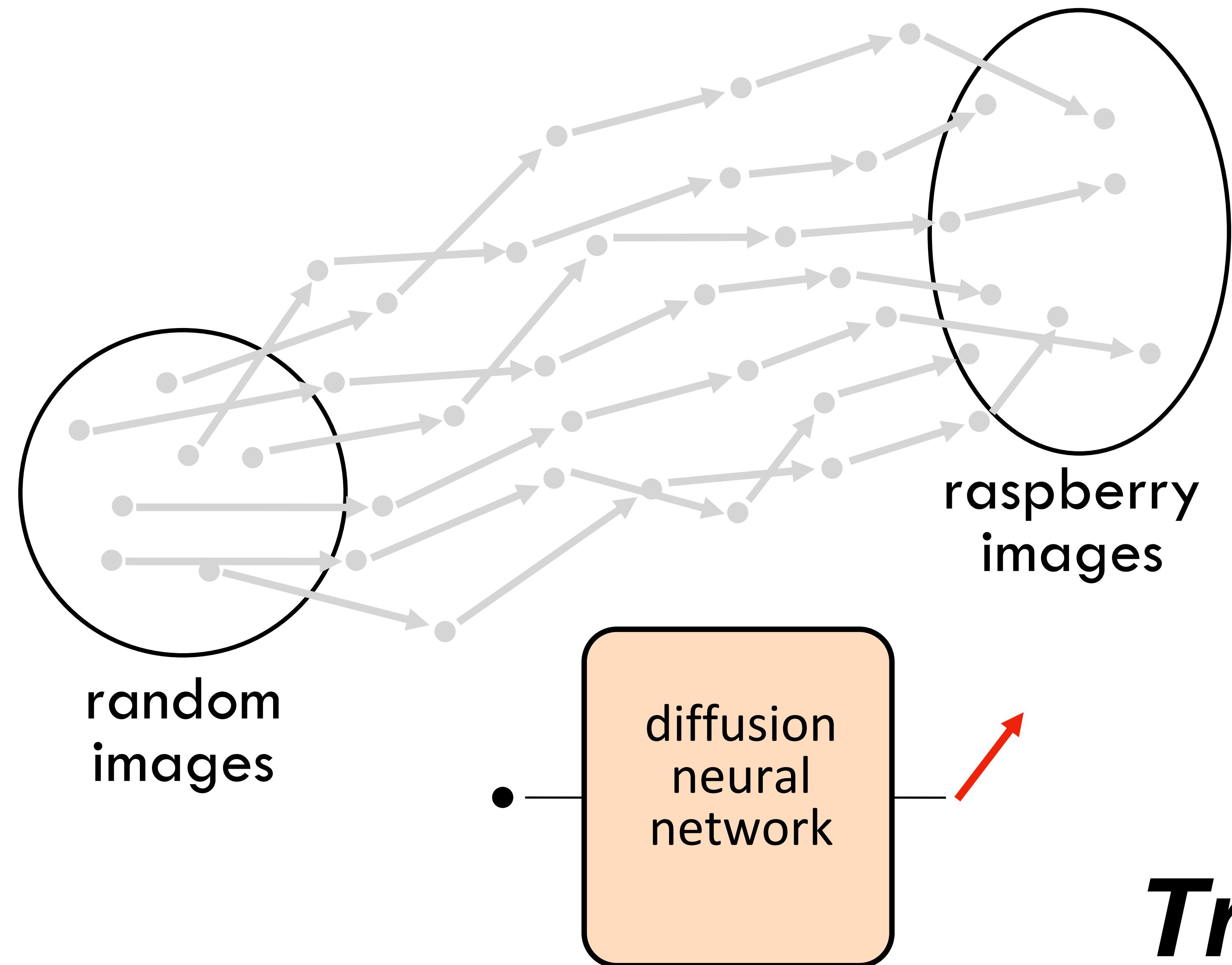
Training

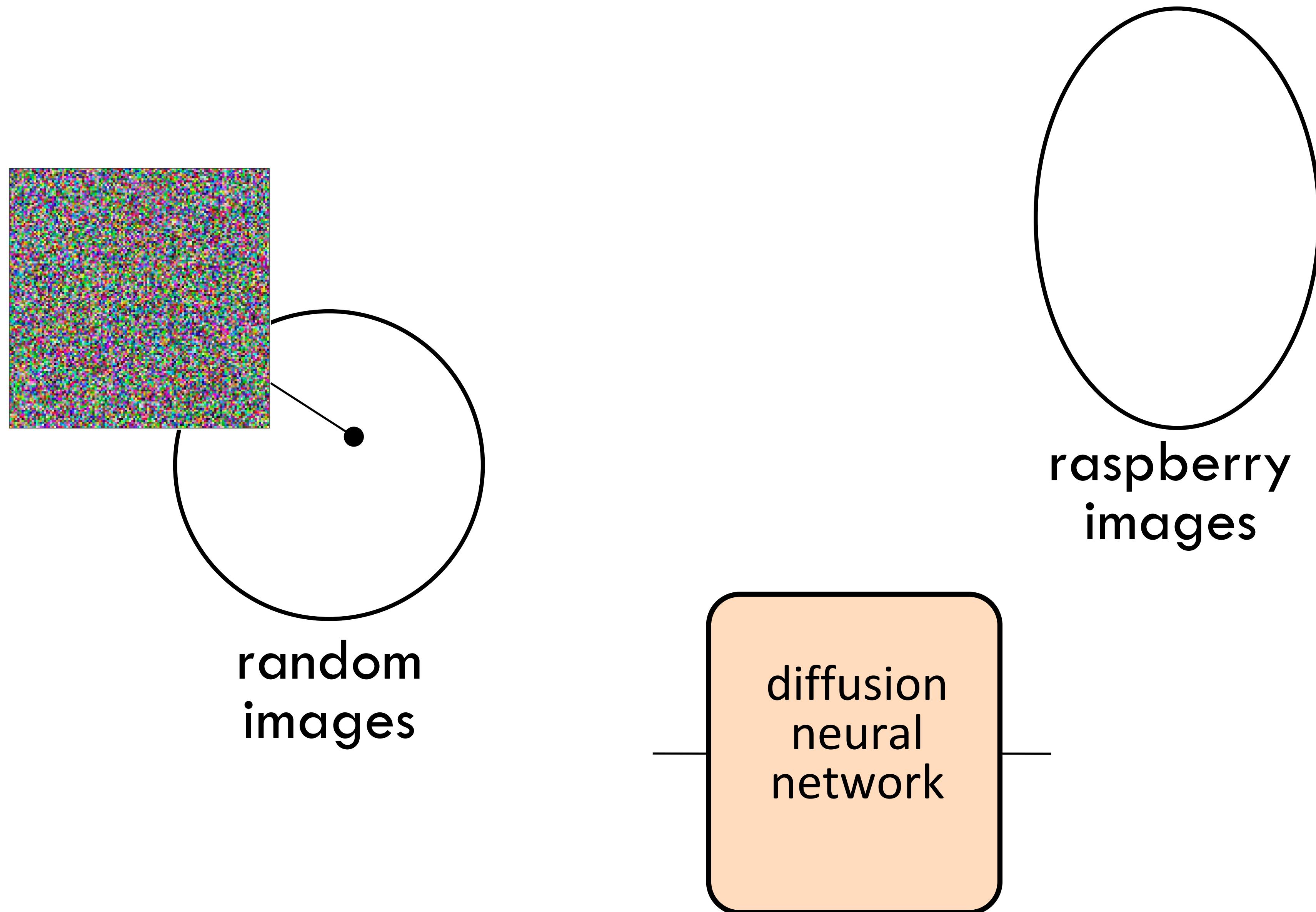
slide from Steve Seitz's [video](#)

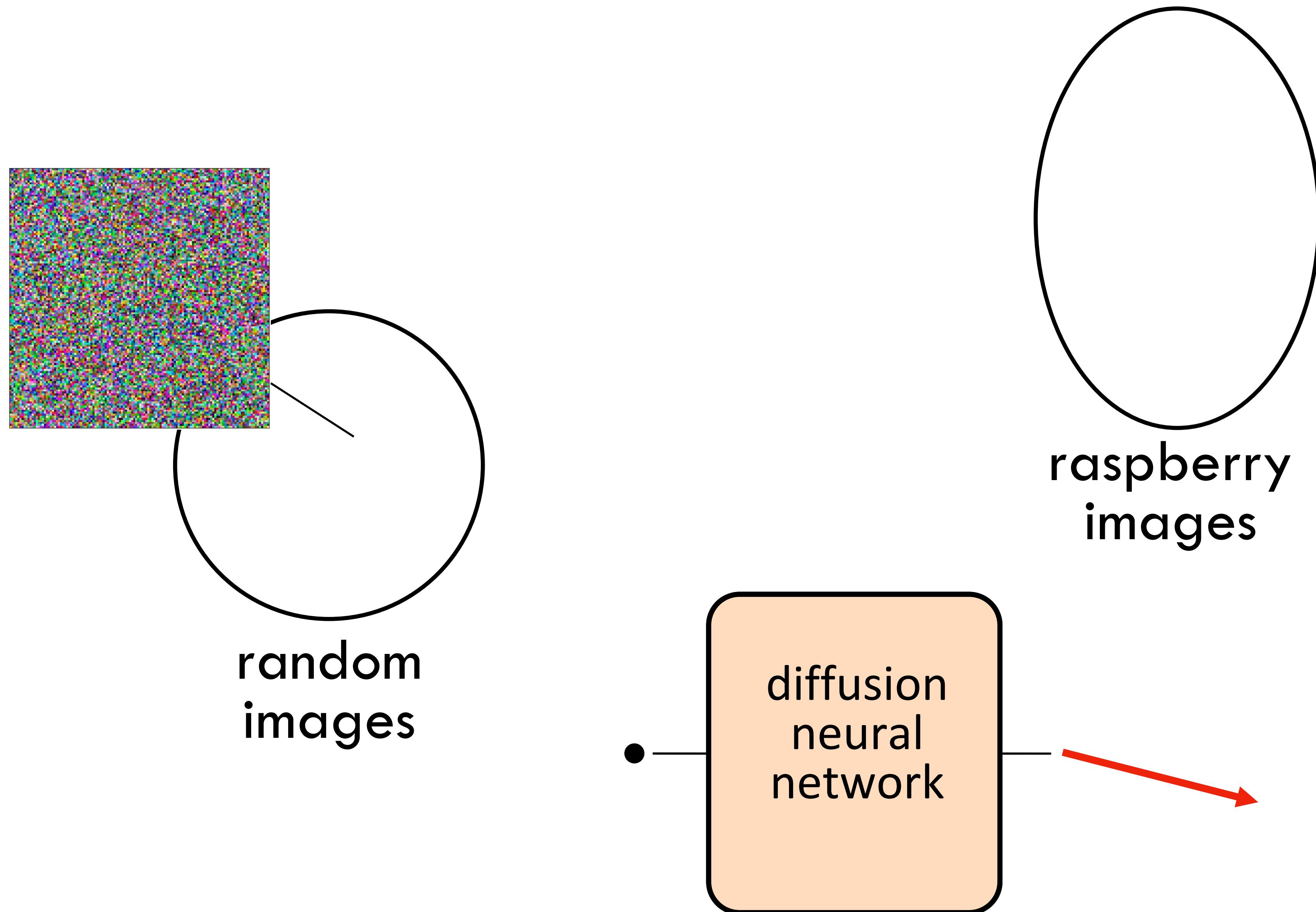


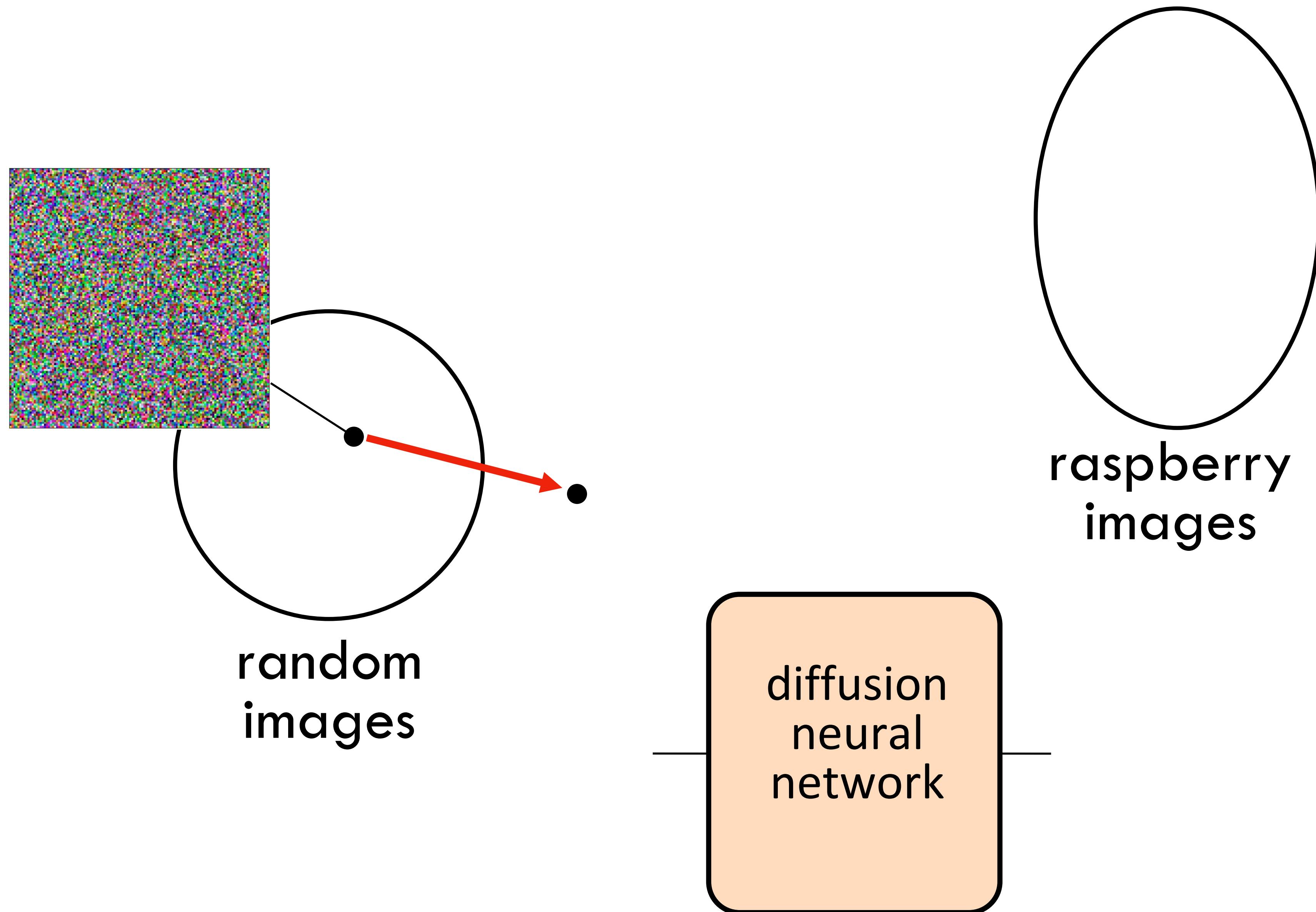
Training

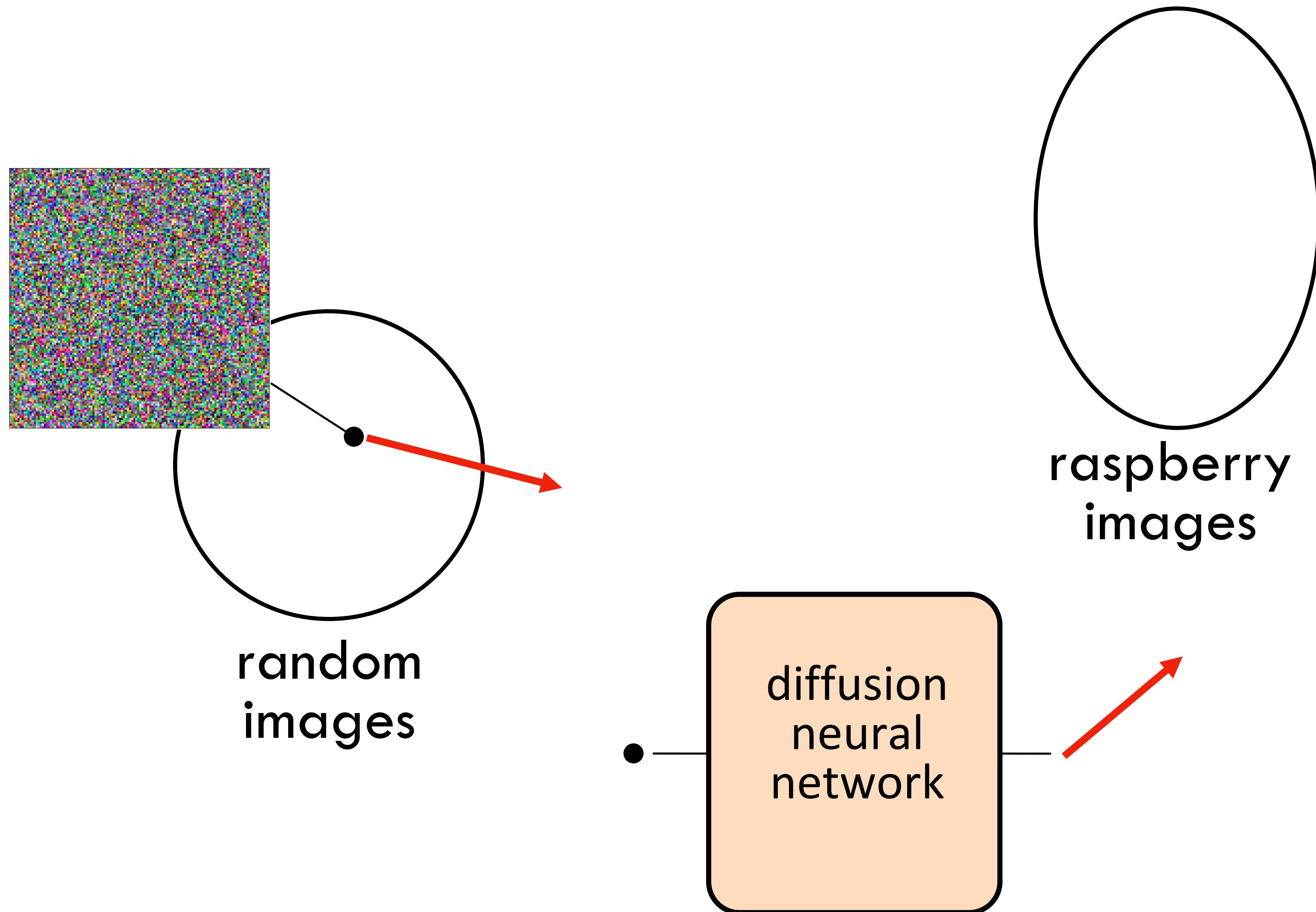
slide from Steve Seitz's [video](#)

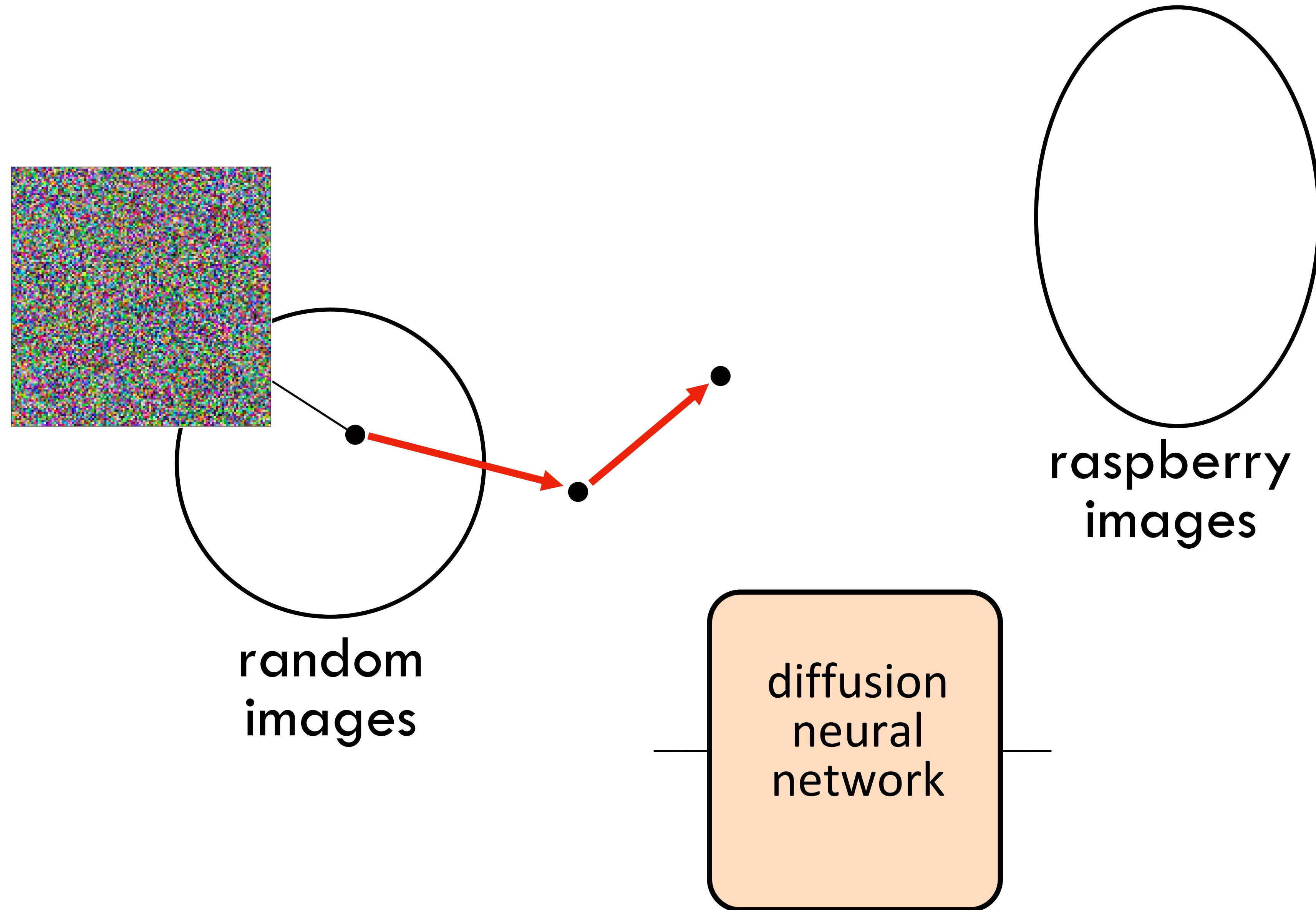


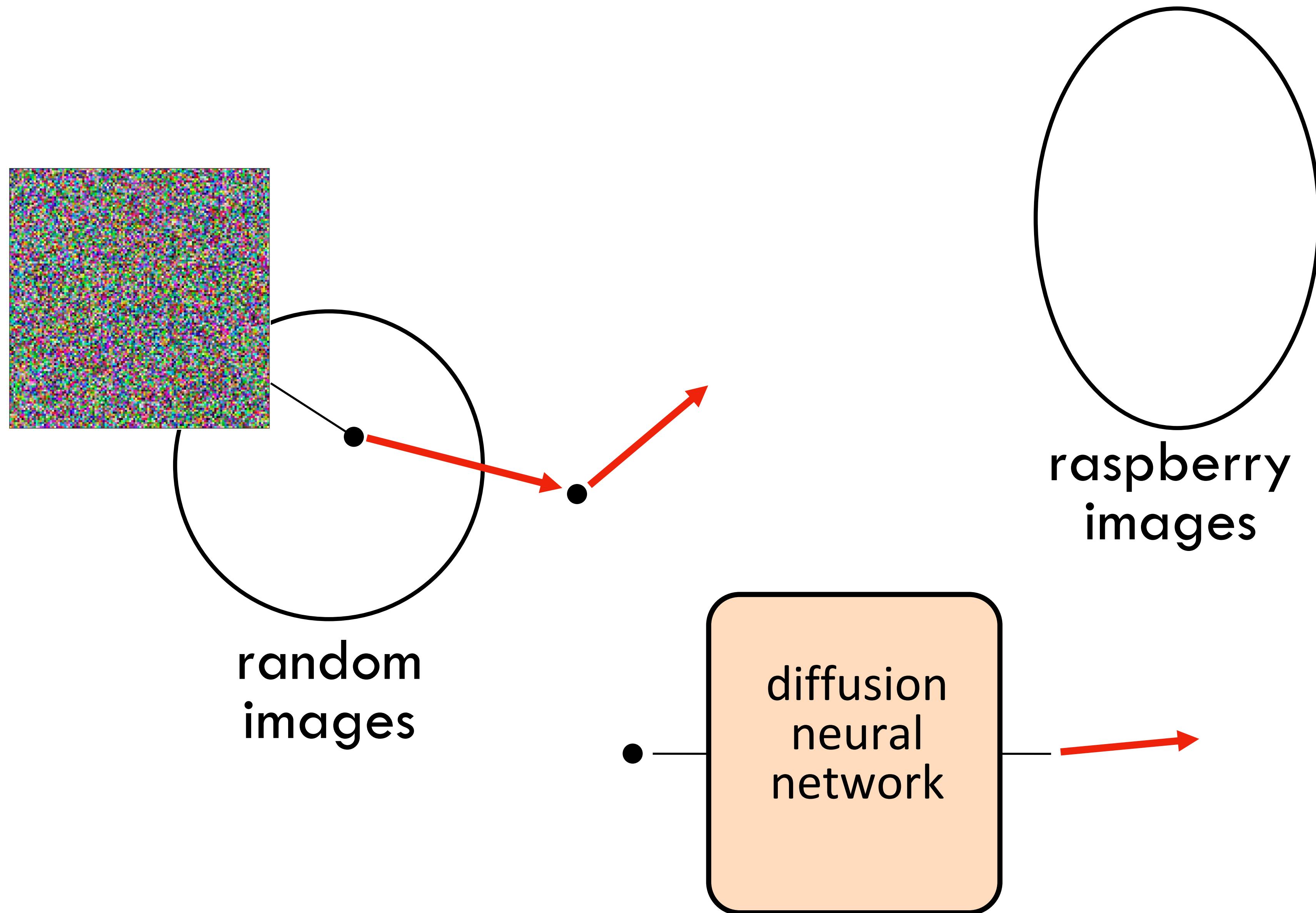


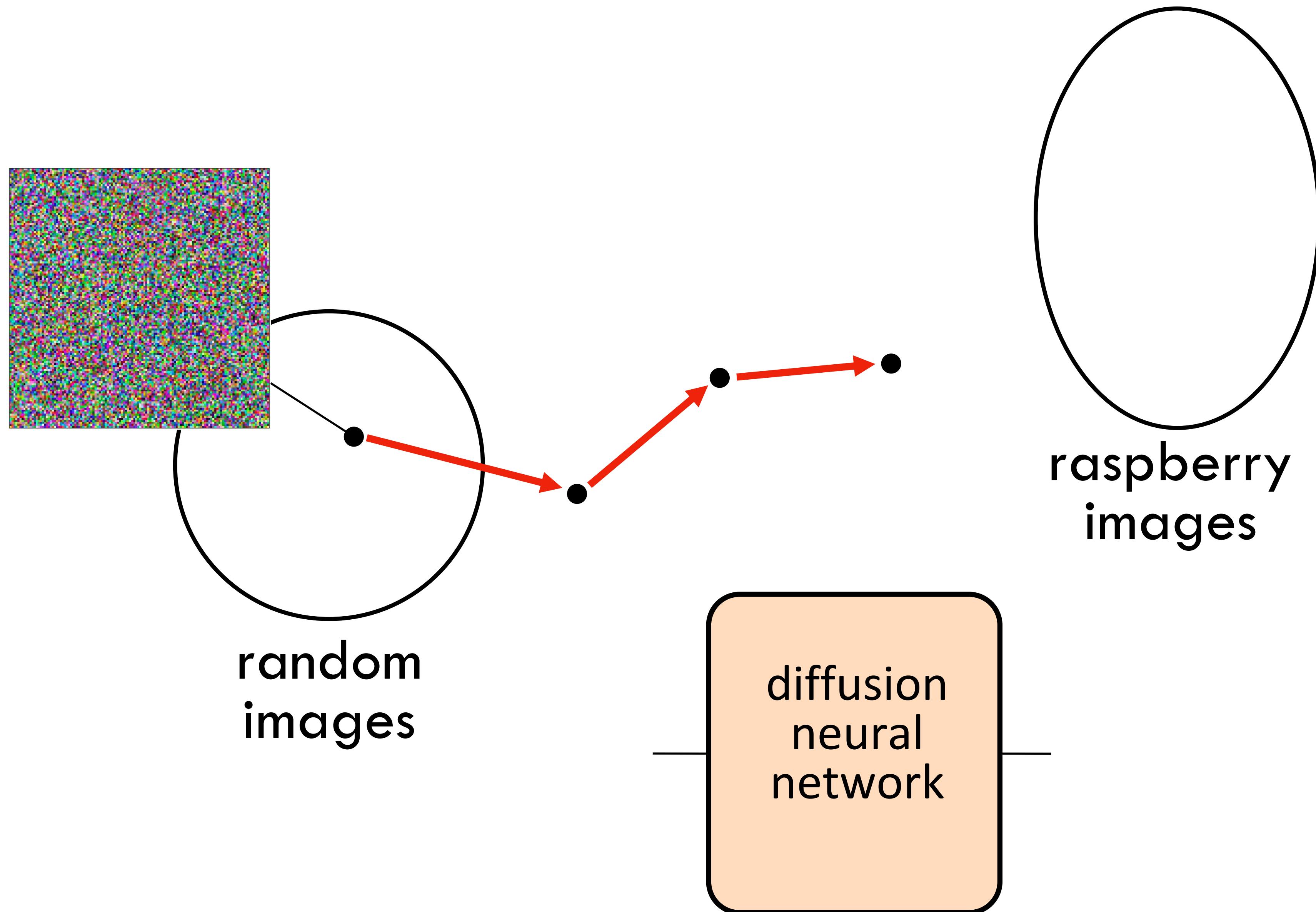


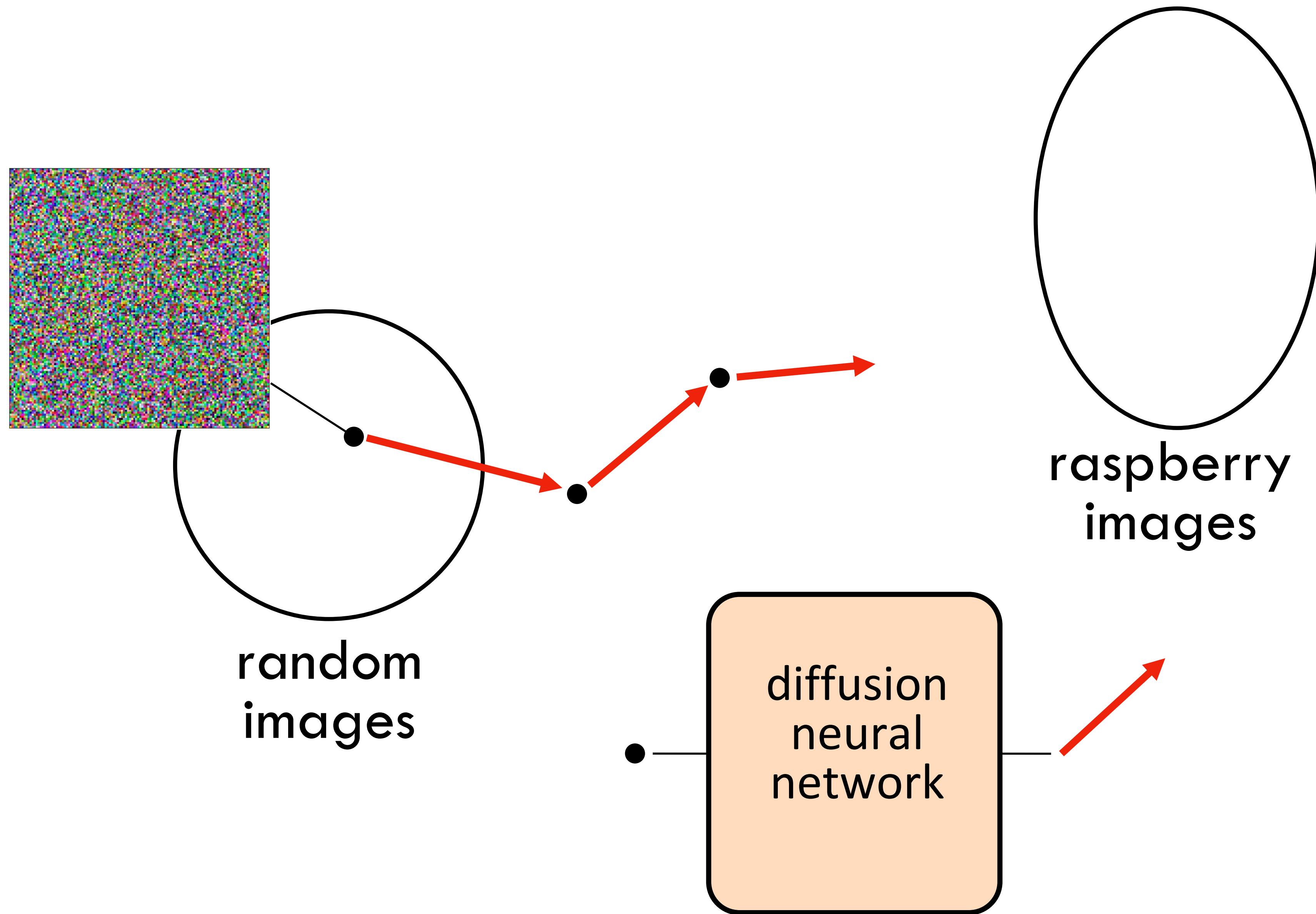


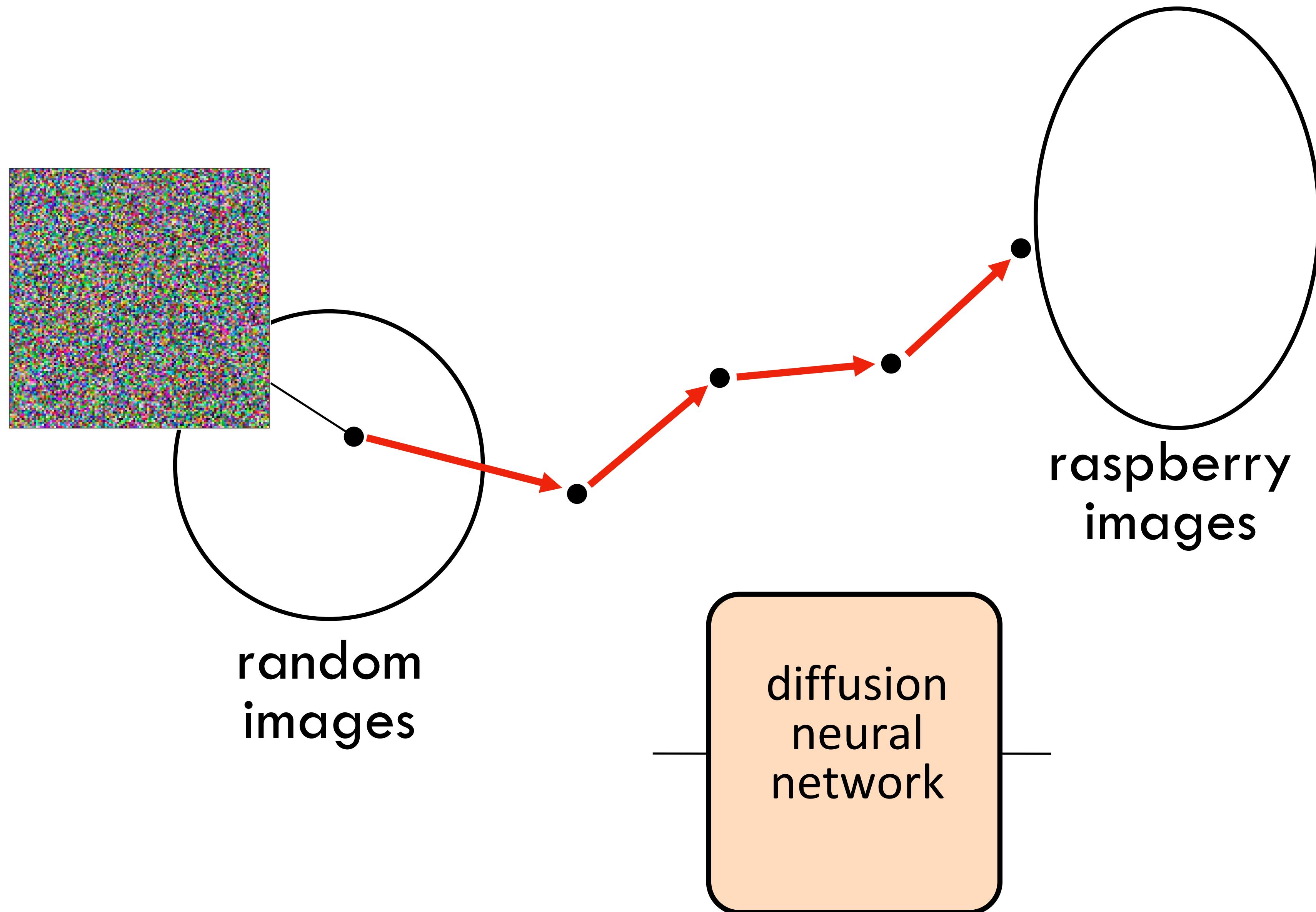


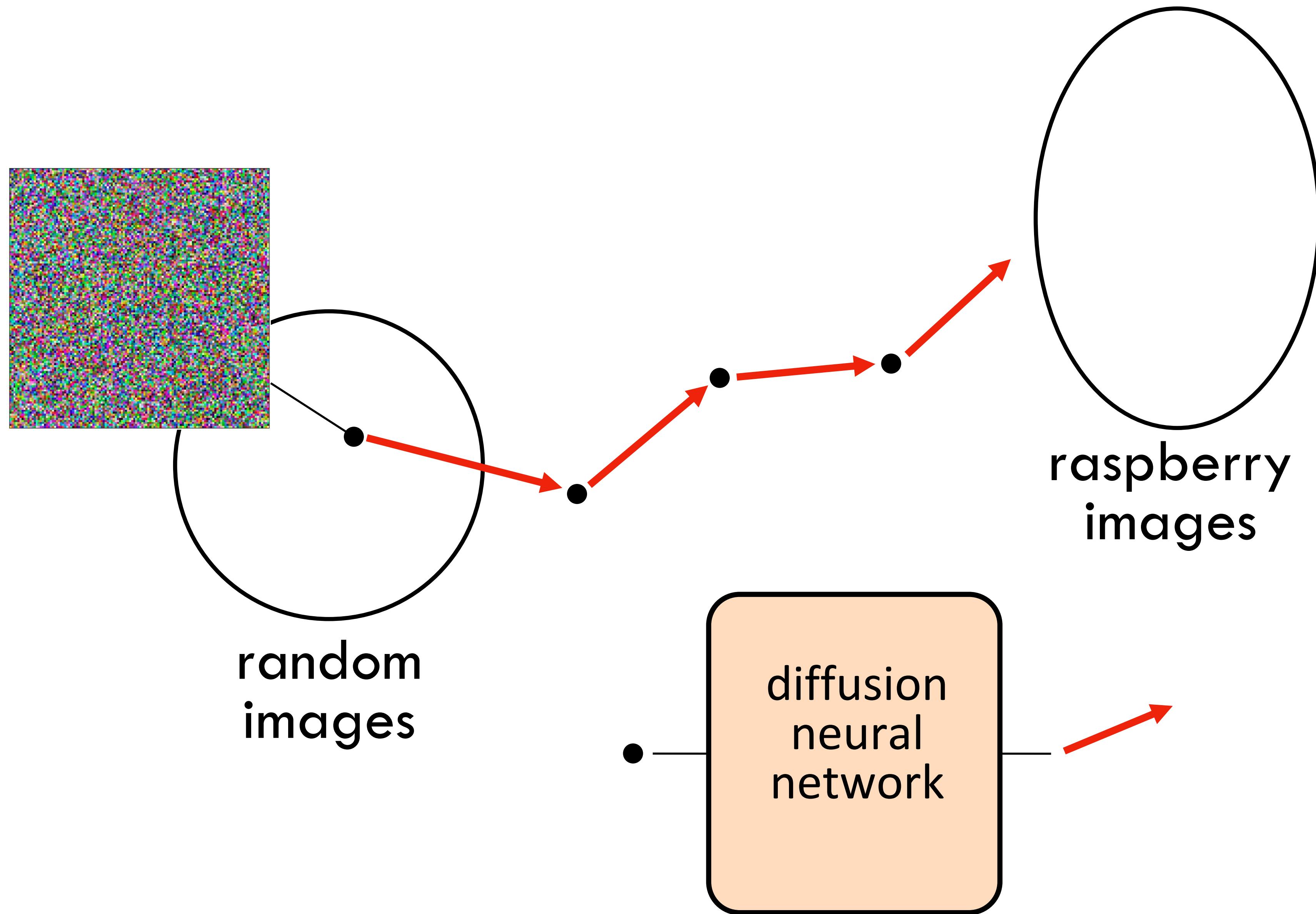


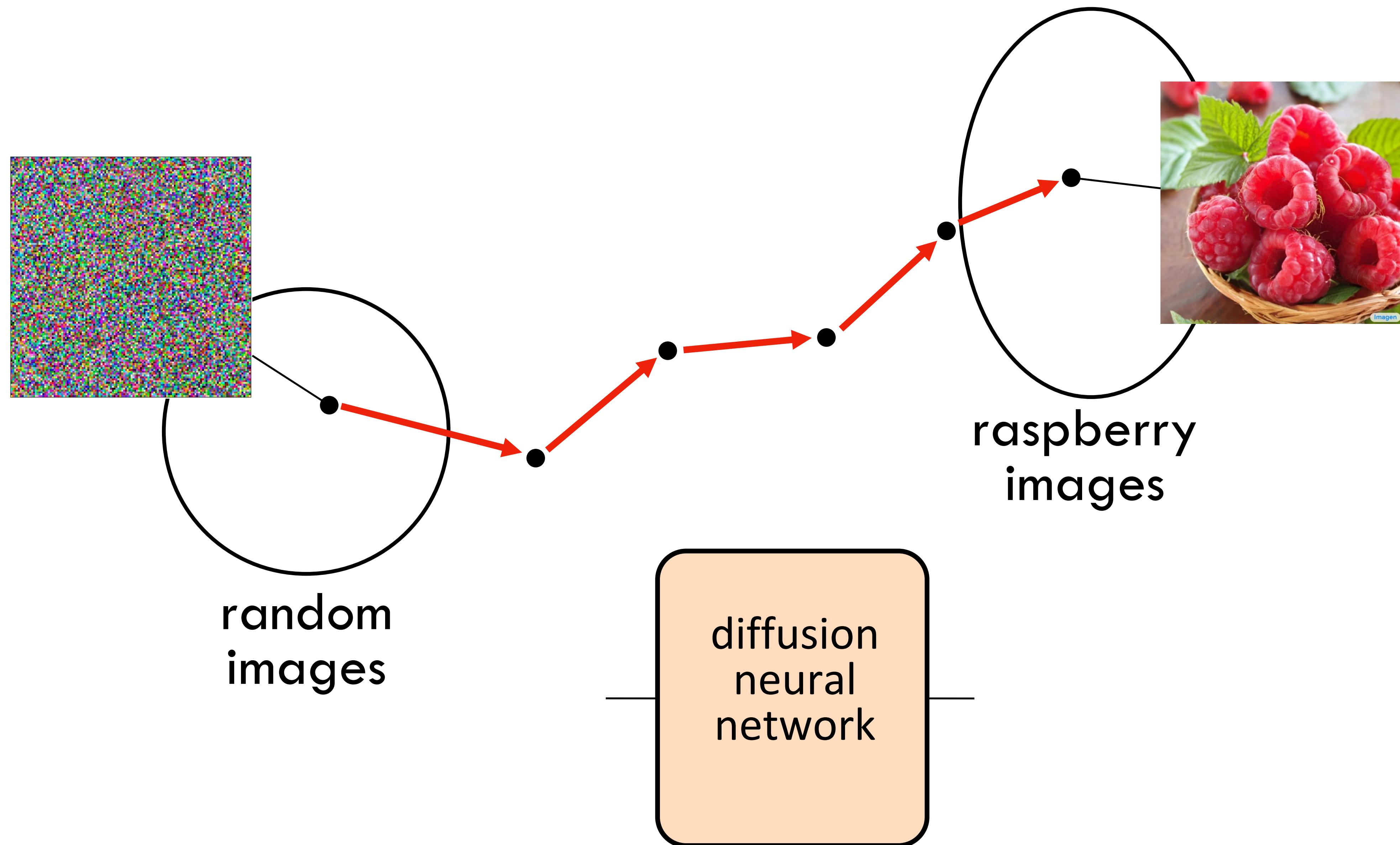


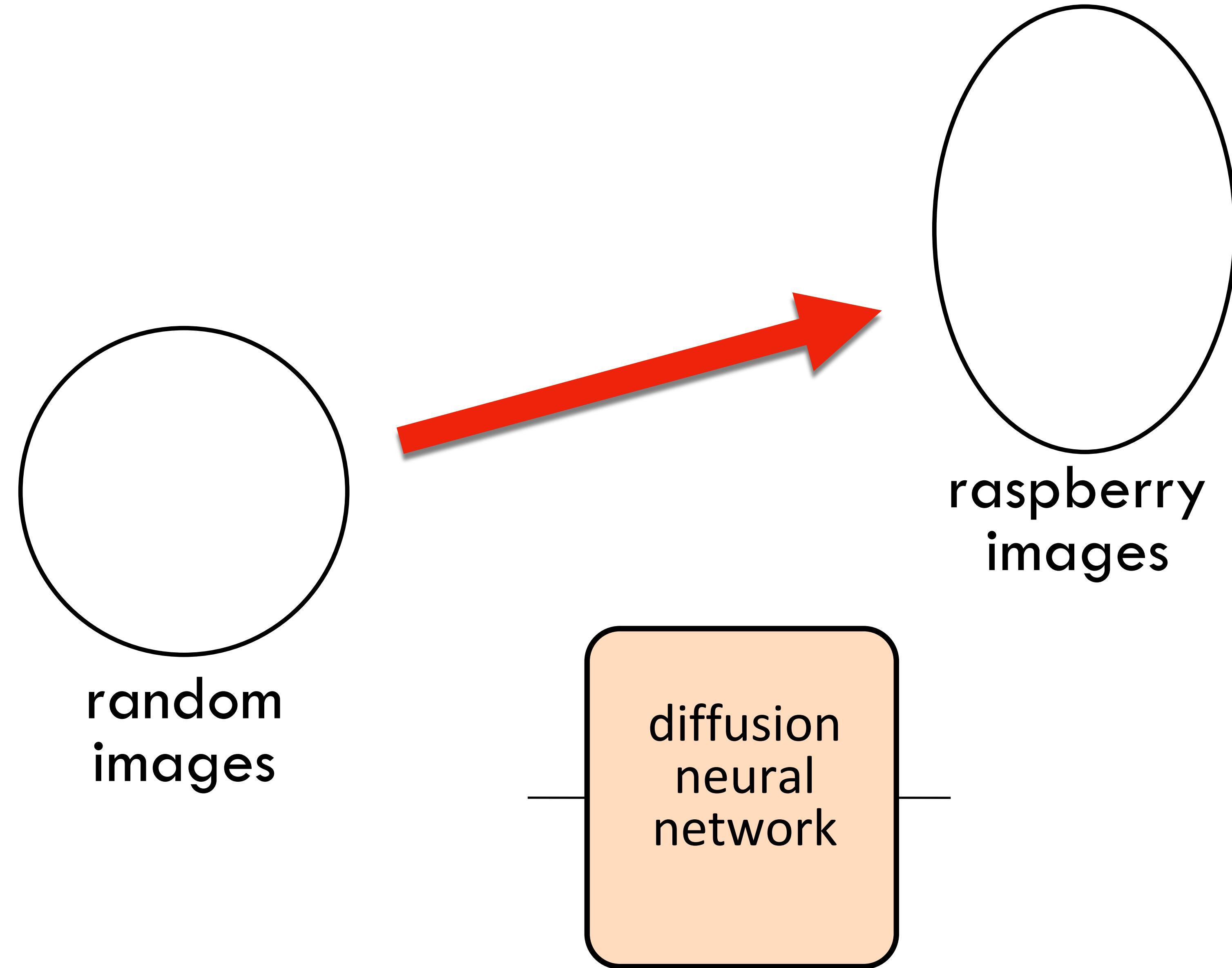


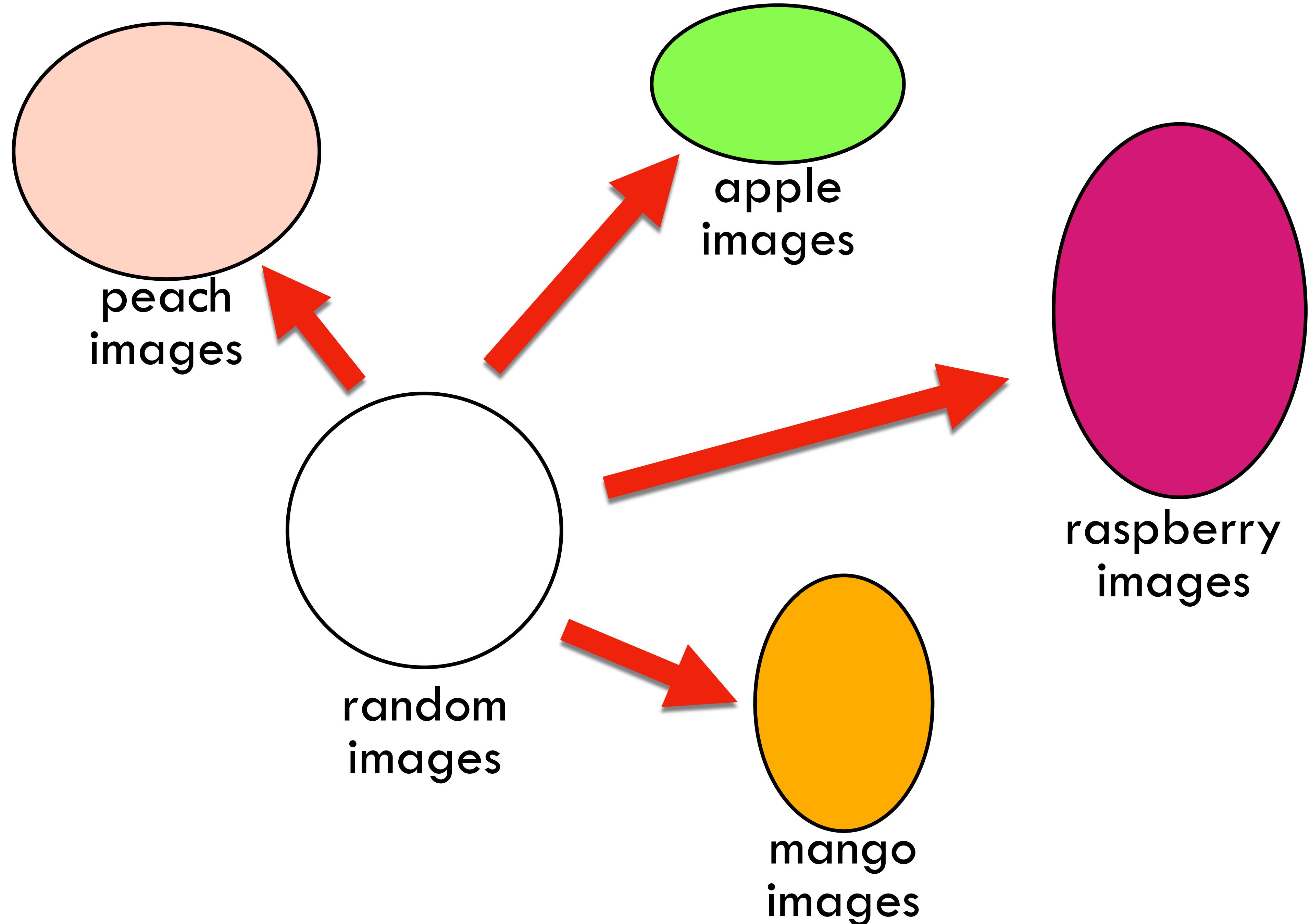


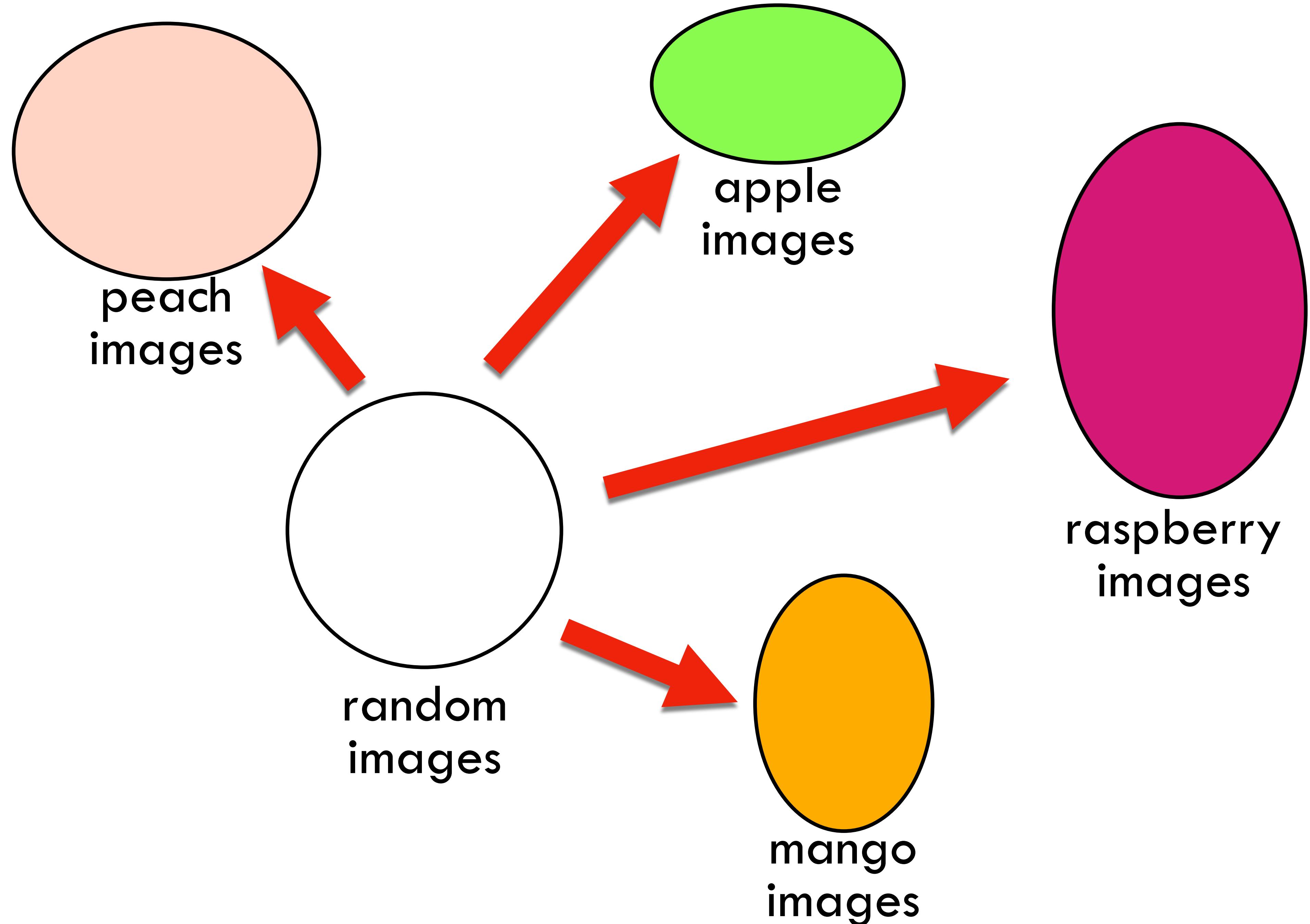


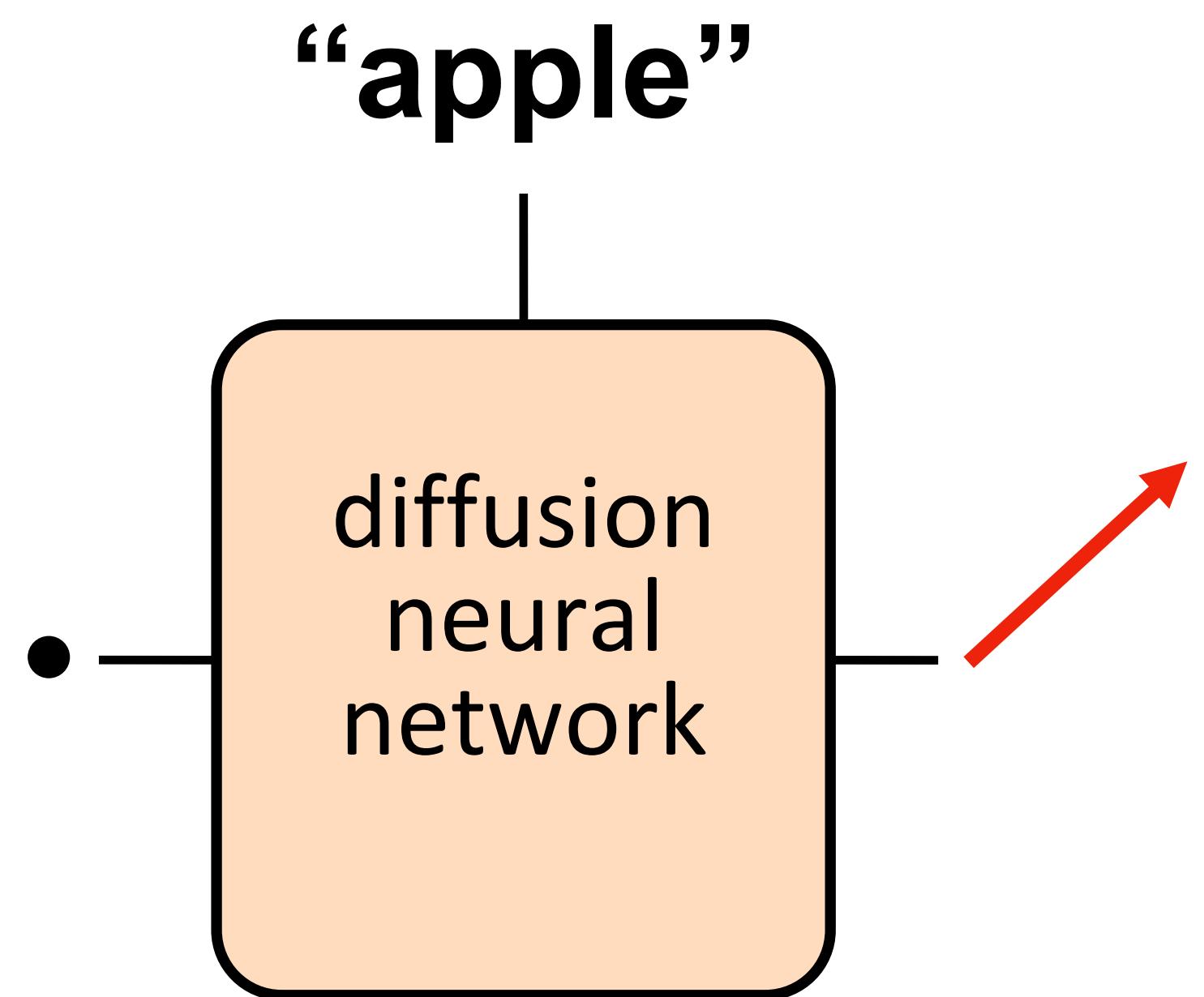
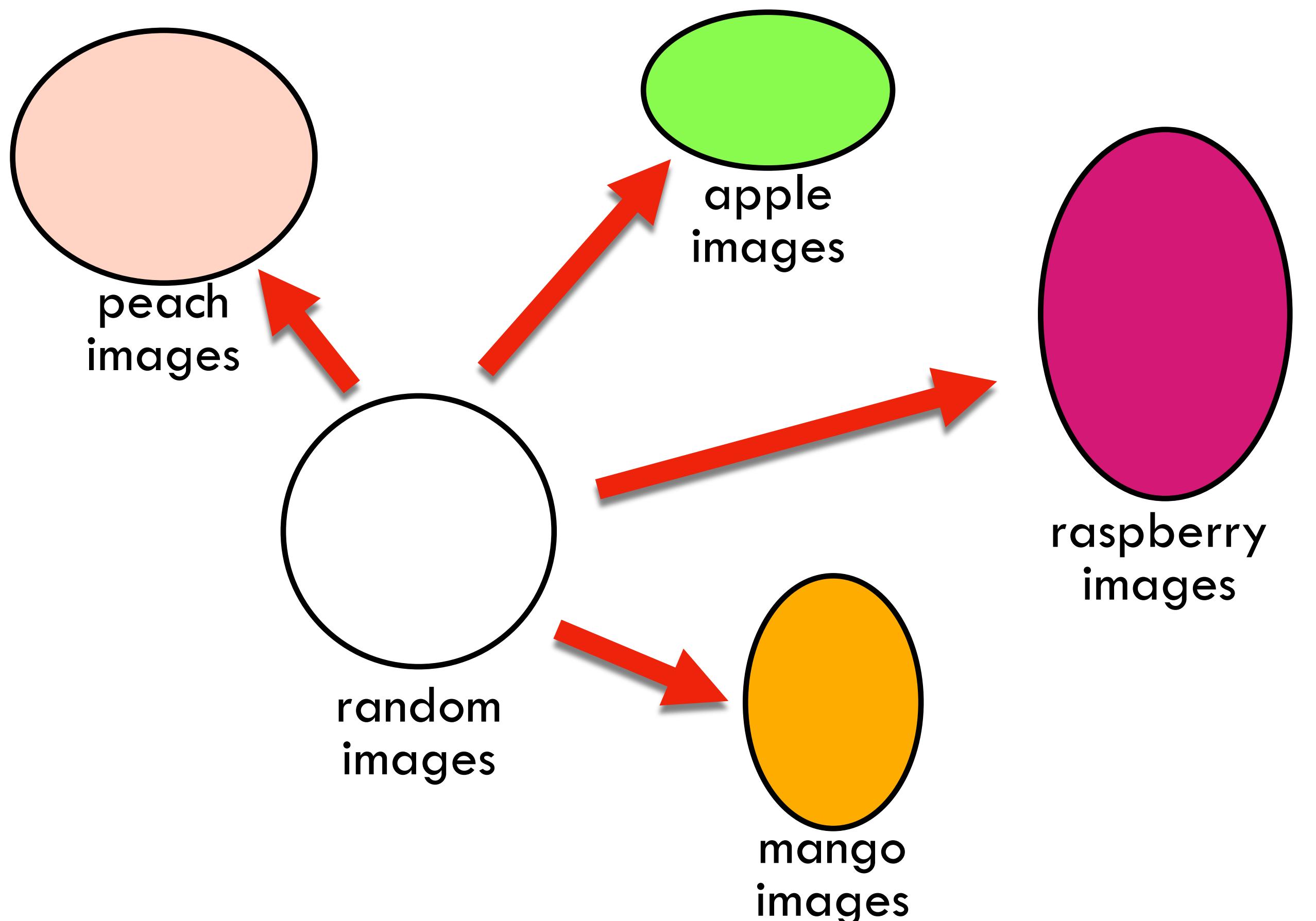


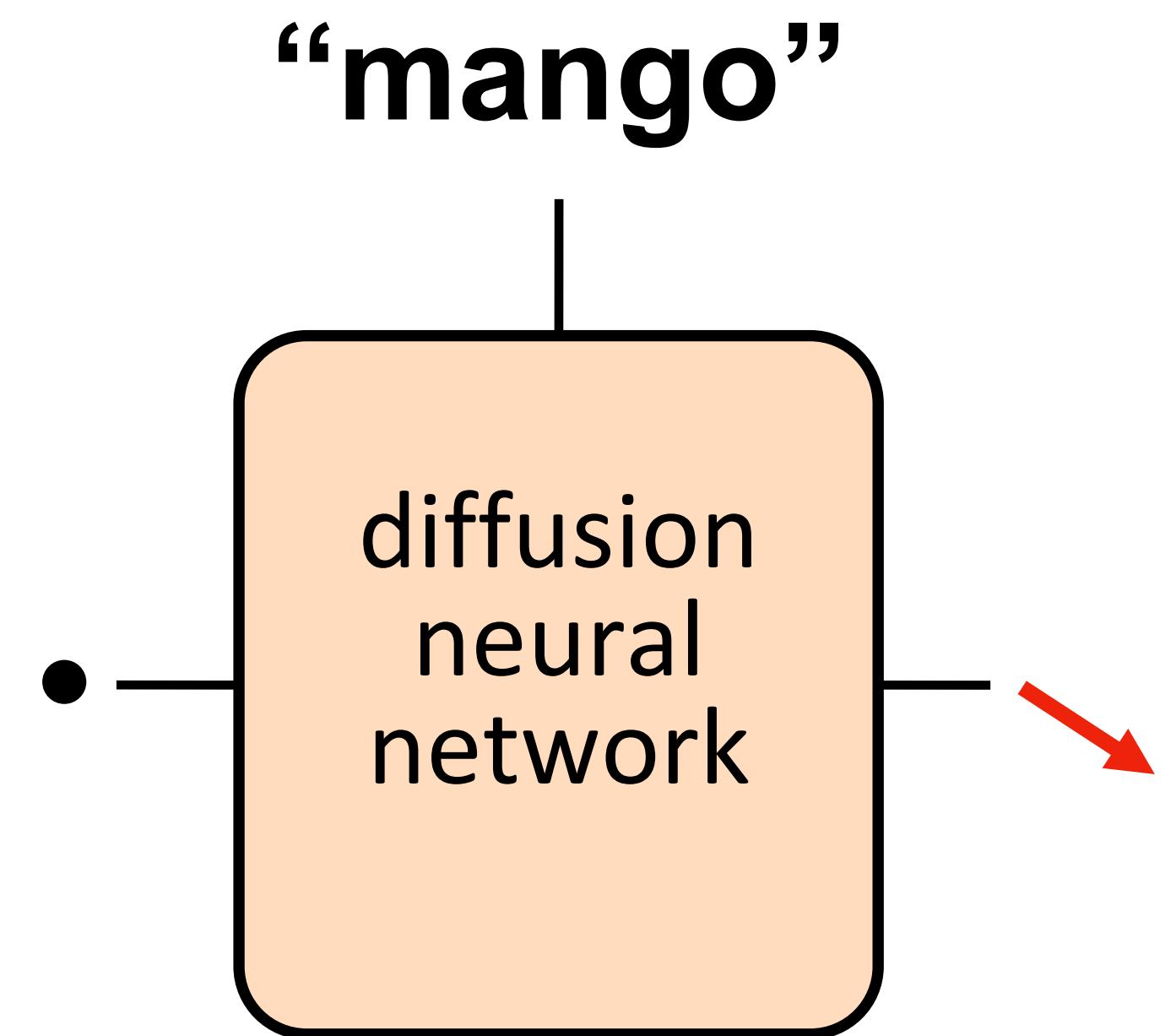
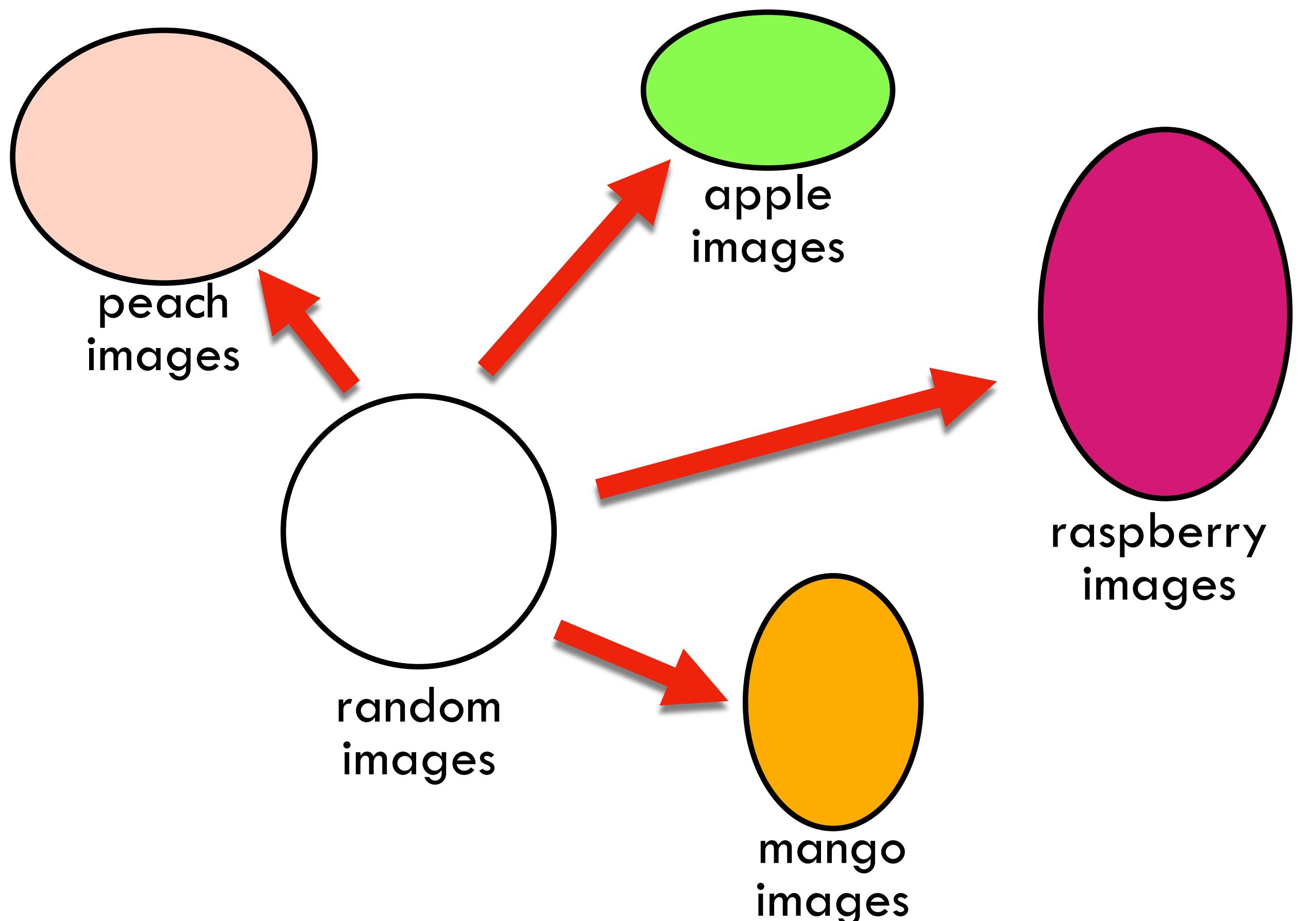




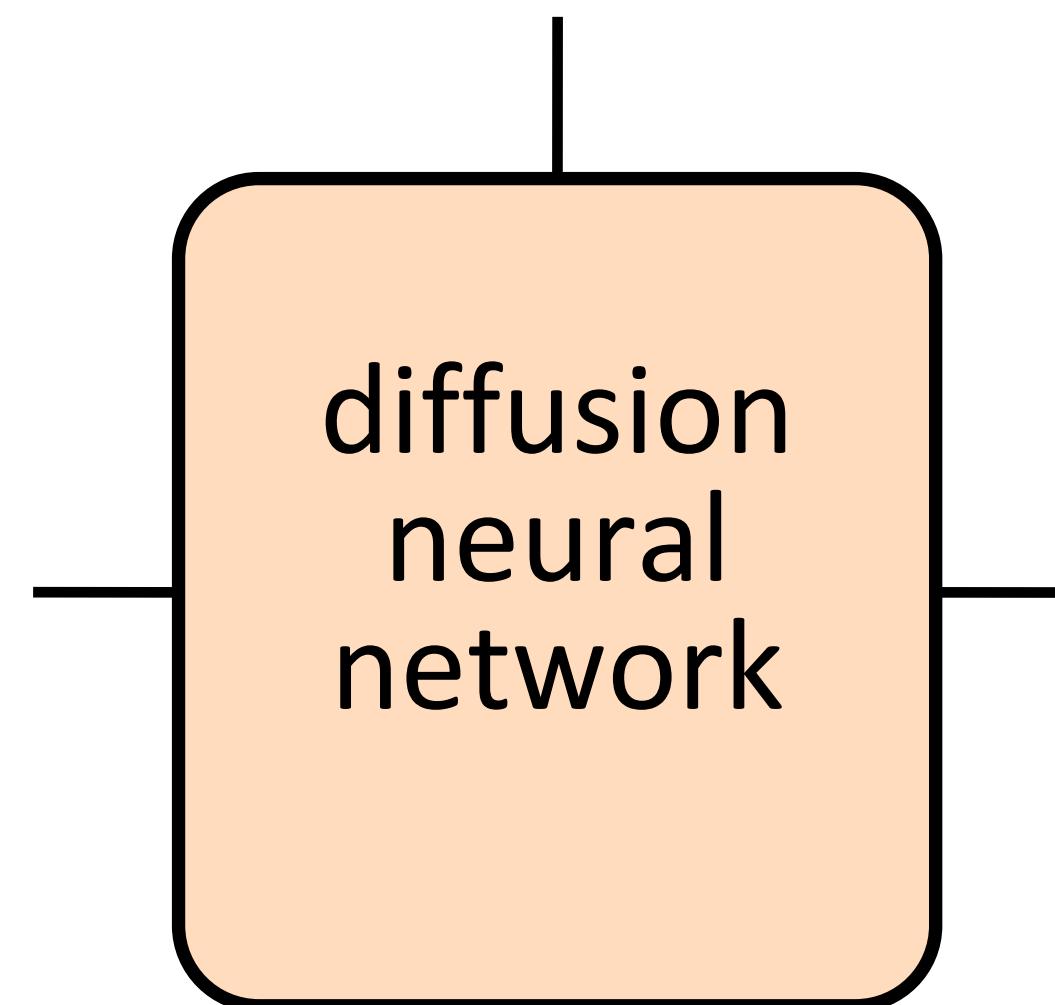


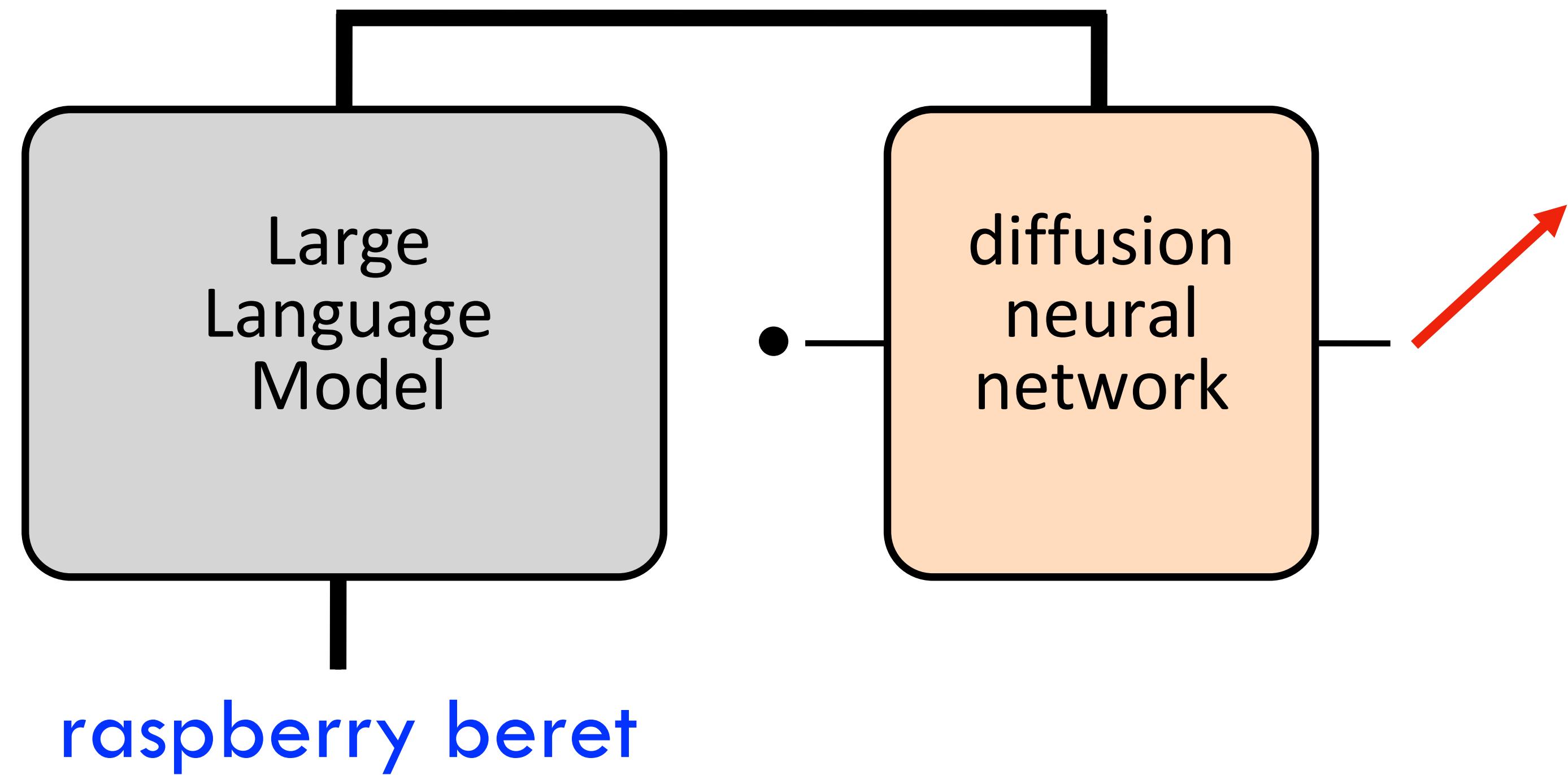






raspberry beret







raspberry beret

slide from Steve Seitz's [video](#)



Imagen

raspberry beret



Imagen

slide from Steve Seitz's [video](#)

beret of raspberries



beret of raspberries

slide from Steve Seitz's [video](#)



beret of raspberries



slide from Steve Seitz's [video](#)



Imagen

chocolate guacamole pancakes



slide from Steve Seitz's [video](#)

squirrel inside
a nutshell



Imagen

slide from Steve Seitz's [video](#)

**Language
Generator**
(+ pixels)

Diffusion
(+ language)

Parti

Imagen

Language Generator

+

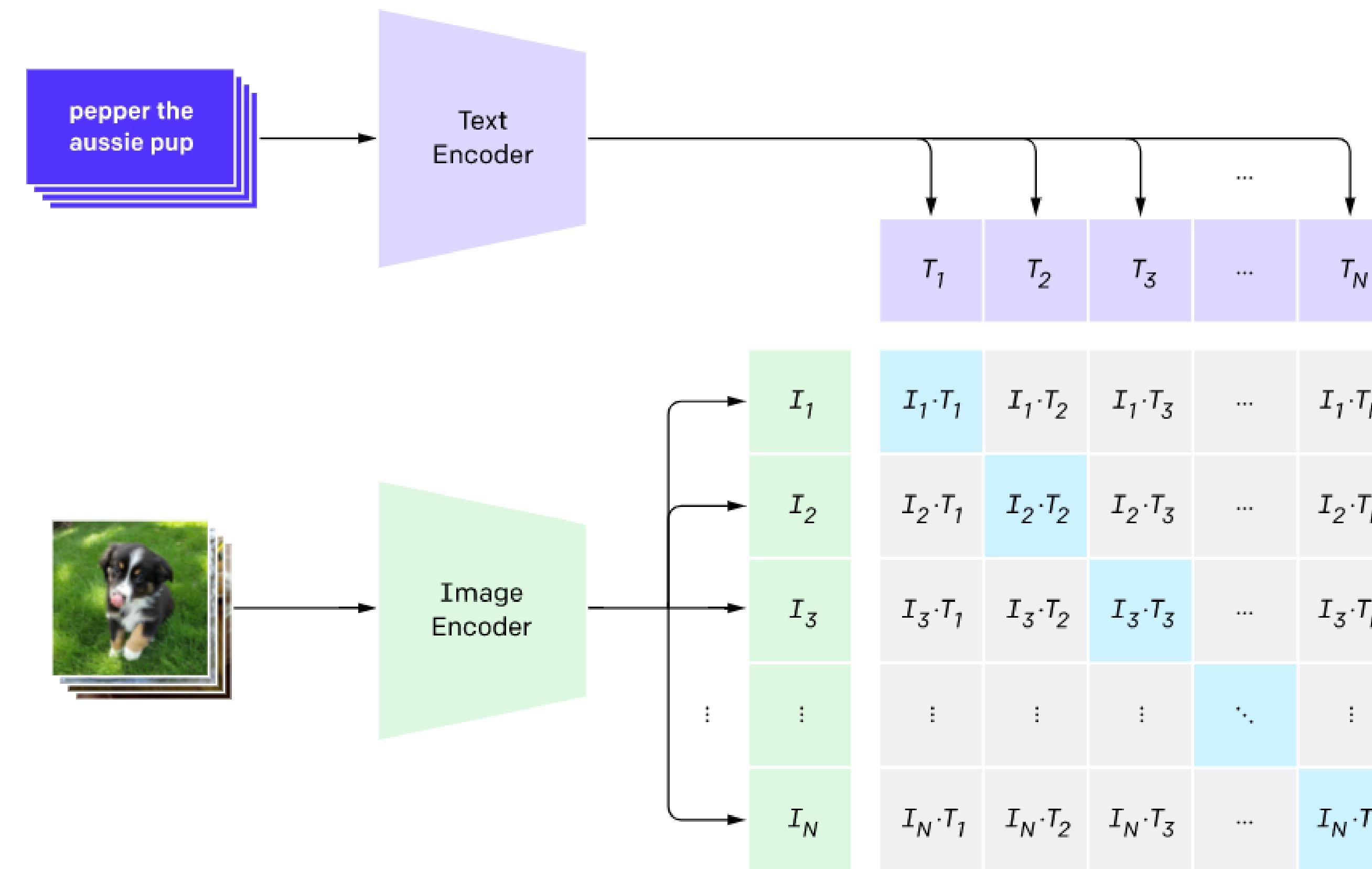
Diffusion

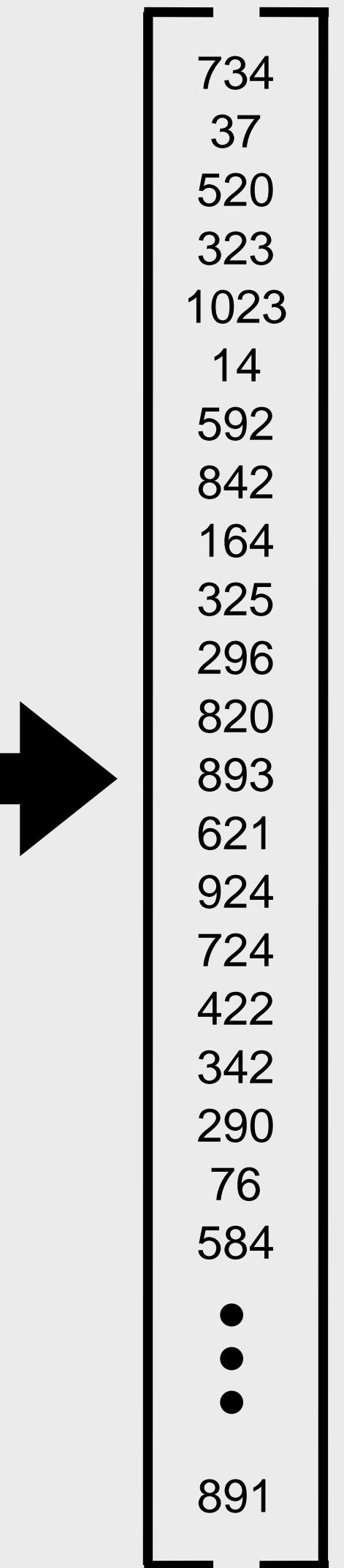
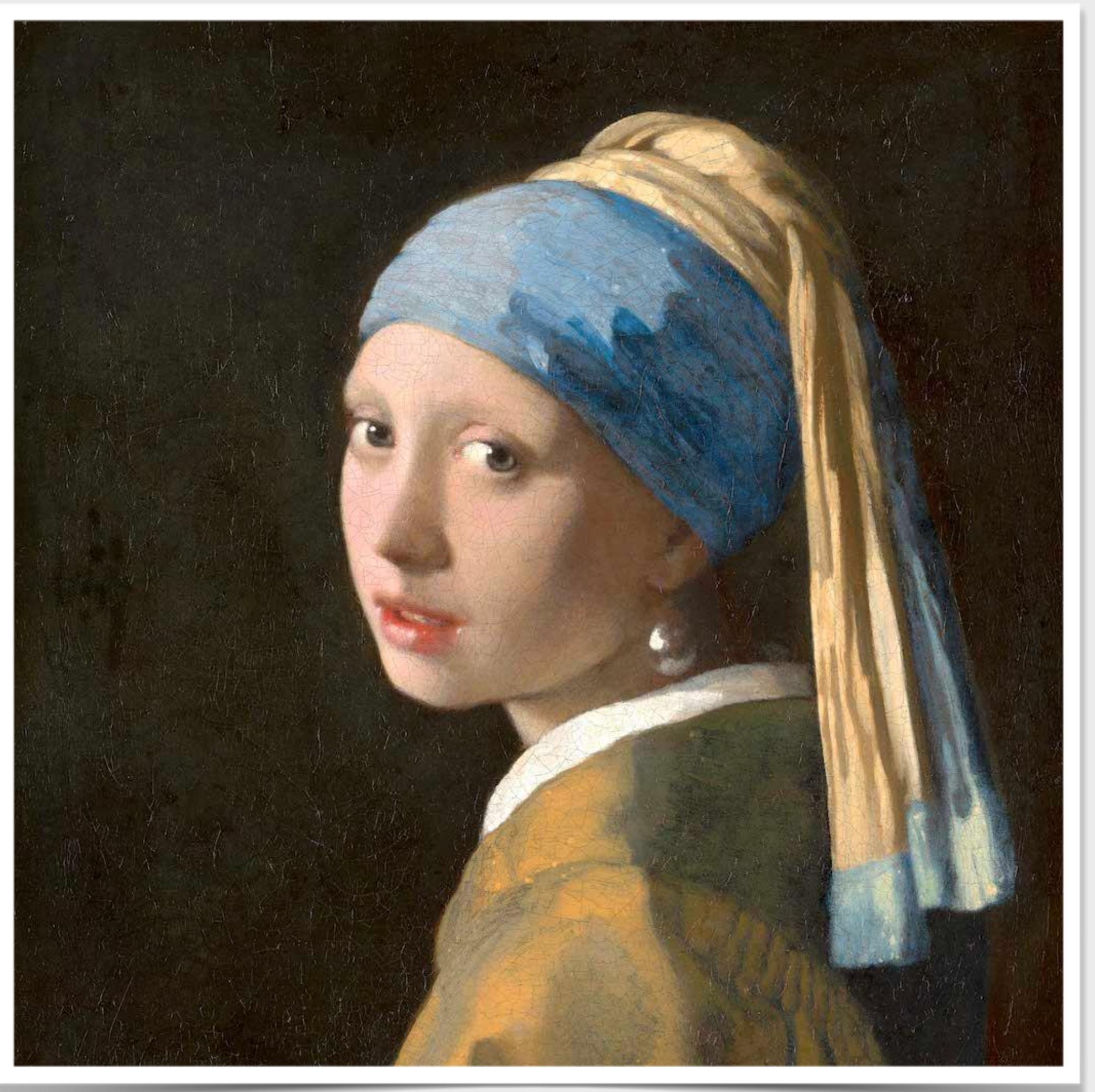
Dall-E 2

Contrastive Language–Image Pre-training (CLIP)

{word+image}2vec

1. Contrastive pre-training







slide from Steve Seitz's [video](#)



slide from Steve Seitz's [video](#)



Squirrel reaching for a nut. Latte art

slide from Steve Seitz's [video](#)



Imagen

A teddy bear making chocolate guacamole pancakes

slide from Steve Seitz's [video](#)

A dog looks curious in the
mirror,
seeing a cat



Imagen

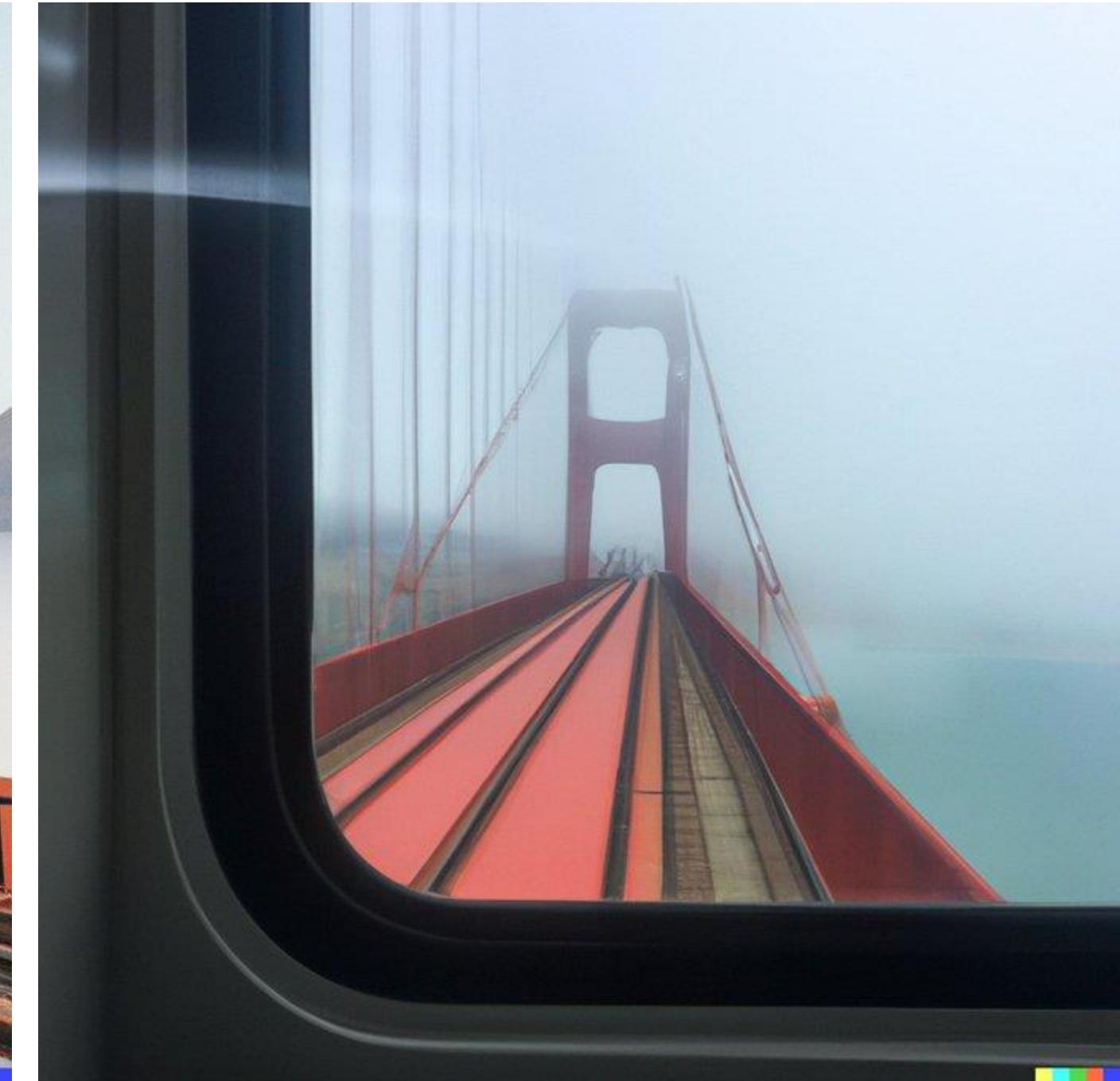
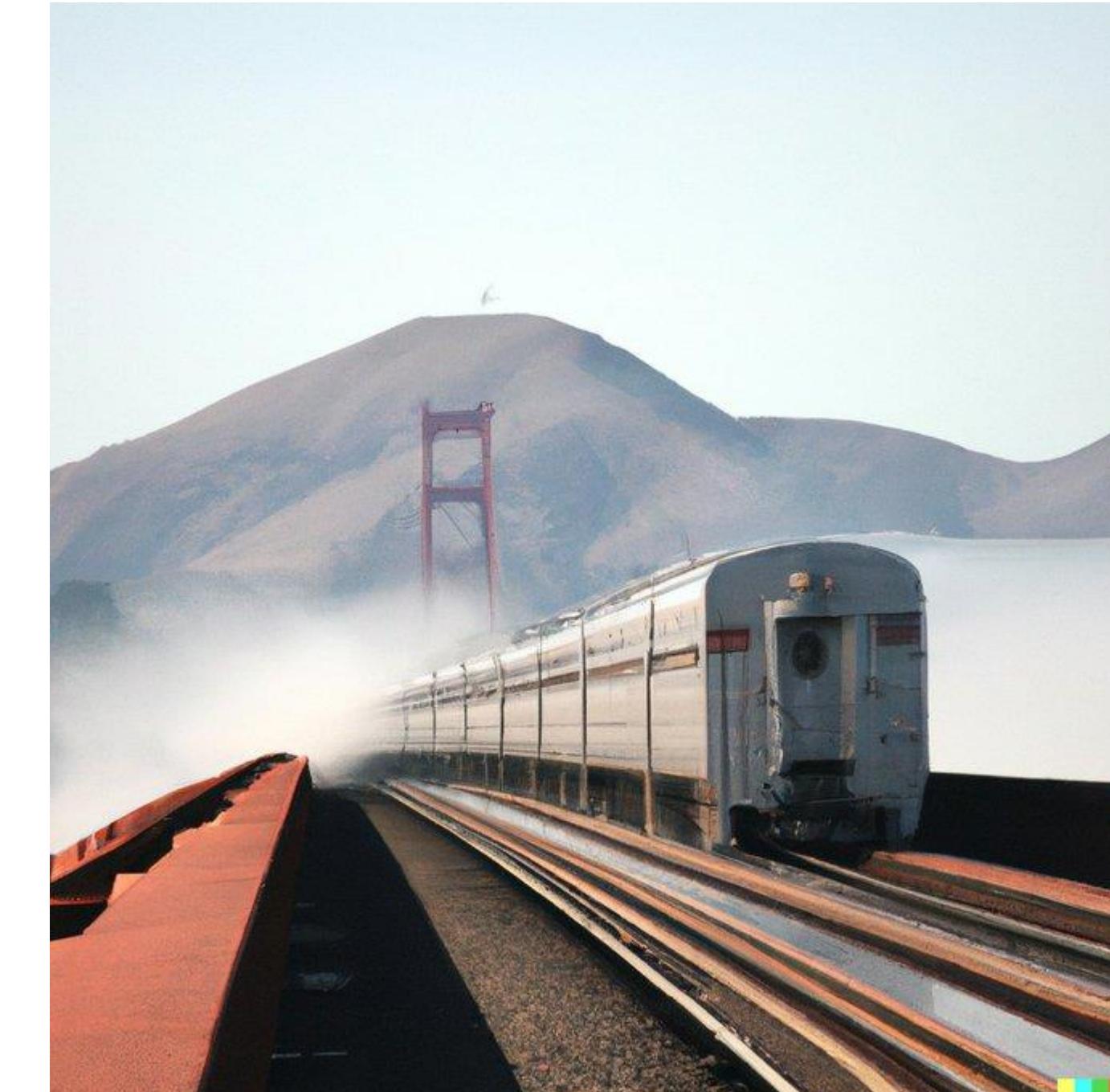
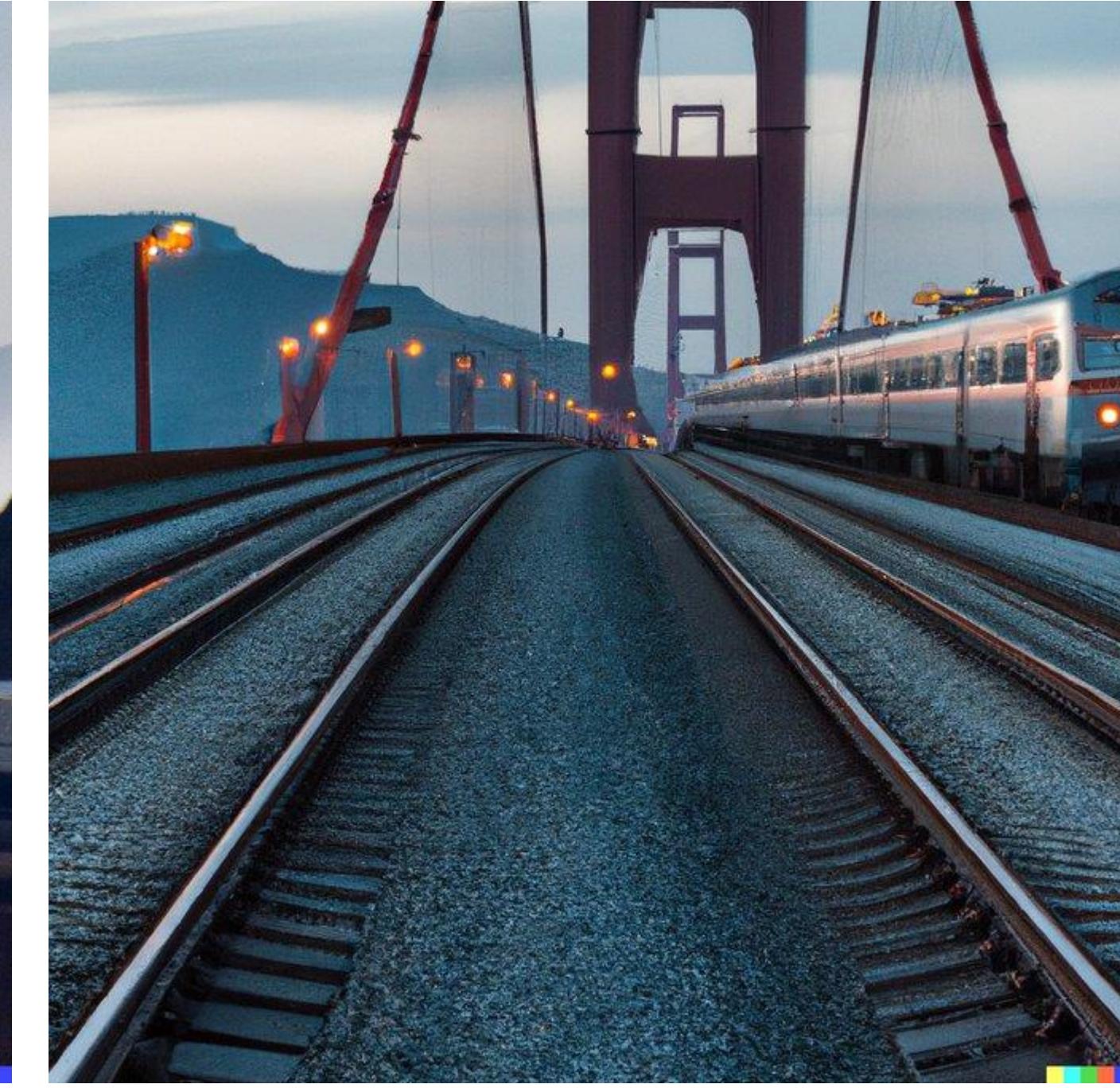
A bowl of soup
that looks like a monster
knitted out of wool



Dall-E 2



Impressive compositionality:



DALL-E + Danielle Baskin





“Person holding a heavy box”



“Person holding a laptop”



“Person holding birthday cake”



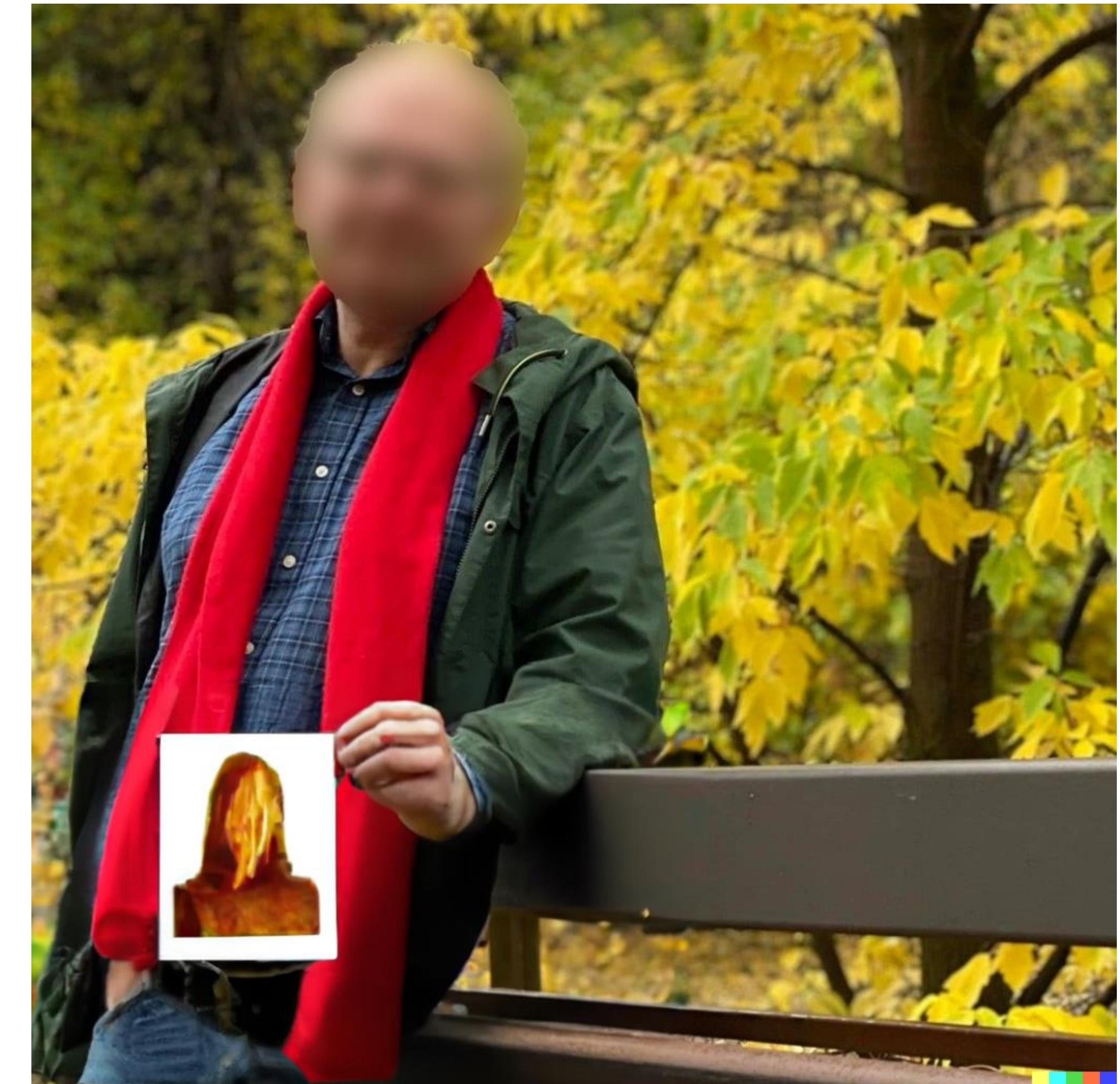
“Person offering tea”



“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...

DALL·E

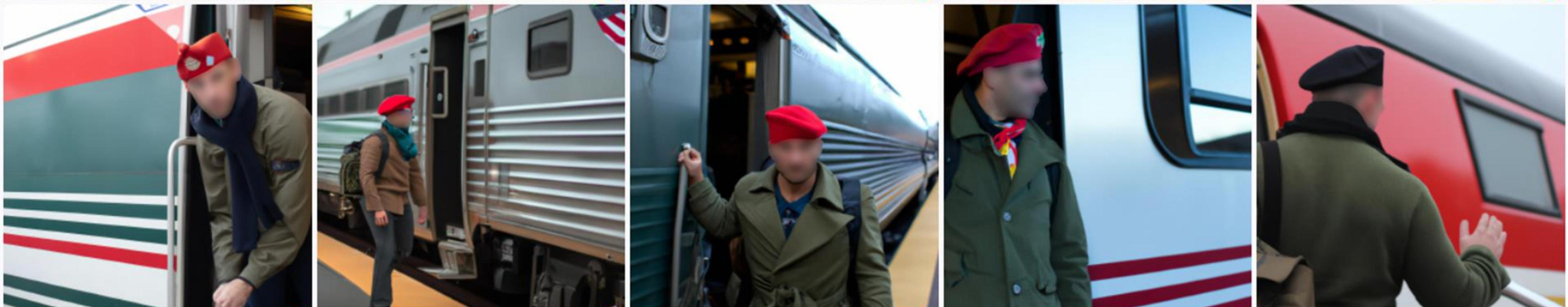
...

A

a man with red scarf and green beret boarding a Amtrak train



Report issue



“Bag of words” seems to work as well

DALL·E

Edit the detailed description

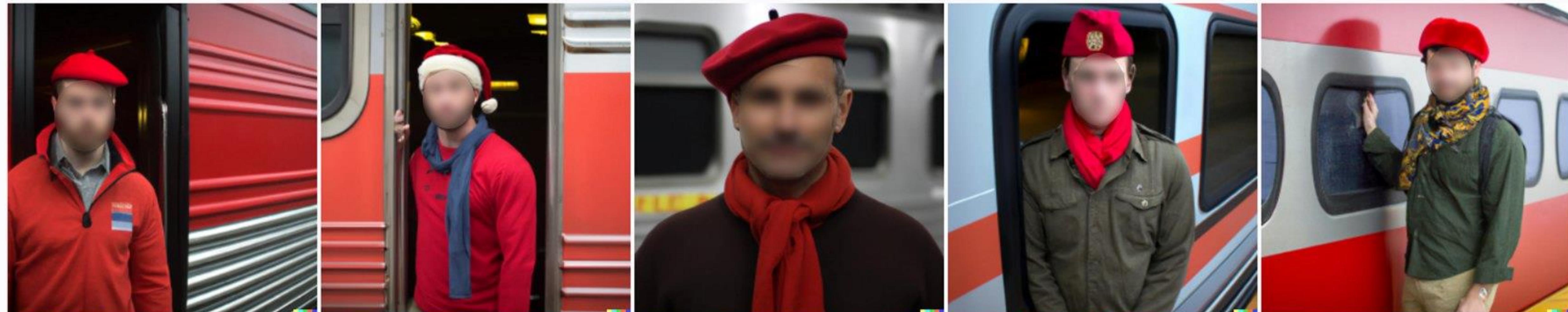
red, man boarding, scarf, Amtrak, green beret, train

Surprise me

Upload



Report issue



Edit the detailed description

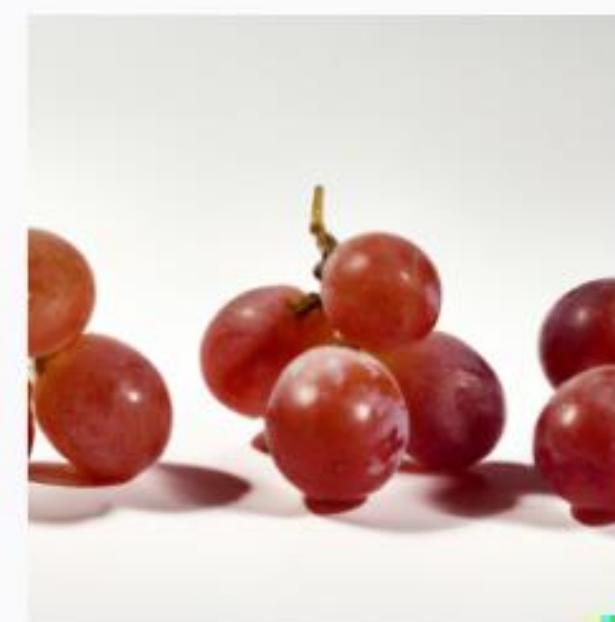
Surprise me

Upload



five grapes

Generate



Report issue



Type here to search



68°F Sunny

2:27 PM
7/15/2022

ENG

Edit the detailed description

[Surprise me](#)[Upload](#)

three grapes

[Generate](#)[Report issue](#)

Type here to search



68°F Sunny

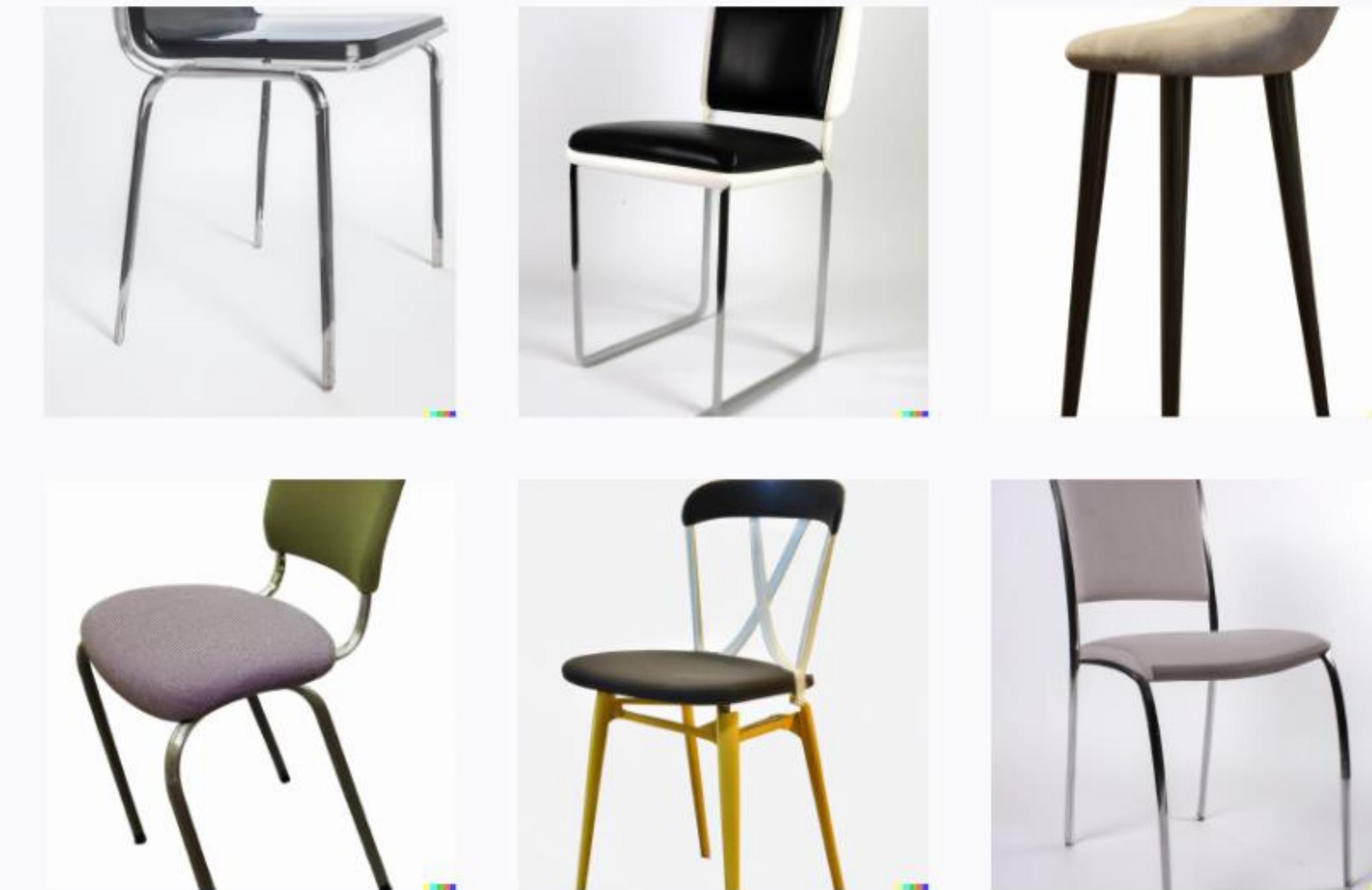
2:27 PM
7/15/2022

labs.openai.com/e/FHeAFqdICN0coL271Y0pFncS[Research](#) [Paris Stuff](#) [teaching](#) [Premium Fare Deals...](#) [Кино онлайн бесп...](#) [PI Portfolio](#) [Inbox \(14,384\) - aae...](#) [Richard bluejeans](#) [Zoom Efros](#) [»](#) [Other bookmarks](#)

Edit the detailed description

[Surprise me](#)[Upload](#)

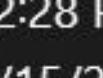
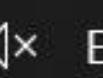
chair with three legs

[Generate](#)[Report issue](#)

Type here to search



68°F Sunny

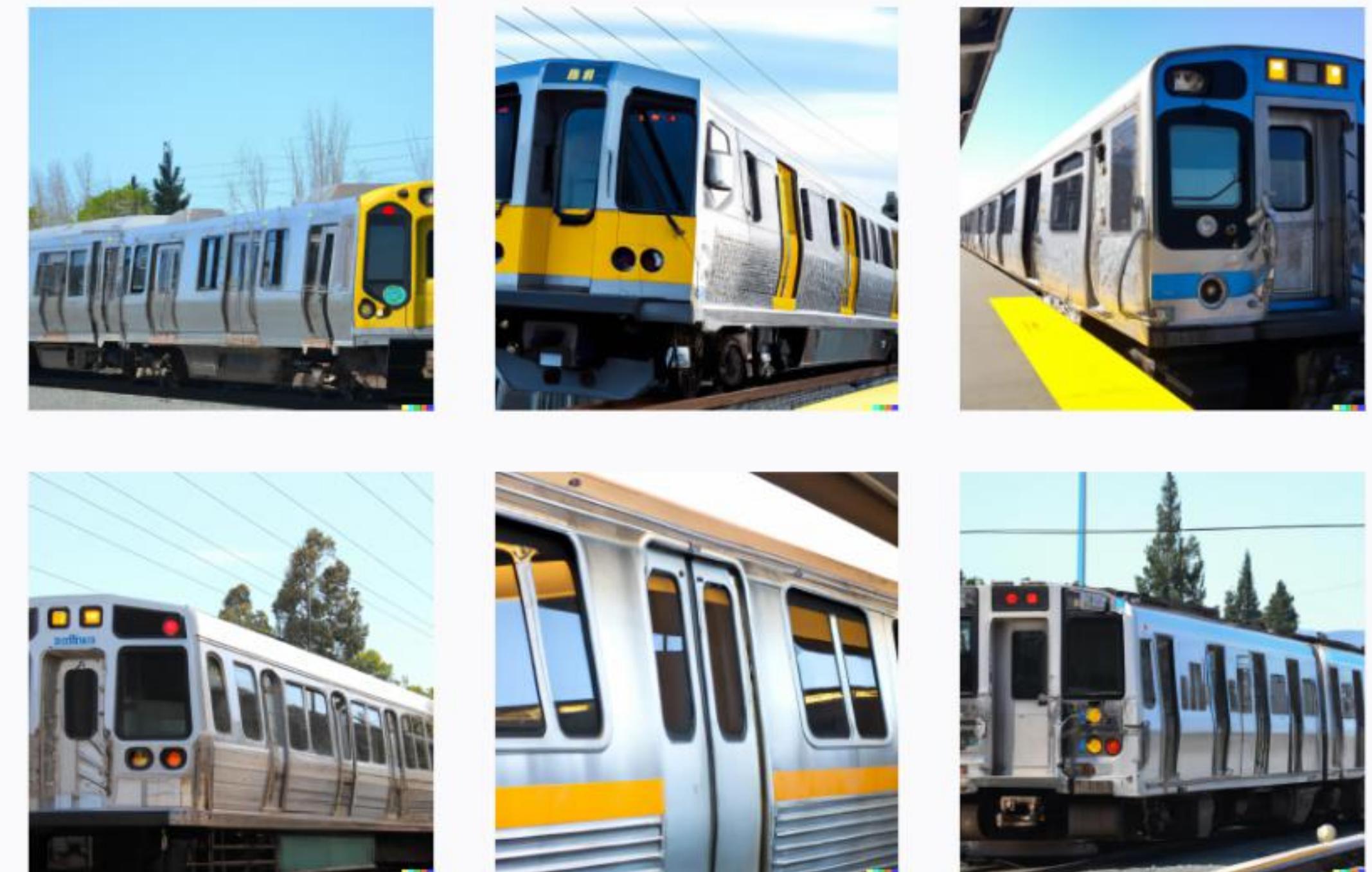
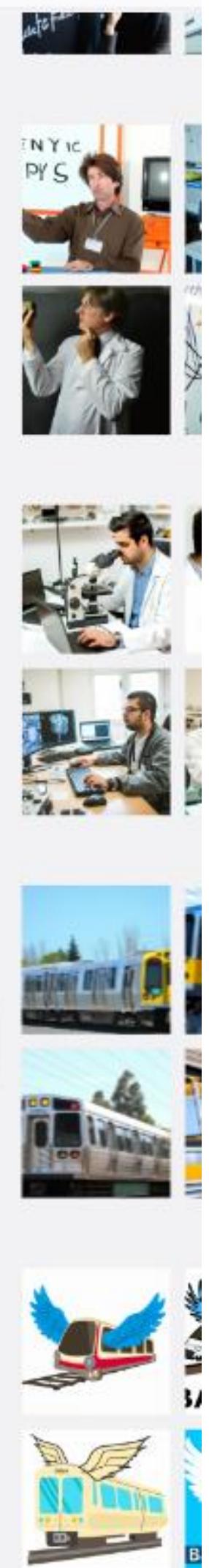
2:28 PM
7/15/2022

<https://labs.openai.com/e/fH5w9uO6p8kuiljN2C7R9HG>[Research](#) [Paris Stuff](#) [teaching](#) [Premium Fare Deals...](#) [Кино онлайн бесп...](#) [PI Portfolio](#) [Inbox \(14,384\) - aae...](#) [Richard bluejeans](#) [Zoom Efros](#) [»](#) [Other bookmarks](#)

Edit the detailed description

[Surprise me](#) [Upload](#)

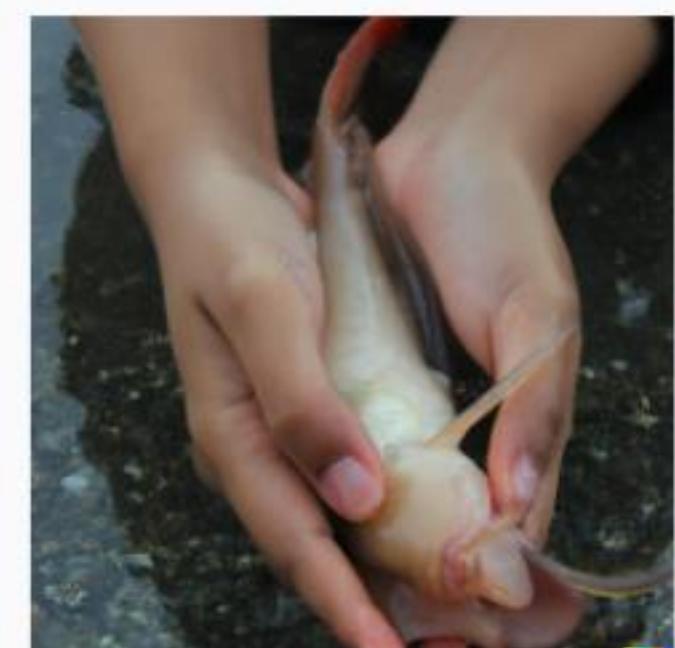
photo of bart train with wings

[Generate](#)[Report issue](#)

Edit the detailed description

[Surprise me](#)[Upload](#)

fish that has human hands

[Generate](#)[Report issue](#)

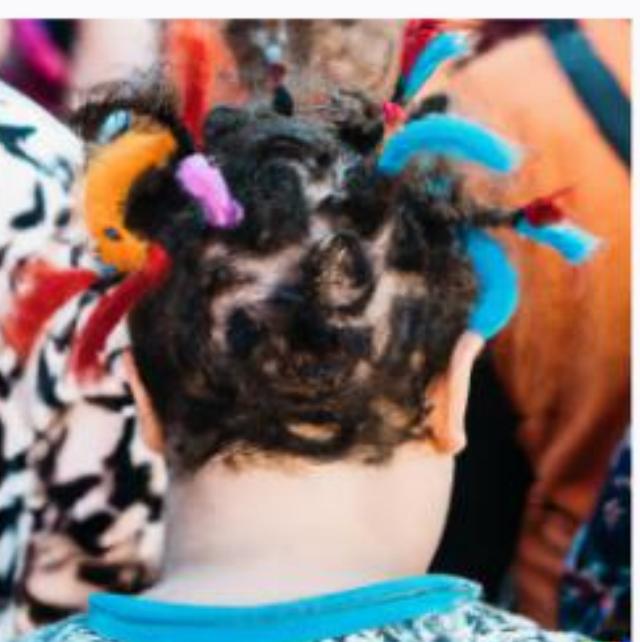
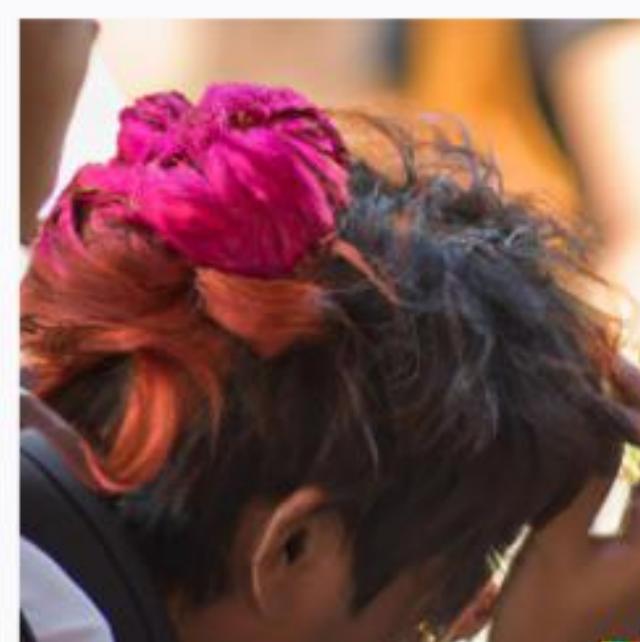
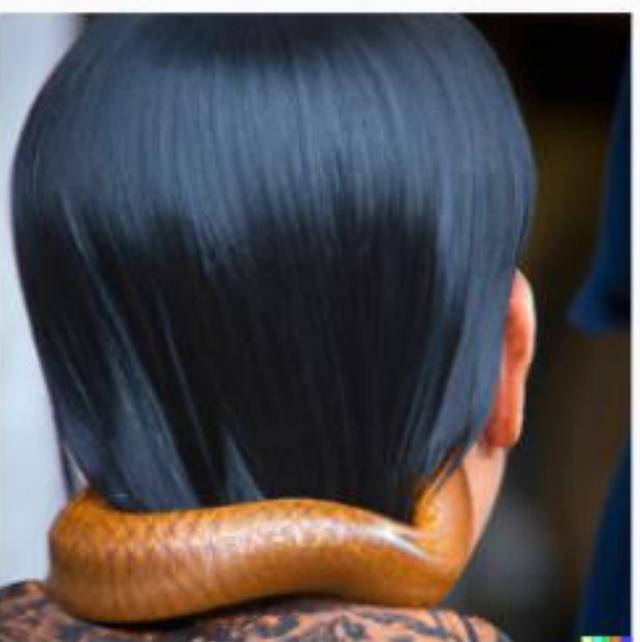
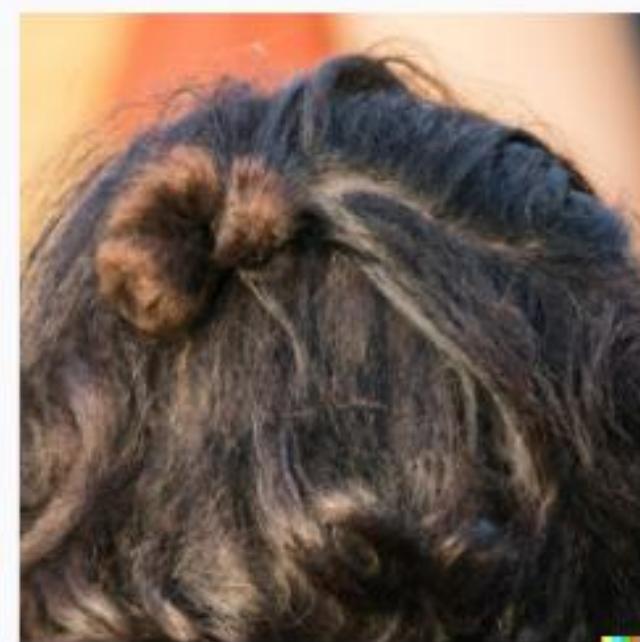
Edit the detailed description

person with hair of snakes

Surprise me

Upload →

Generate



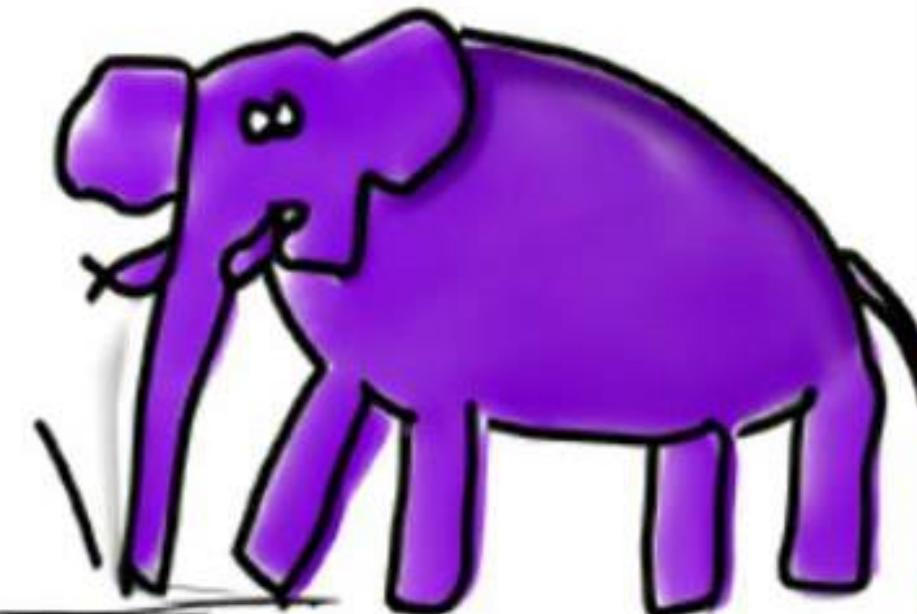
Report issue

whatever you do, don't draw a purple elephant



Report issue

DO DELLENT WHAT



BE VONT DEAT WAT

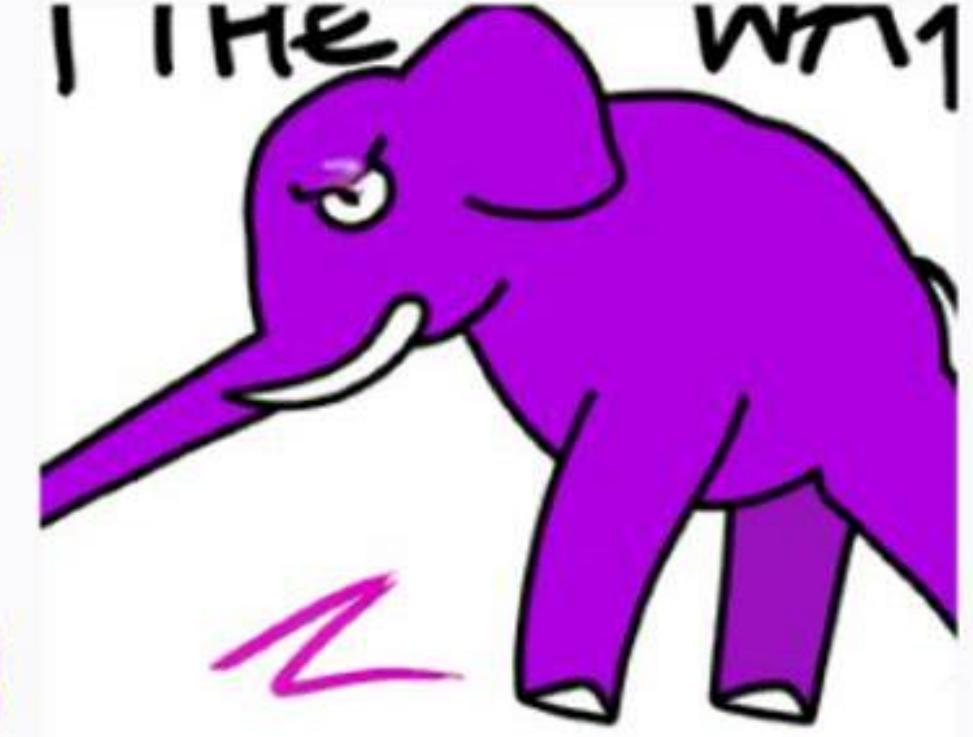
VAN' YOU



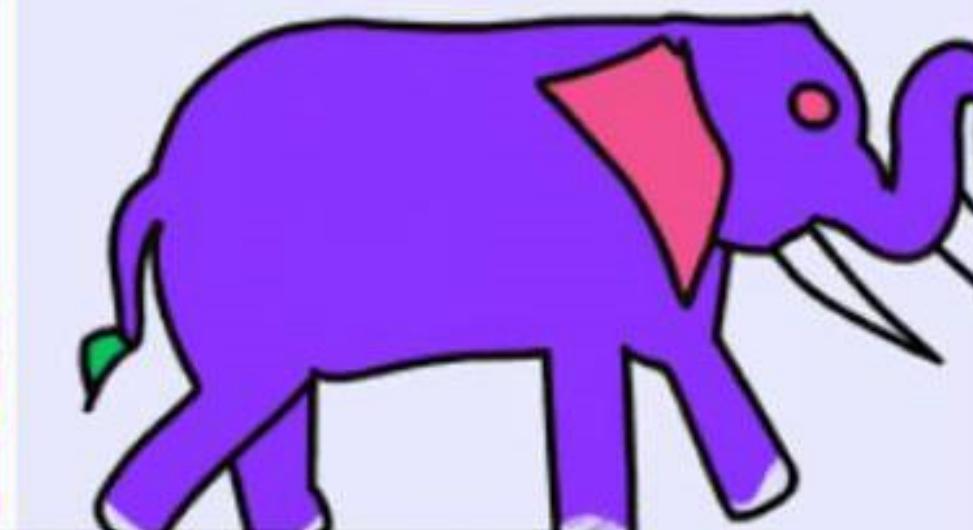
DON'T ANT WHAR
WHAT DET WAT?

DOI WH'T WAT
DENT DE WHAT
DAUL....

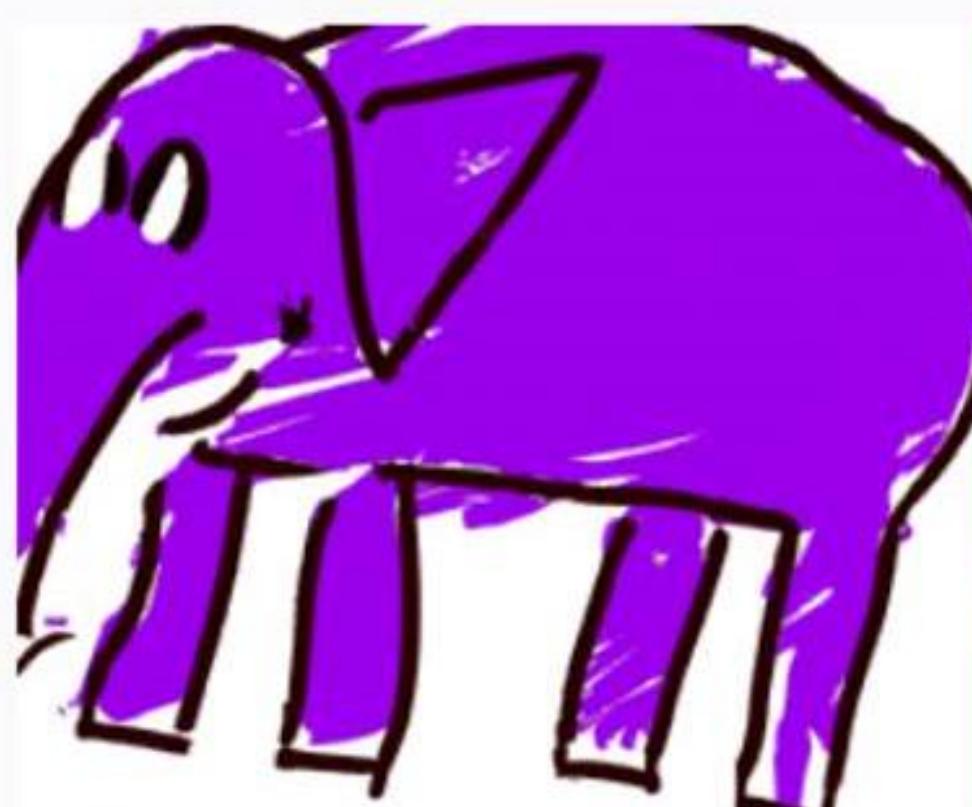
I THE WAT



DEWANTREA' N'T DEF DAHANT'S WAT



NHAT WT WAT DIOU



DET' WNT DON'T
DANT
DEFUEPWART!

