媒体智能实验室
Media Intelligence LAB

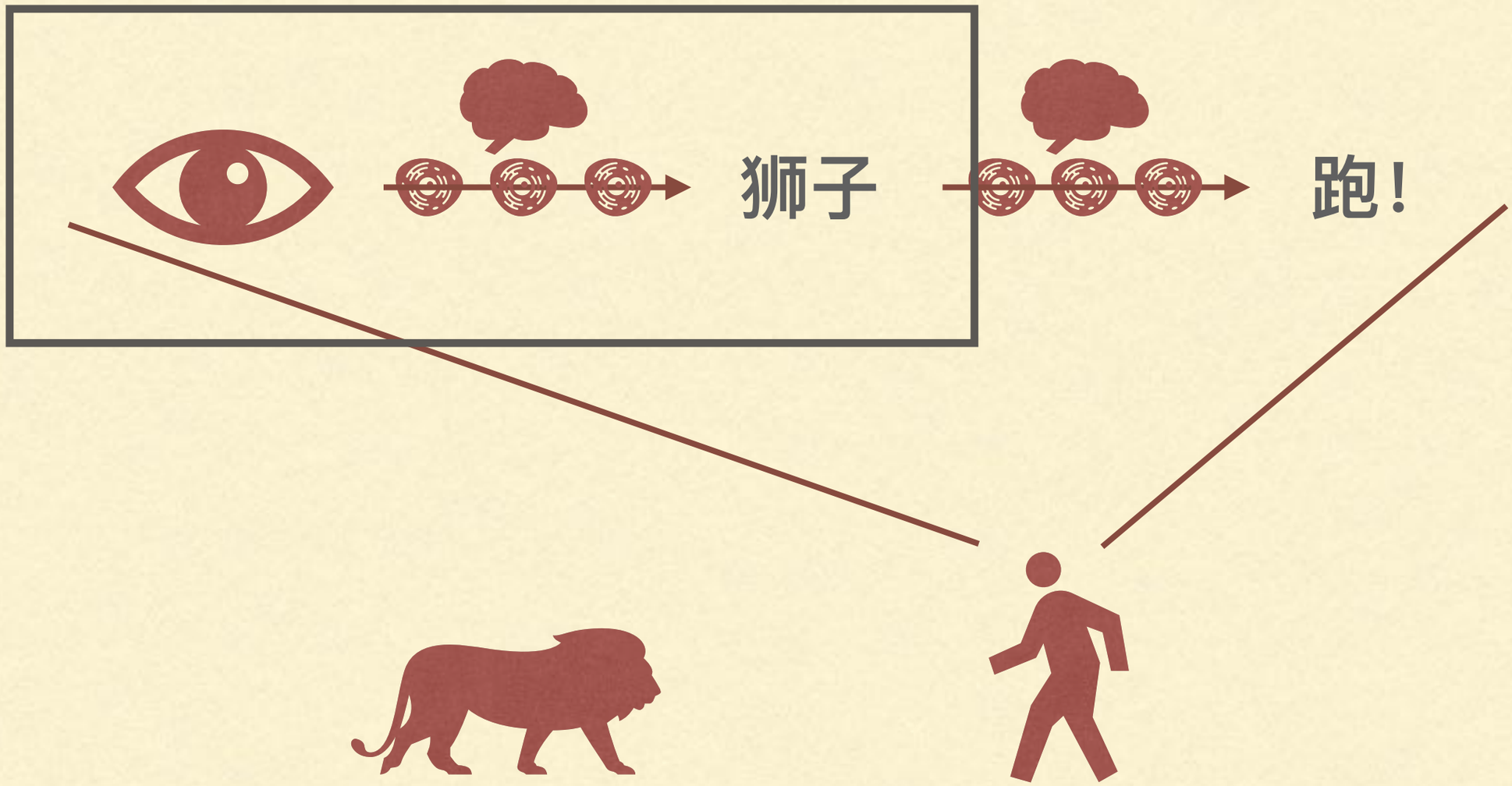
卷积神经网络
从相识、相知到相爱

罗宇矗

2017 年 11 月 10 日

Intelligent Agent Group

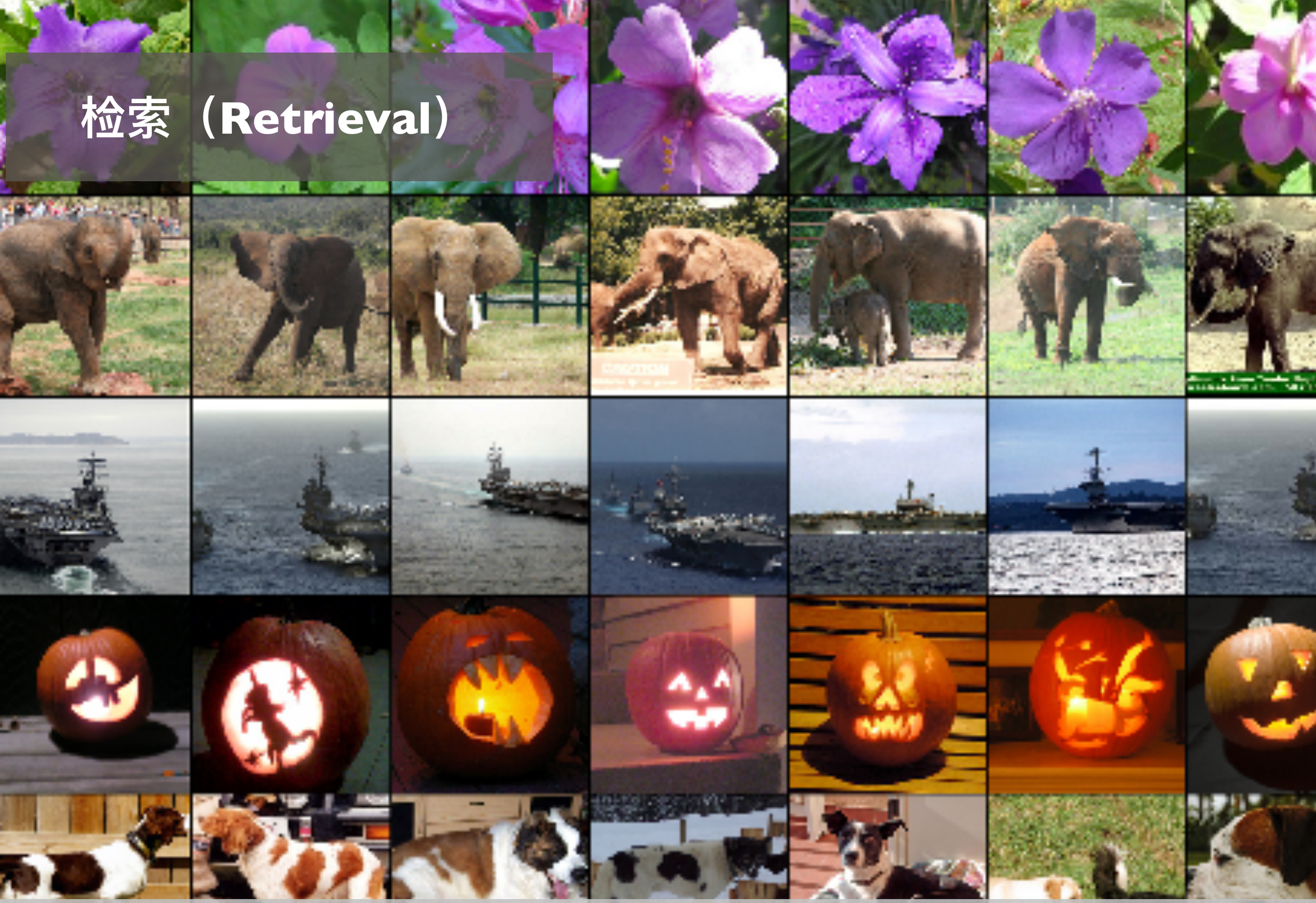
如何根据视觉信息进行识别?



分类（Classification）

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017.
ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (May 2017)

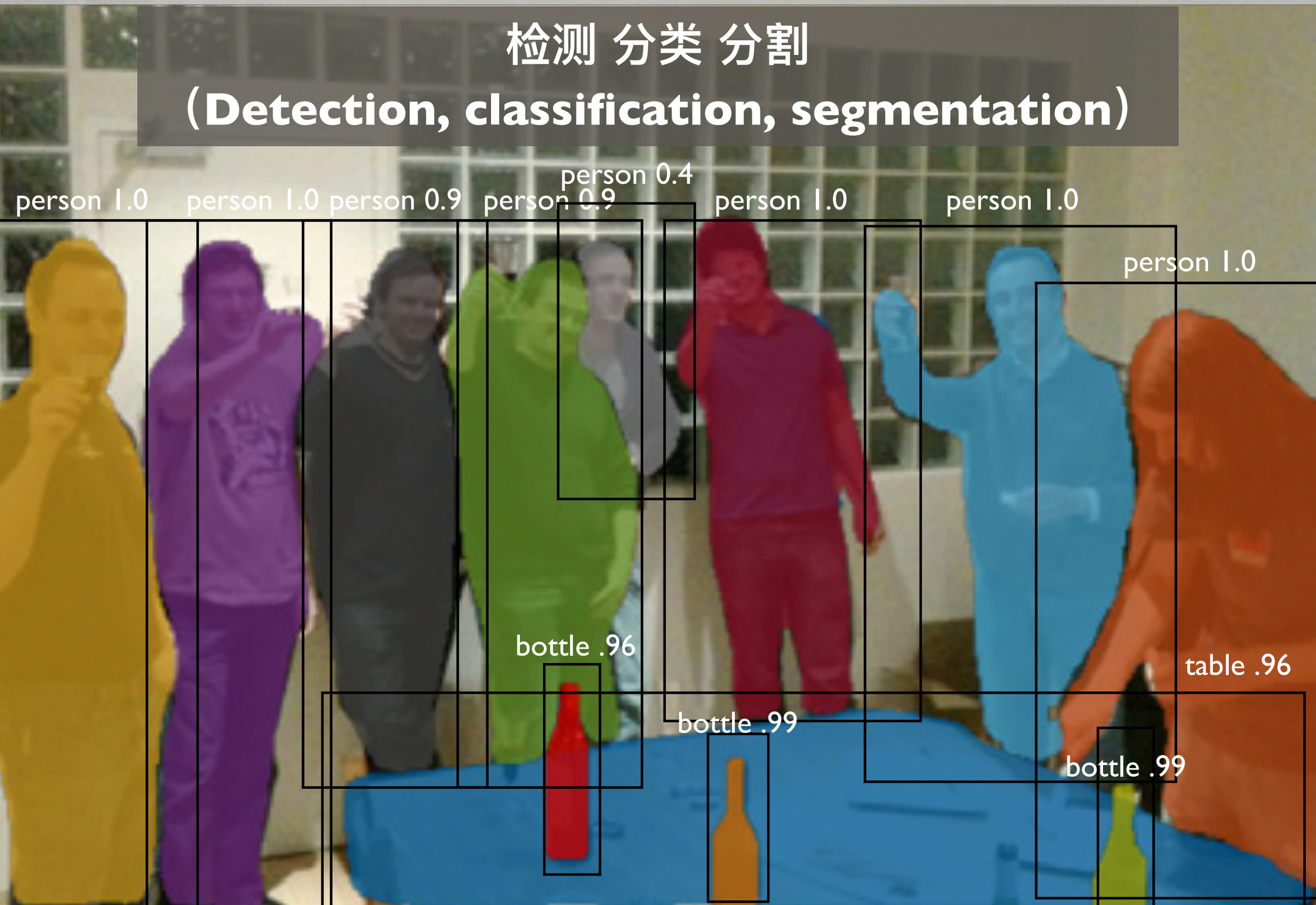
検索（Retrieval）



Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017.
ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (May 2017)

Kaiming He, Georgia Gkioxari, Piotr Dolla´r and Ross Girshick
(2017). Mask R-CNN. *CoRR, abs/1703.06870*, .

检测 分类 分割
（Detection, classification, segmentation）



检测 分类 分割
（Detection, classification, segmentation）

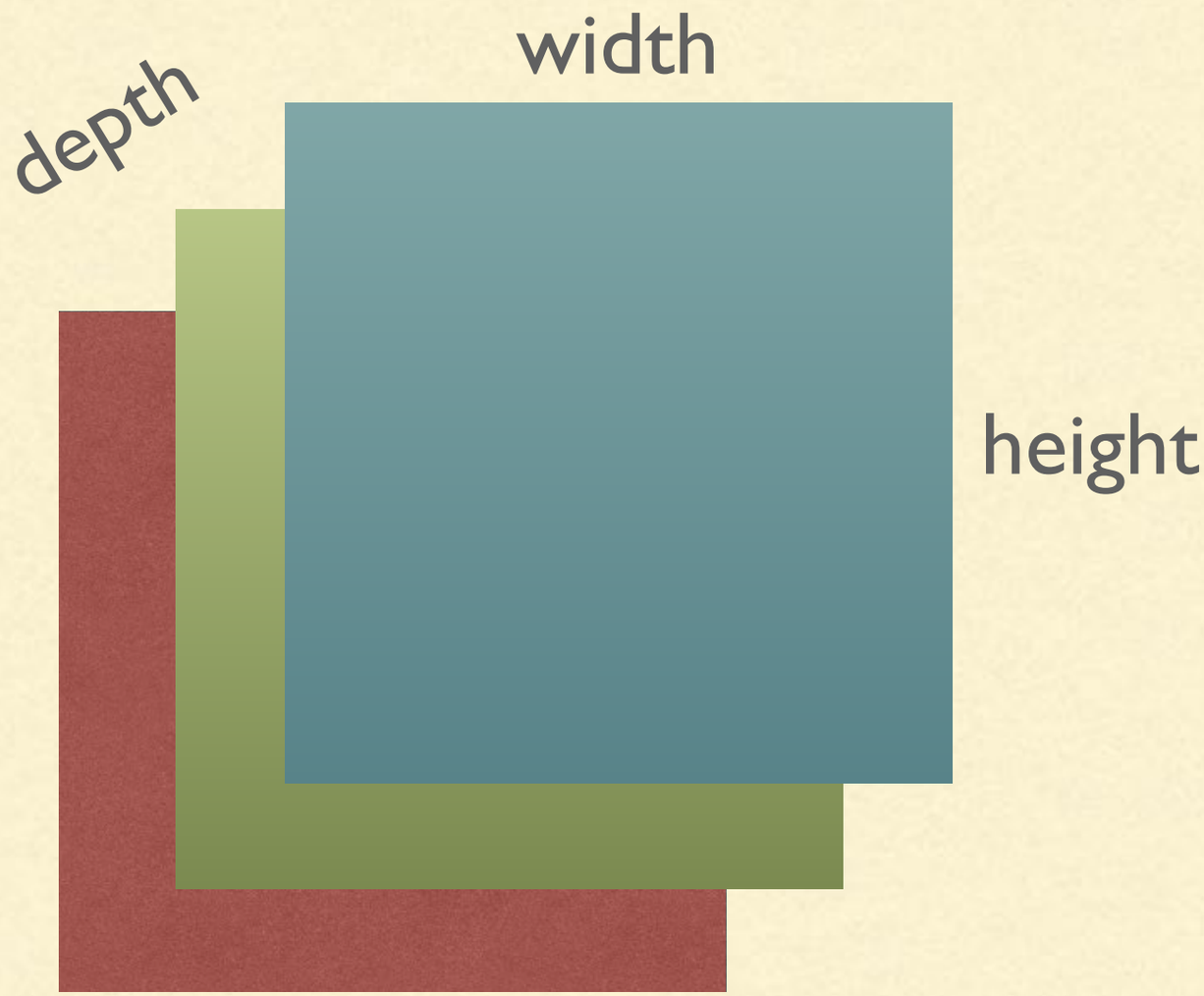
Kaiming He, Georgia Gkioxari, Piotr Dolla´r and Ross Girshick
(2017). Mask R-CNN. *CoRR, abs/1703.06870*, .

Kaiming He, Georgia Gkioxari, Piotr Dolla´r and Ross Girshick (2017). Mask R-CNN. *CoRR, abs/1703.06870*, .

风格迁移（Style transfer）

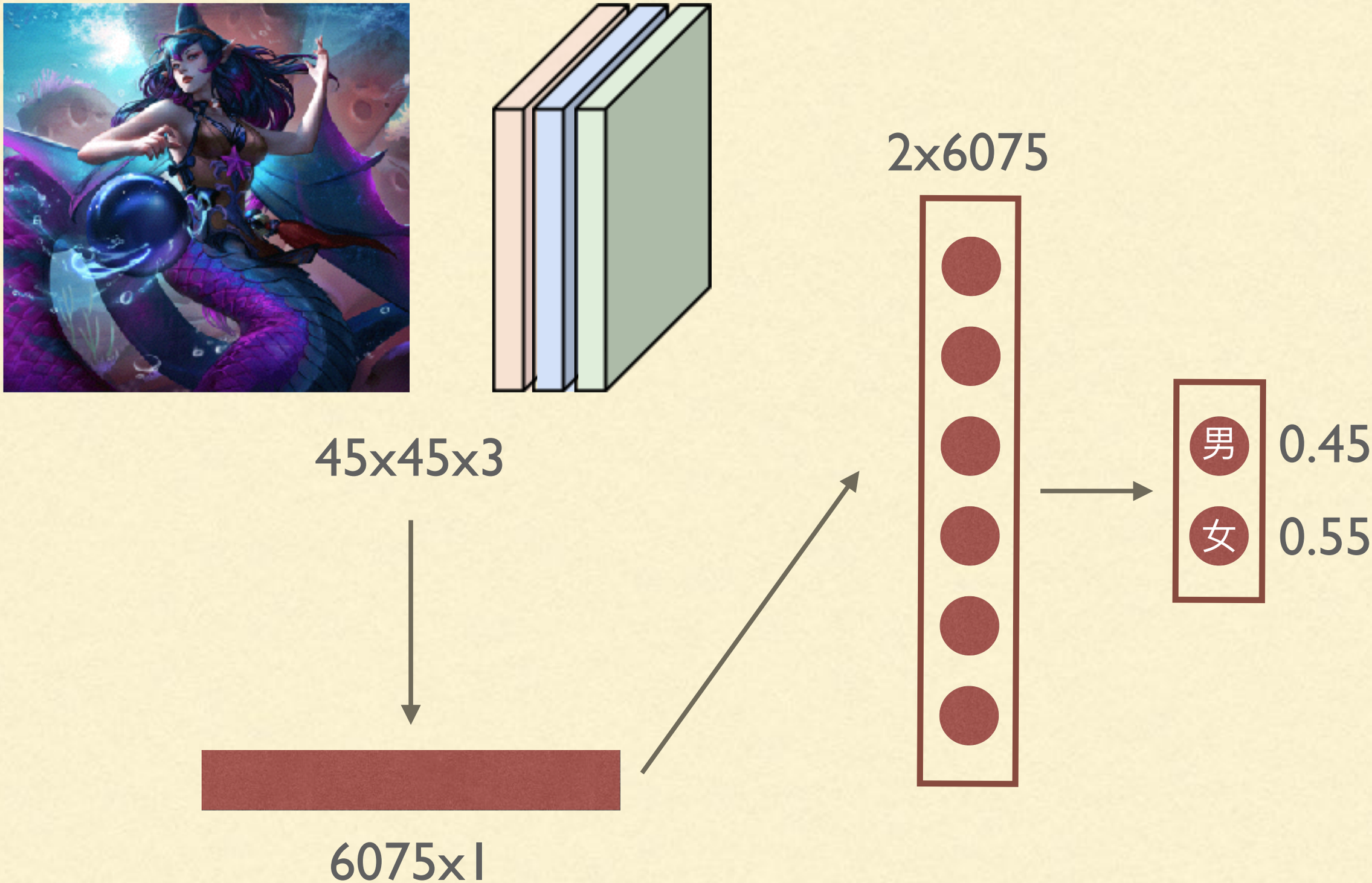
Dongdong Chen and (2017).
StyleBank: An Explicit Representation for Neural Image Style Transfer. *CoRR, abs/1703.09210,* .

图像表达



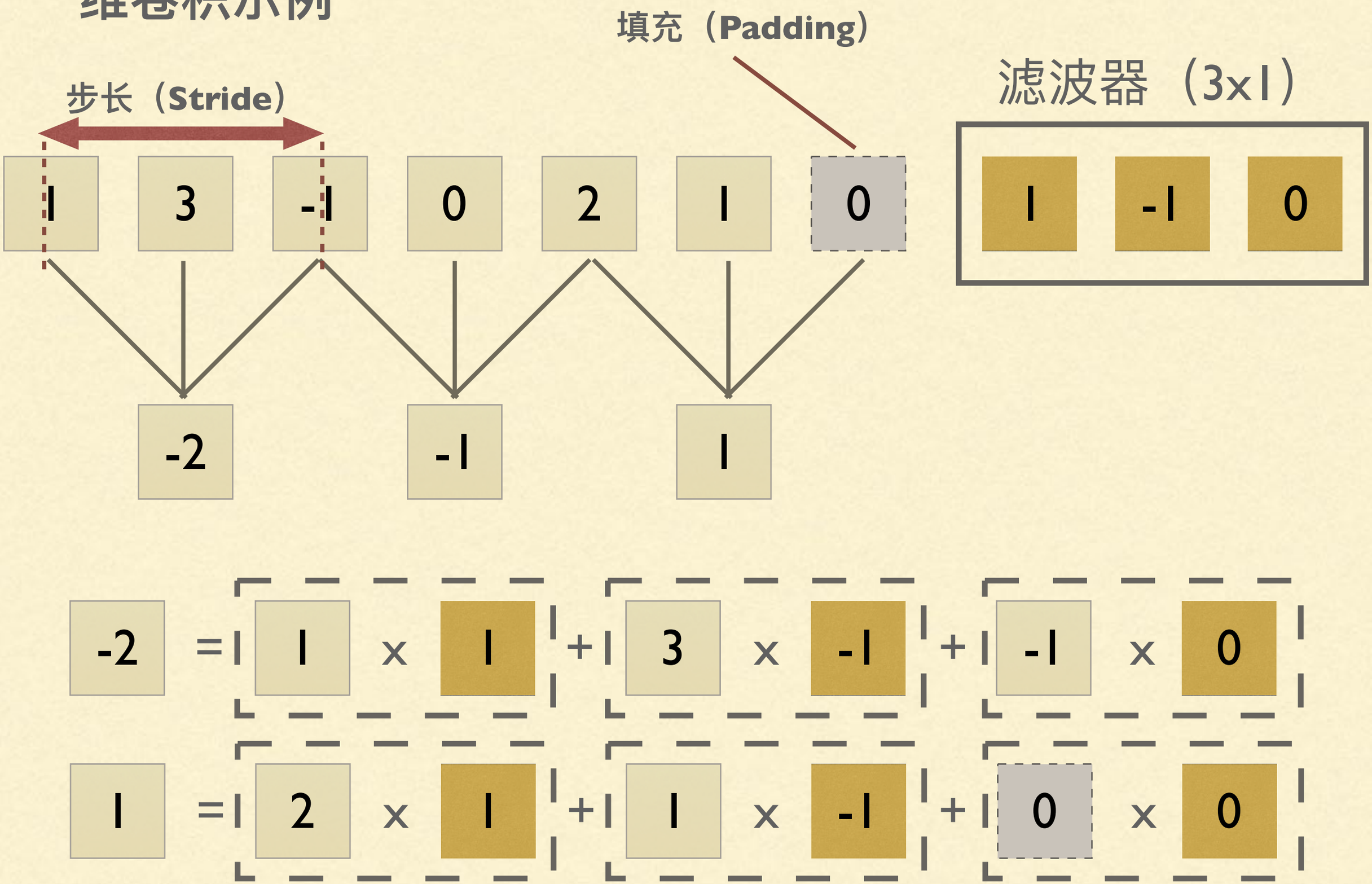
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	12	0	11	39	137	37	0	152	147	84	0	0	0
0	0	1	0	0	0	41	160	250	255	235	162	255	238	206	11	13	0
0	0	0	16	9	9	150	251	45	21	184	159	154	255	233	40	0	0
10	0	0	0	0	0	145	146	3	10	0	11	124	253	255	107	0	0
0	0	3	0	4	15	236	216	0	0	38	109	247	240	169	0	11	0
1	0	2	0	0	0	253	253	23	62	224	241	255	164	0	5	0	0
6	0	0	4	0	3	252	250	228	255	255	234	112	28	0	2	17	0
0	2	1	4	0	21	255	253	251	255	172	31	8	0	1	0	0	0
0	0	4	0	163	225	251	255	229	120	0	0	0	0	0	11	0	0
0	0	21	162	255	255	254	255	126	6	0	10	14	6	0	0	9	0
3	79	242	255	141	66	255	245	189	7	8	0	0	5	0	0	0	0
26	221	237	98	0	67	251	255	144	0	8	0	0	7	0	0	11	0
125	255	141	0	87	244	255	208	3	0	0	13	0	1	0	1	0	0
145	248	228	116	235	255	141	34	0	11	0	1	0	0	0	1	3	0
85	237	253	246	255	210	21	1	0	1	0	0	6	2	4	0	0	0
6	23	112	157	114	32	0	0	0	0	2	0	8	0	7	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

全连接层（Fully Connected Layer）

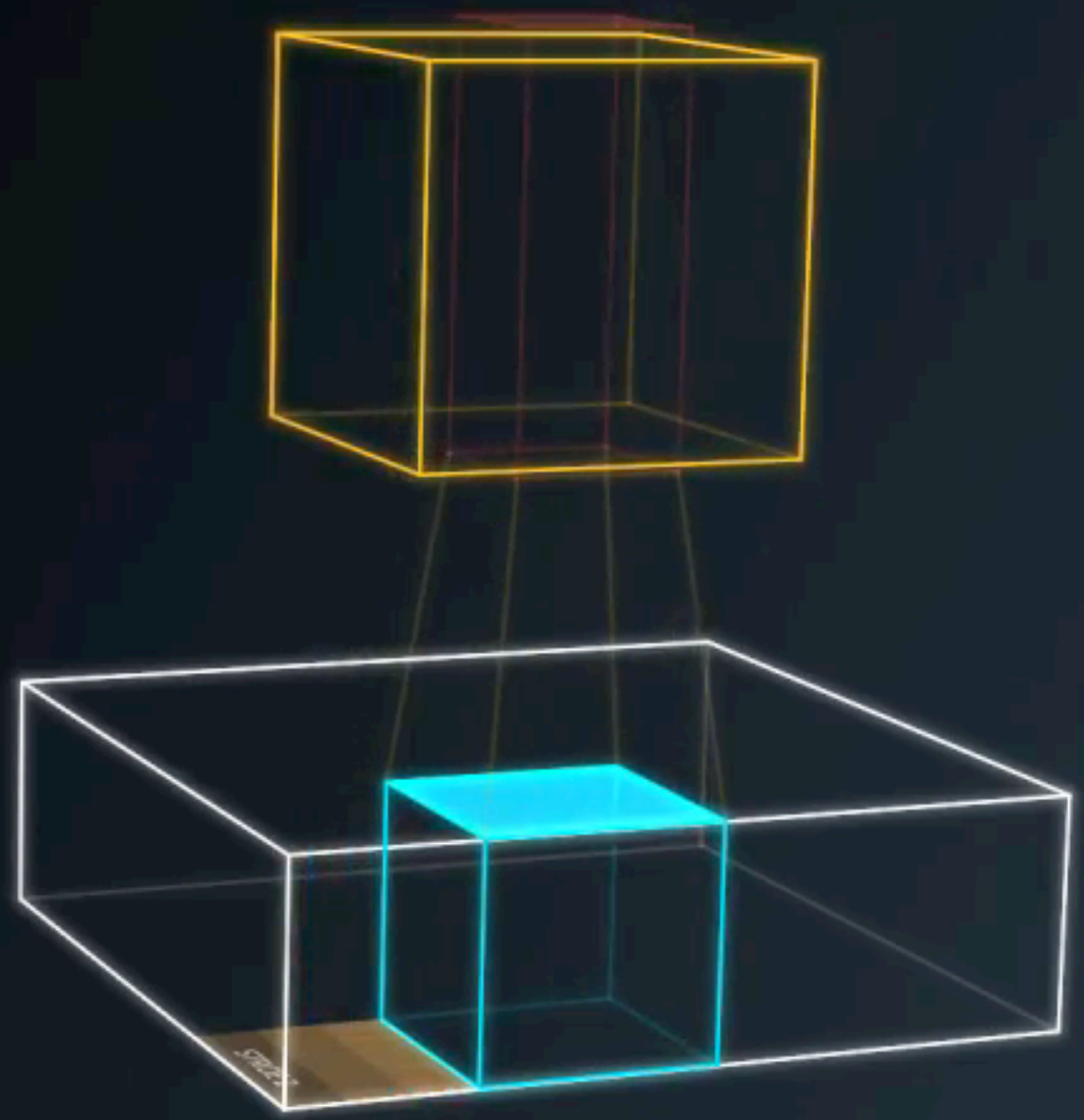


如何处理画面不变性和局部相关性?

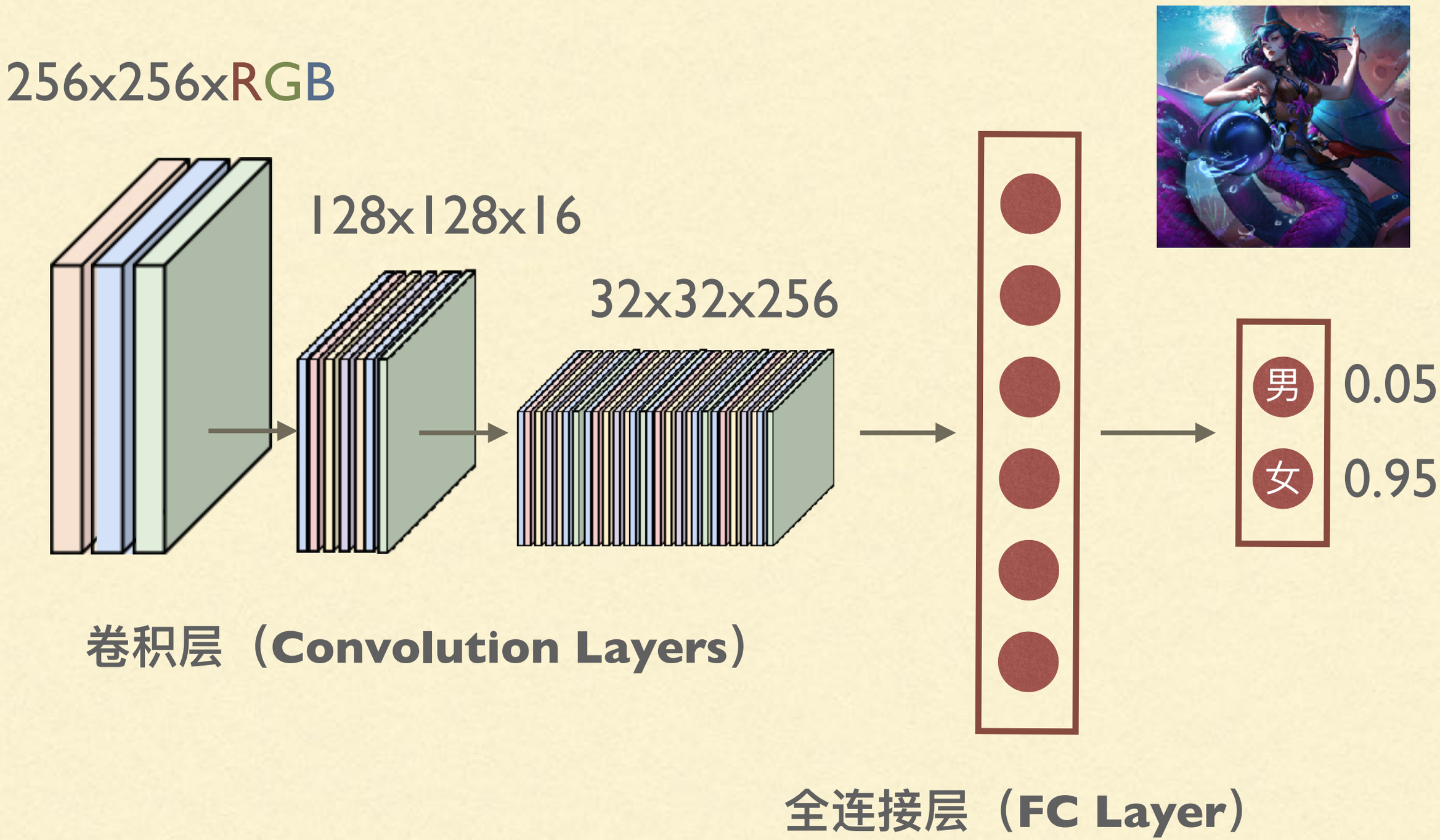
一维卷积示例

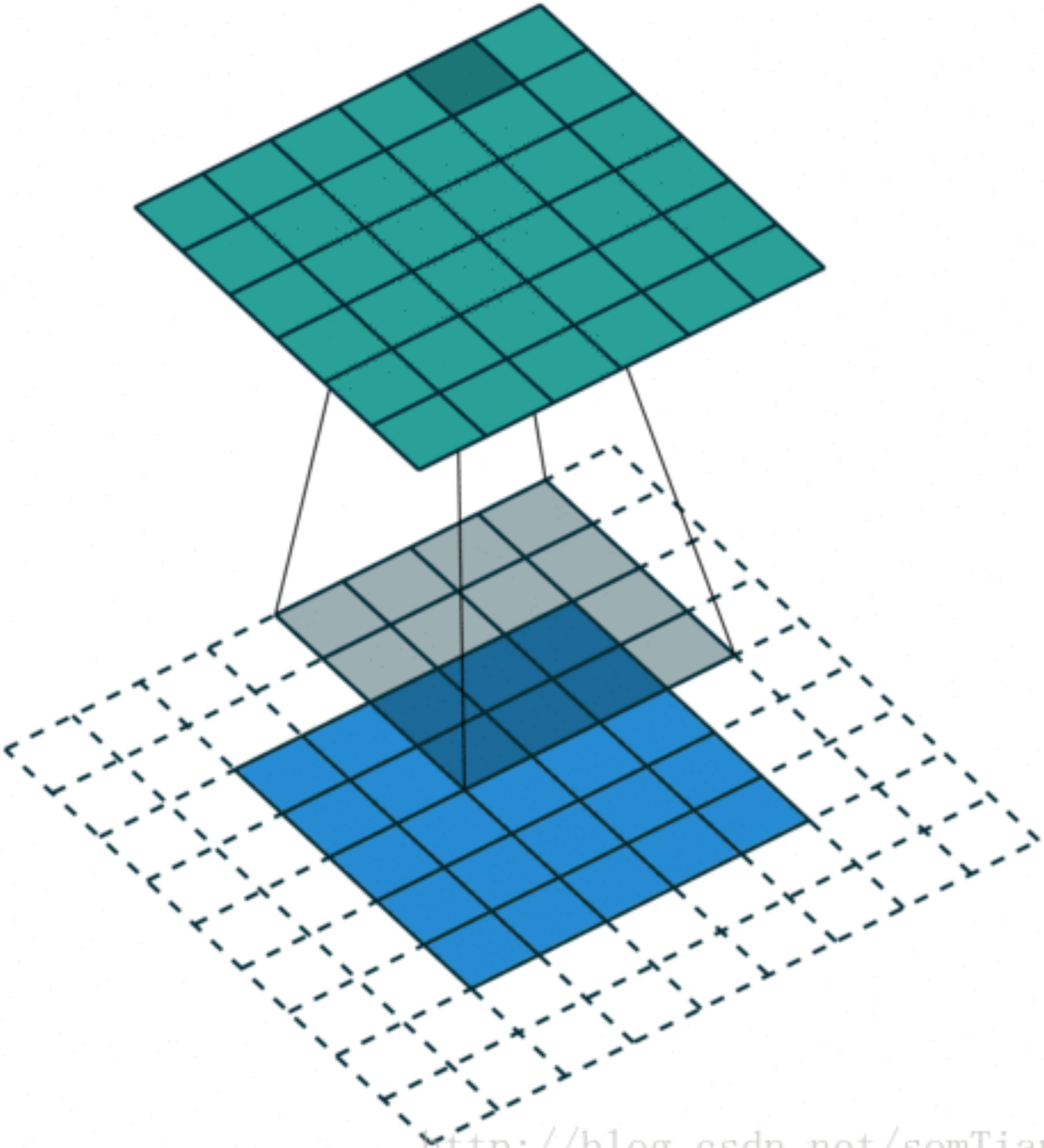

填充（Padding）
步长（Stride）
滤波器（3x1）
1 3 -1 0 2 1 0
1 -1 0
-2 -1 1
-2 = 1 x 1 + 3 x -1 + -1 x 0
1 = 2 x 1 + 1 x -1 + 0 x 0

VIDEO

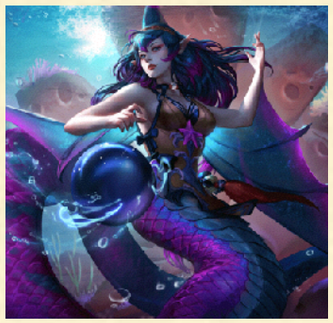


使用卷积网络的图像识别

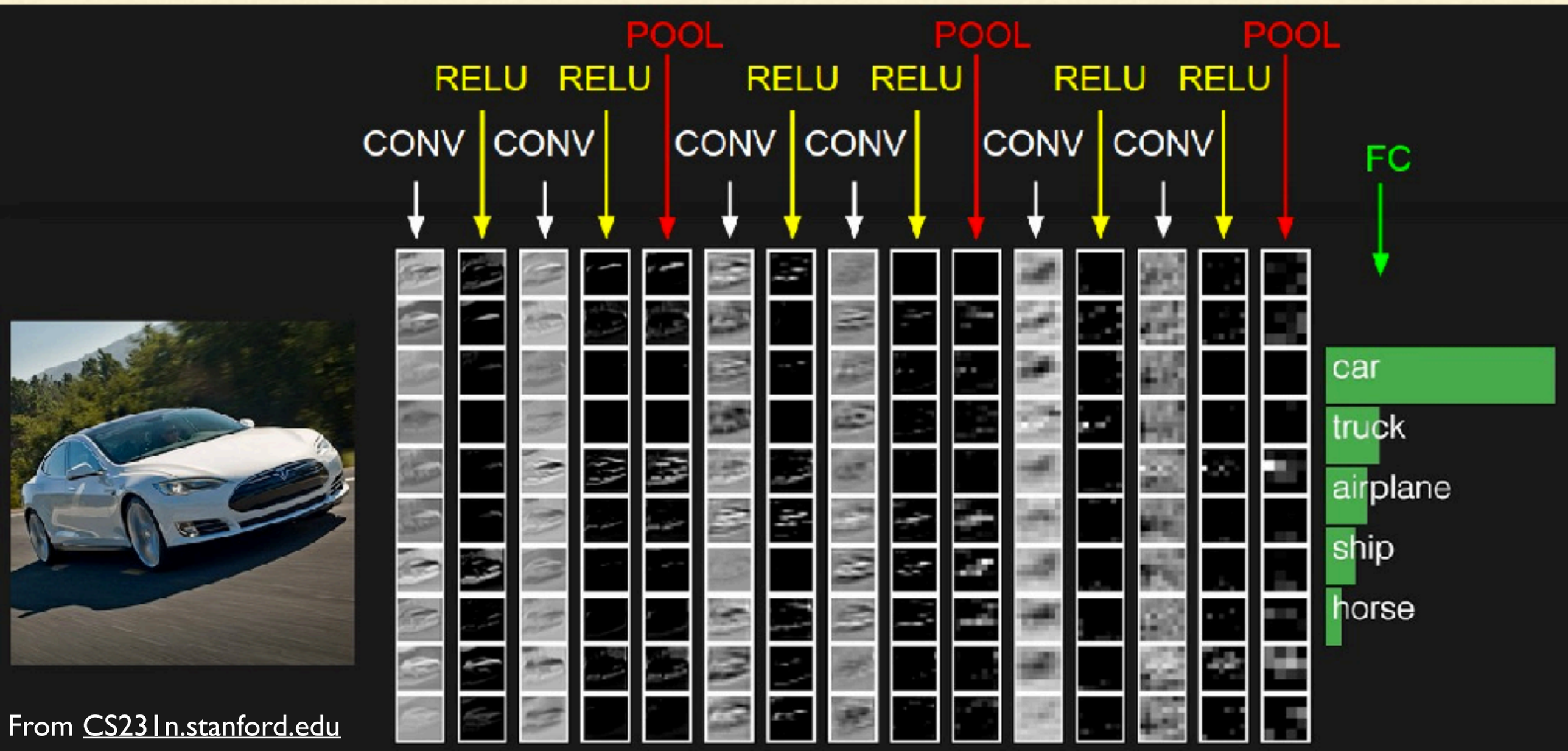



http://blog.csdn.net/somTian

为什么卷积是有效的?



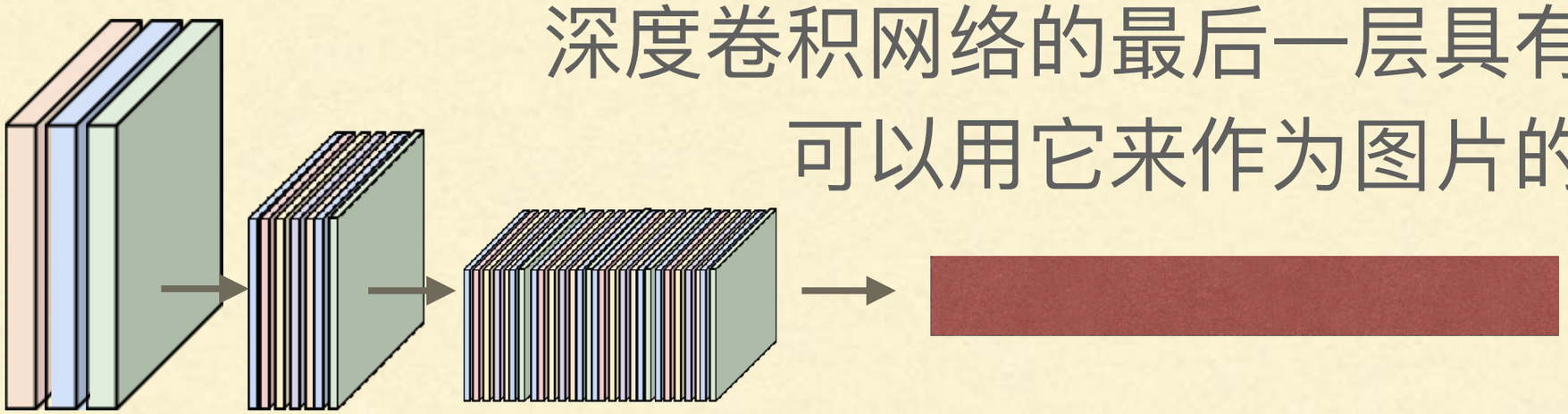
→ 低级图像特征 → 中级图像特征 → 复杂语义特征



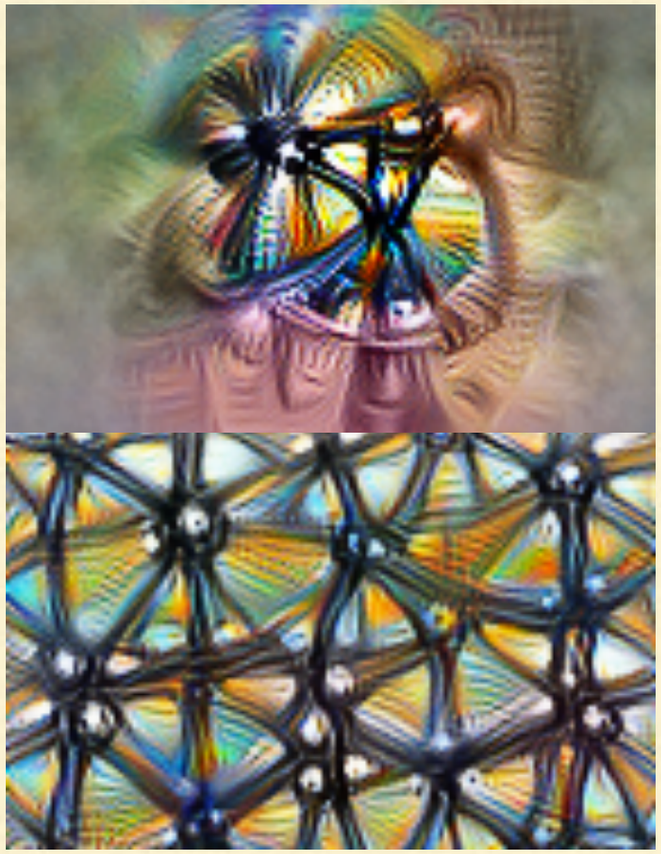
From CS231n.stanford.edu

深度卷积网络的最后一层具有丰富的语义信息

可以用它来作为图片的一种表达

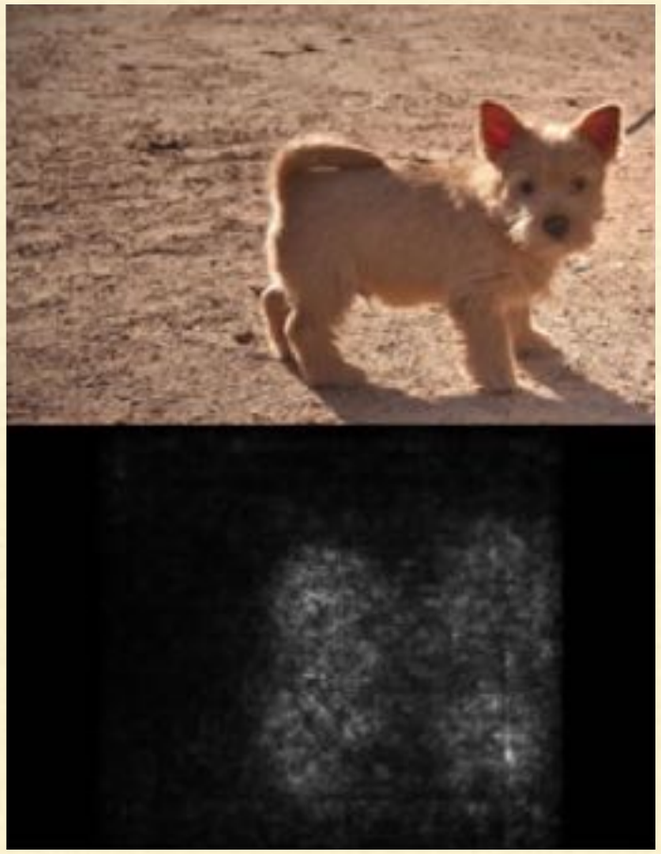


可视化卷积网络的两种方式



Feature visualization

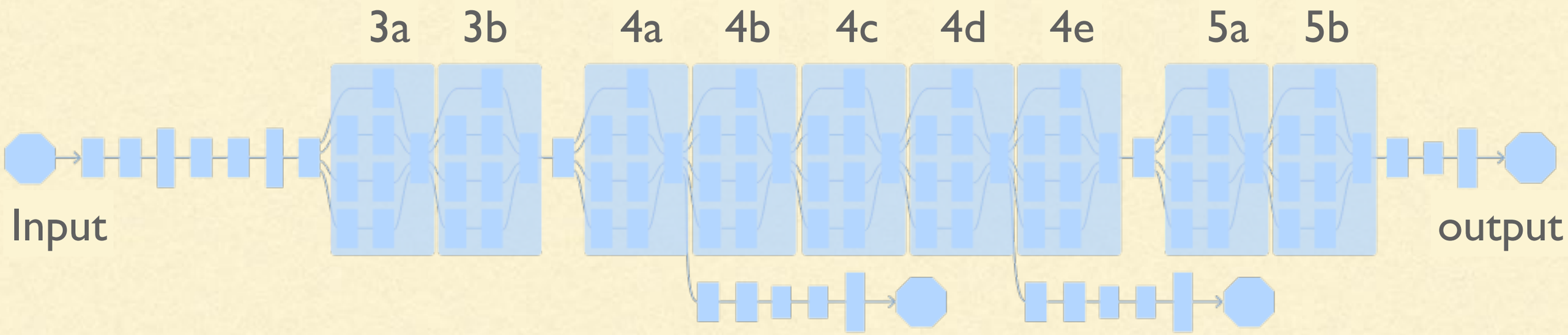
通过生成样本
确定神经网络对什么产生响应

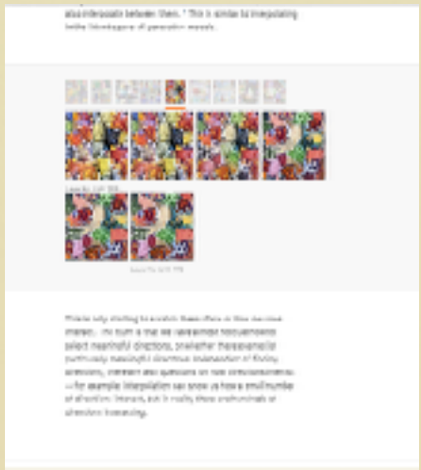


Attribution

研究样本中的哪一部分
会令神经网络产生响应

GoogLeNet 可视化
Feature visualization

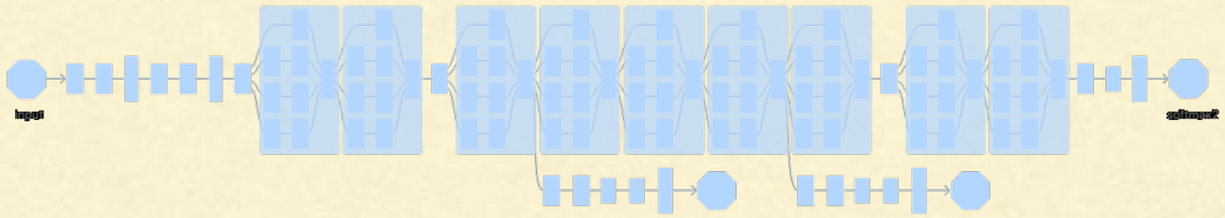


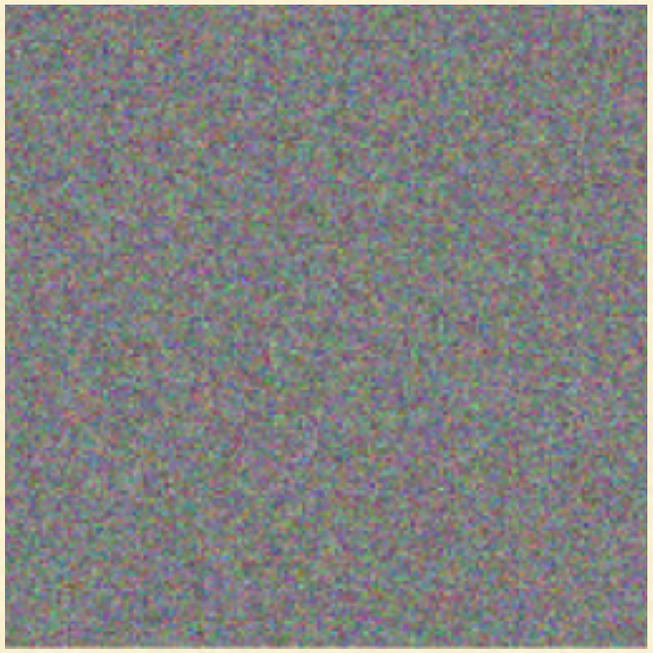


演示以及论文地址： https://**distill.pub**/2017/feature-visualization/

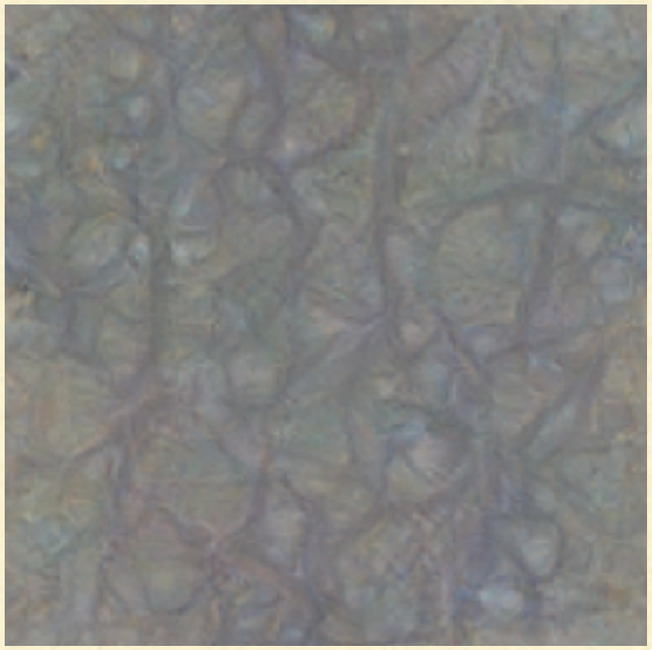
Olah, et al., "Feature Visualization", Distill, 2017.

GoogLeNet 可视化
Feature visualization

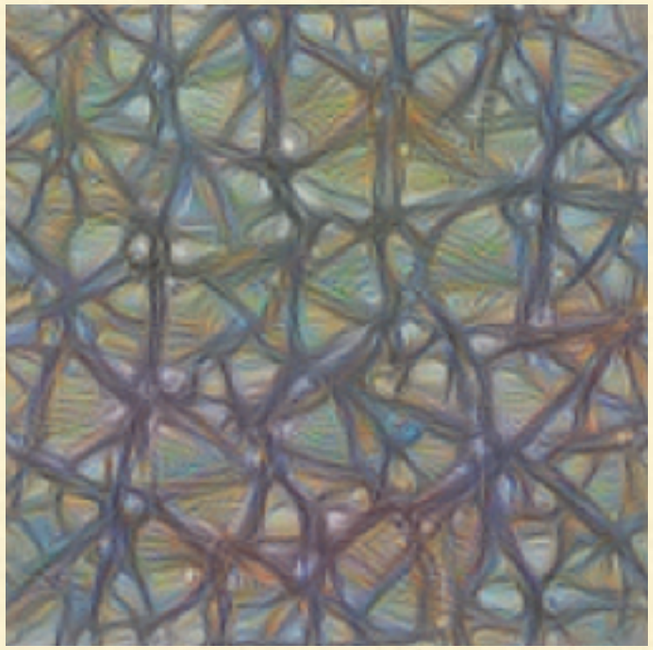




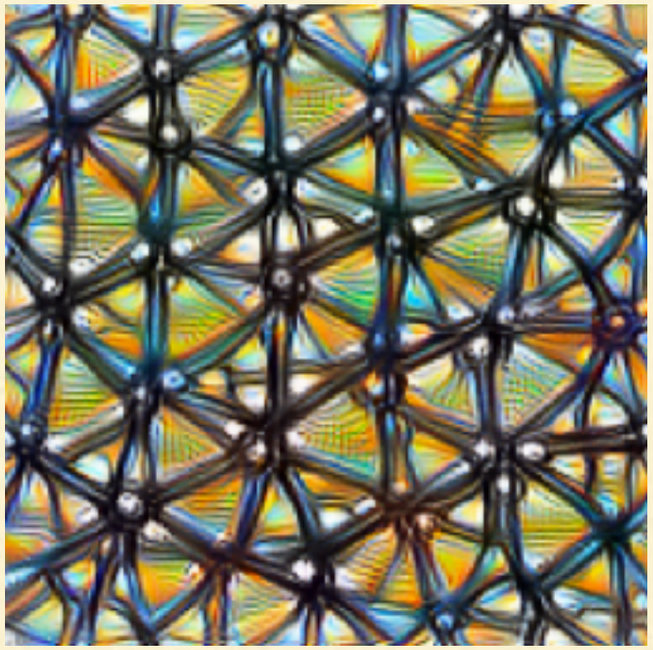
Step 0



Step 4



Step 48

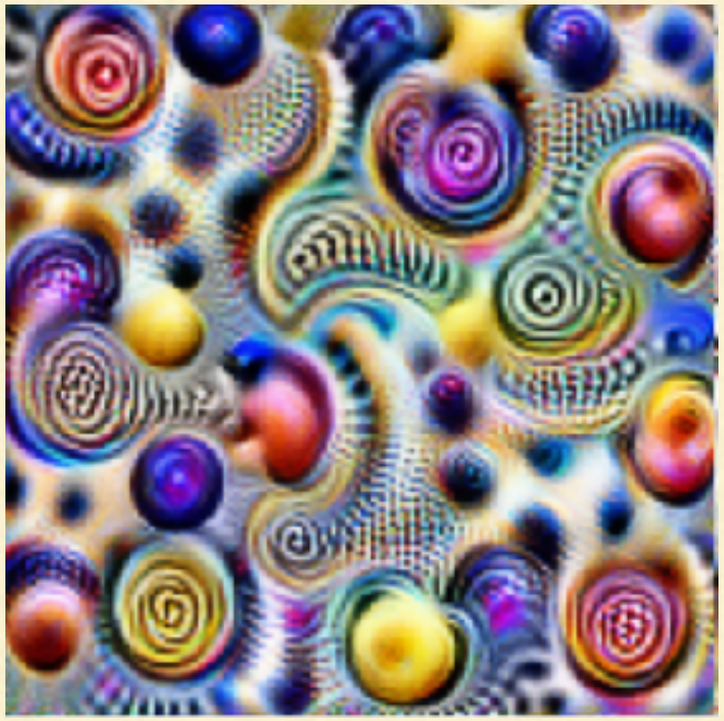


Step 2048

从随机 noise 开始，
优化 (optimize) 图片以激活特定的神经元

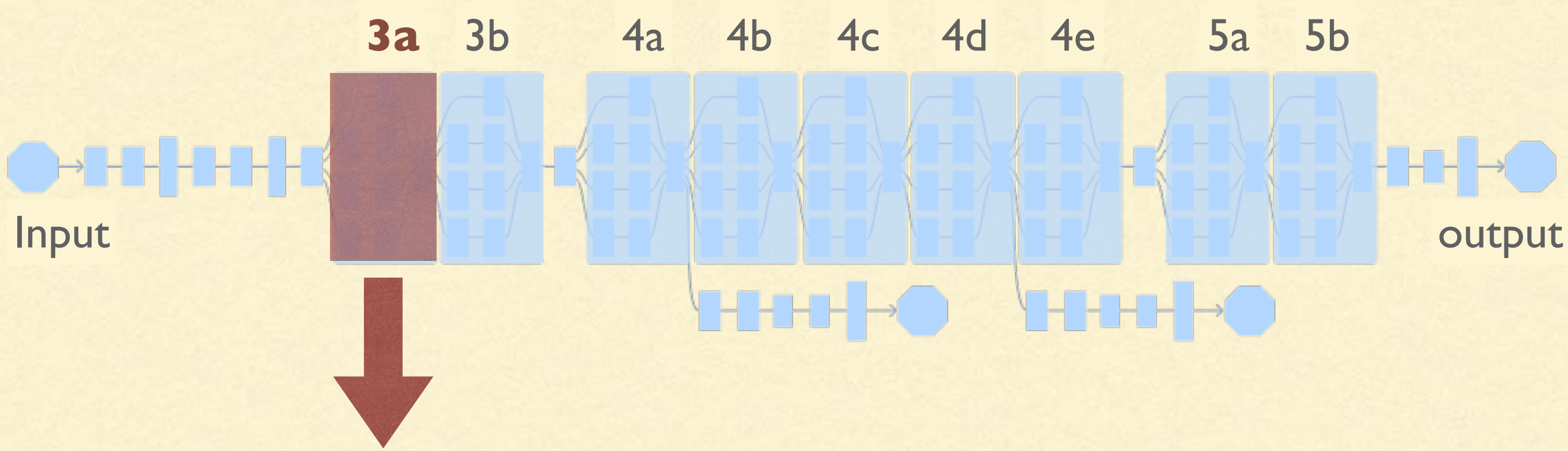
Olah, et al., "Feature Visualization", Distill, 2017.

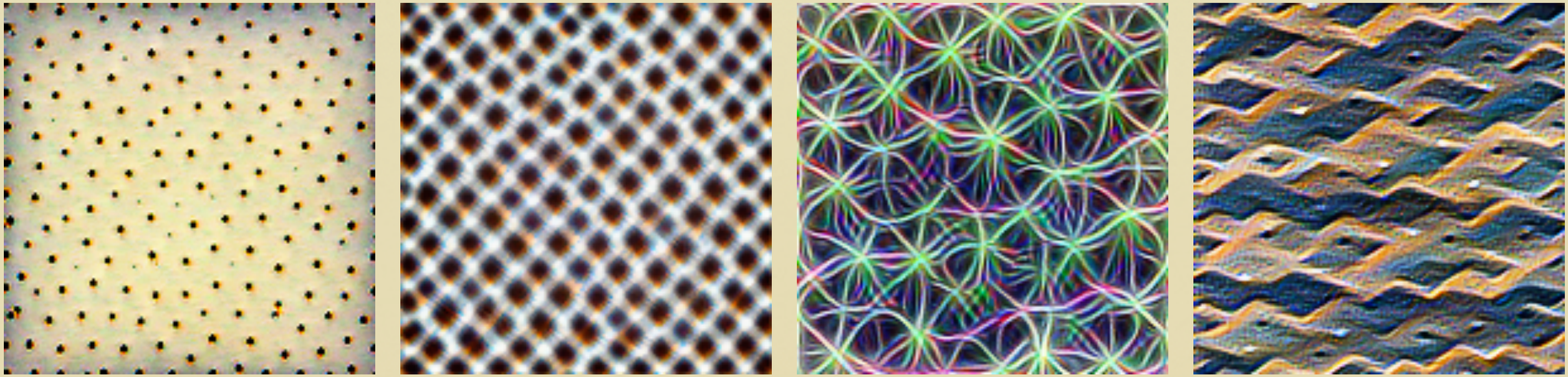
可视化效果



激活正样本

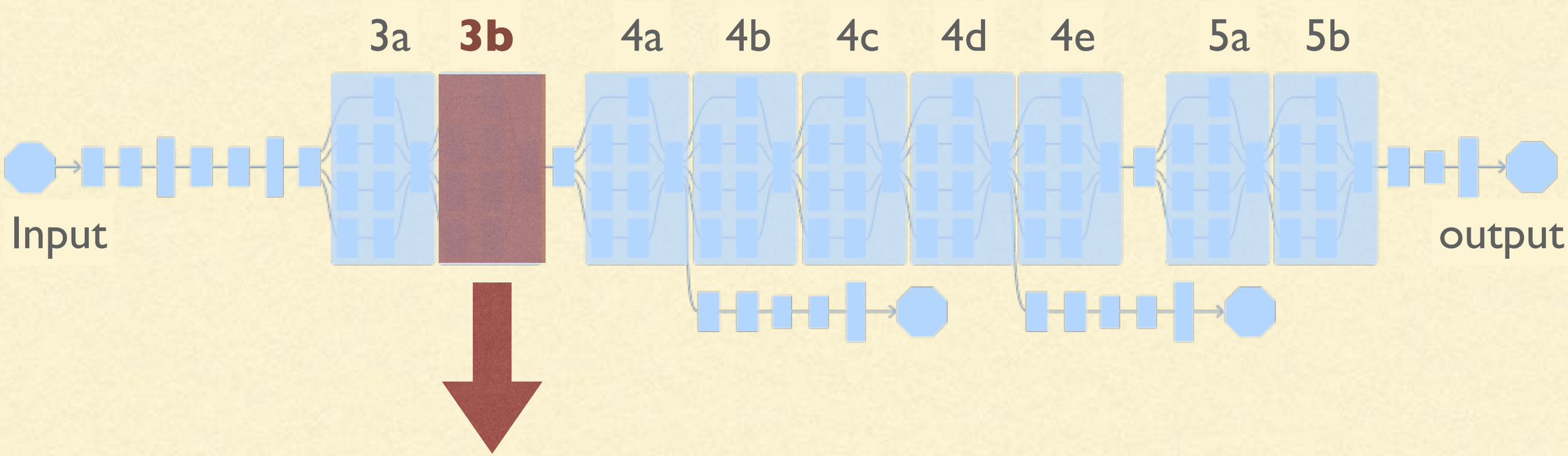
Olah, et al., "Feature Visualization", Distill, 2017.

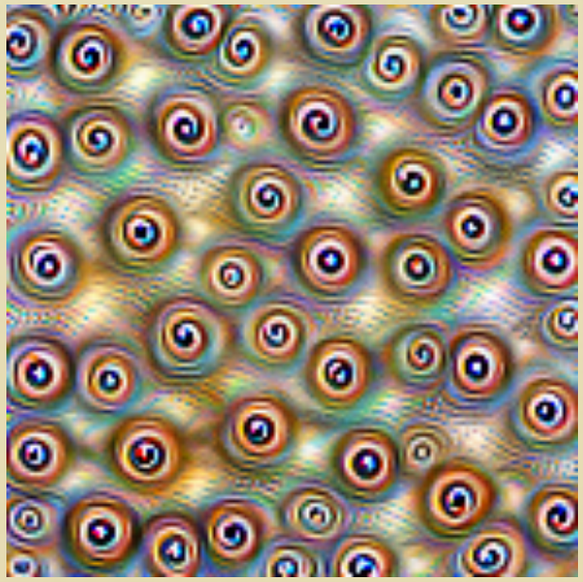
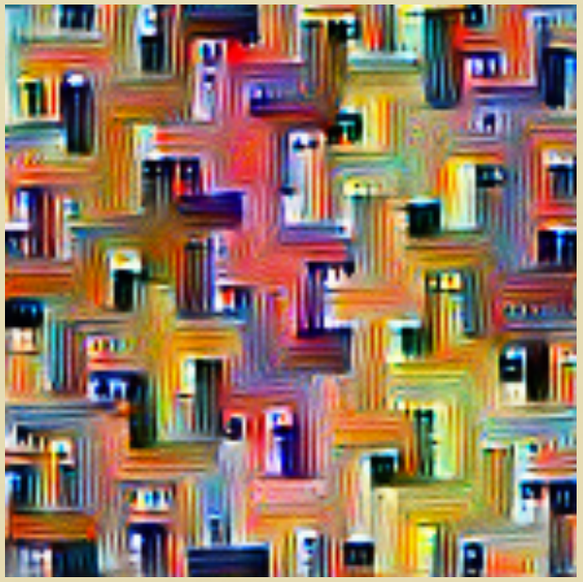
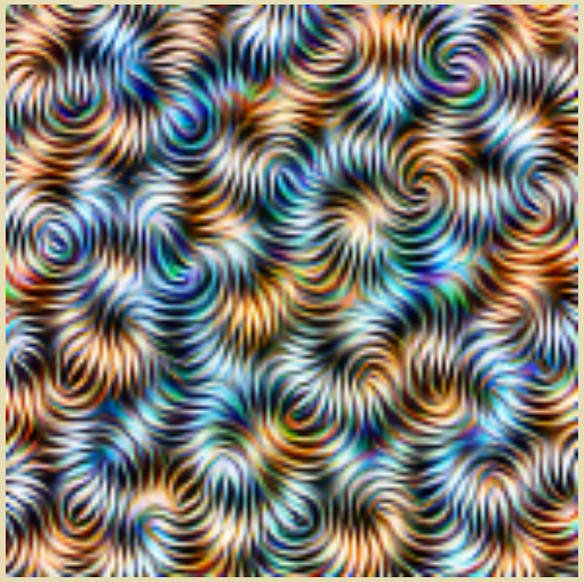




检测出了简单的纹理信息（texture），且都是局部（local）纹理

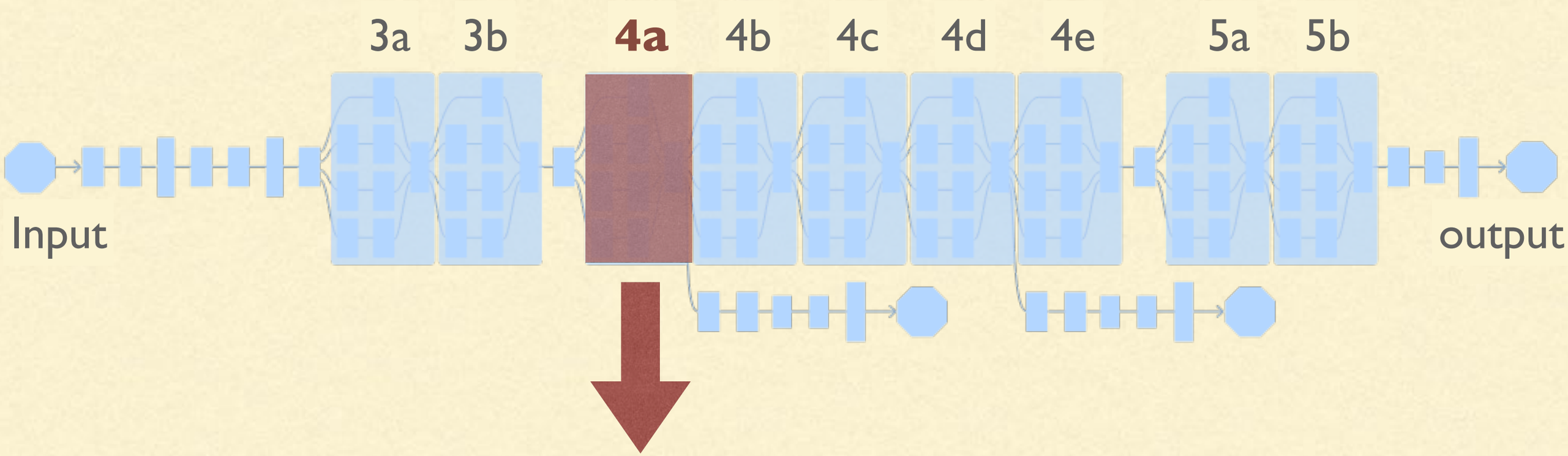
Olah, et al., "Feature Visualization", Distill, 2017.



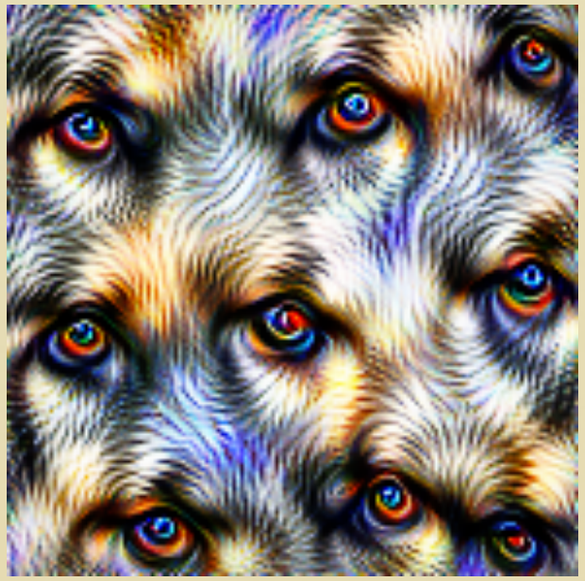
纹理开始变得复杂，仍然都是局部纹理

Olah, et al., "Feature Visualization", Distill, 2017.



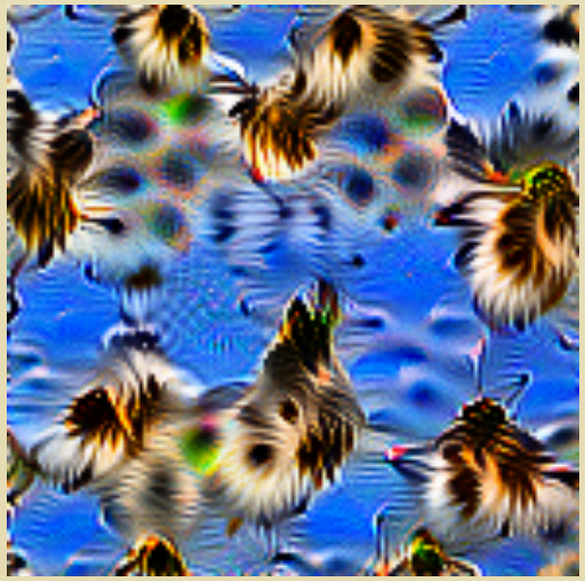
复杂度显著提升，出现了复杂的模式（pattern）

书架



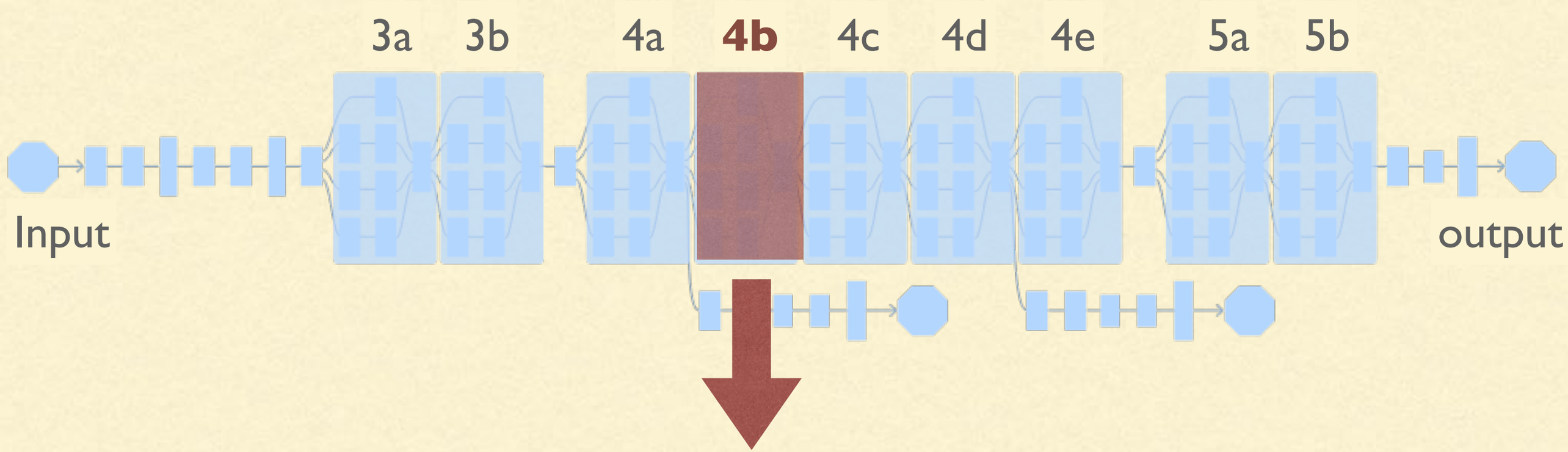
狗眼睛

文本，柳钉



鸟

Olah, et al., "Feature Visualization", Distill, 2017.

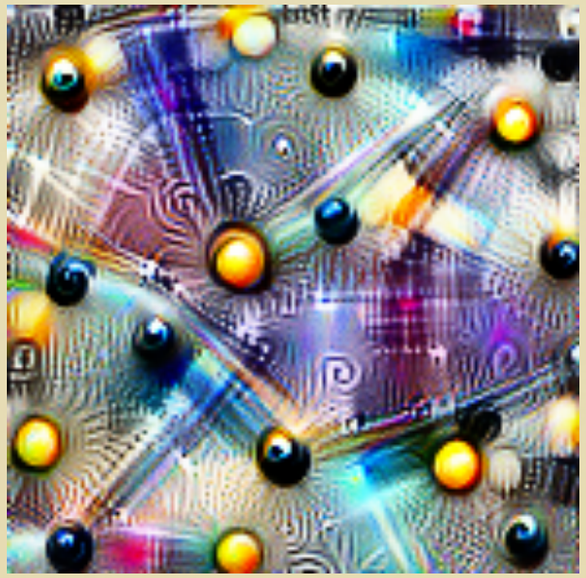


可以辨别出部分物体了，更多的 context 开始出现

建筑结构

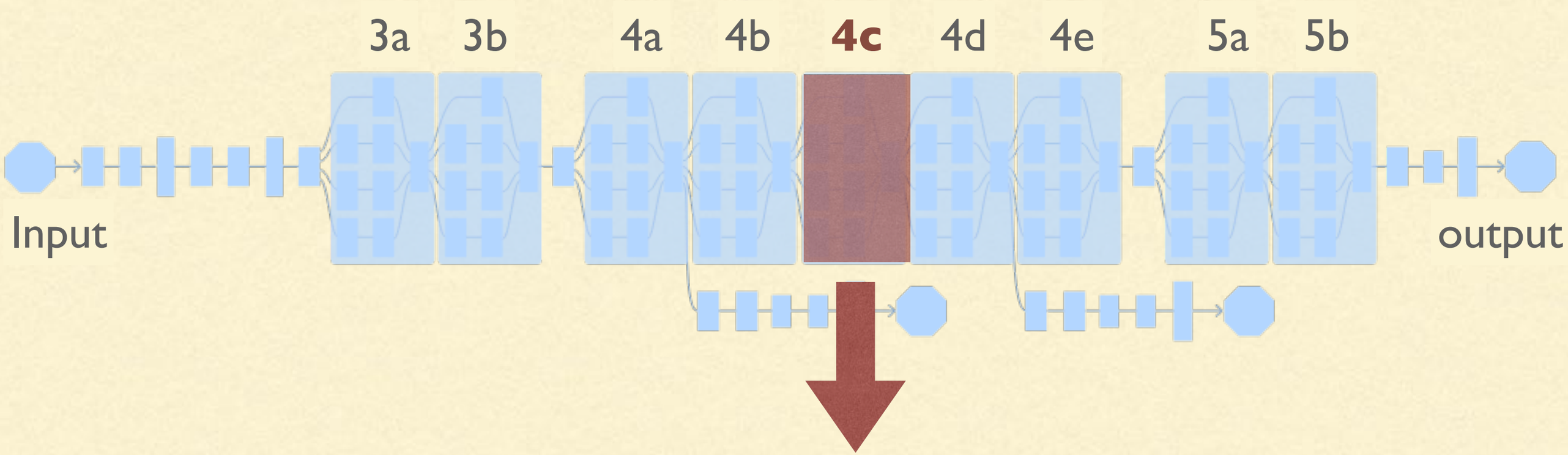
蓬松的绳

树

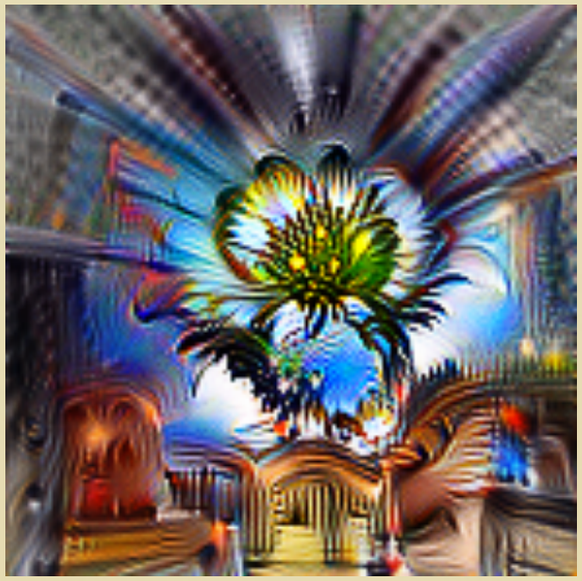


台球

Olah, et al., "Feature Visualization", Distill, 2017.



网络开始响应具体的对象（这也许是最值得探索的一层）



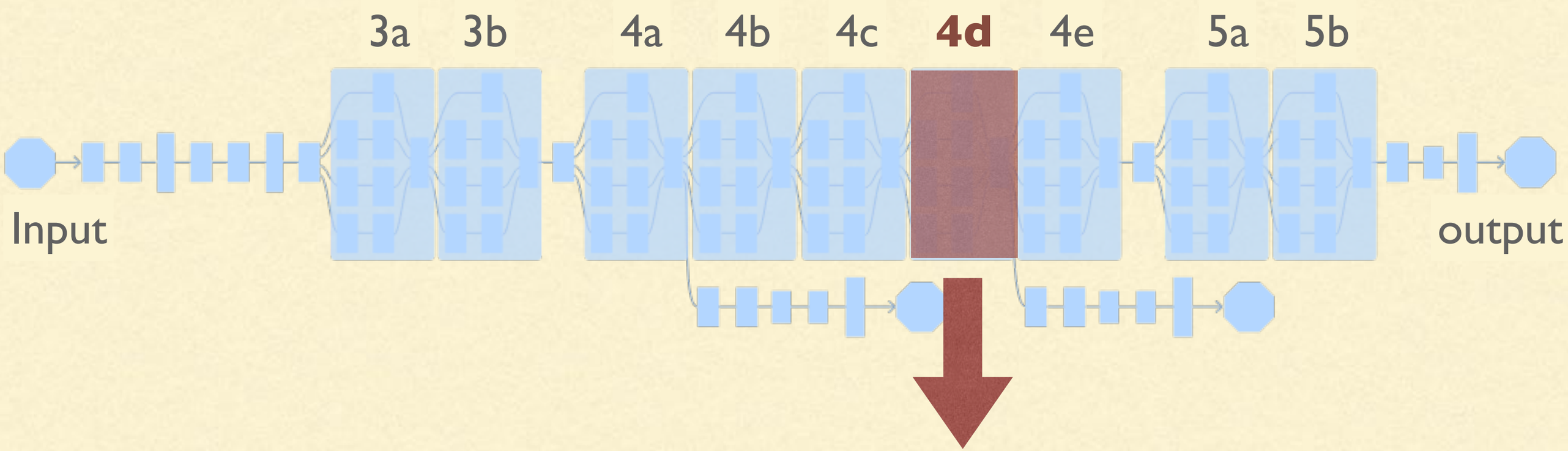
棕榈树

轮子

带项圈的狗

房子

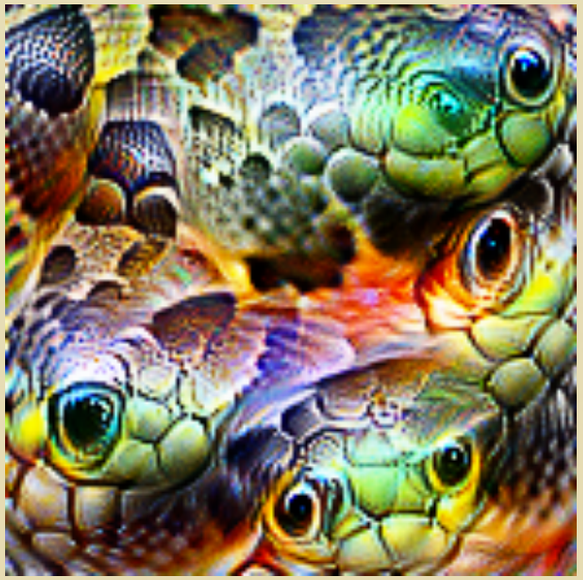
Olah, et al., "Feature Visualization", Distill, 2017.



出现更复杂的概念，开始看到神经元对多个不相关的概念产生反应

狗嘴

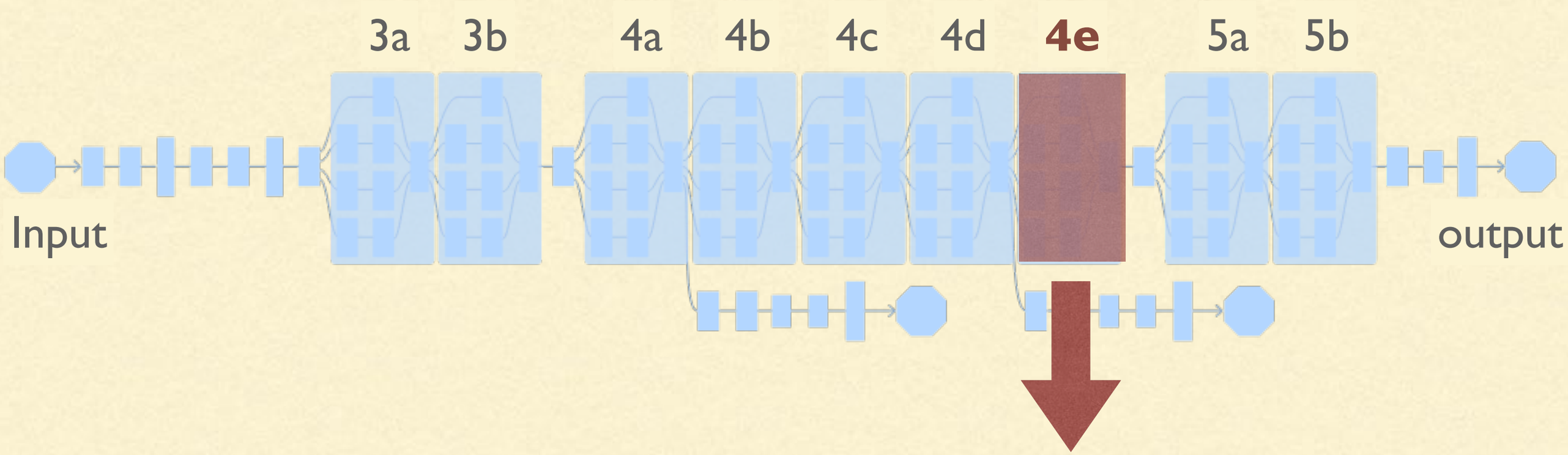
灵长目动物



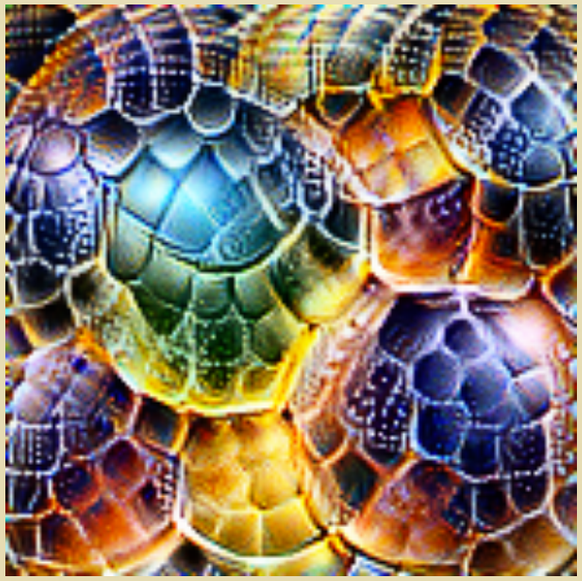
蛇头

餐厅的菜肴

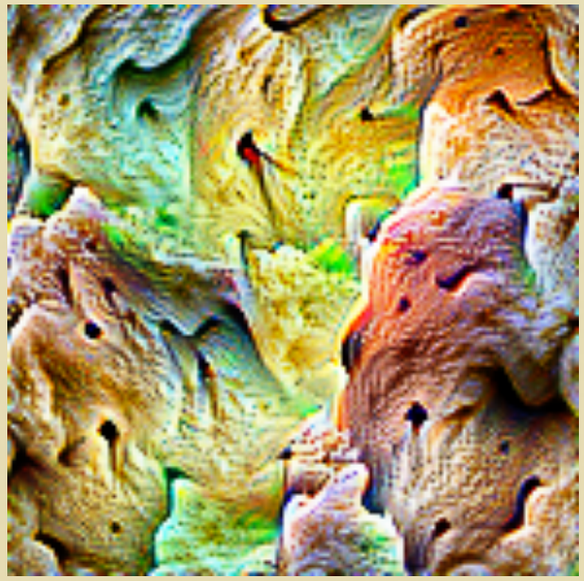
Olah, et al., "Feature Visualization", Distill, 2017.



区分特定的物种，对多个视觉相关的概念产生响应



龟壳

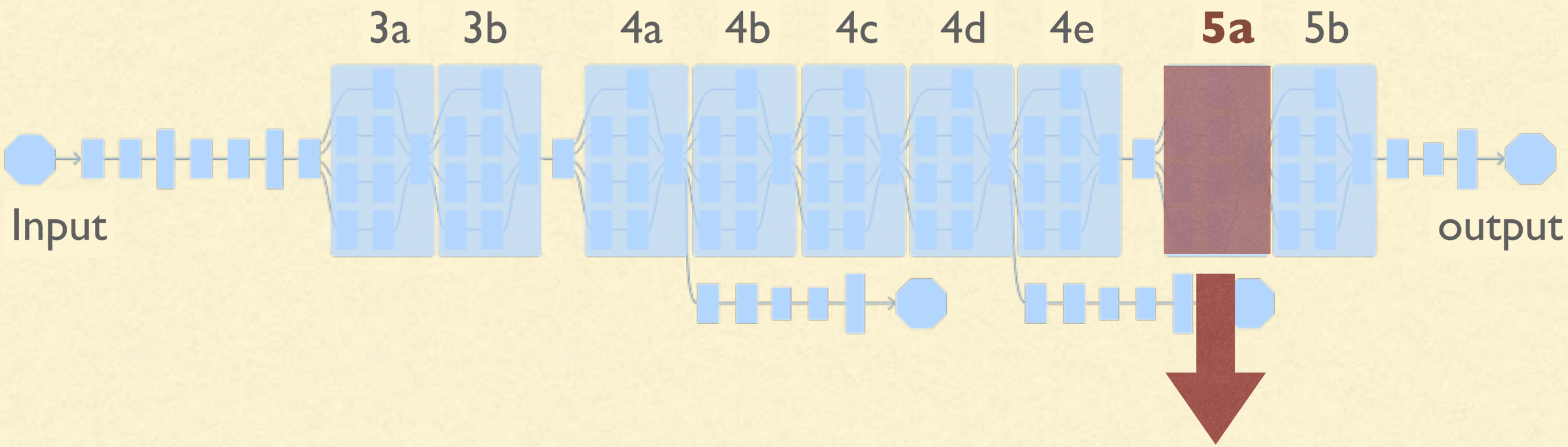


冰淇淋 & 面包

猫的皮肤

宽边帽

Olah, et al., "Feature Visualization", Distill, 2017.



可视化开始变得抽象，但所包含的语义信息还是较为具体

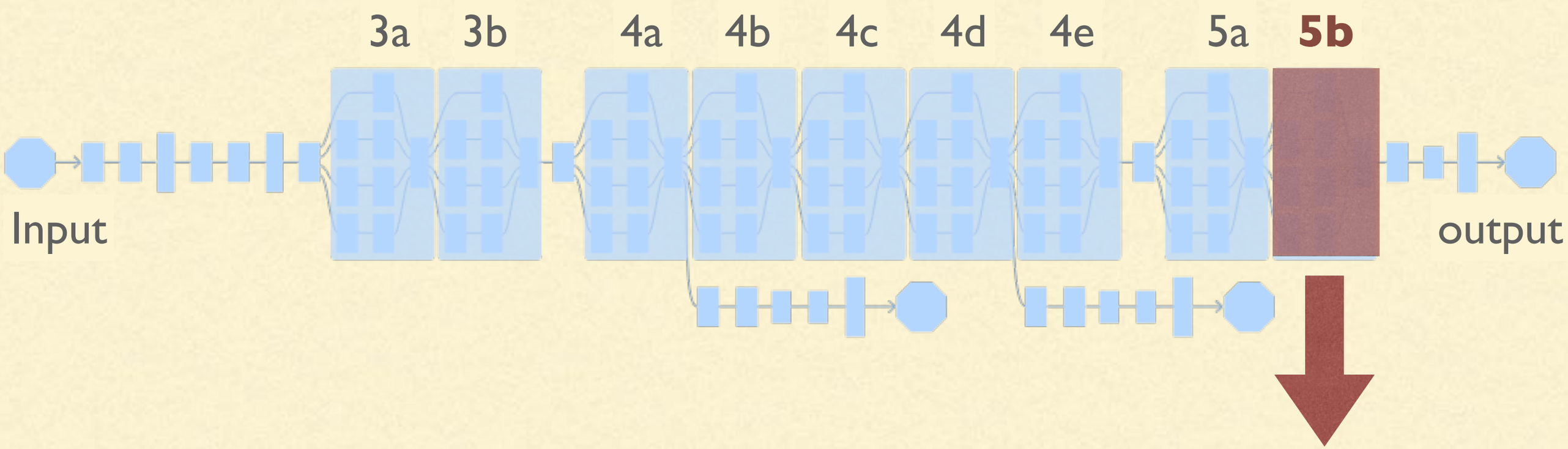
糖果

球

铜管乐器

交通灯

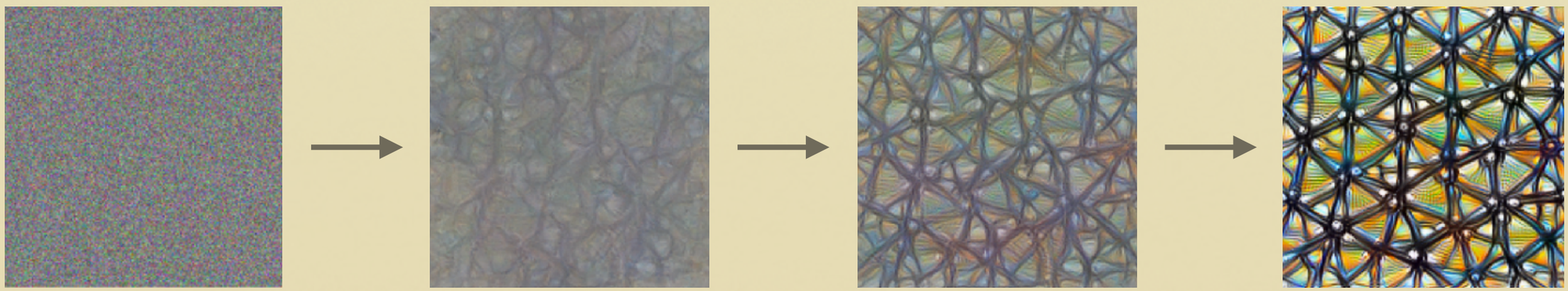
Olah, et al., "Feature Visualization", Distill, 2017.



像是没有意义的大杂烩，神经元似乎不再对应特别的语义概念

Olah, et al., "Feature Visualization", Distill, 2017.

DeepDream: 放大已存在的特征

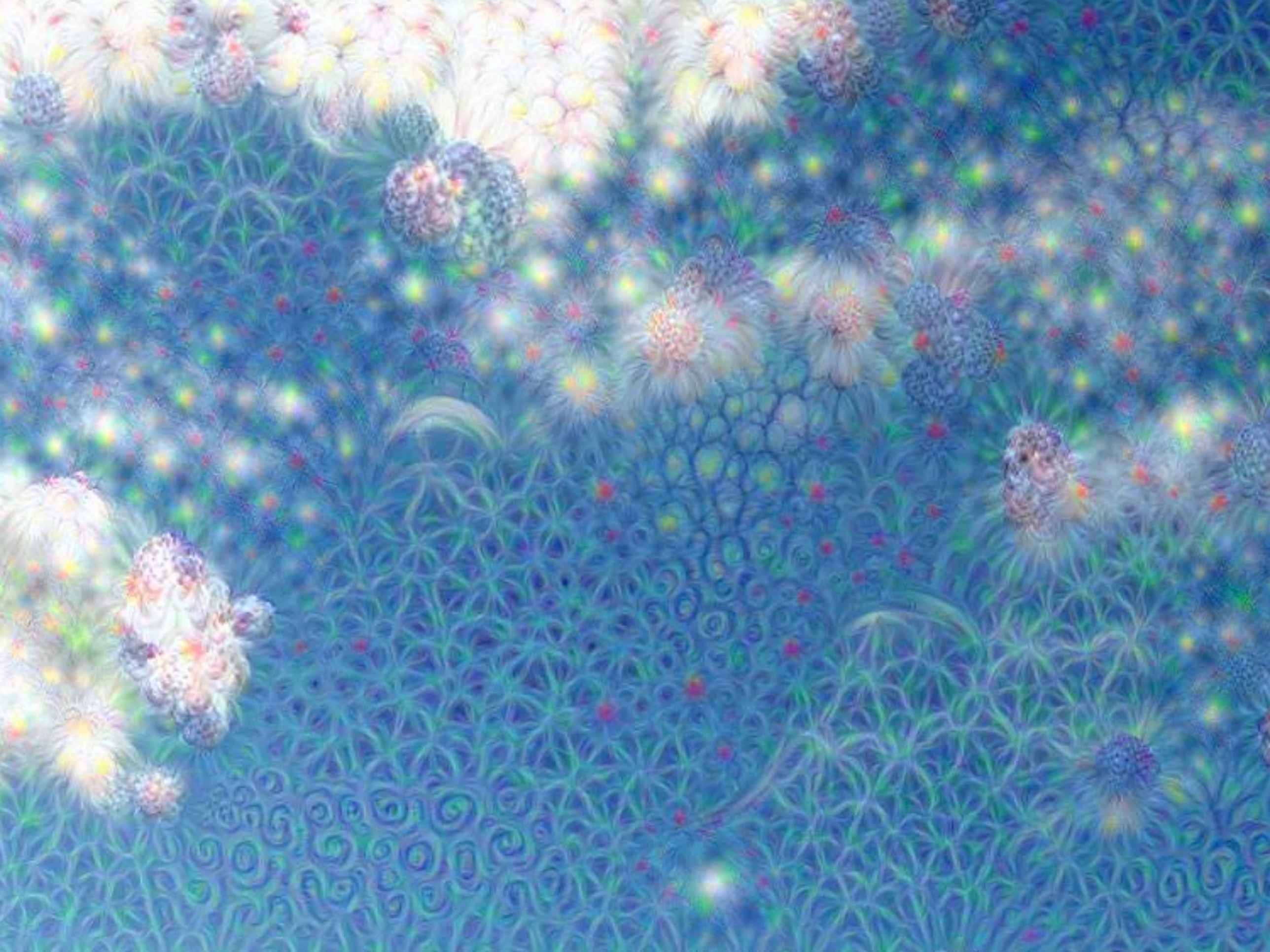


从随机 noise 开始，优化 (optimize) 图片以激活特定的神经元（网络层）

如果我们输入一张正常的图片替代噪声（noise），会发生什么?

项目地址：https://github.com/google/deepdream

Mordvintsev, Olah, and Tyka,
"Inceptionism: Going Deeper into Neural Networks", Google Research Blog.



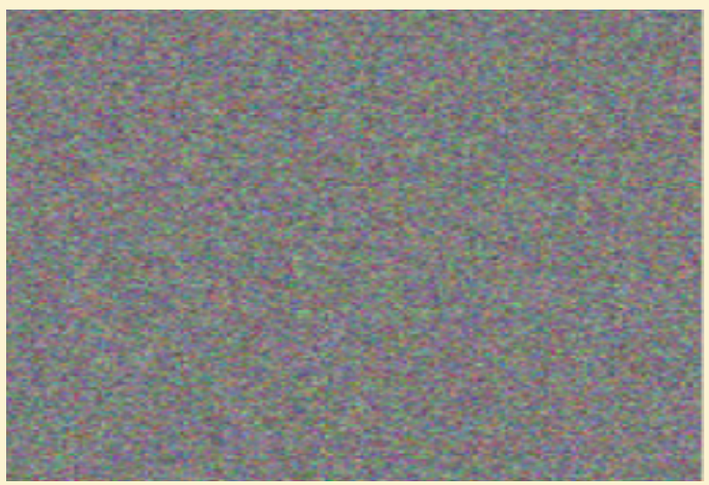
VIDEO

https://www.youtube.com/watch?v=DgPaCWJL7XI

风格迁移（Style transfer）

Leon A. Gatys and (2015).
A Neural Algorithm of Artistic Style.
CoRR, abs/1508.06576, .

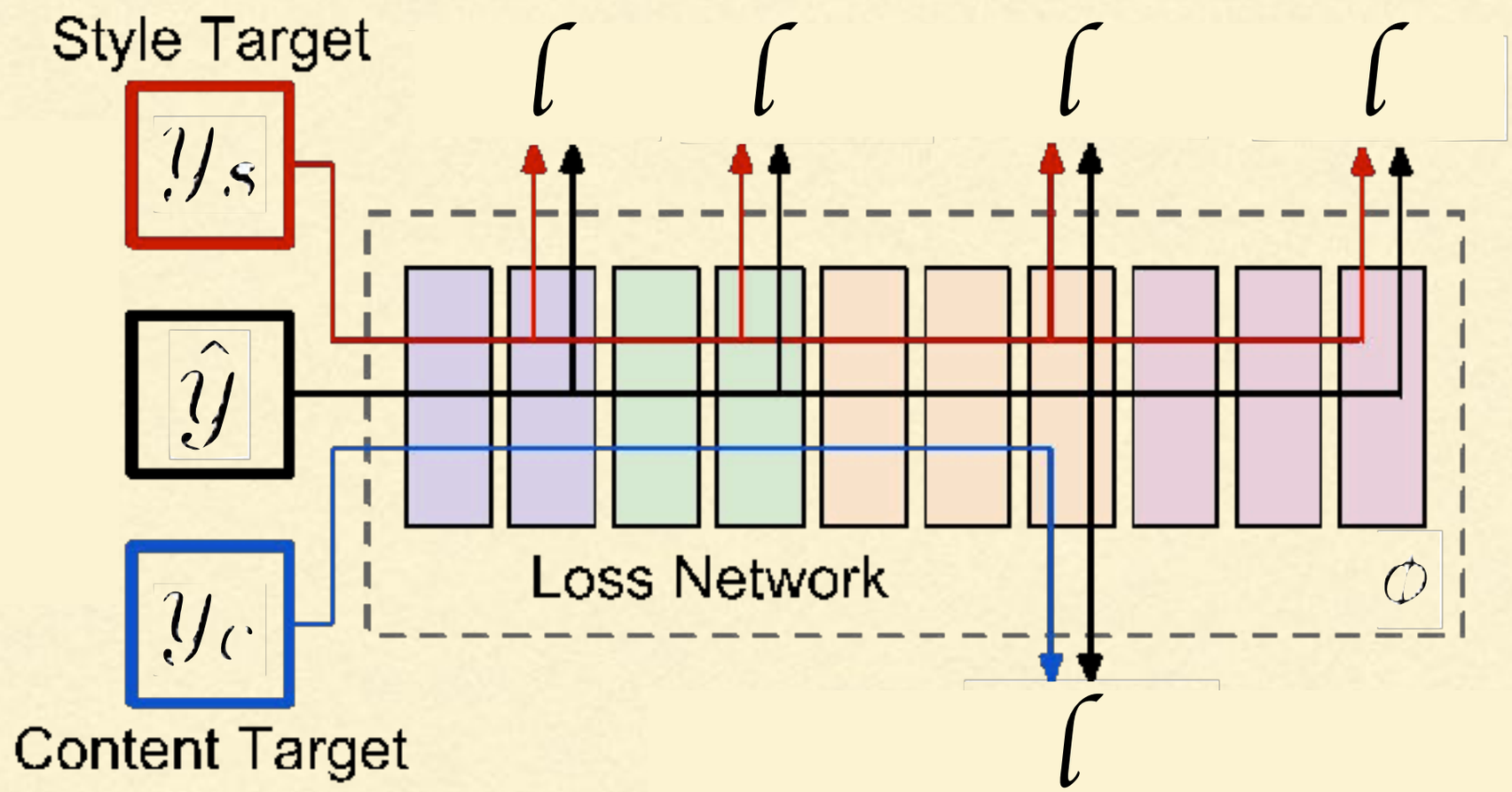
风格图片



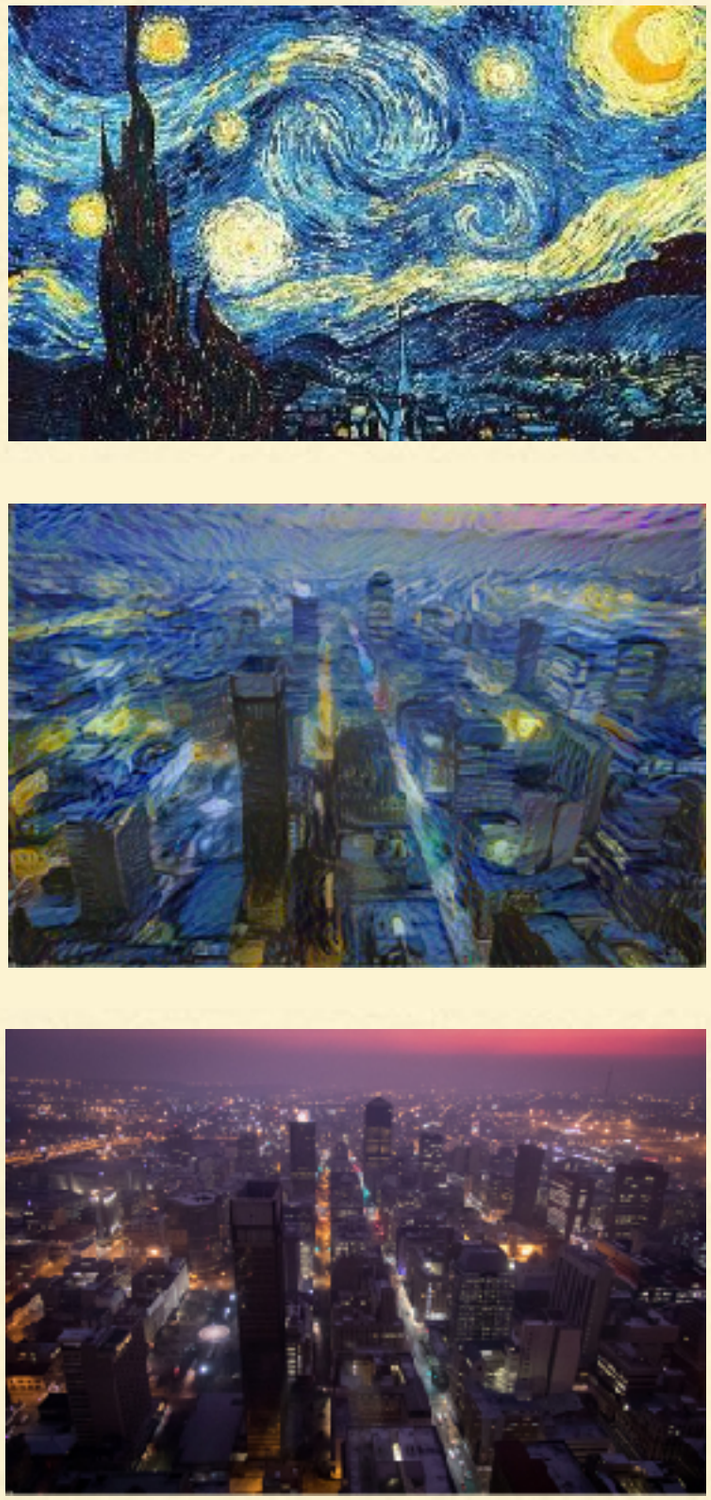
内容图片

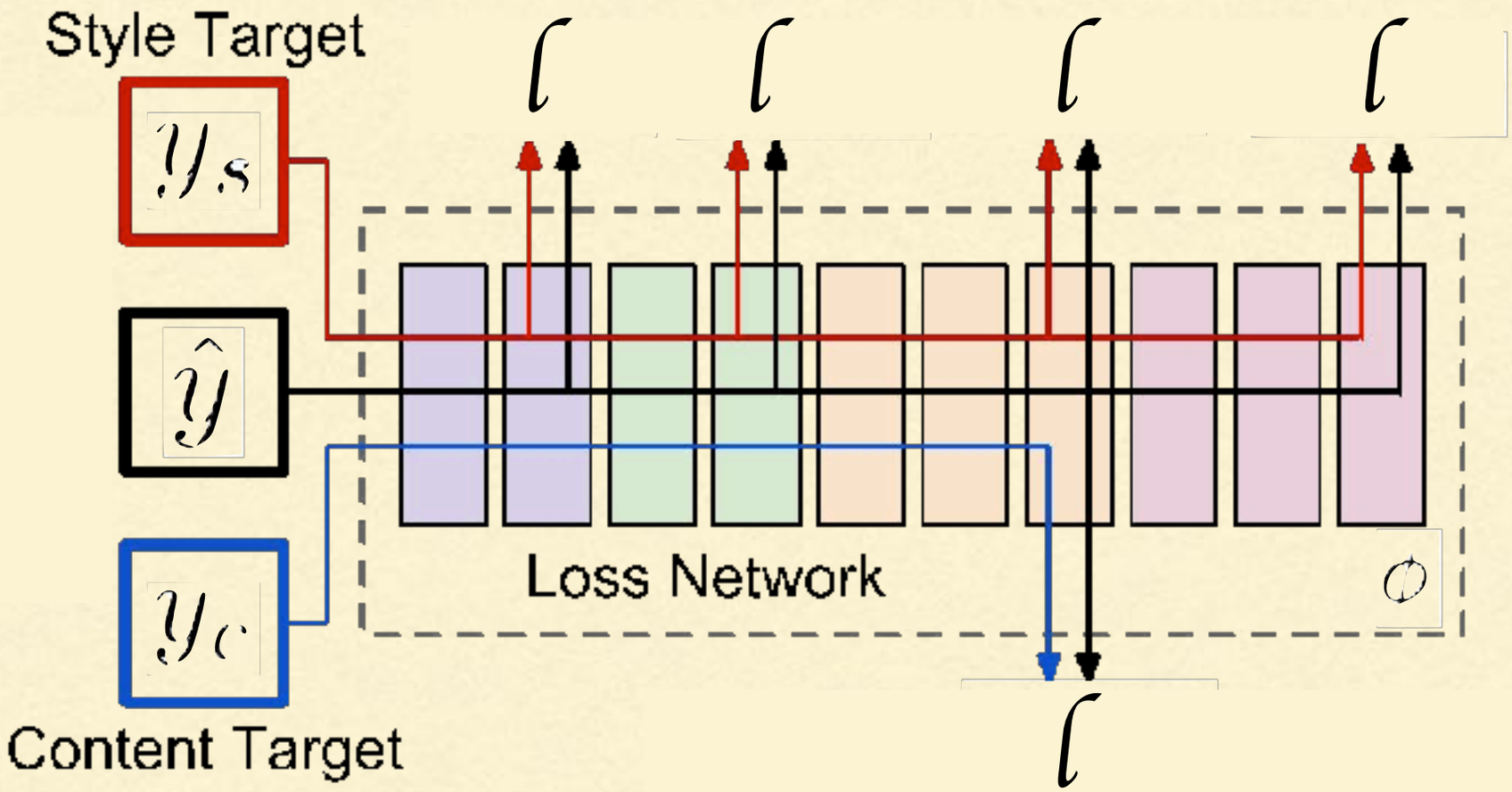
优化目标： 风格 Loss + 内容 Loss

底层纹理特征　　深层语义特征



Gatys, Ecker, and Bethge, "Image style transfer using convolutional neural networks", CVPR 2016 Figure copyright Justin Johnson, 2015.

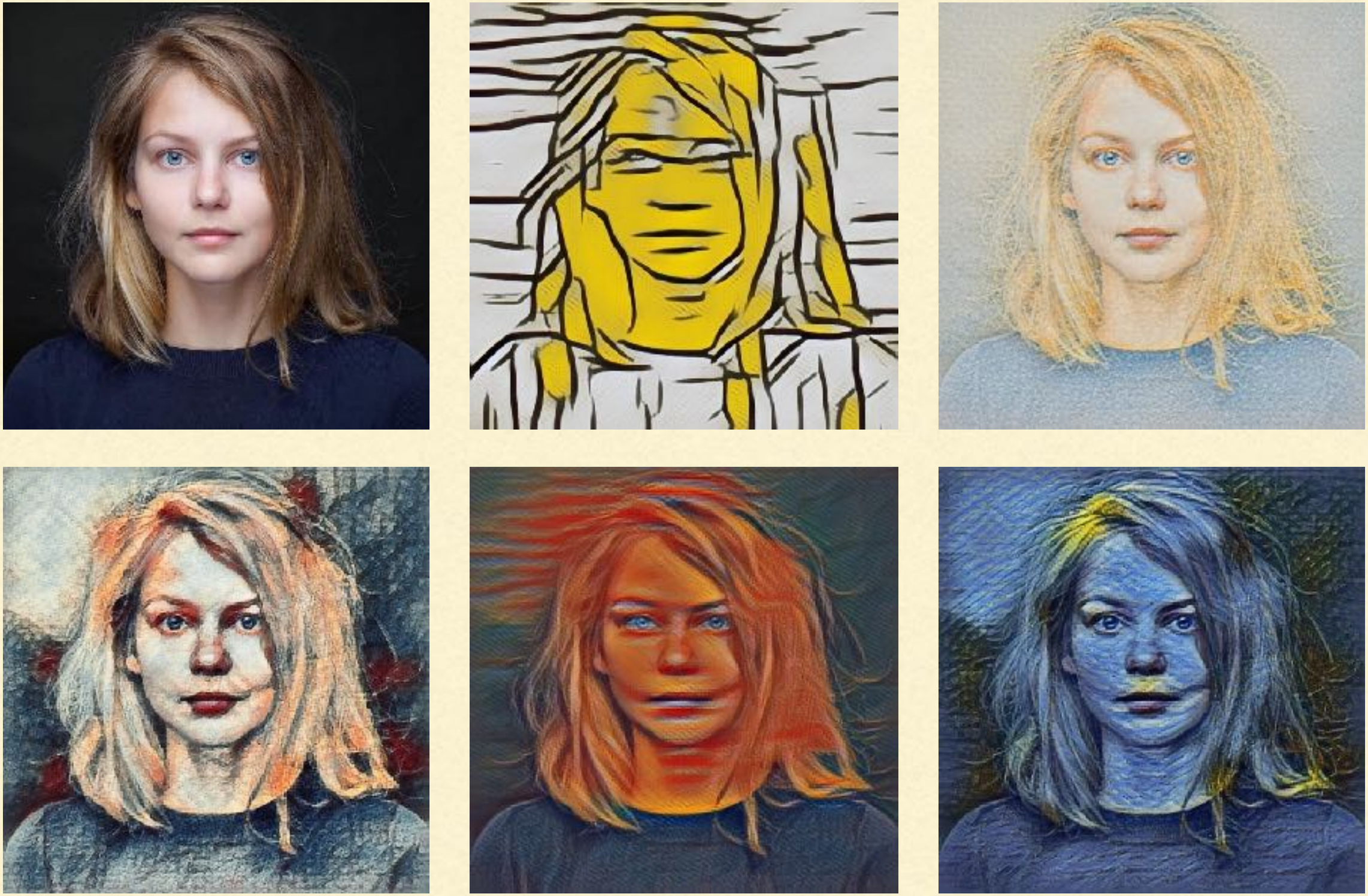




Gatys, Ecker, and Bethge, "Image style transfer using convolutional neural networks", CVPR 2016 Figure copyright Justin Johnson, 2015.

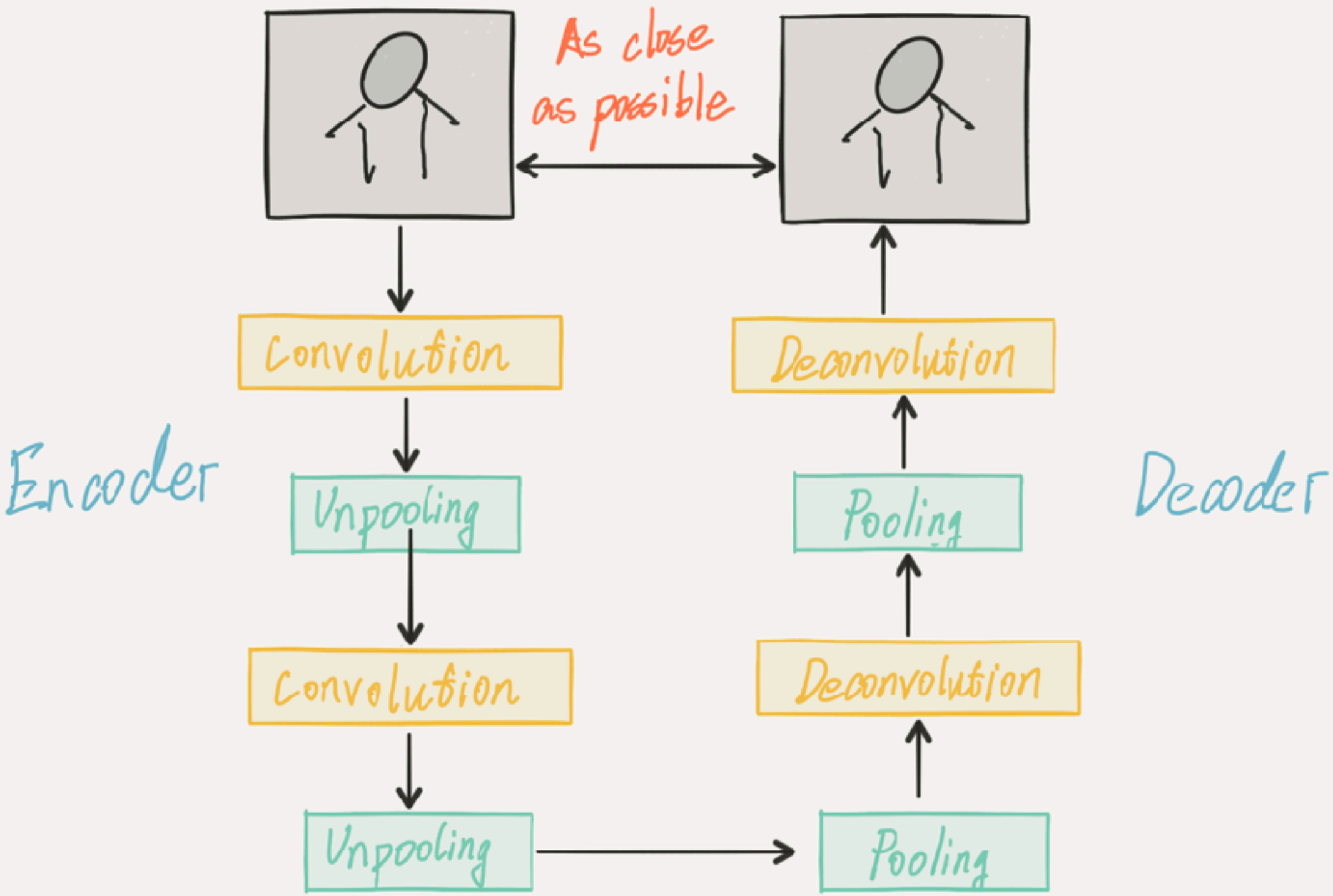
一个模型，多种风格

Dumoulin, Shlens, and Kudlur,
"A Learned Representation for Artistic Style", ICLR 2017.

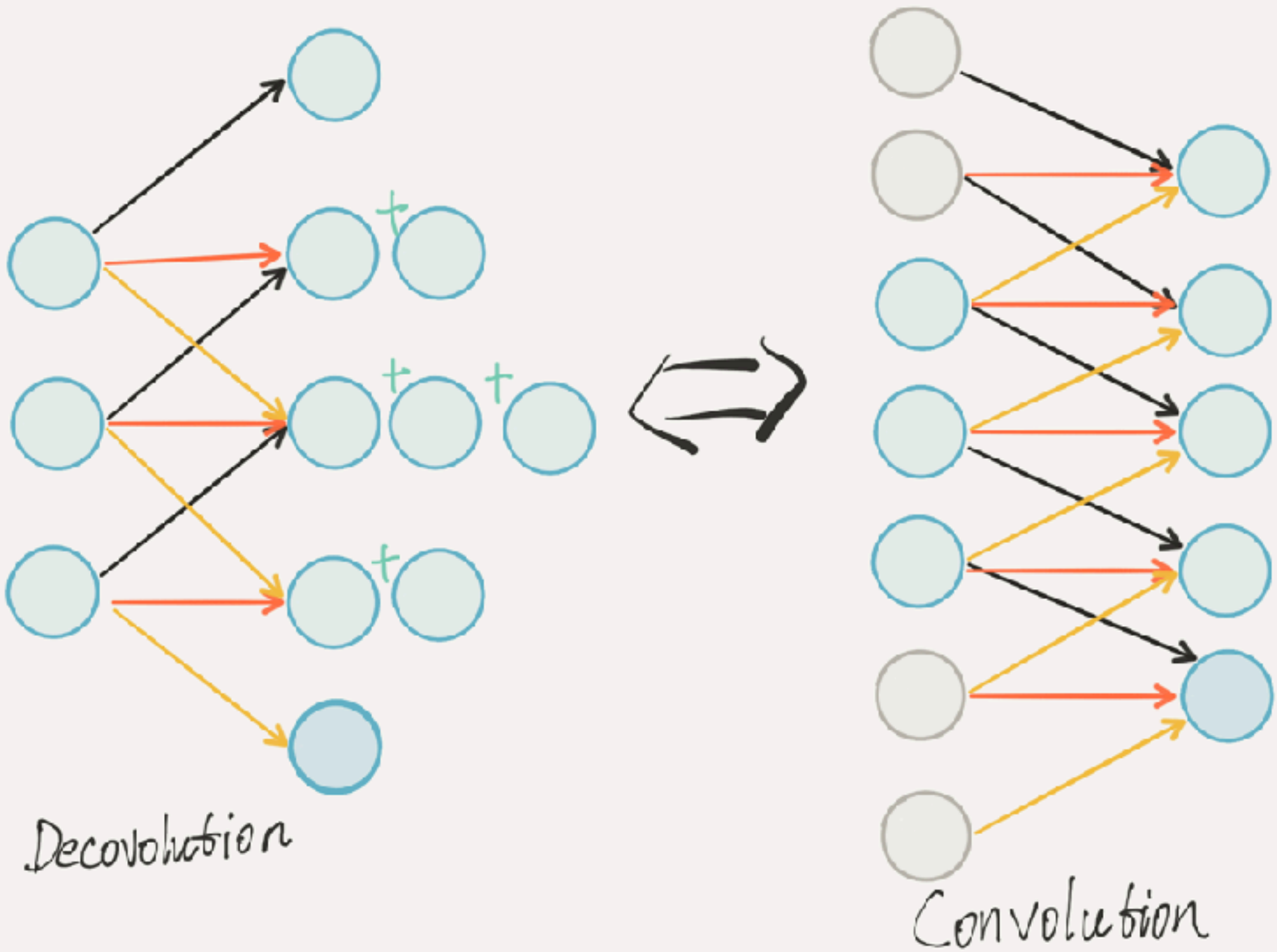


Dumoulin, Shlens, and Kudlur,
"A Learned Representation for Artistic Style", ICLR 2017.

Auto-encoder for CNN (基于 CNN 的自动编码器)



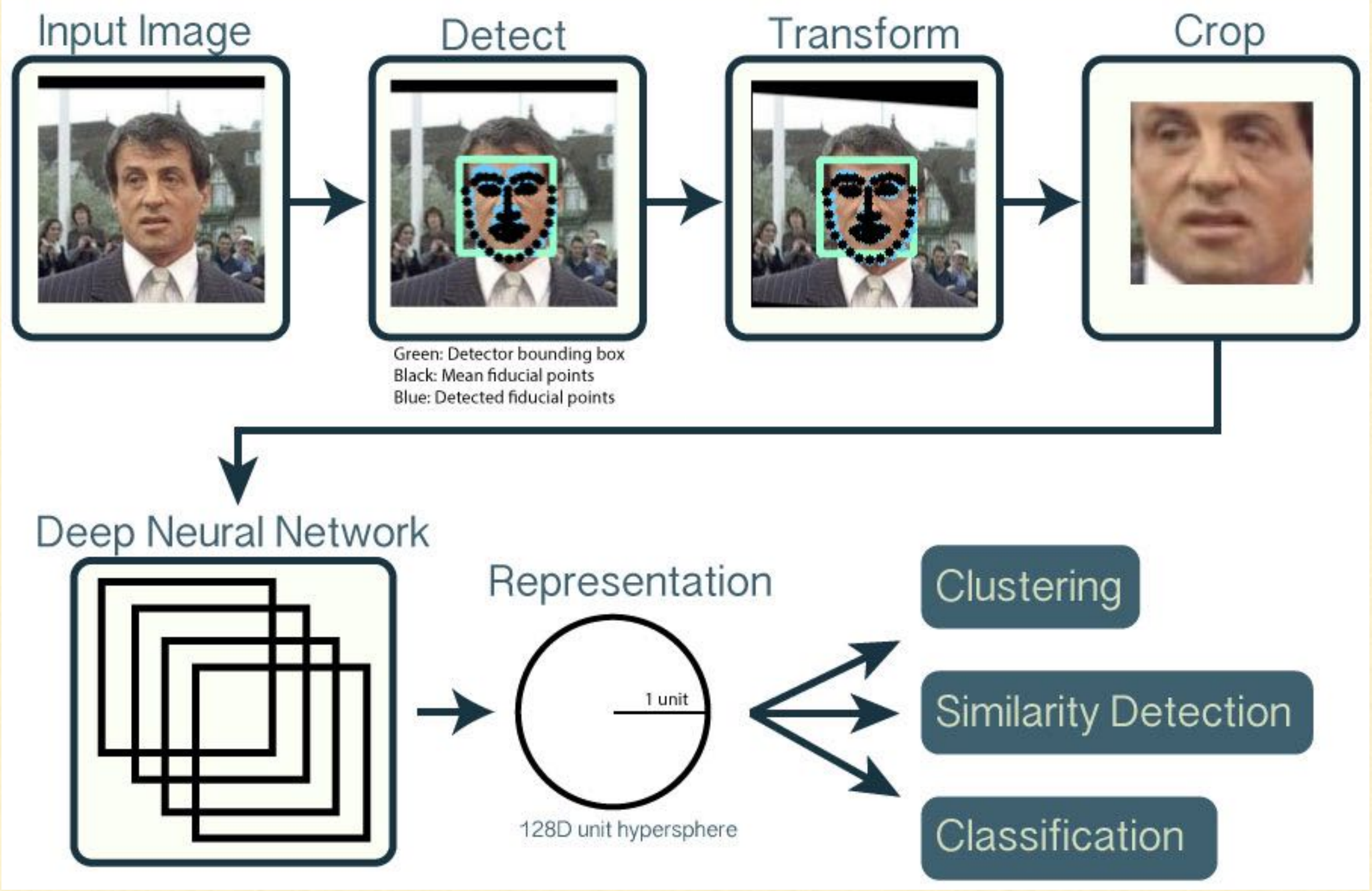
反卷积（Deconvolution）



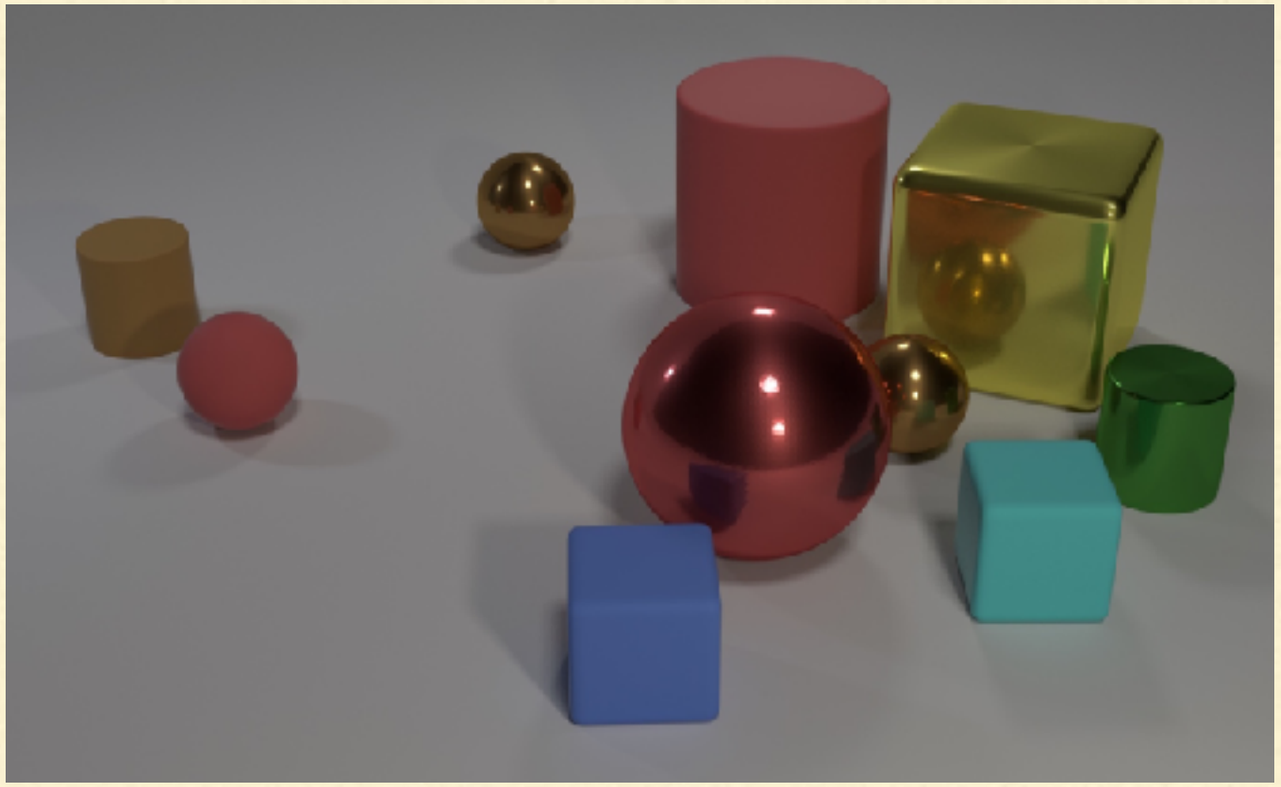
VIDEO

GAN 人脸生成

人脸识别


Input Image
Detect
Transform
Crop
Green: Detector bounding box
Black: Mean fiducial points
Blue: Detected fiducial points
Deep Neural Network
Representation
1 unit
128D unit hypersphere
Clustering
Similarity Detection
Classification

High-level Task: 视觉推理

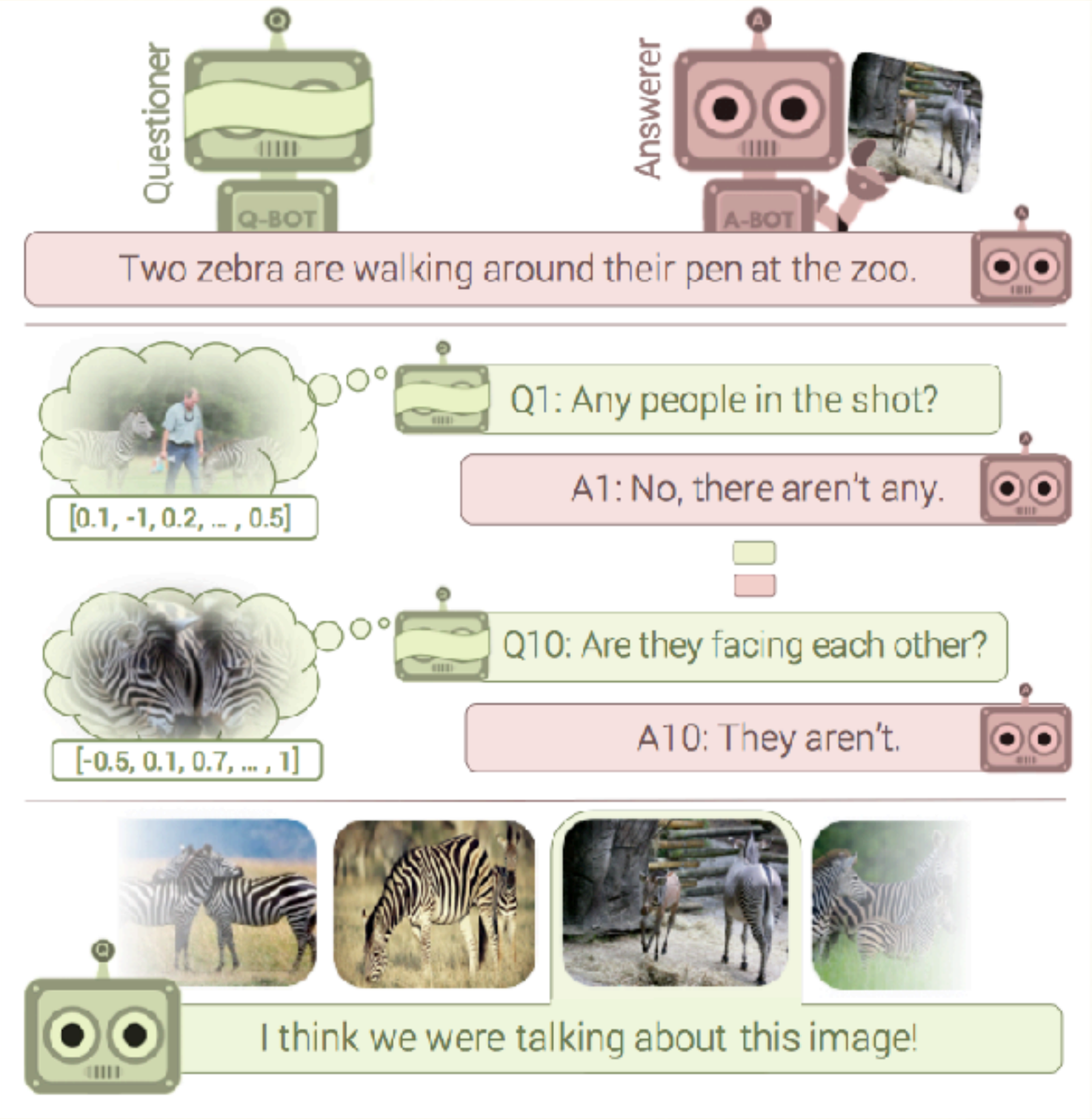


Q: Are there an **equal number** of **large things** and **metal spheres**?
Q: **What size** is the **cylinder** **that is left of** the **brown metal** thing **that is left of** the **big sphere**?
Q: **How many** objects are **either small cylinders** or **red** things?

AI 不仅需要识别出图像中的物体，
还需要有推理、判断能力

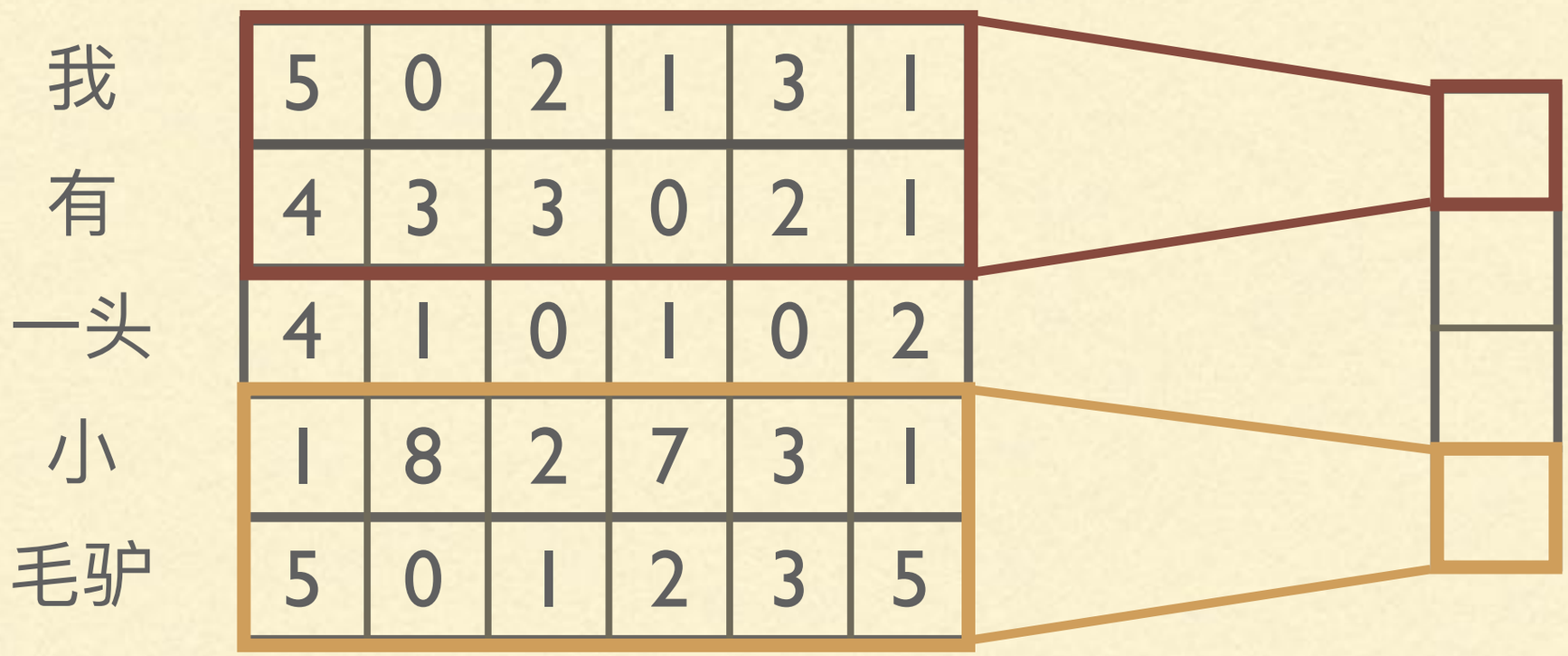
Justin Johnson and (2016).
CLEVR: A Diagnostic Dataset for Compositional Language and Elementary. CoRR, abs/1612.06890, .

High-level Task: 视觉会话



Abhishek Das and (2017).
Learning Cooperative Visual Dialog Agents with Deep Reinforcement.
CoRR, abs/1703.06585, .

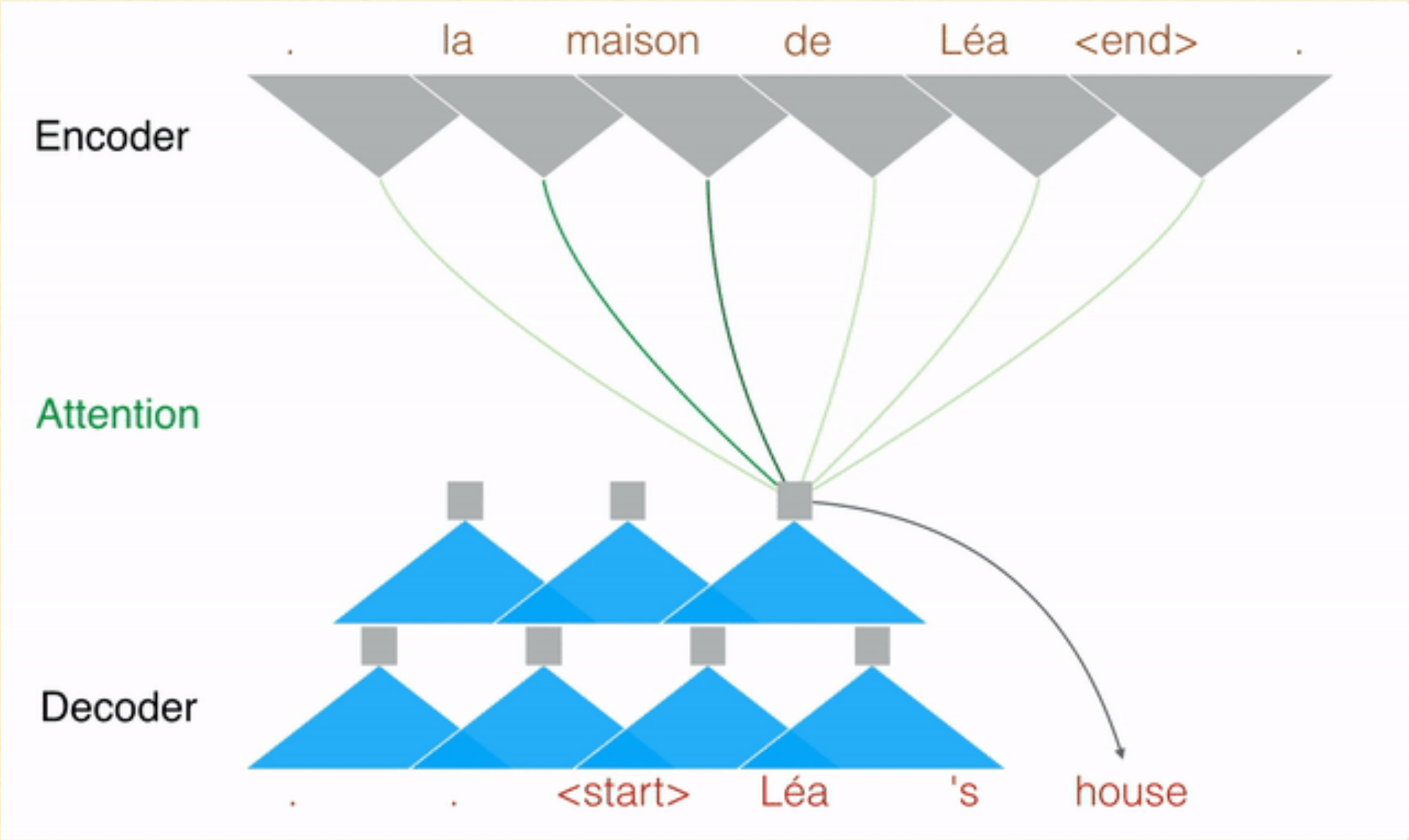
卷积神经网络（CNN）与自然语言处理（NLP）



语言具有局部相关性

Yoon Kim (2014).
Convolutional Neural Networks for Sentence Classification. CoRR, abs/1408.5882, .

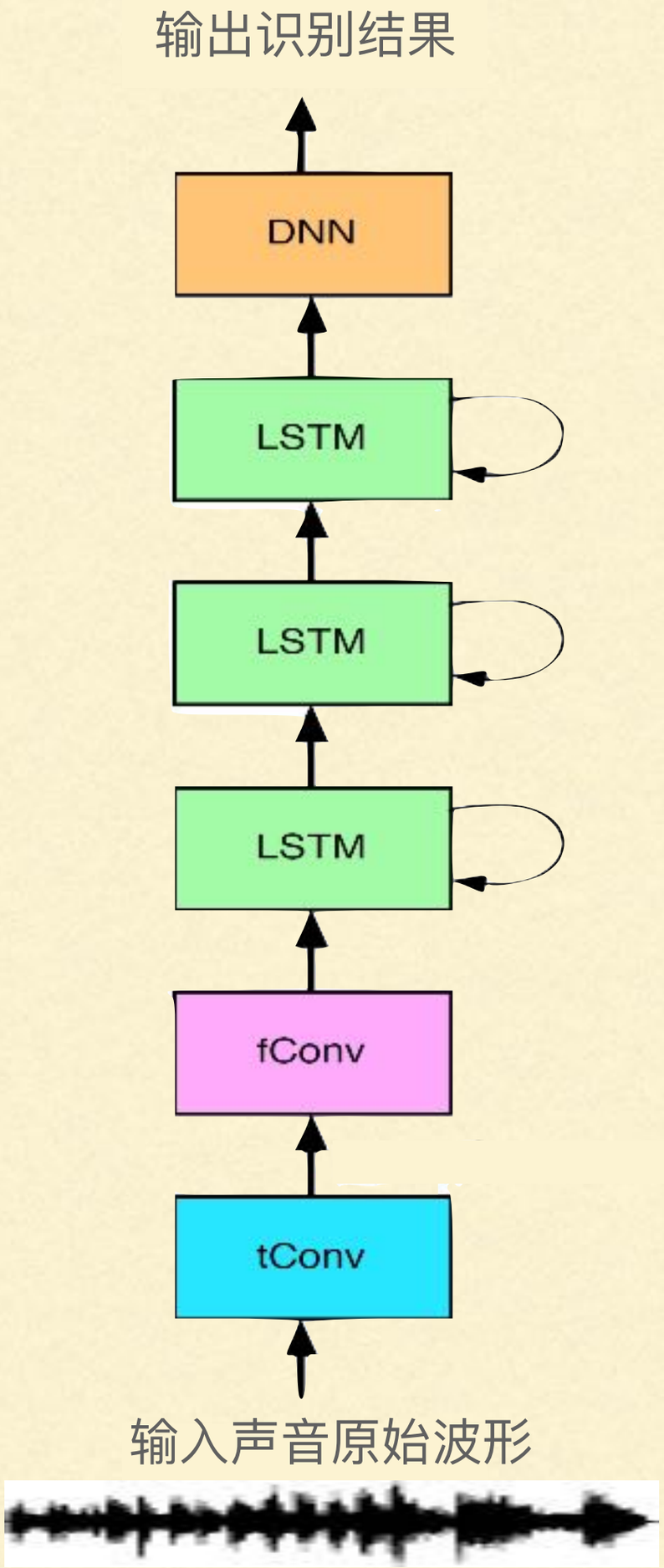
使用卷积神经网络（CNN）的机器翻译模型

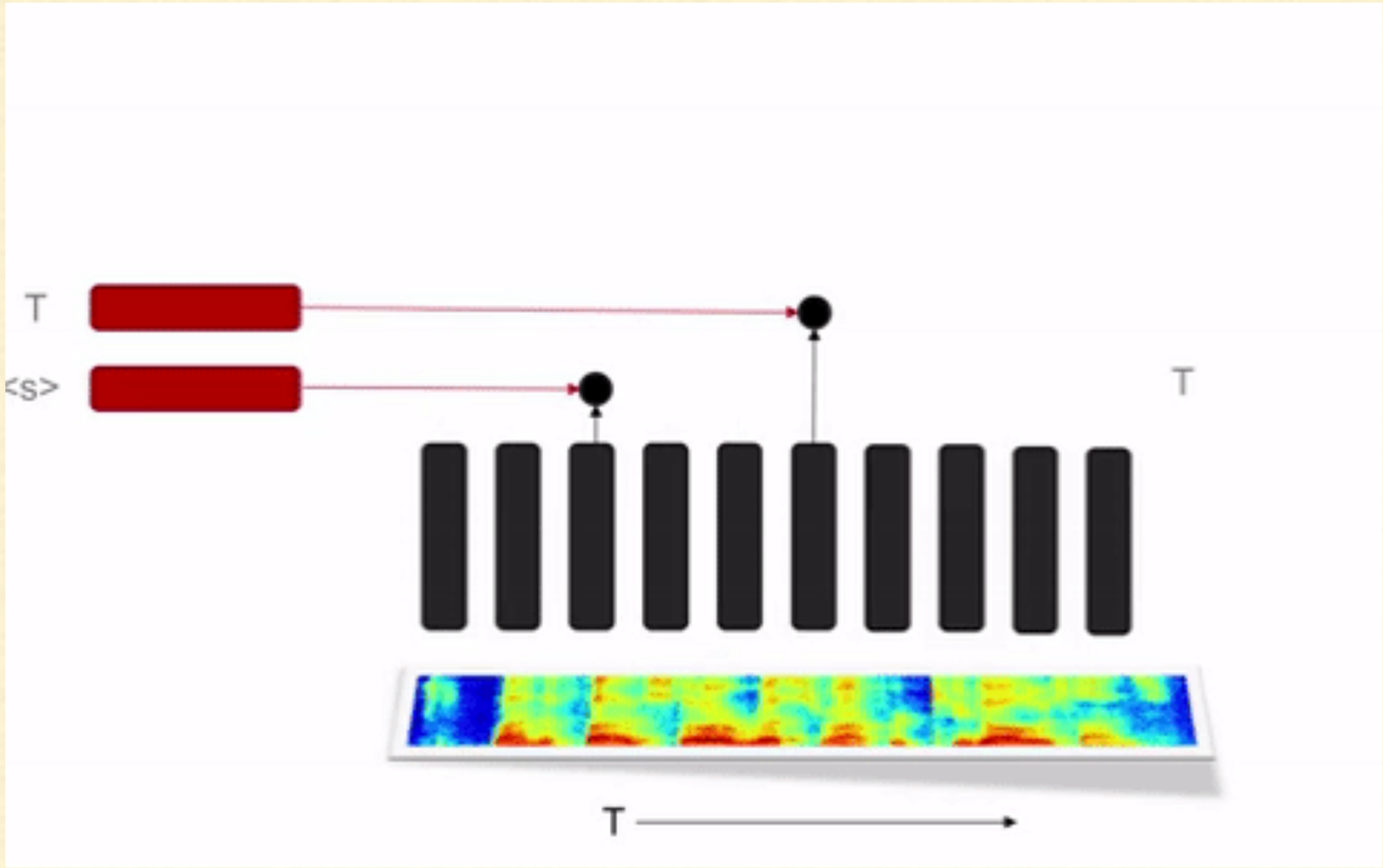


Gehring, Jonas,, Auli, Michael, Grangier, David, Yarats, Denis & Dauphin, Yann N (2017).
"Convolutional Sequence to Sequence Learning". ArXiv e-prints

https://github.com/facebookresearch/fairseq

卷积神经网络（CNN）
与语音识别（Speech Recognition）





Eric Battenberg and (2017).
Exploring Neural Transducers for End-to-End Speech Recognition.
CoRR, abs/1707.07413, .

Tara N. Sainath, Ron J. Weiss, Andrew Senior, Kevin W. Wilson, Oriol Vinyals,
"Learning the Speech Front-endWith Raw Waveform CLDNNs,"
In INTERPSEECH 2015

媒体智能实验室
Media Intelligence LAB

卷积神经网络

从相识、相知到相爱

再到相杀

QA