计算机视觉

概述

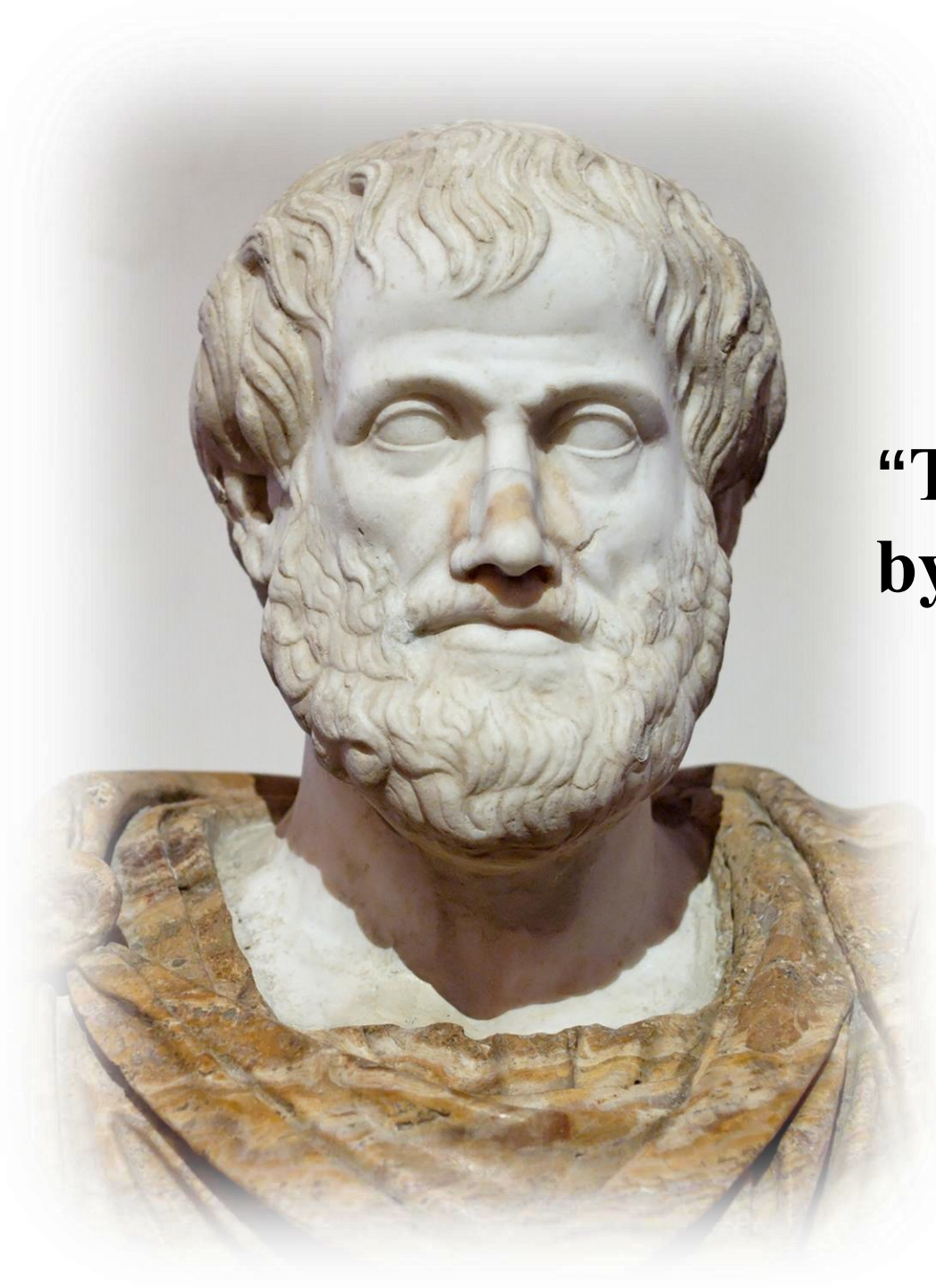
中国传媒大学
COMMUNICATION UNIVERSITY OF CHINA

任课教师：方力
联 系 信 息 ： lifang8902@cuc.edu.cn
办公地址：国重大楼801c

特别鸣谢：

Prof. Kosta Derpanis
York University
Toronto, ON, CA

什么是计算机视觉？



“To know what is where by looking.”

—亚里士多德

（300 BC）

To know what is where by looking

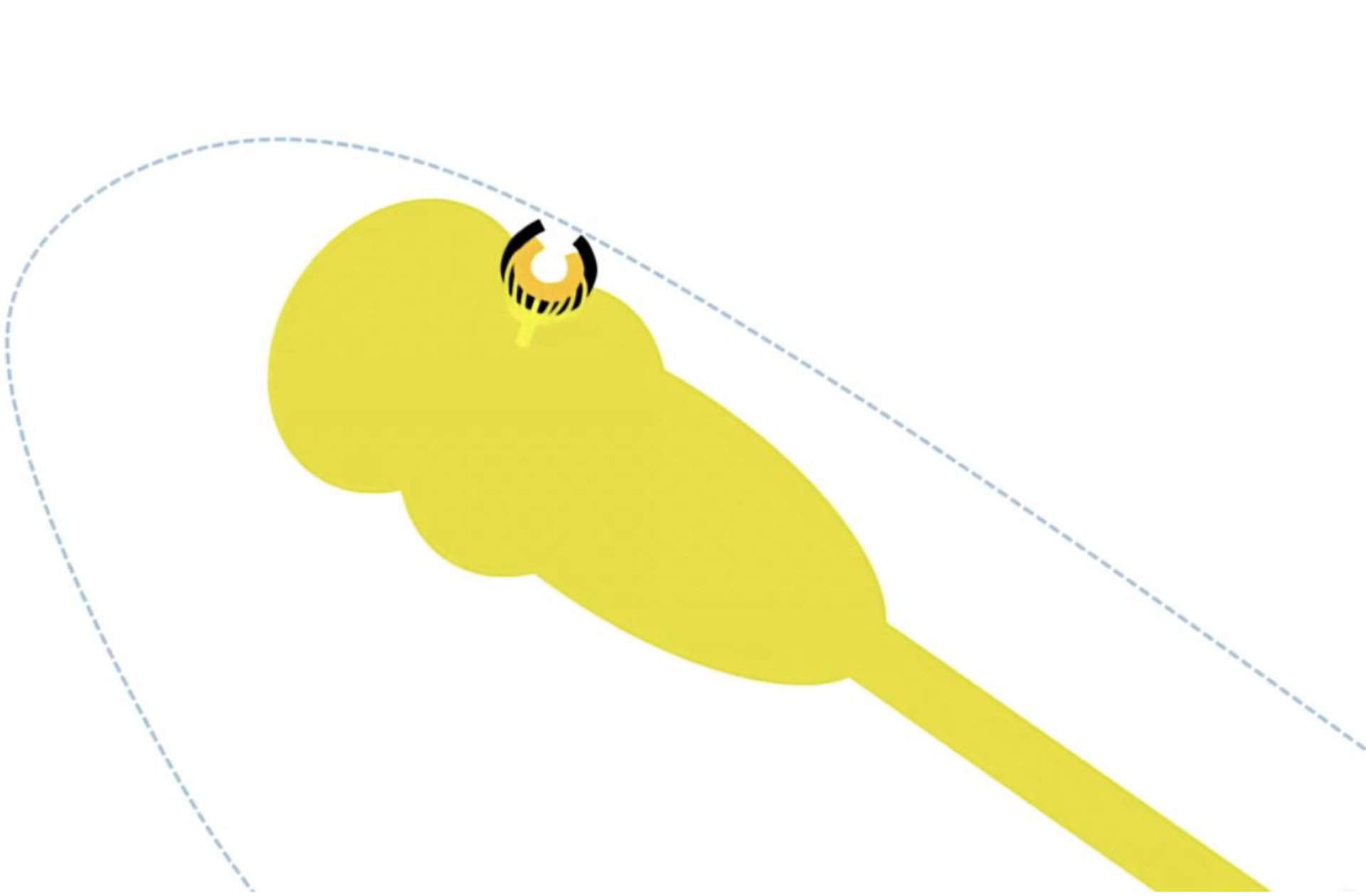
通过观察
知道什么东西在什么地方

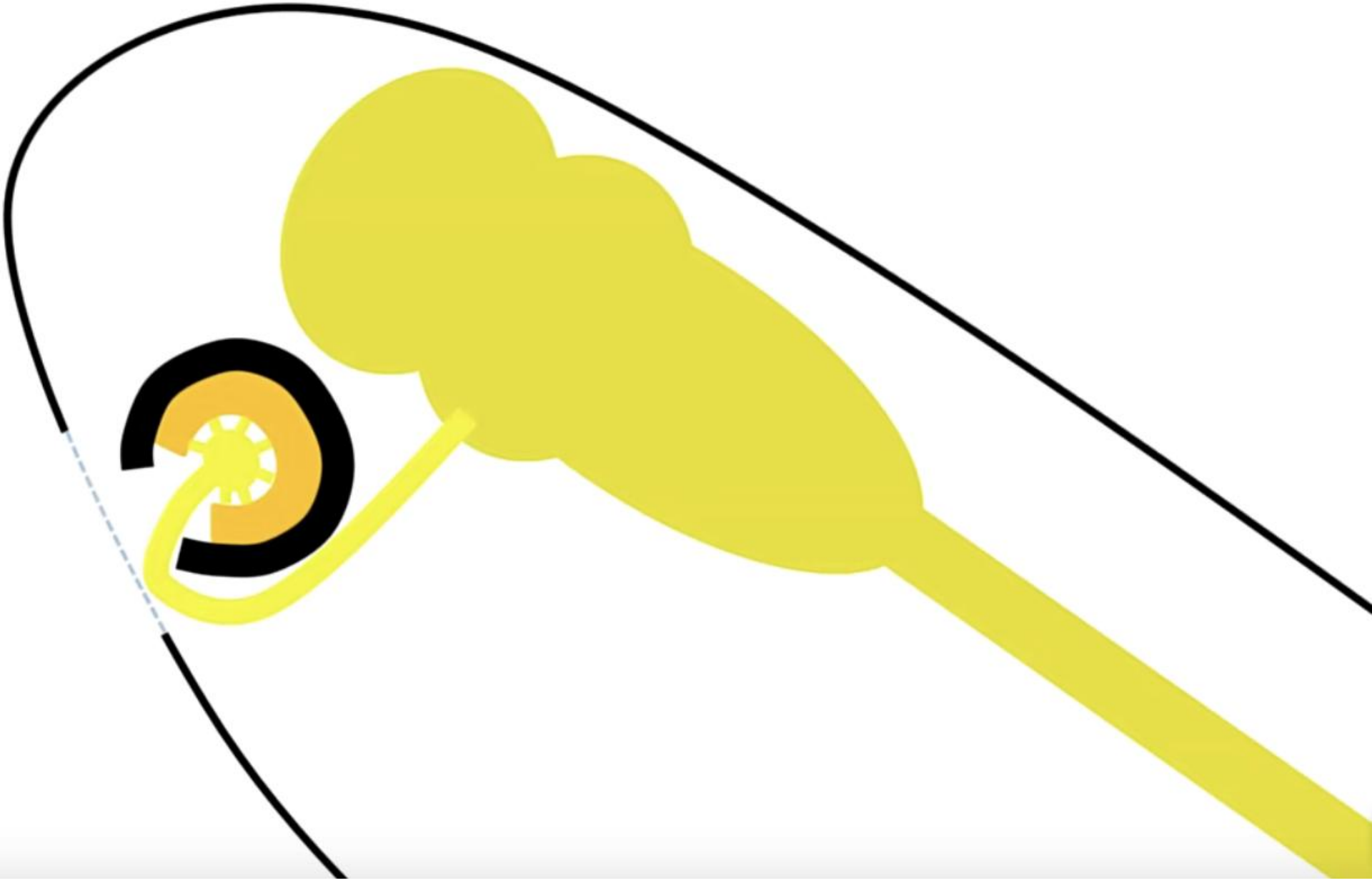
通过一张或更多图像
知道什么东西在什么地方

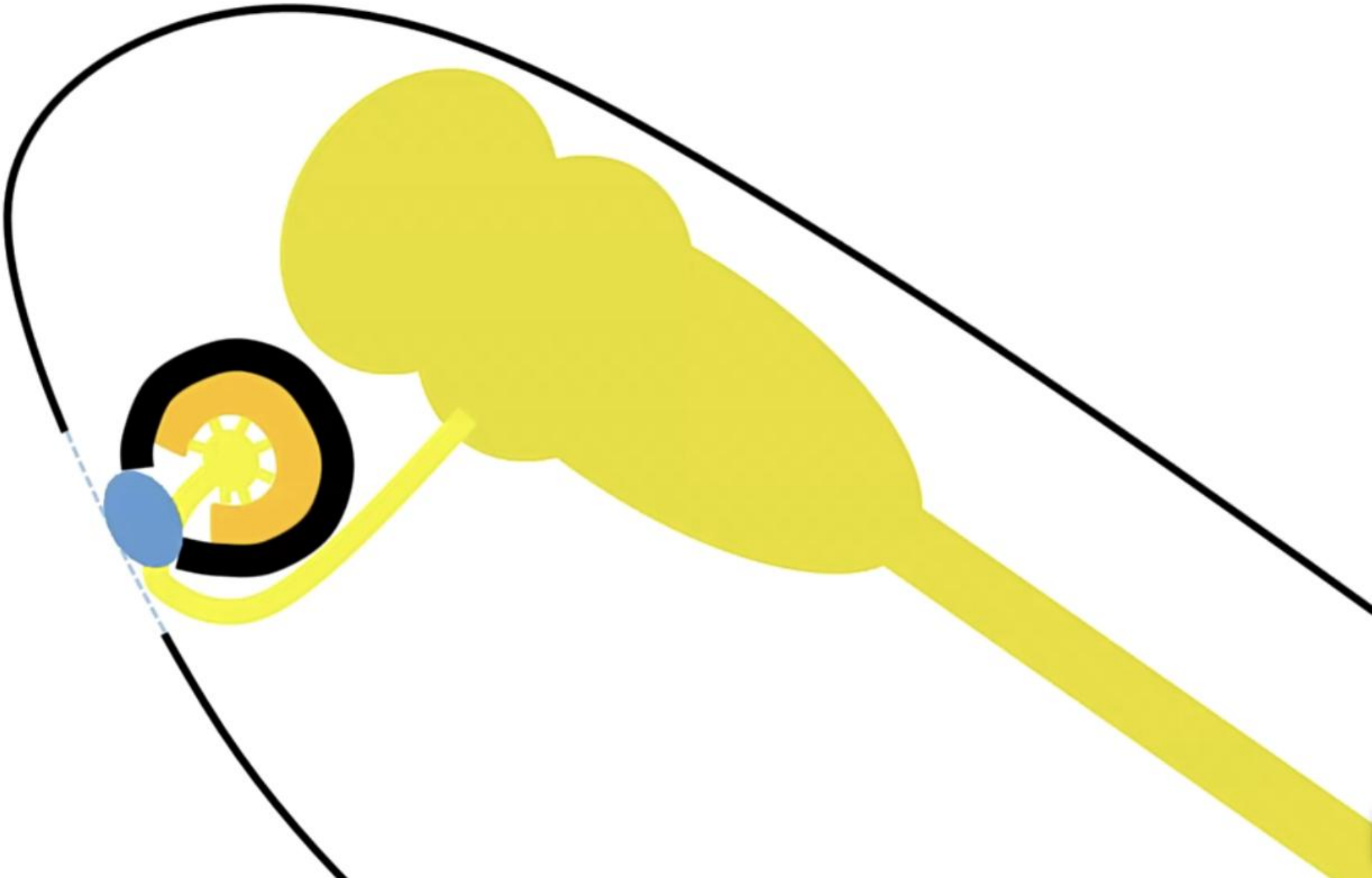
通过一张或更多图像
研究如何恢复世界有用特性

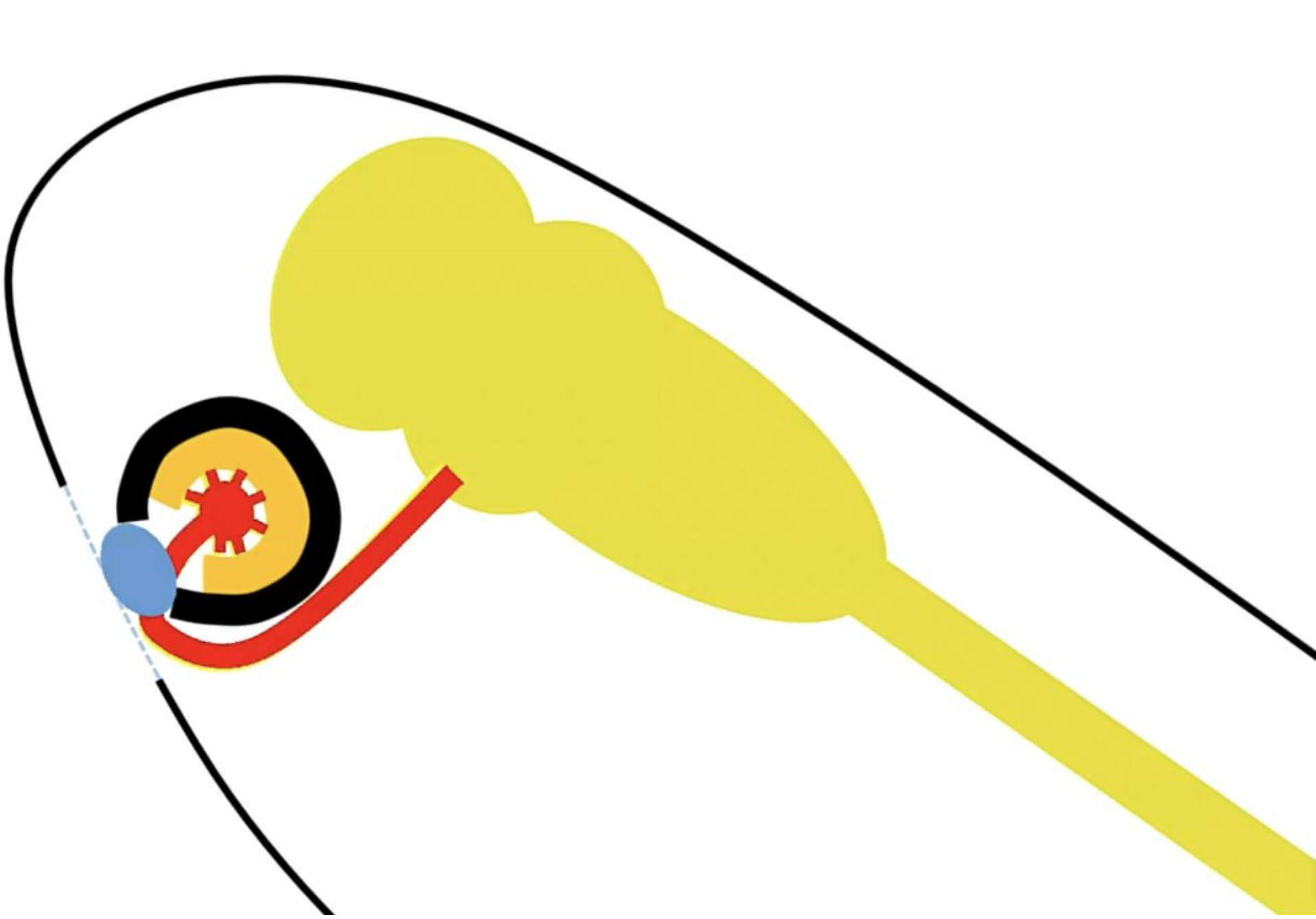
人类通过视觉所获得的信息
占获取信息总量的70%~80%，

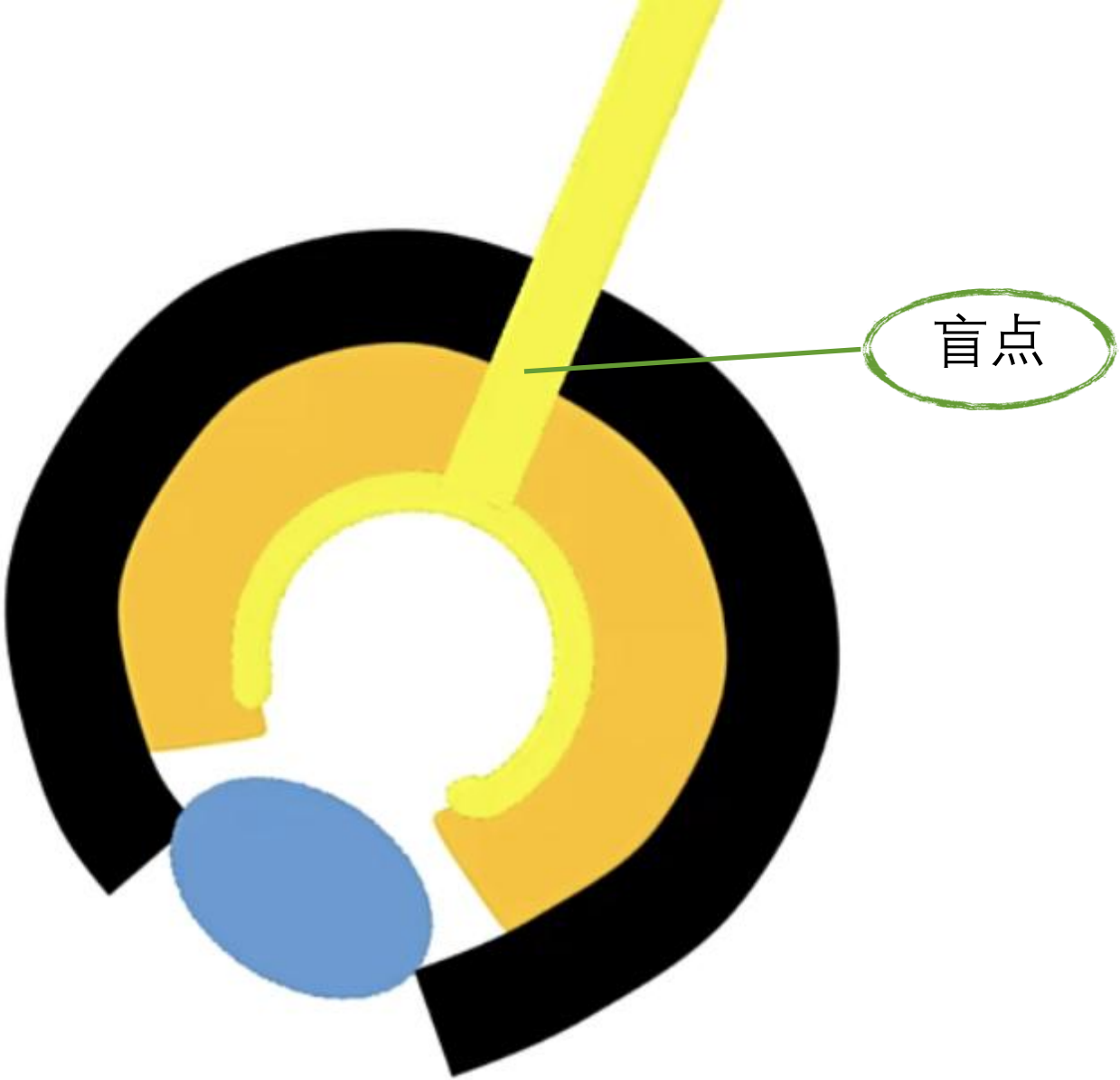

脊索动物

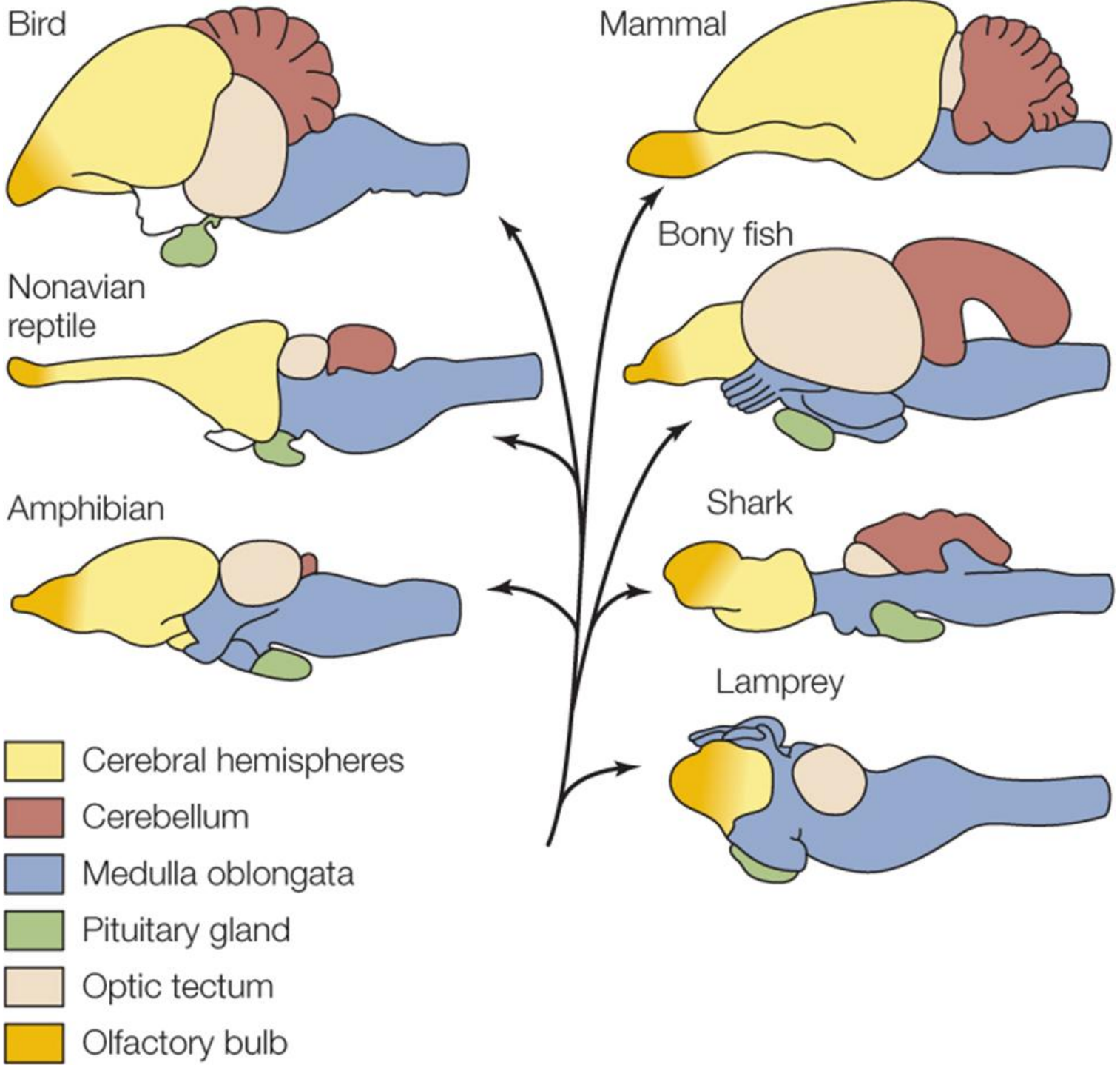









盲点


Bird
Mammal
Bony fish
Nonavian reptile
Amphibian
Shark
Lamprey
Cerebral hemispheres
Cerebellum
Medulla oblongata
Pituitary gland
Optic tectum
Olfactory bulb

人类通过视觉所获得的信息占获取信息总量的70%~80%，而大脑有大约50%的能力都用于处理视觉信息

人类通过视觉所获得的信息占获取信息总量的70%~80%，而大脑有大约50%的能力都用于处理视觉信息

视觉感知和认知是人类智能的核心

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group

July 7, 1966

Vision Memo. No. 100.

THE SUMMER VISION PROJECT

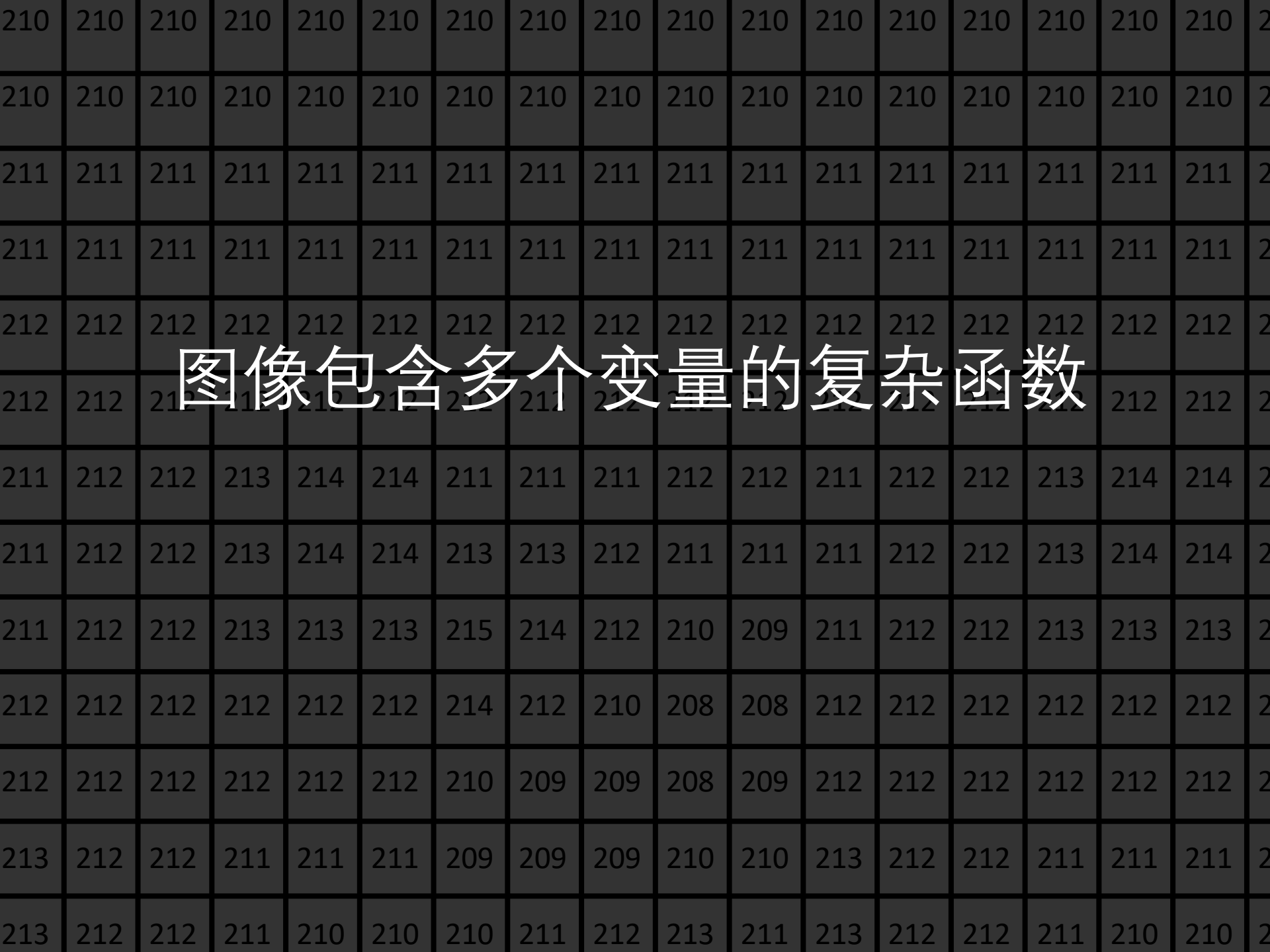
Seymour Papert

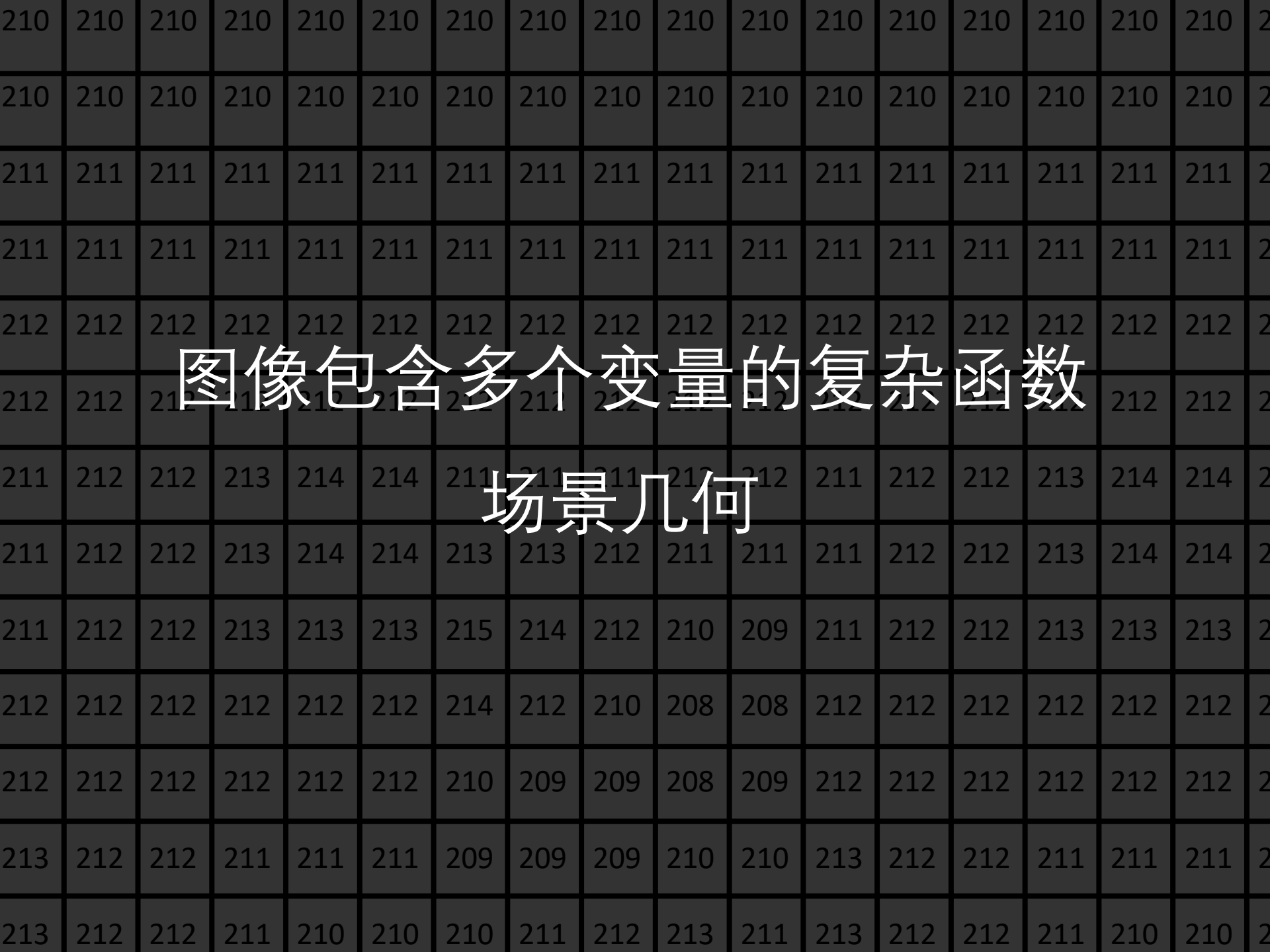
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

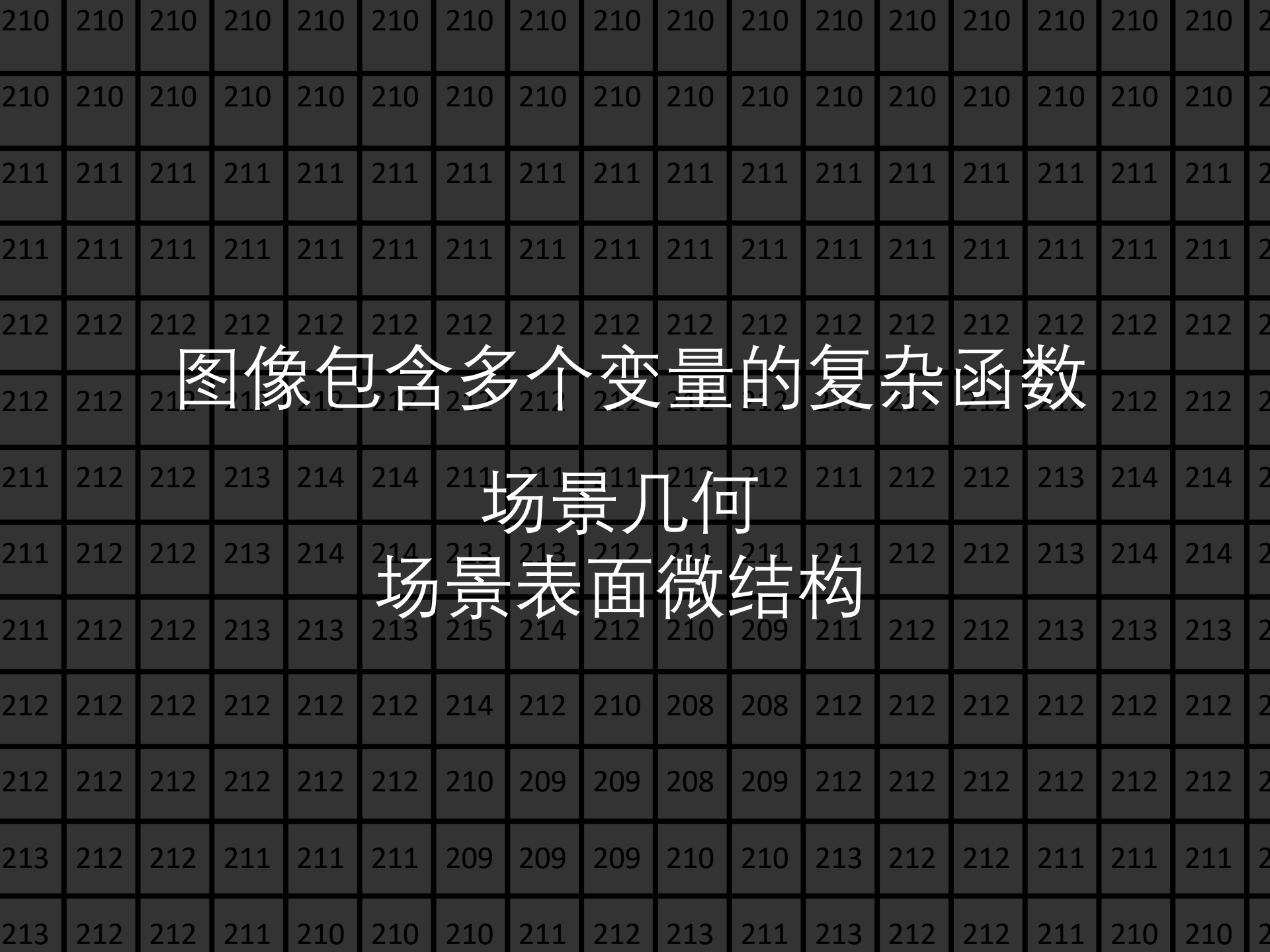
是什么让计算机视觉

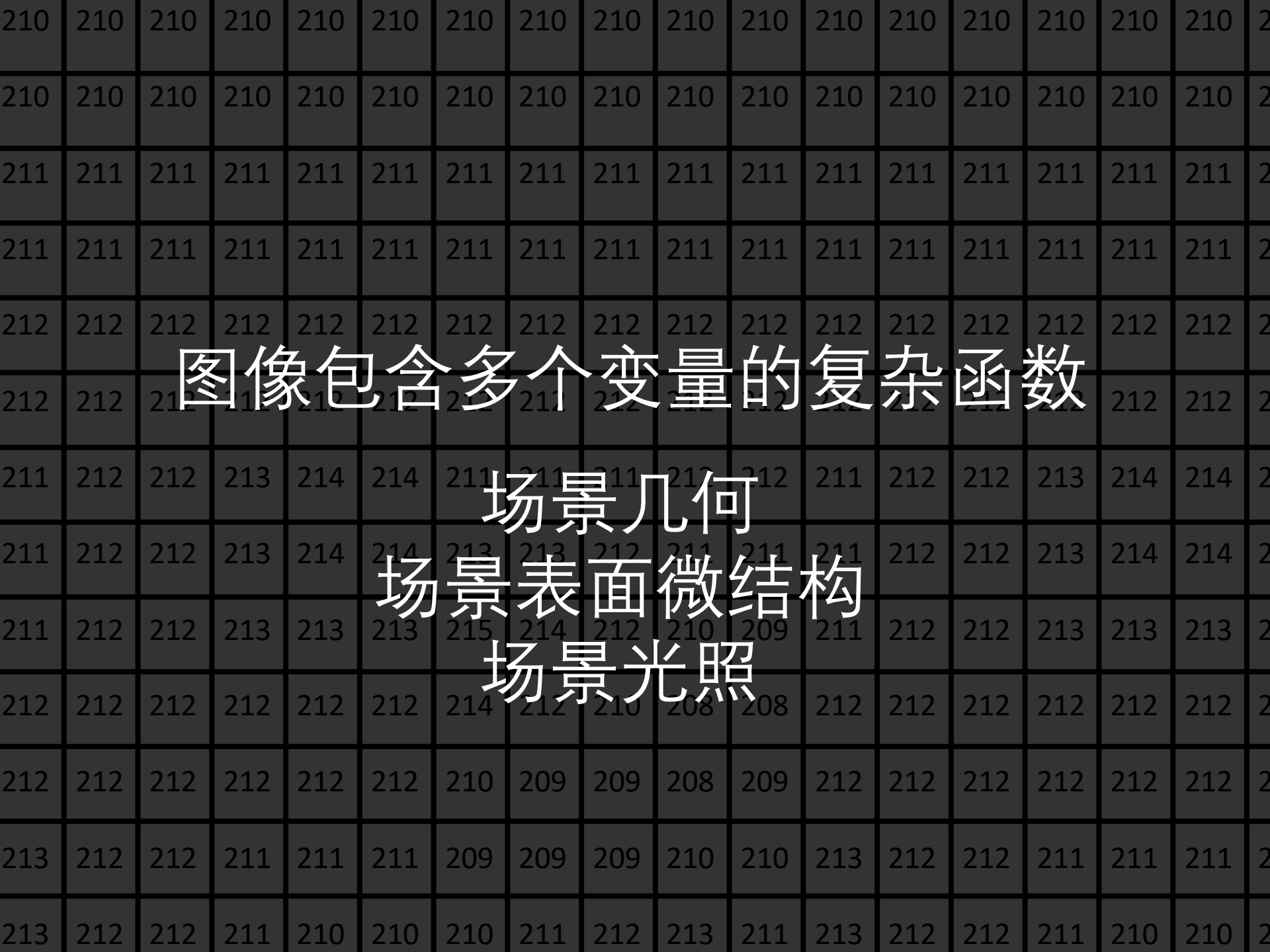
困难？

210	210	210	210	210	210	210	210	210	210	210	210	210	210	210	210	210
210	210	210	210	210	210	210	210	210	210	210	210	210	210	210	210	210
211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211
211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211	211
212	212	212	212	212	212	212	212	212	212	212	212	212	212	212	212	212
212	212	212	212	212	212	212	212	212	212	212	212	212	212	212	212	212
211	212	212	213	214	214	211	211	211	212	212	211	212	212	213	214	214
211	212	212	213	214	214	213	213	212	211	211	211	212	212	213	214	214
211	212	212	213	213	213	215	214	212	210	209	211	212	212	213	213	213
212	212	212	212	212	212	214	212	210	208	208	212	212	212	212	212	212
212	212	212	212	212	212	210	209	209	208	209	212	212	212	212	212	212
213	212	212	211	211	211	209	209	209	210	210	213	212	212	211	211	211
213	212	212	211	210	210	210	211	212	213	211	213	212	212	211	210	210

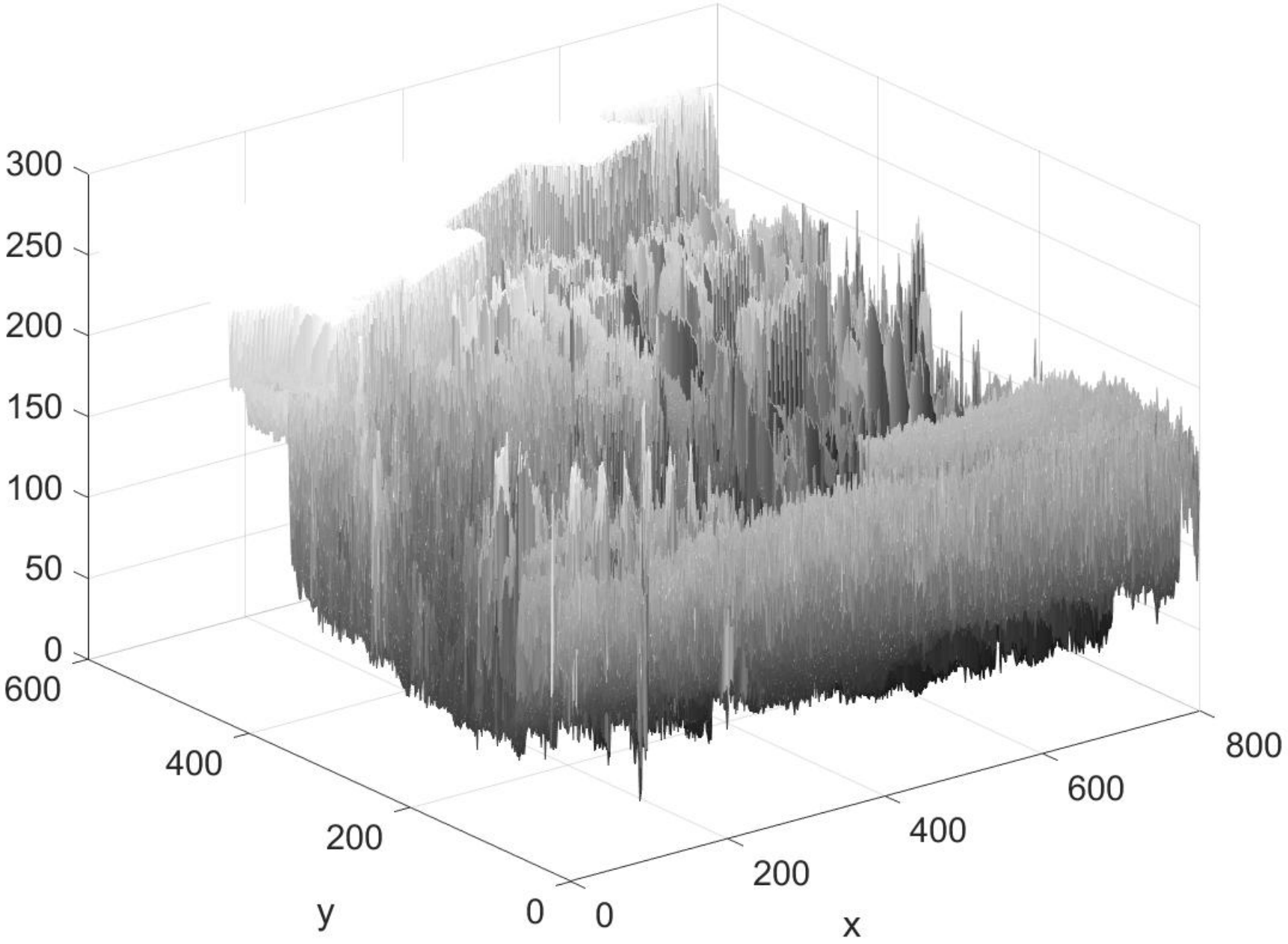

图像包含多个变量的复杂函数

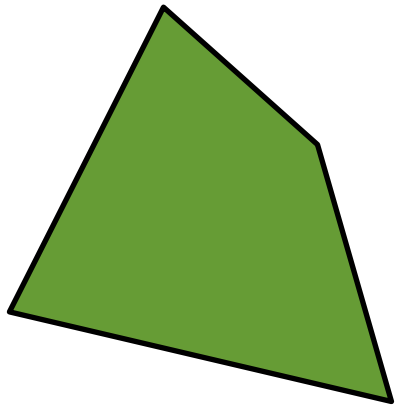

图像包含多个变量的复杂函数
场景几何

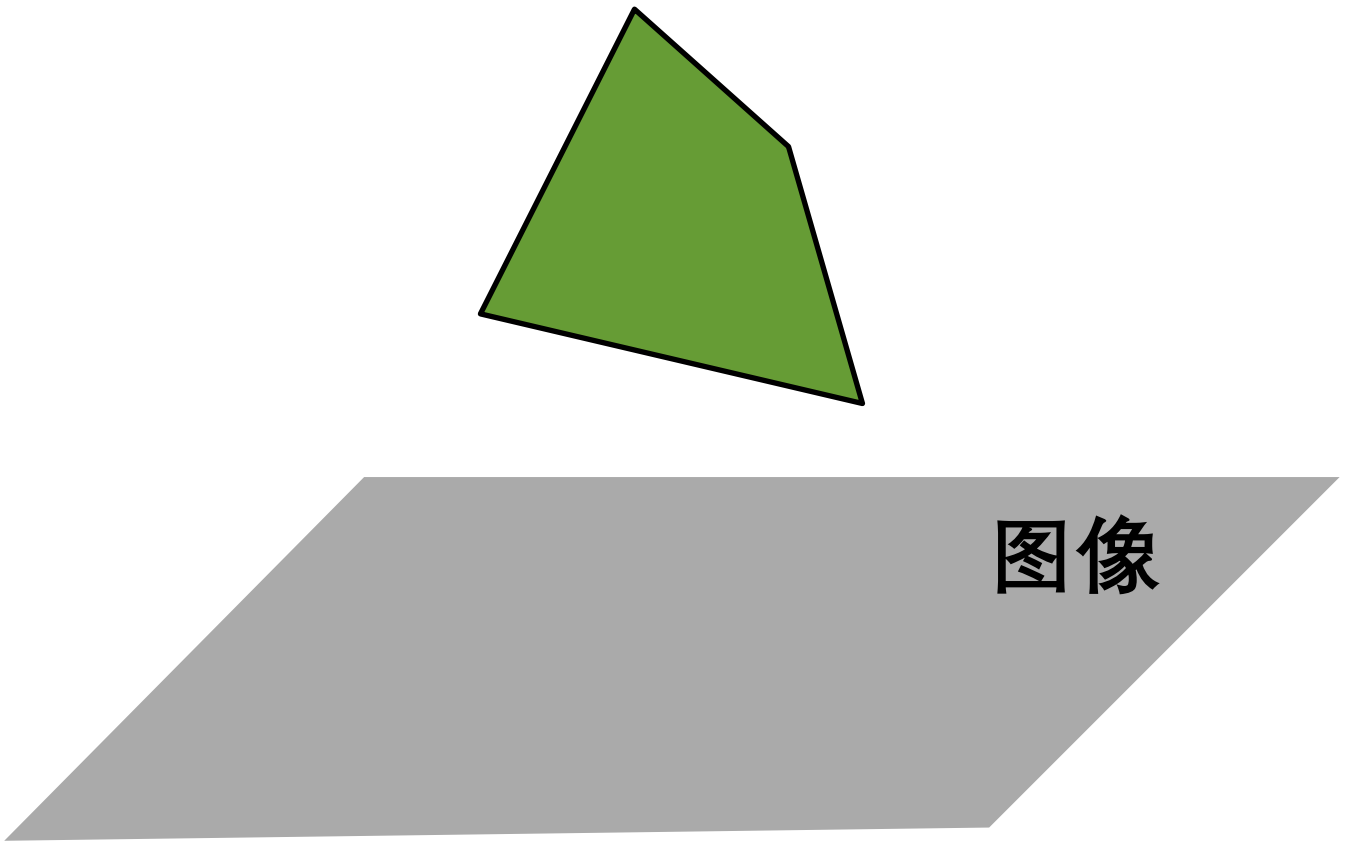

图像包含多个变量的复杂函数
场景几何
场景表面微结构

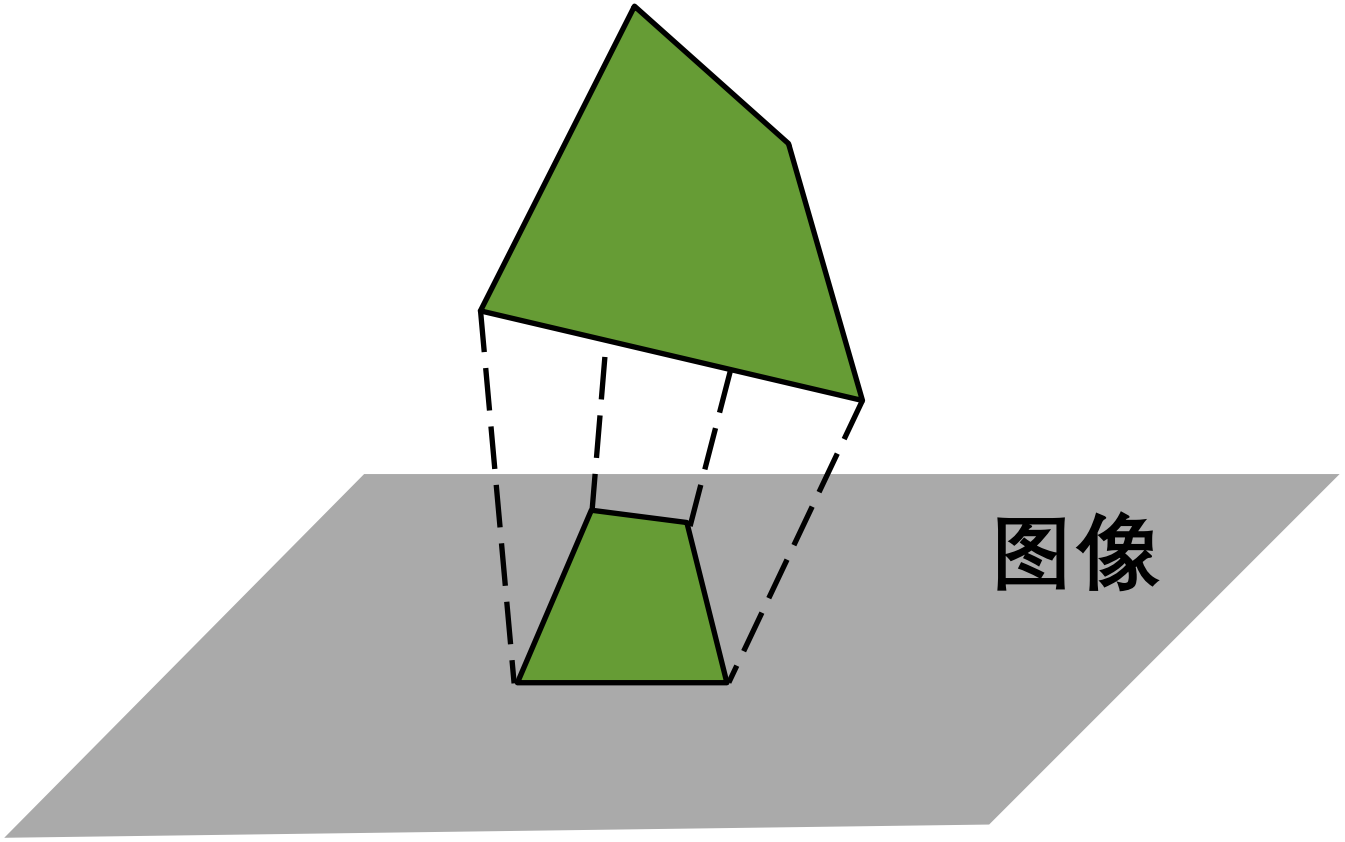

图像包含多个变量的复杂函数
场景几何
场景表面微结构
场景光照

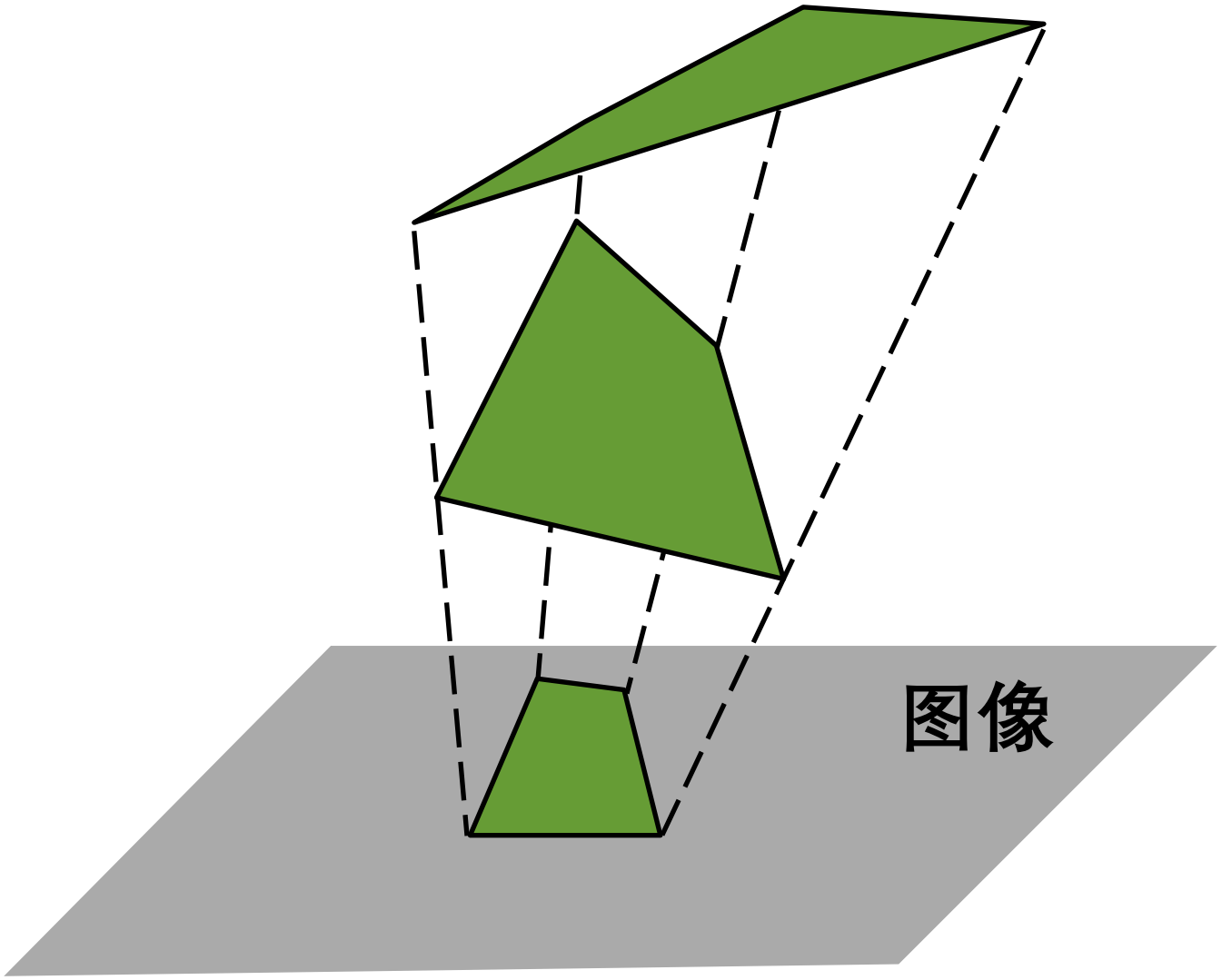

y
x

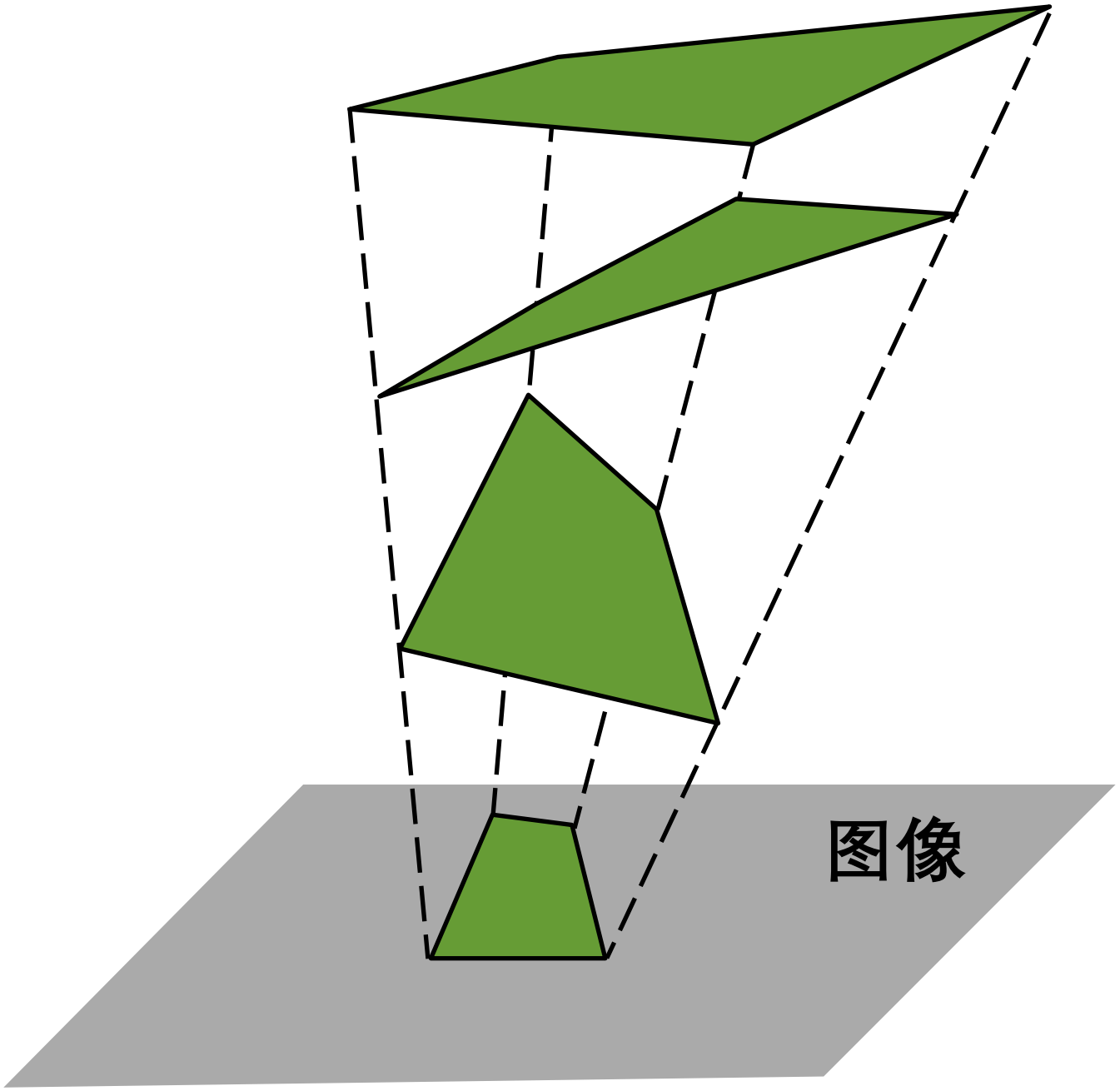

300
250
200
150
100
50
0
600
400
200
0
y
0
200
400
600
800
x

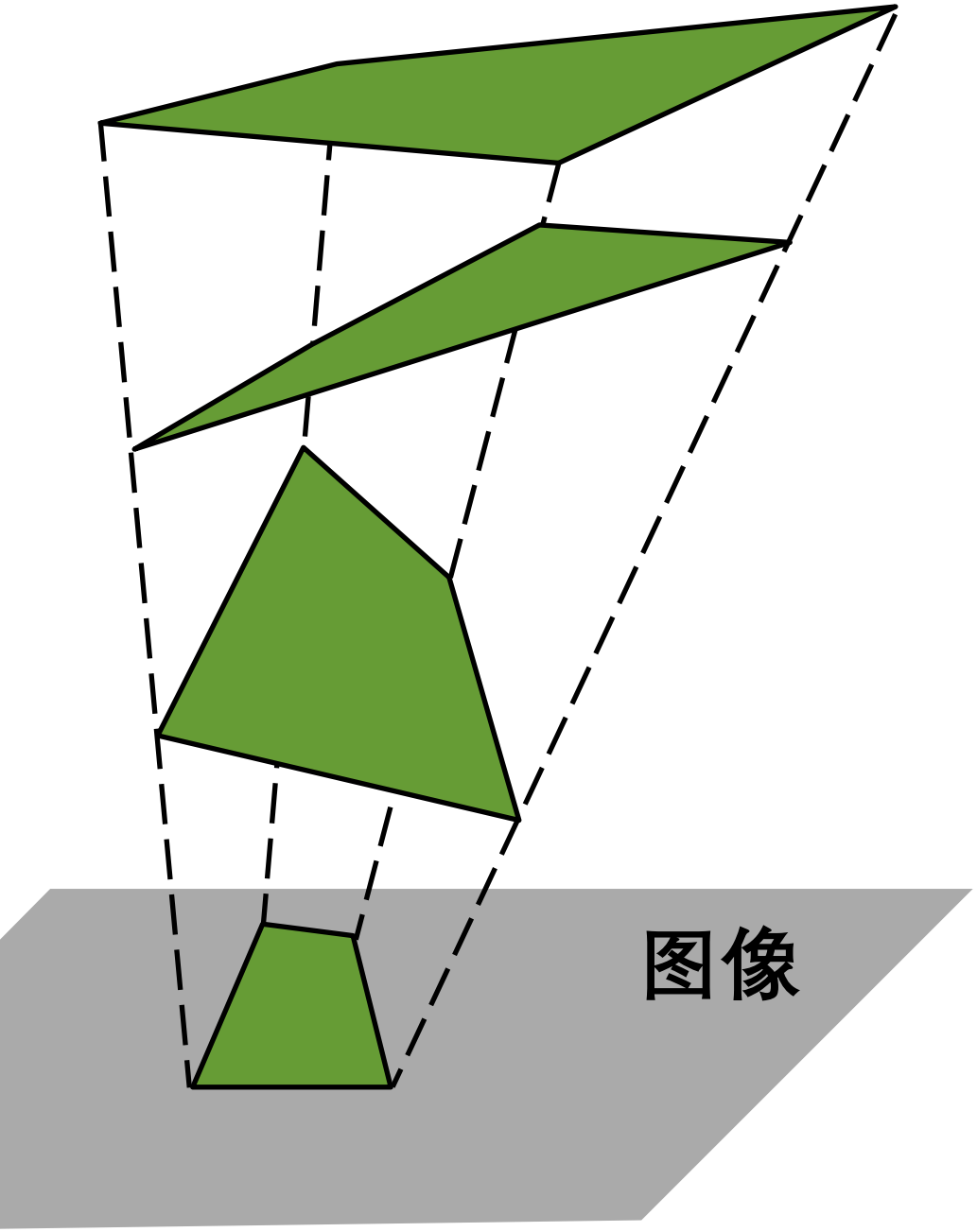



图像


图像


图像

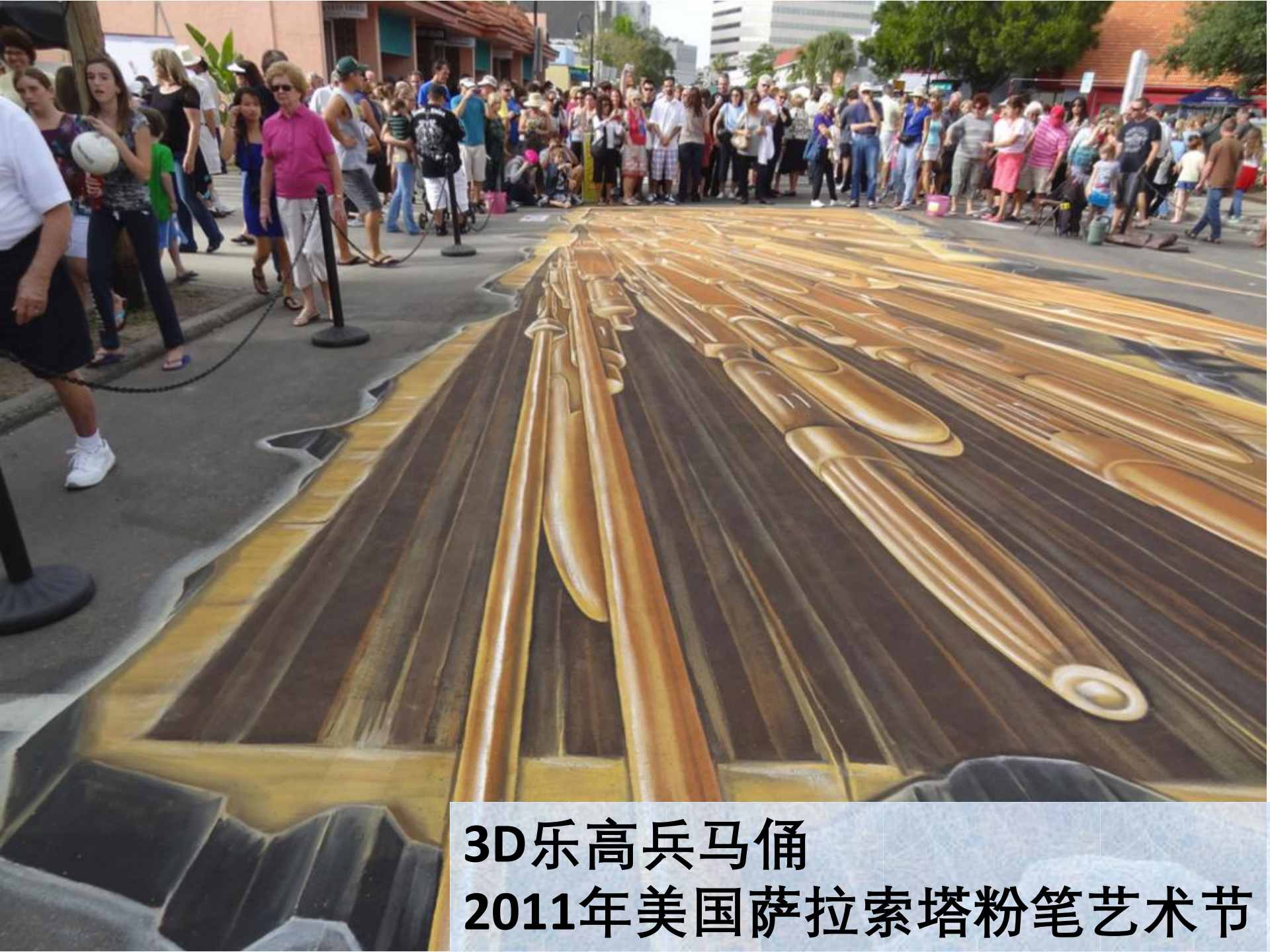

图像



无限多个可能的形状

北斗七星3D动画

3D乐高兵马俑
2011年美国萨拉索塔粉笔艺术节



3D乐高兵马俑
2011年美国萨拉索塔粉笔艺术节

本田汽车广告《不可能成就可能》

《不可能成就可能》幕后花絮

物体具有高度可变的外观

视角的差异

不同的大小

不同的实例

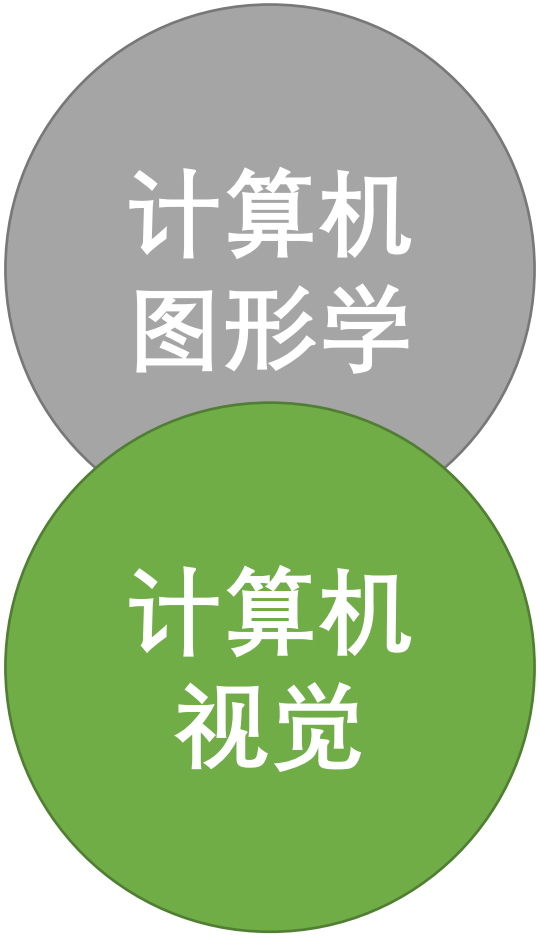

遮挡

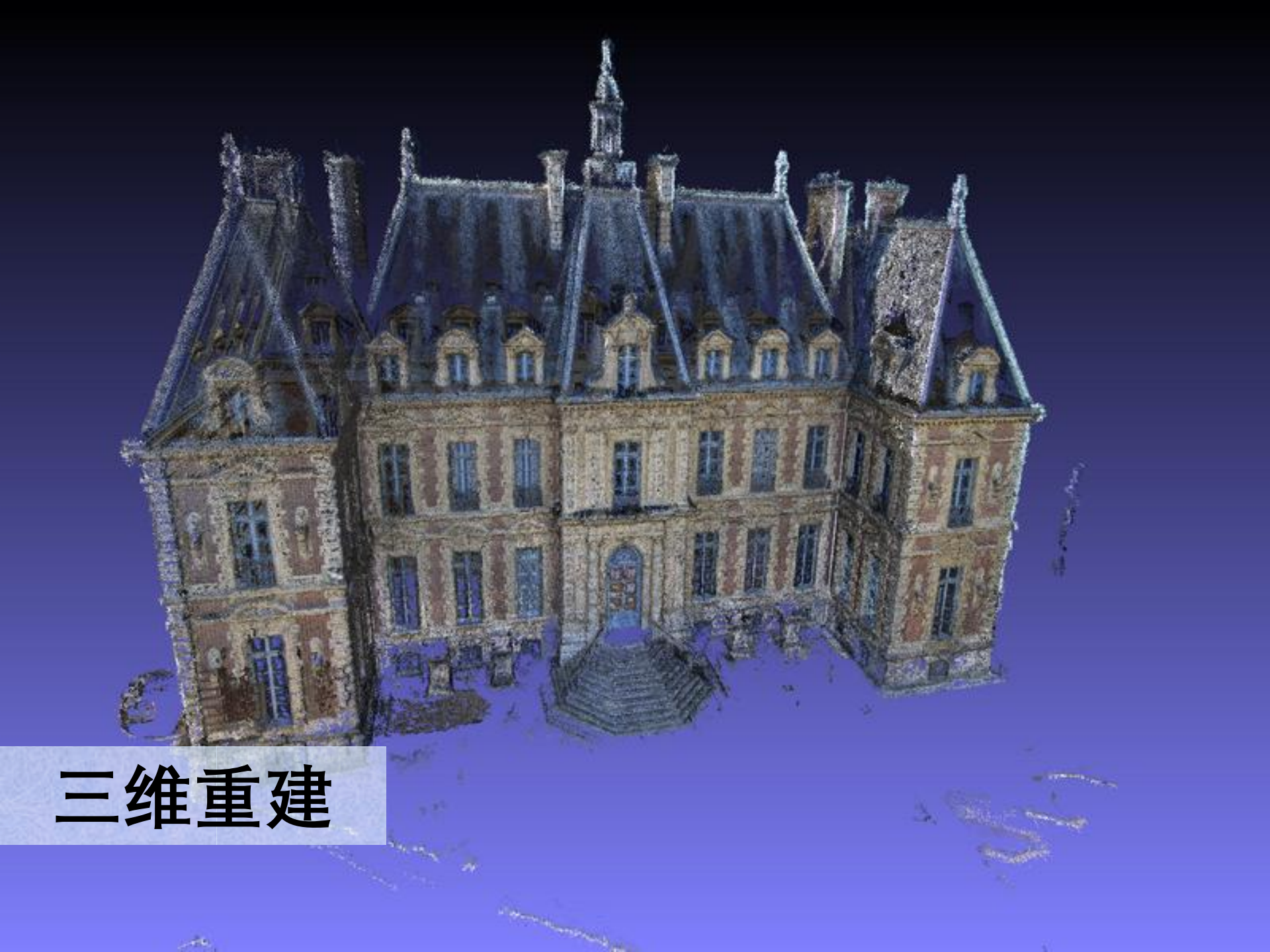
形变与铰接

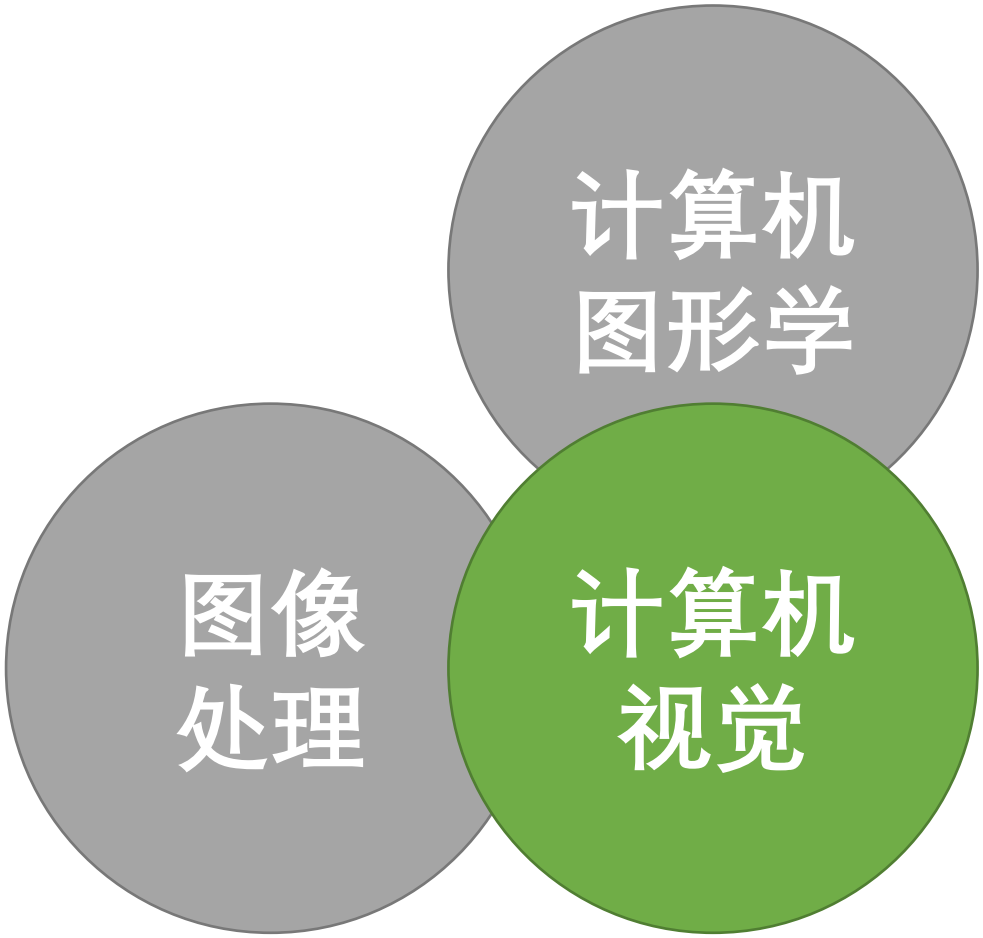

背景干扰

光照

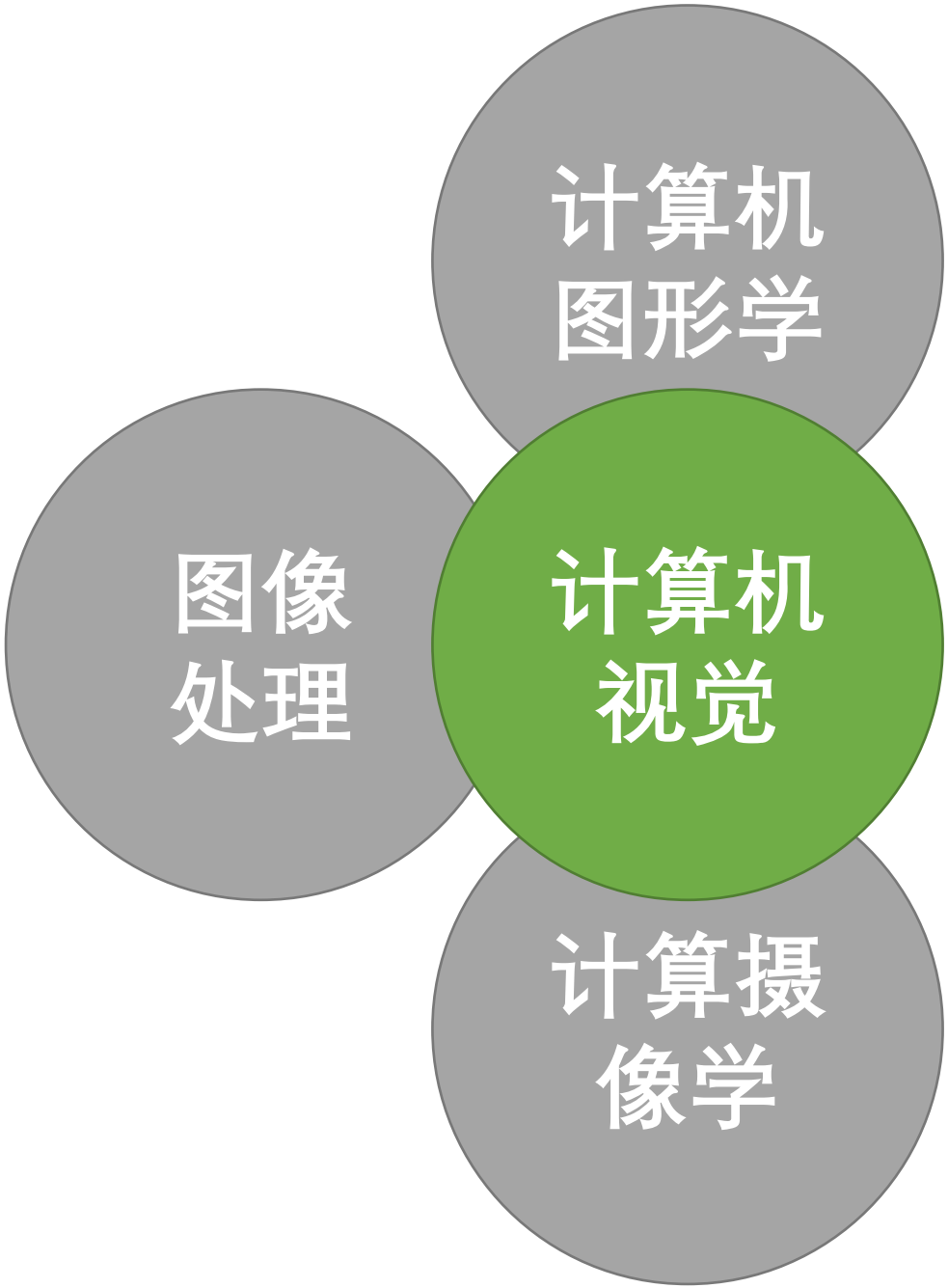
计算机视觉


计算机
图形学
计算机
视觉


三维重建


计算机
图形学
图像
处理
计算机
视觉


图像去噪


计算机
图形学
图像
处理
计算机
视觉
计算摄
像学

HDR

曝光值 (EV) : -7

EV : 0

EV : +5

HDR

全景照片


计算机
图形学
图像
处理
计算机
视觉
机器
学习
计算摄
像学

输入 \ 输出	图像	知识
图像	图像处理	计算机视觉
知识	计算机图形学	机器学习

研究领域


特征检测

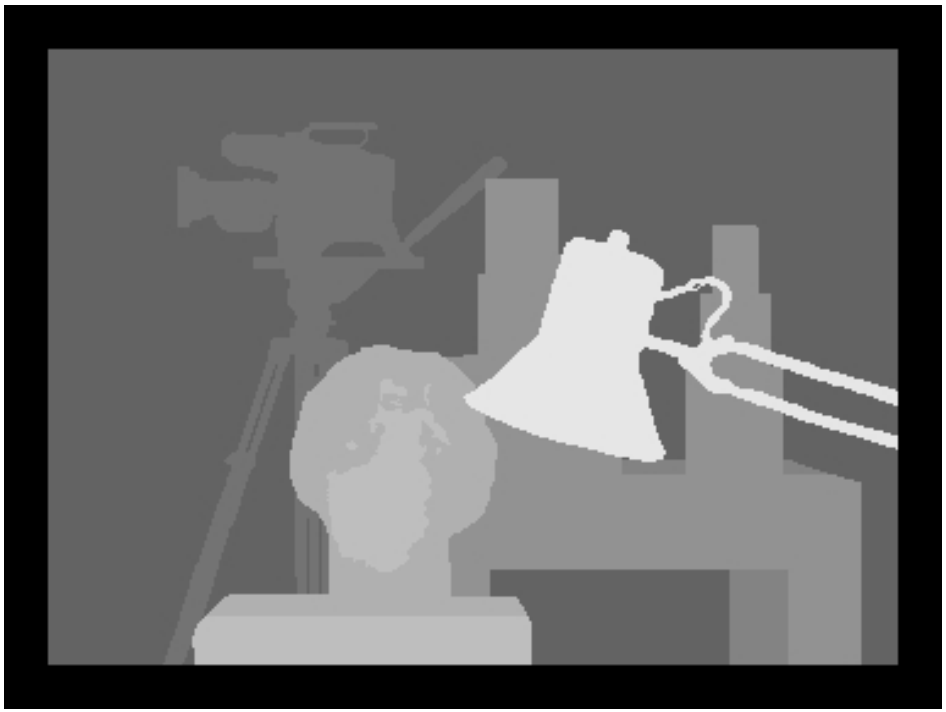

特征检测


Shape-from-X

立体视觉

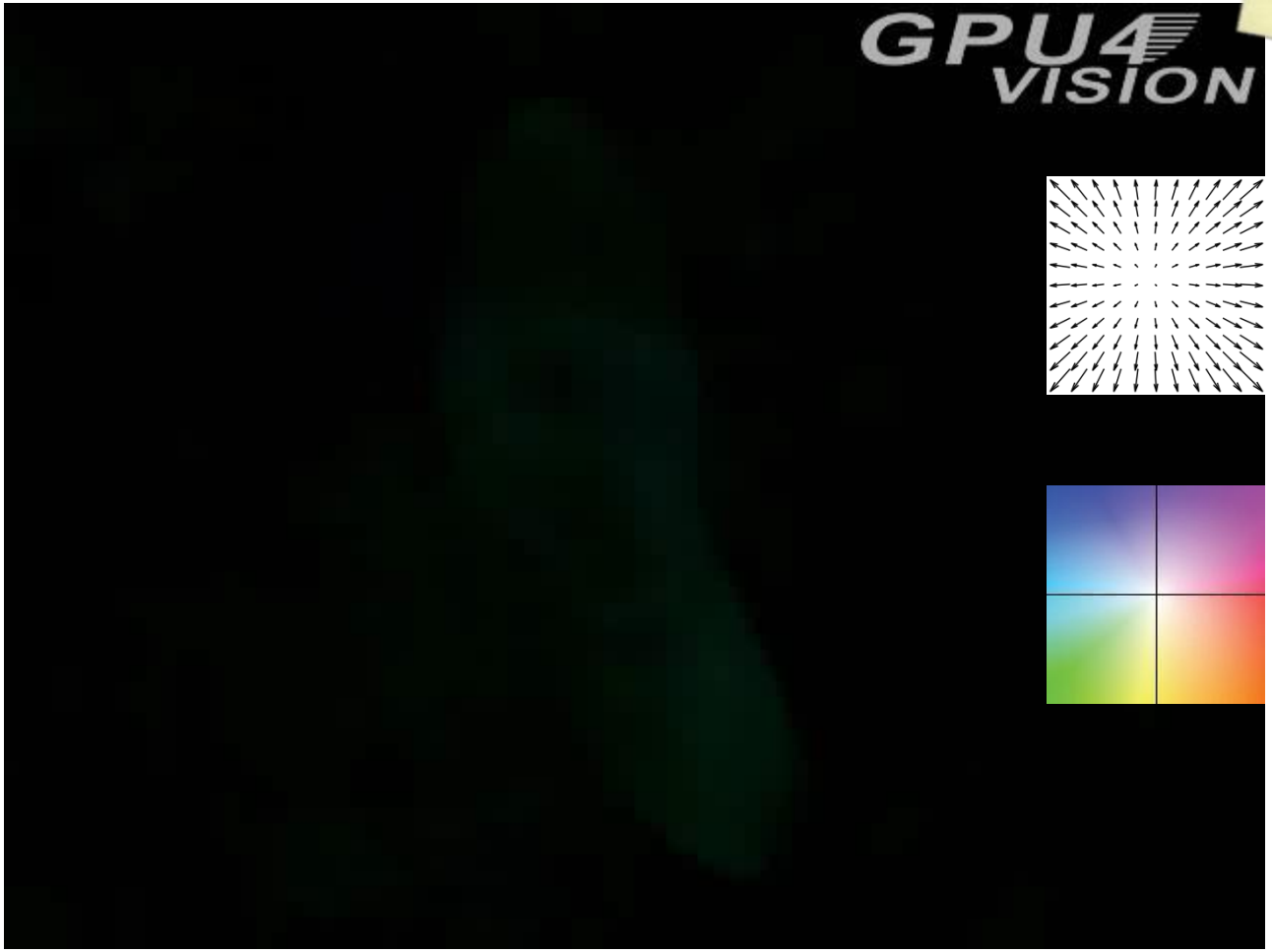
立体视觉

立体视觉



深度图

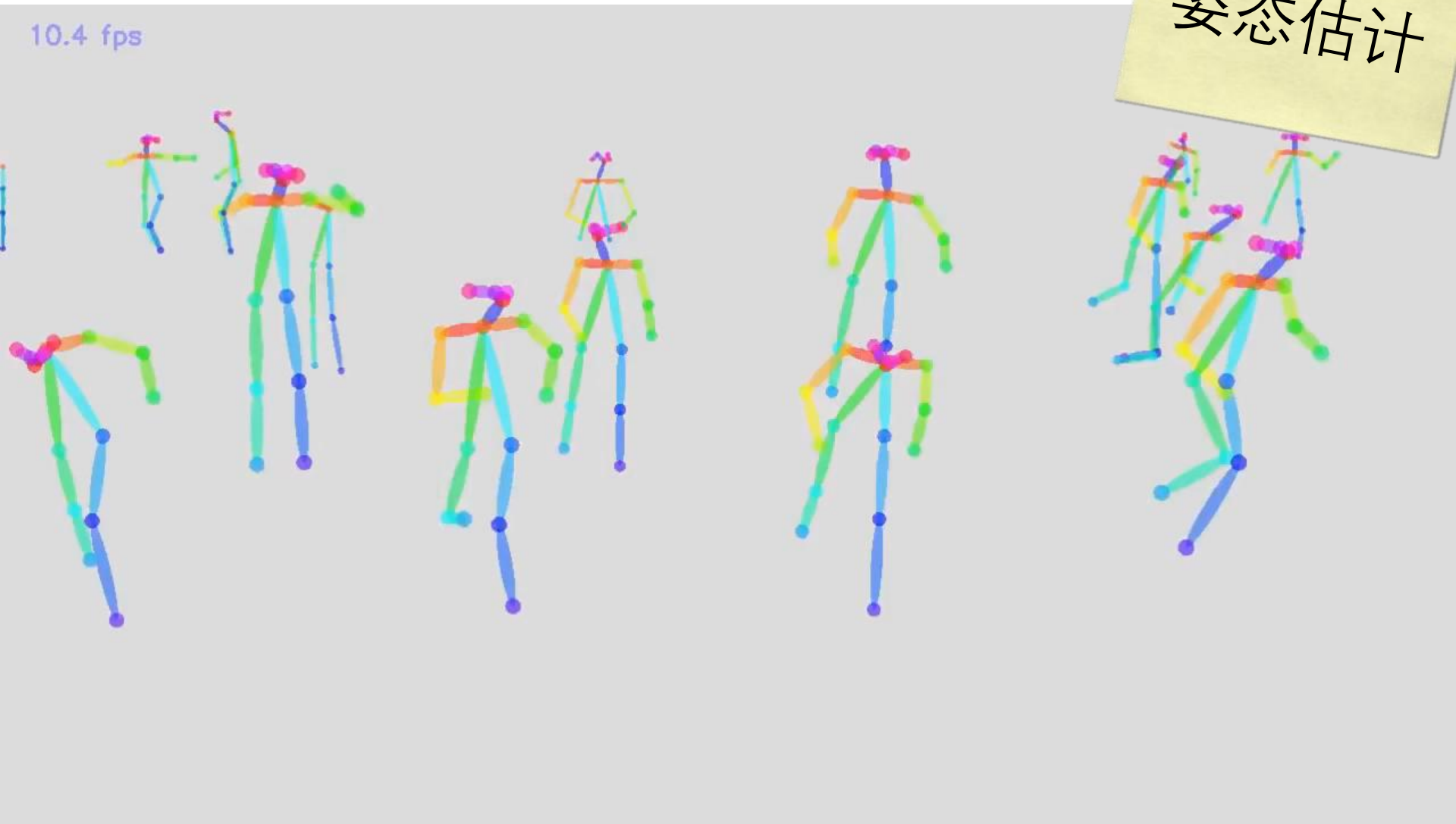
运动分析



鸣谢：gpu4vision

鸣谢：Anton Andriyenko

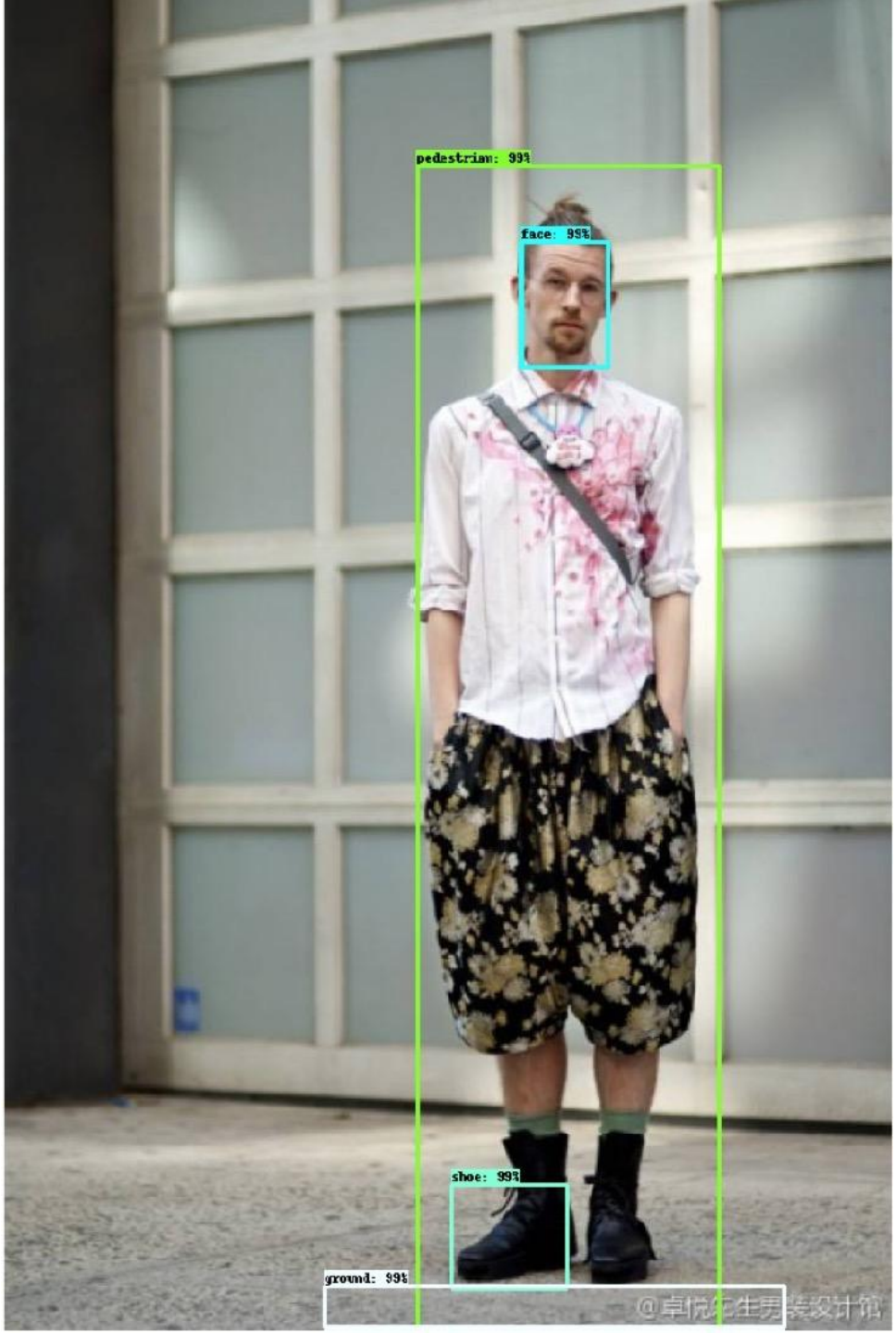
鸣谢：Akshay Asthana



姿态估计

鸣谢：Zhe Cao

物体识别



细粒度
物体识别

Gender:

male 85.89%

Shoes:

leather shoe 99.94%

Ground:

cement court 99.35%

鸣谢：夏开阳


当用户上传照片时，我们的app要能判断出用户是否身处一个国家公园


没问题，只需要简单的地理信息系统查找即可实现，给我几个小时


还要判断照片里有没有鸟


那我需要一个研究团队，至少得5年时间

在计算机科学中，可能很难解释简单和几乎不可能的区别

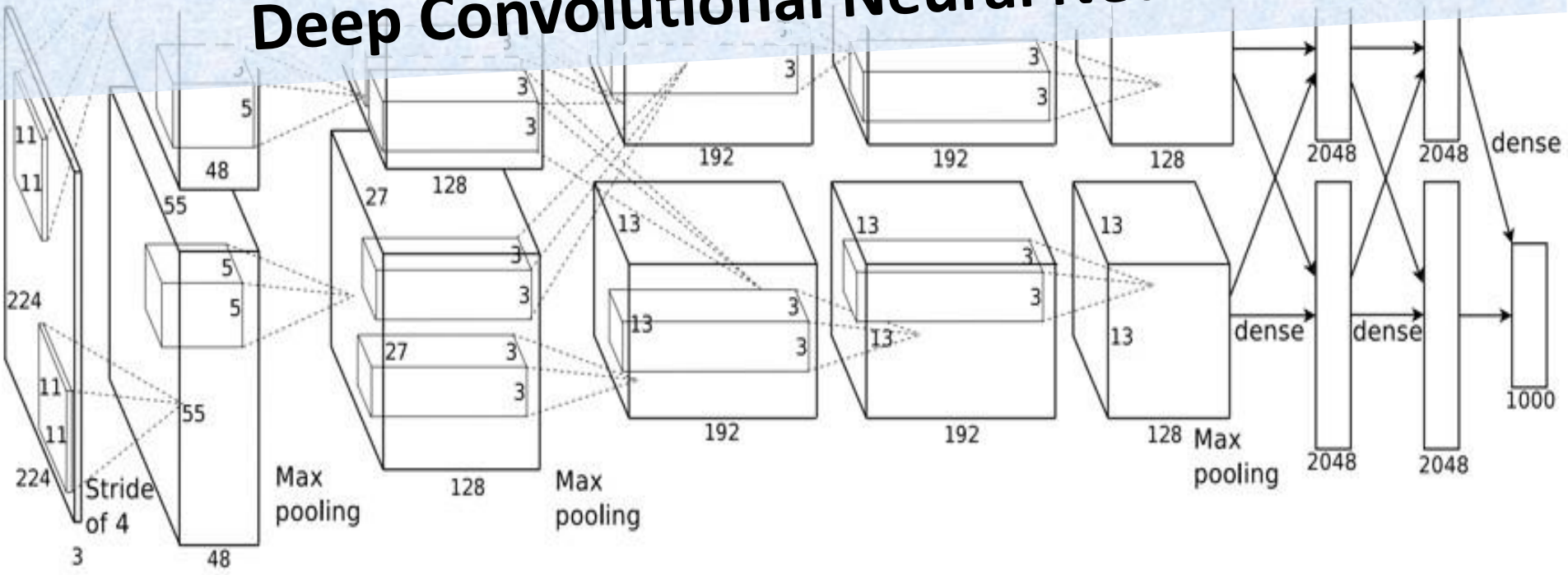

We need to go deeper


深度学习学派


Geoff
Yann
深度学习学派

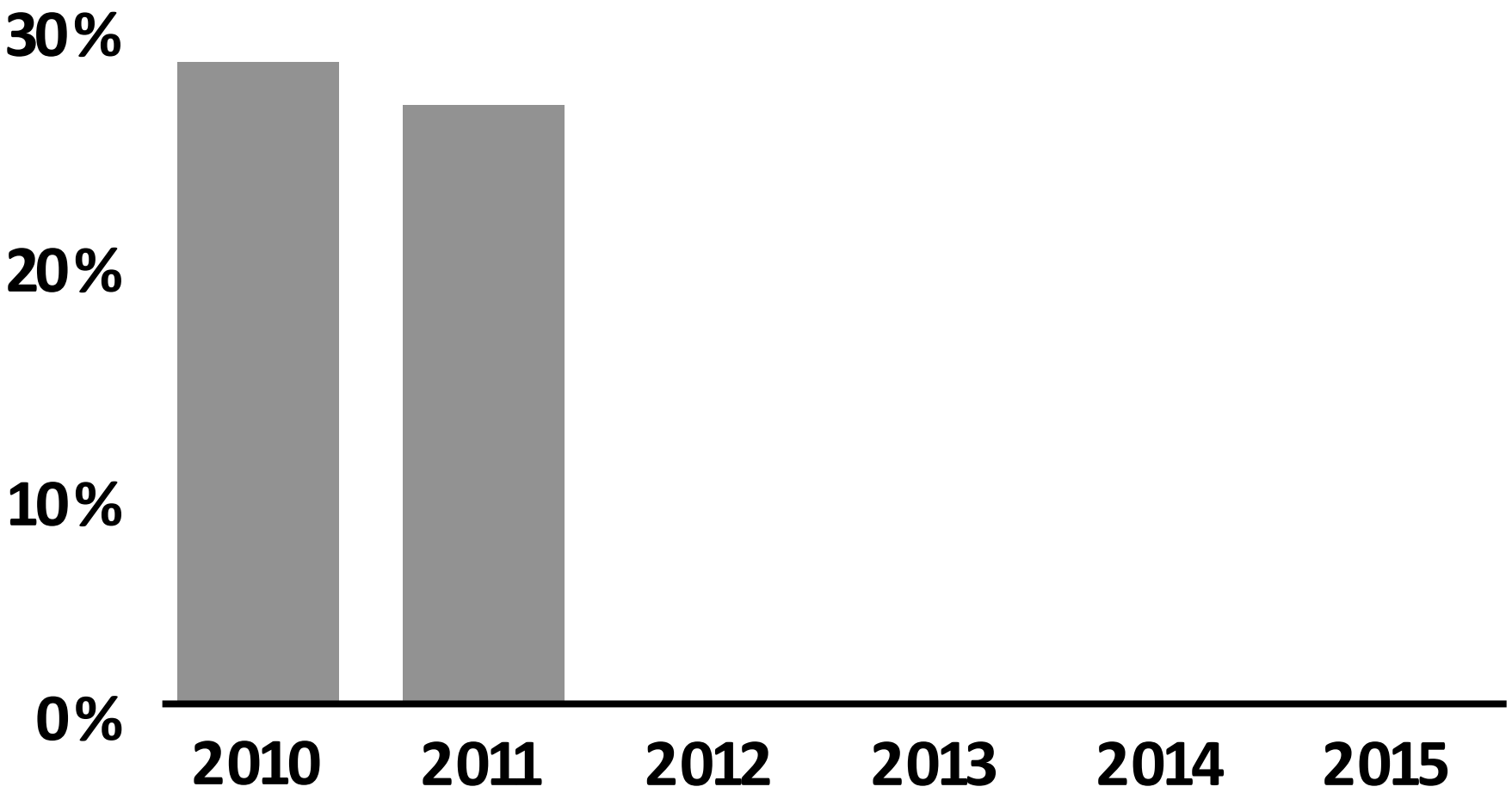

2012

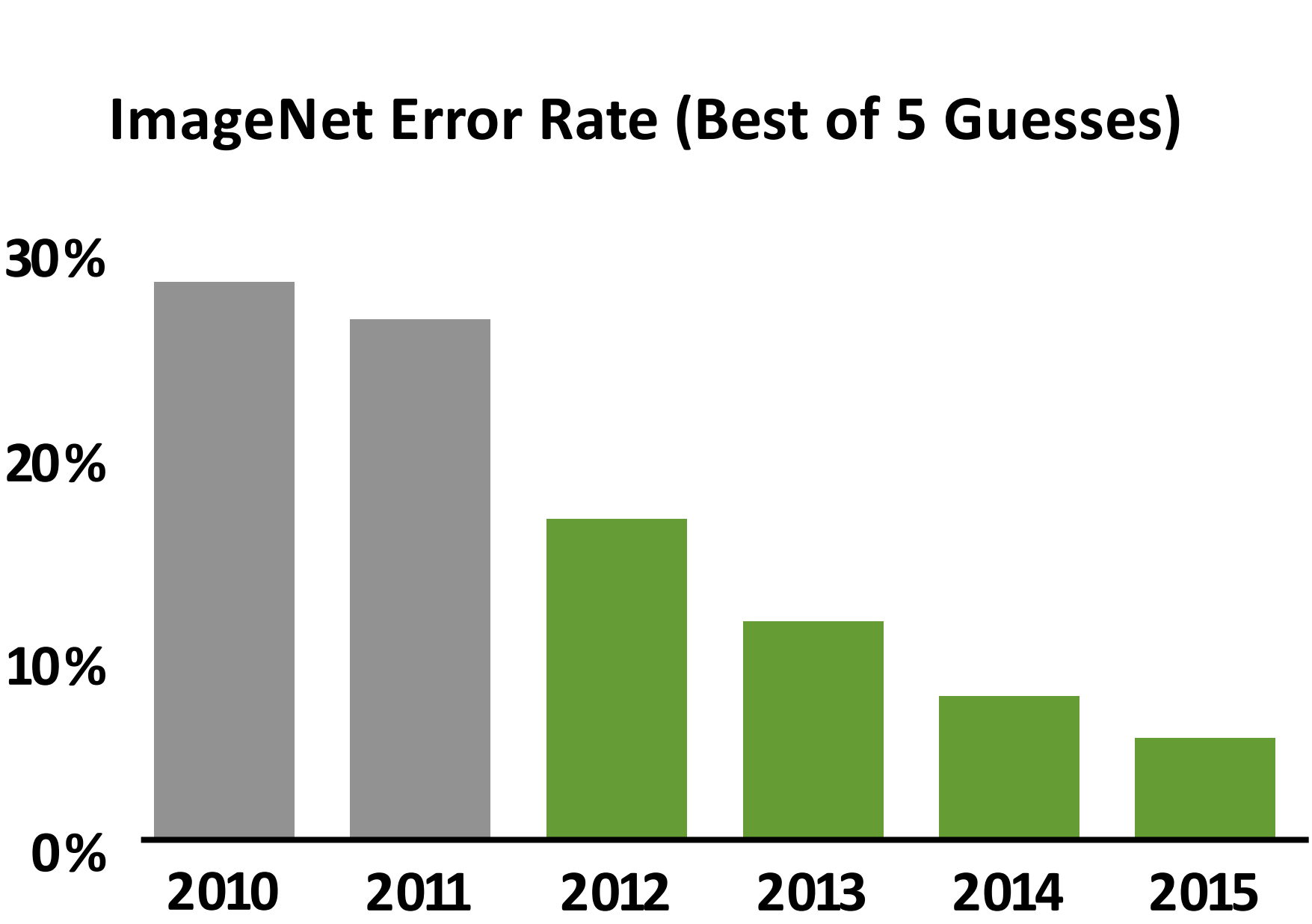
Deep Convolutional Neural Network




ImageNet
1000类，100万图像

ImageNet Error Rate (Best of 5 Guesses)




ImageNet Error Rate (Best of 5 Guesses)
30%
20%
10%
0%
2010
2011
2012
2013
2014
2015

Park or Bird?

parkorbird.flickr.com

Reader

Park or Bird?

code.flickr.com

flickr.com Flickr jobs!

To play, drag an image from the examples or from your desktop.

EXAMPLE PHOTOS

Photo credits

PARK or BIRD

Want to know if your photo is from a U.S. national park? Want to know if it contains a bird? Just drag it into the box to the left, and we'll tell you. We'll use the GPS embedded in your photo (if it's there) to see whether it's from a park, and we'll use our super-cool computer vision skills to try to see whether it's a bird (which is a hard problem, but we do a pretty good job at it).

To try it out, just drag any photo from your desktop into the upload box, or try dragging any of our example images. We'll give you your answers below!

Want to know more about PARK or BIRD, including why the heck we did this? Just click here for more info →

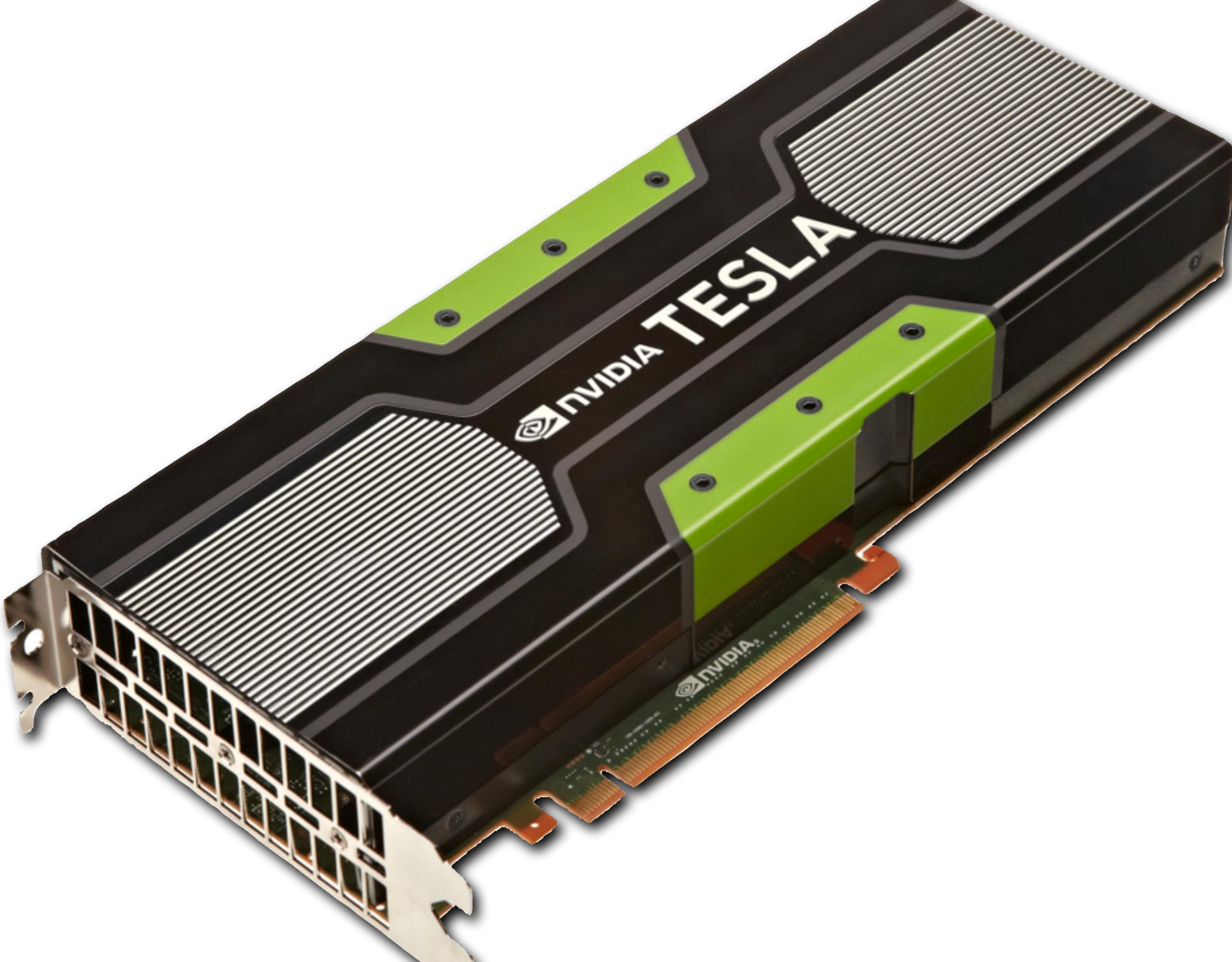
PARK?

BIRD?

“我们以前来过这里！”


BACK
TO THE FUTURE
TM

“发生了什么变化？”


NVIDIA TESLA


Use the GeForce, Luke

The New York Times

ROBOTICS

Computer Eyesight Gets a Lot More Accurate

By John Markoff

August 18, 2014 8:01 pm

Just as the Big Bad Wolf promised Little Red Riding Hood that his bigger eyes were “the better to see you with,” a machine’s ability to see the world around it is benefiting from bigger computers and more accurate mathematical calculations.

The improvement was visible in contest results released Monday evening by computer scientists and companies that sponsor an annual challenge to measure improvements in the state of machine vision technology.

Started in 2010 by Stanford, Princeton and Columbia University scientists, the Large Scale Visual Recognition Challenge this year drew 38 entrants from 13 countries. The groups use advanced software, in most cases modeled loosely on the biological vision systems, to detect, locate and classify a huge set of images taken from Internet sources like Twitter. The contest was sponsored this year by Google, Stanford, Facebook and the University of North Carolina.

Contestants run their recognition programs on high-performance computers based in many cases on specialized processors called G.P.U.s, for graphic processing units.

This year there were six categories based on object detection, locating objects and classifying them. Winners included the National University of Singapore, the Oxford University, Adobe Systems, the Center for Intelligent Perception and Computing at the Chinese Academy of Sciences, as well as Google in two separate categories.

Accuracy almost doubled in the 2014 competition and error rates were cut in

The New

ROBOTICS

Comp

By John Ma

August

MIT Technology Review

Facebook Creates Software That Matches Faces Almost as Well as You Do

Facebook's new AI research group reports a major improvement in face-processing software.

By Tom Simonite on March 17, 2014

Asked whether two unfamiliar photos of faces show the same person, a human being will get it right 97.53 percent of the time. New software developed by researchers at Facebook can score 97.25 percent on the same challenge, regardless of variations in lighting or whether the person in the picture is directly facing the camera.

That's a significant advance over previous face-matching software, and it demonstrates the power of a new approach to artificial intelligence known as deep learning, which Facebook and its competitors have bet heavily on in the past year (see "Deep Learning"). This area of AI involves software that uses networks of simulated neurons to learn to recognize patterns in large amounts of data.

"You normally don't see that sort of improvement," says Yaniv Taigman, a member of Facebook's AI team, a research group created last year to explore how deep learning might help the company (see "Facebook Launches Advanced AI Effort"). "We closely approach human performance," says Taigman of the new software. He notes that the error rate has been reduced by more than a quarter relative to earlier software that can take on the same task.

The New

ROBOTICS

Comp

By John Ma

August

MIT Techno Review

Facebook Faces Alm

Facebook's new processing softv

By Tom Simonite on Marc

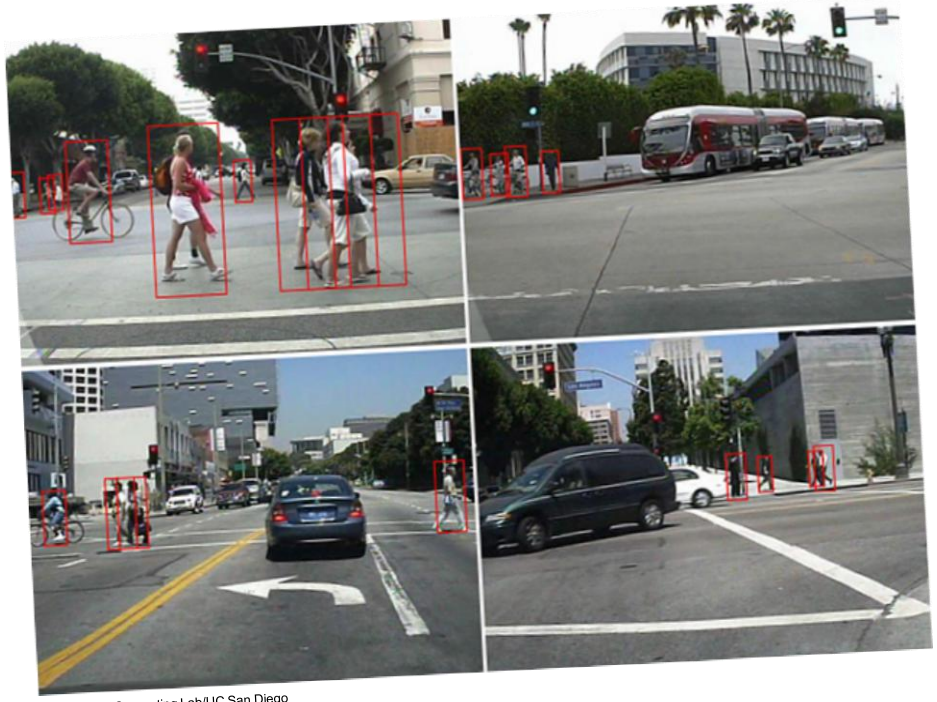
Asked whether two unf 97.53 percent of the tim percent on the same ch directly facing the came

That's a significant advar new approach to artificia bet heavily on in the past networks of simulated neu

"You normally don't see tha team, a research group cre "Facebook Launches Advar the new software. He notes earlier software that can tak

Deep Learning Makes Driverless Cars Better at Spotting Pedestrians

By Jeremy Hsu
Posted 9 Feb 2016 | 17:00 GMT



Images: Statistical Visual Computing Lab/UC San Diego

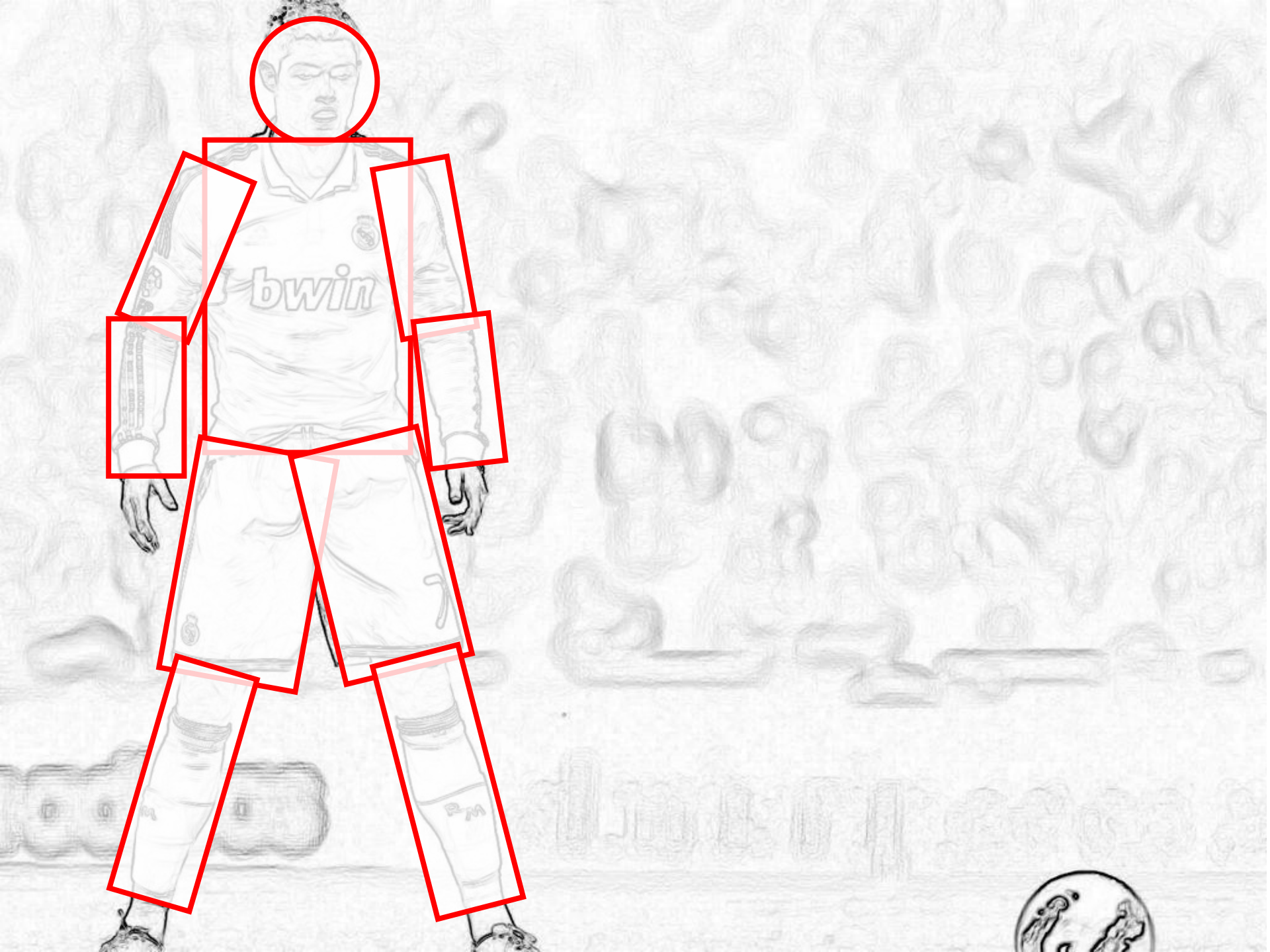
Today's car crash-avoidance systems and experimental driverless cars rely on radar and other sensors to detect pedestrians on the road. The next improvement may come from engineers at the University of California, San Diego (UCSD), who have developed a pedestrian detection system that can perform in close to real-time based on visual cues alone. This video-only detection could make systems for spotting pedestrians both cheaper and more effective.

Such a vision-based safety system has remained elusive in cars because computers typically face a tradeoff between analyzing video images quickly and drawing the right conclusions. On the one hand, a simple "cascade detection" computer vision algorithm can quickly detect many pedestrians in certain images, but lacks the sophistication to distinguish between pedestrians and similar-looking objects in the toughest cases. On the other hand, machine learning algorithms called deep plex pattern recognition, but work too slowly for real-time pedestrian detection.

手工打造的算法


bwin
7


bwin

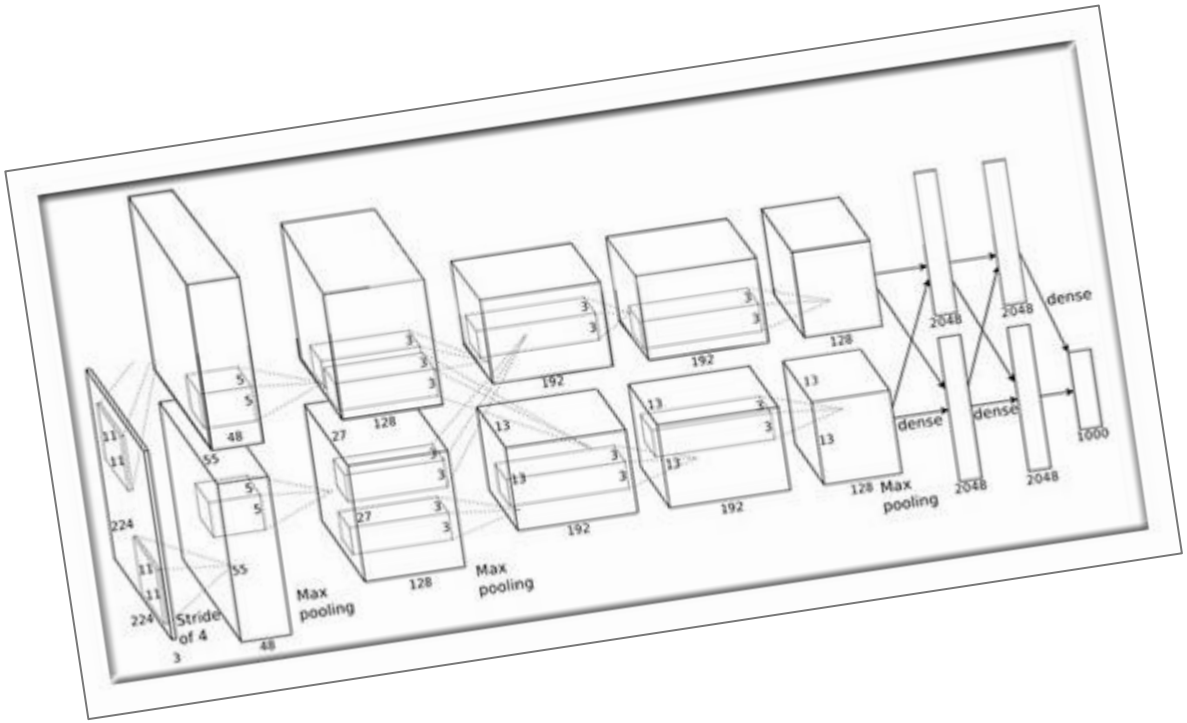

bwin

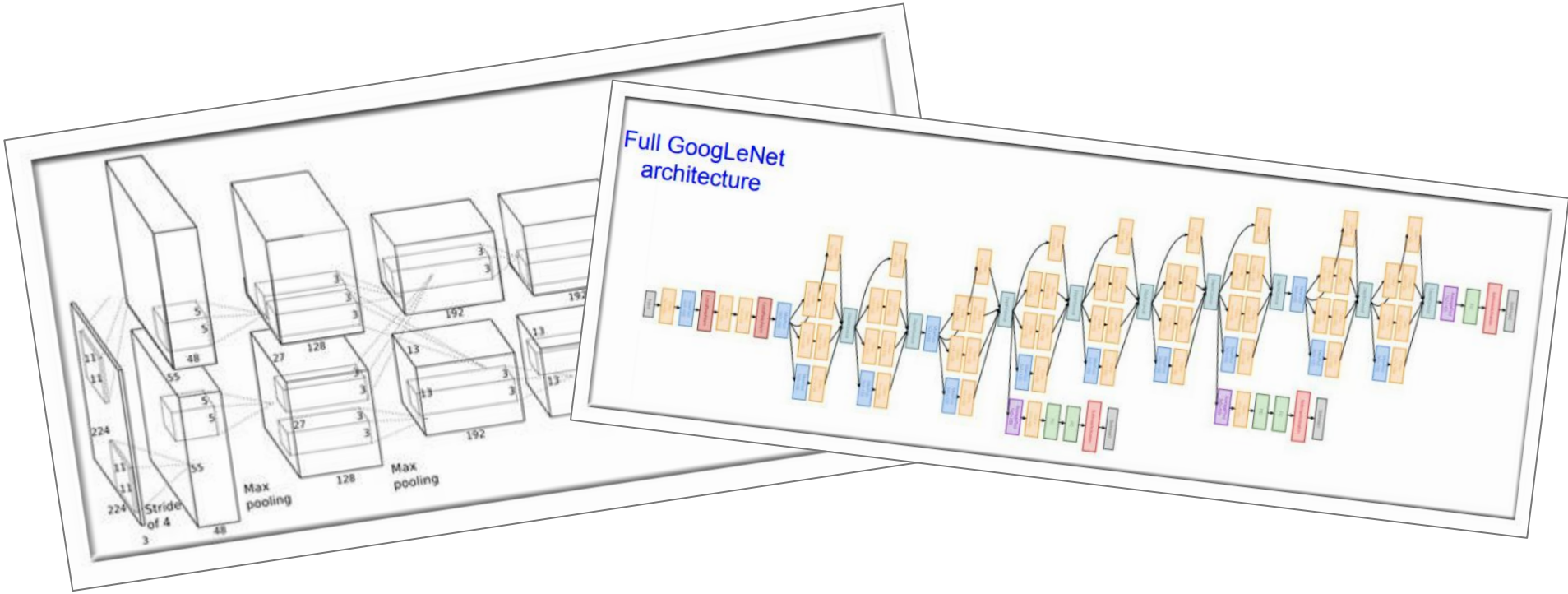
特征工程

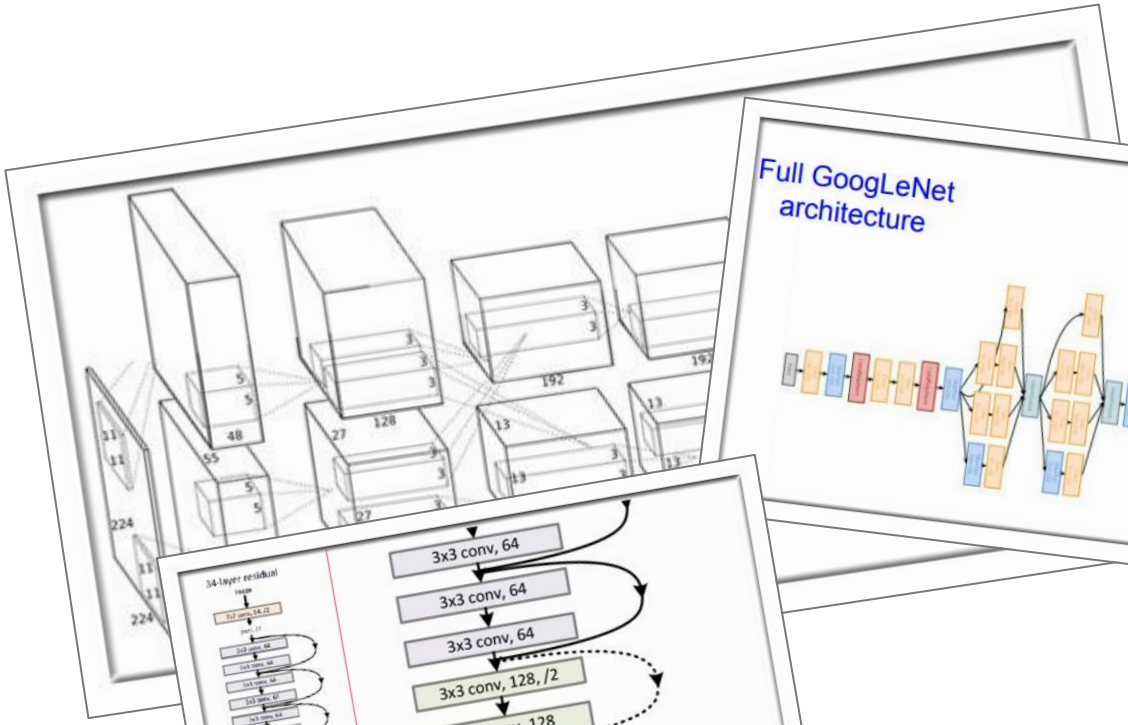

黑盒子

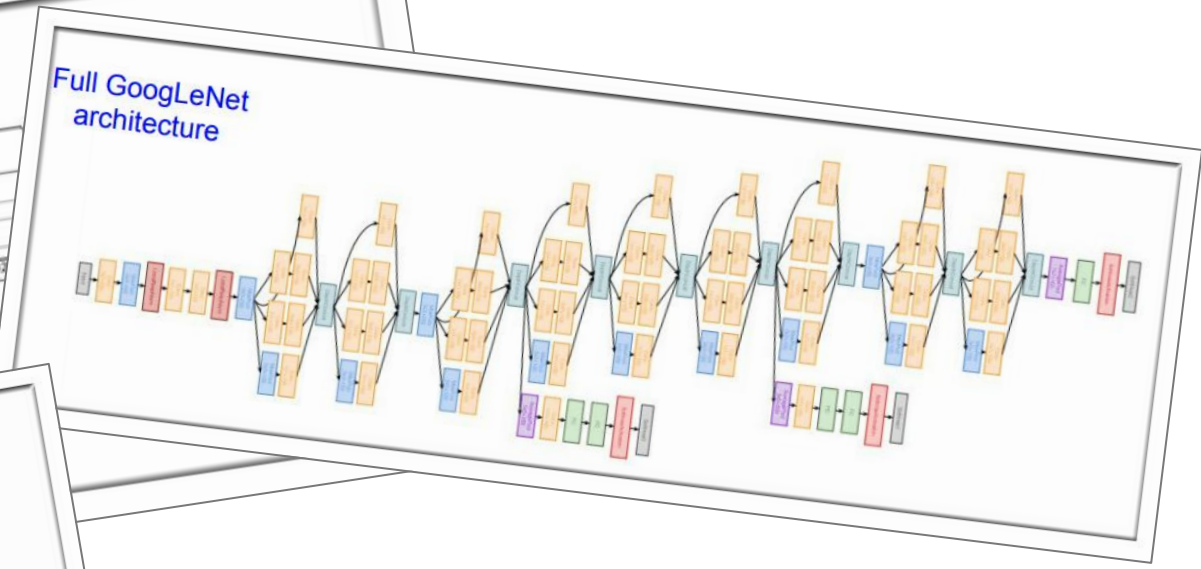
深度卷积
神经网络

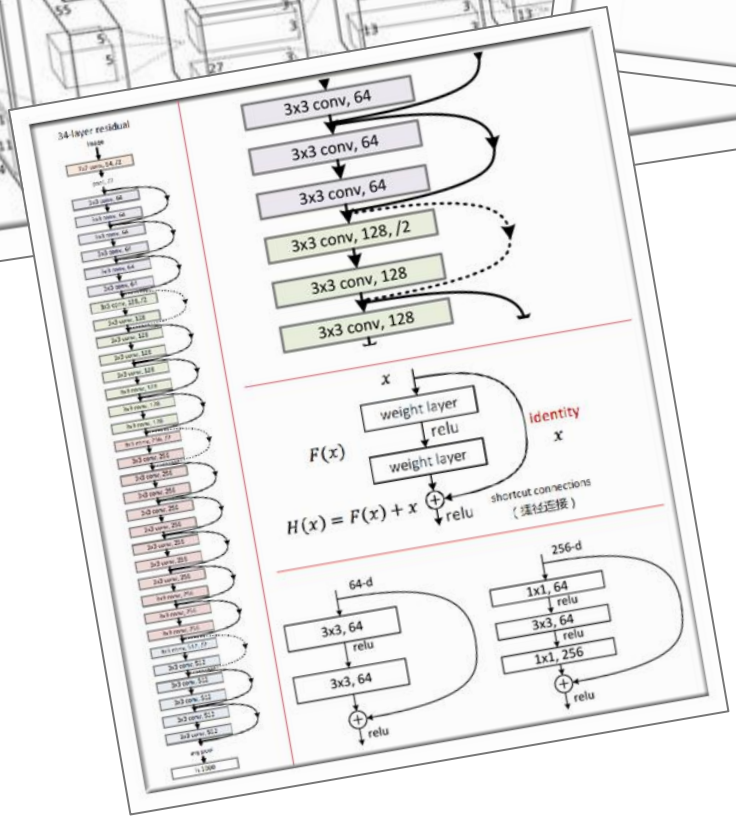
answer

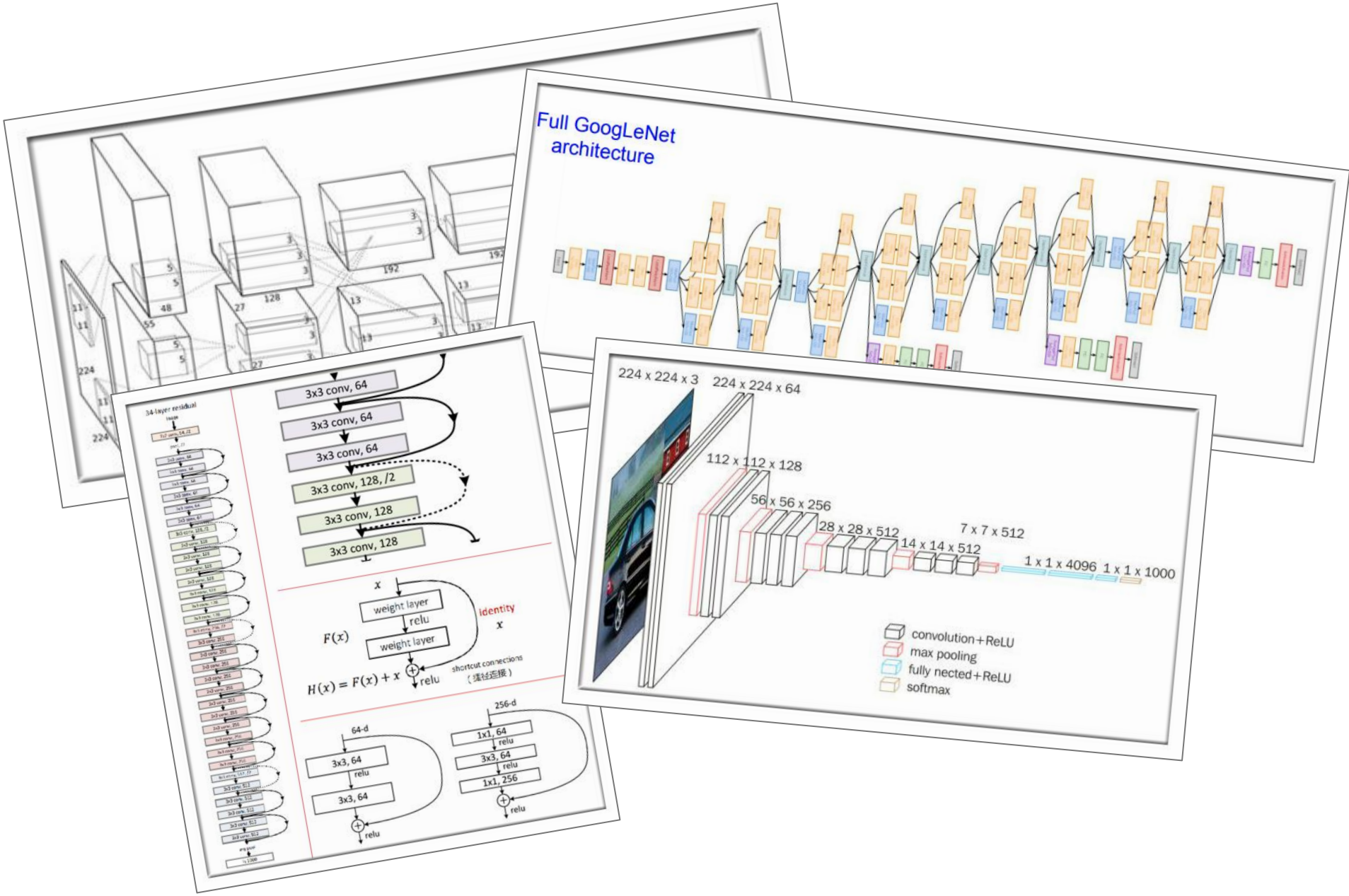

11
11
3
224
224
Stride
of 4
5
5
48
48
55
55
Max
pooling
3
3
128
27
27
Max
pooling
3
3
192
13
13
3
3
192
3
3
192
3
3
128
13
13
Max
pooling
dense
2048
2048
dense
2048
2048
dense
1000

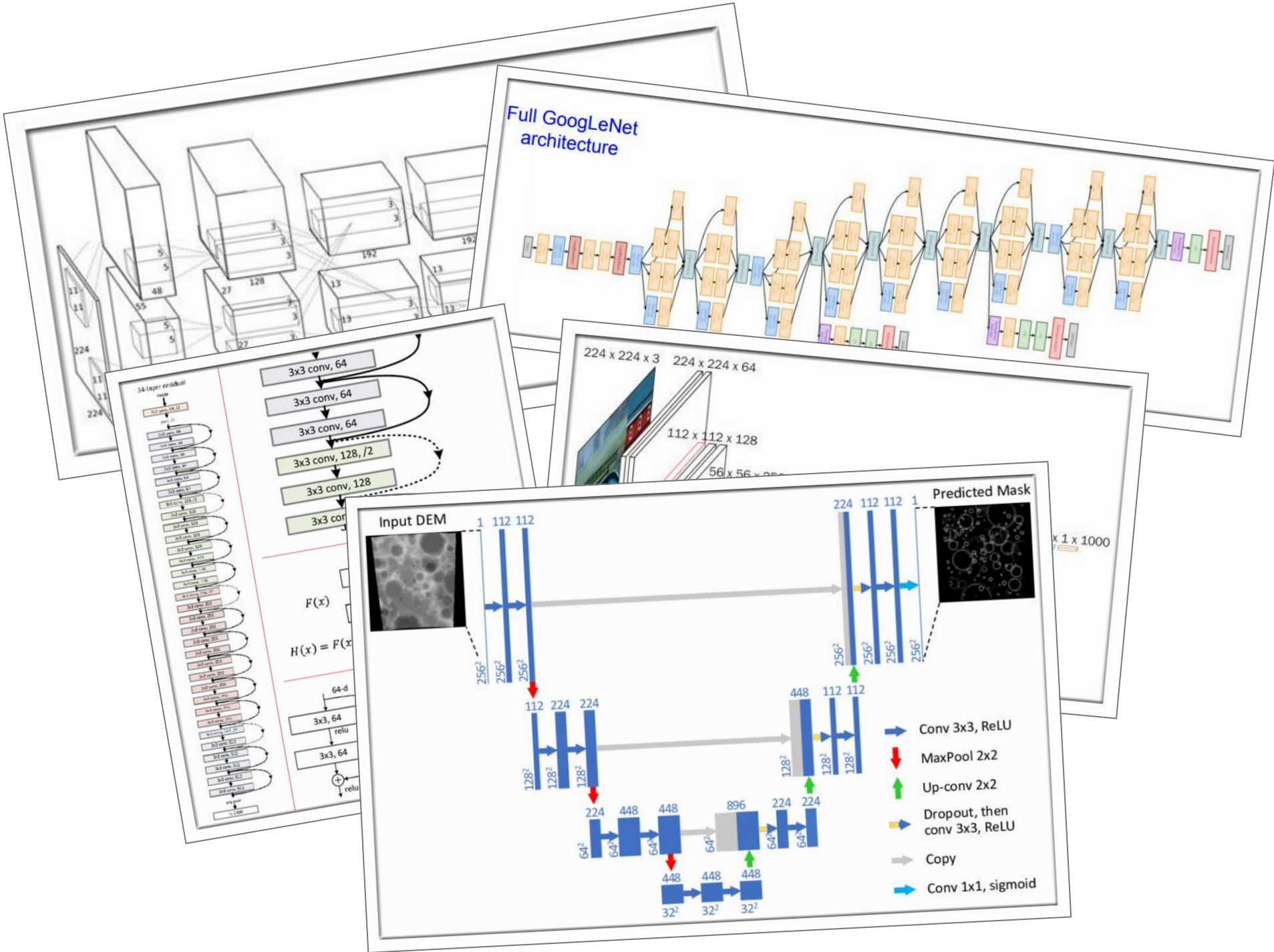

Full GoogLeNet
architecture
224
11
11
3
Stride
of 4
55
48
5
5
Max
pooling
27
128
3
3
Max
pooling
13
192

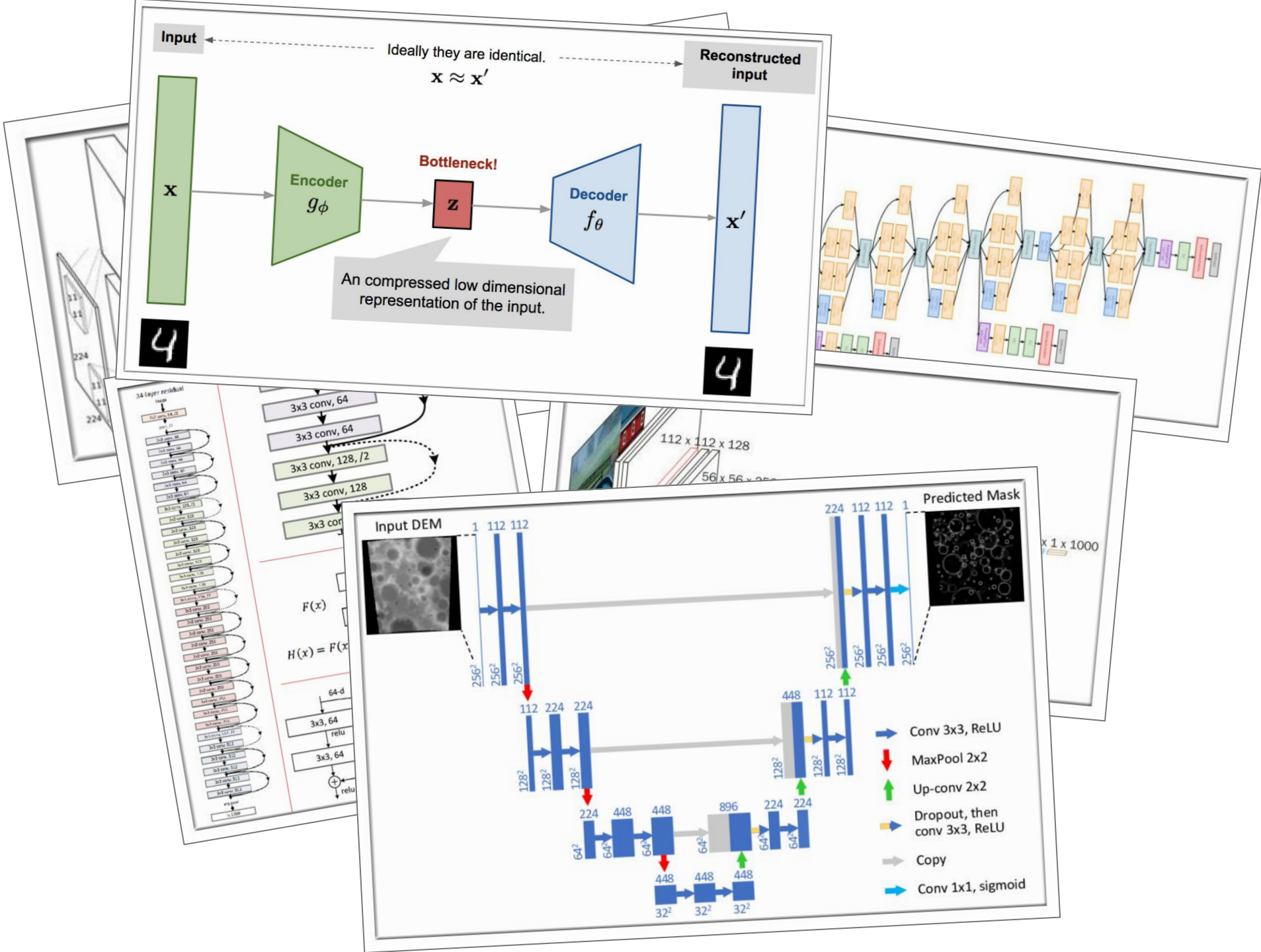


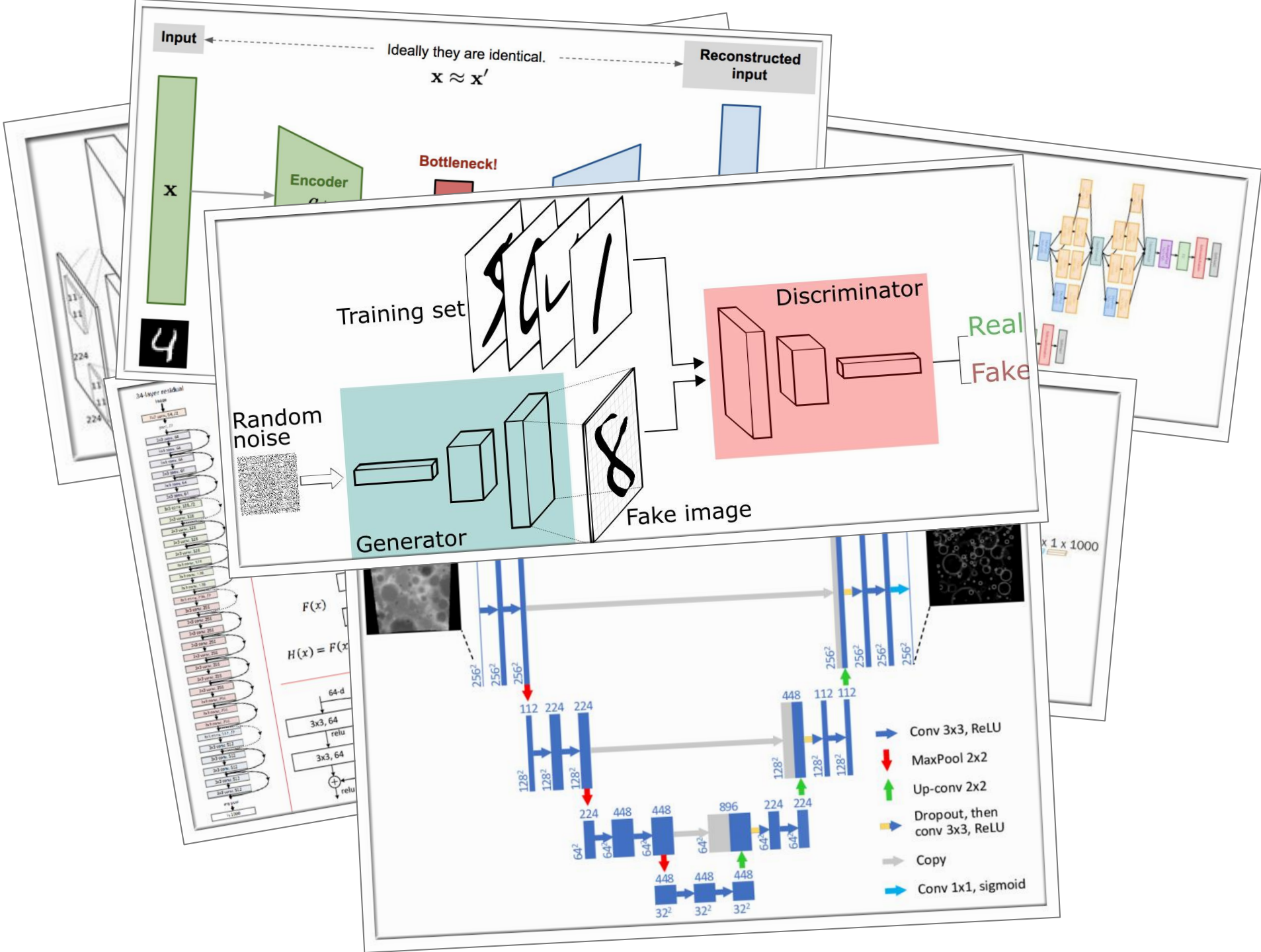

Full GoogLeNet
architecture

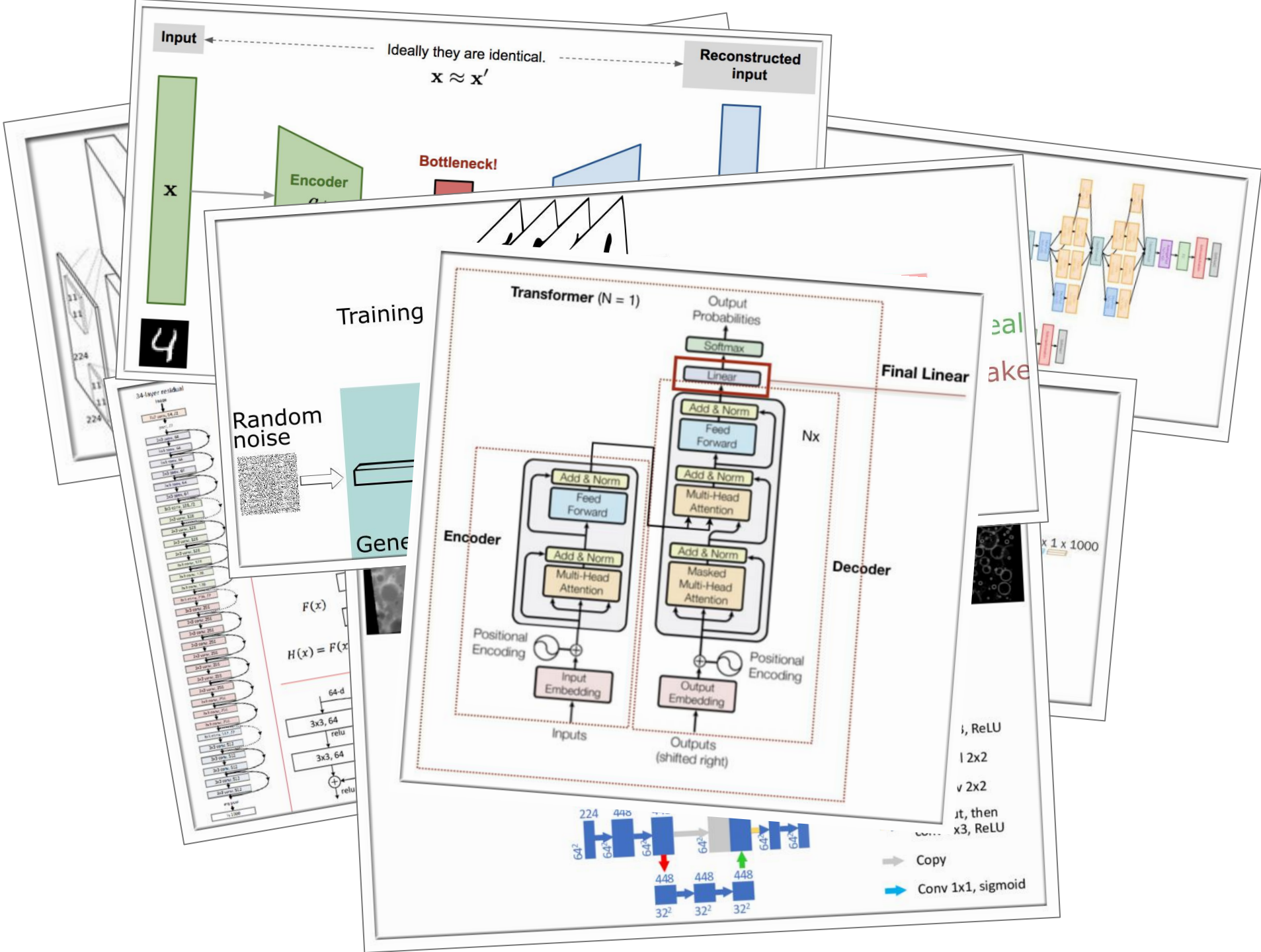

34-layer residual
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 128
x
weight layer
relu
weight layer
F(x)
identity
x
relu
shortcut connections
(捷径连接)
H(x) = F(x) + x
64-d
3x3, 64
relu
3x3, 64
relu
256-d
1x1, 64
relu
3x3, 64
relu
1x1, 256
relu


Full GoogLeNet
architecture
34-layer residual
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 128
x
weight layer
relu
weight layer
F(x)
identity
x
H(x) = F(x) + x
relu
shortcut connections
64-d
3x3, 64
relu
3x3, 64
relu
256-d
1x1, 64
relu
3x3, 64
relu
1x1, 256
relu
224 x 224 x 3
224 x 224 x 64
112 x 112 x 128
56 x 56 x 256
28 x 28 x 512
14 x 14 x 512
7 x 7 x 512
1 x 1 x 4096
1 x 1 x 1000
convolution+ReLU
max pooling
fully nected+ReLU
softmax


Full GoogLeNet architecture
3x3 conv, 64
3x3 conv, 64
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
$F(x)$
$H(x) = F(x$
64-d
3x3, 64
relu
3x3, 64
relu
224 x 224 x 3
224 x 224 x 64
112 x 112 x 128
x 1 x 1000
Input DEM
Predicted Mask
Conv 3x3, ReLU
MaxPool 2x2
Up-conv 2x2
Dropout, then conv 3x3, ReLU
Copy
Conv 1x1, sigmoid

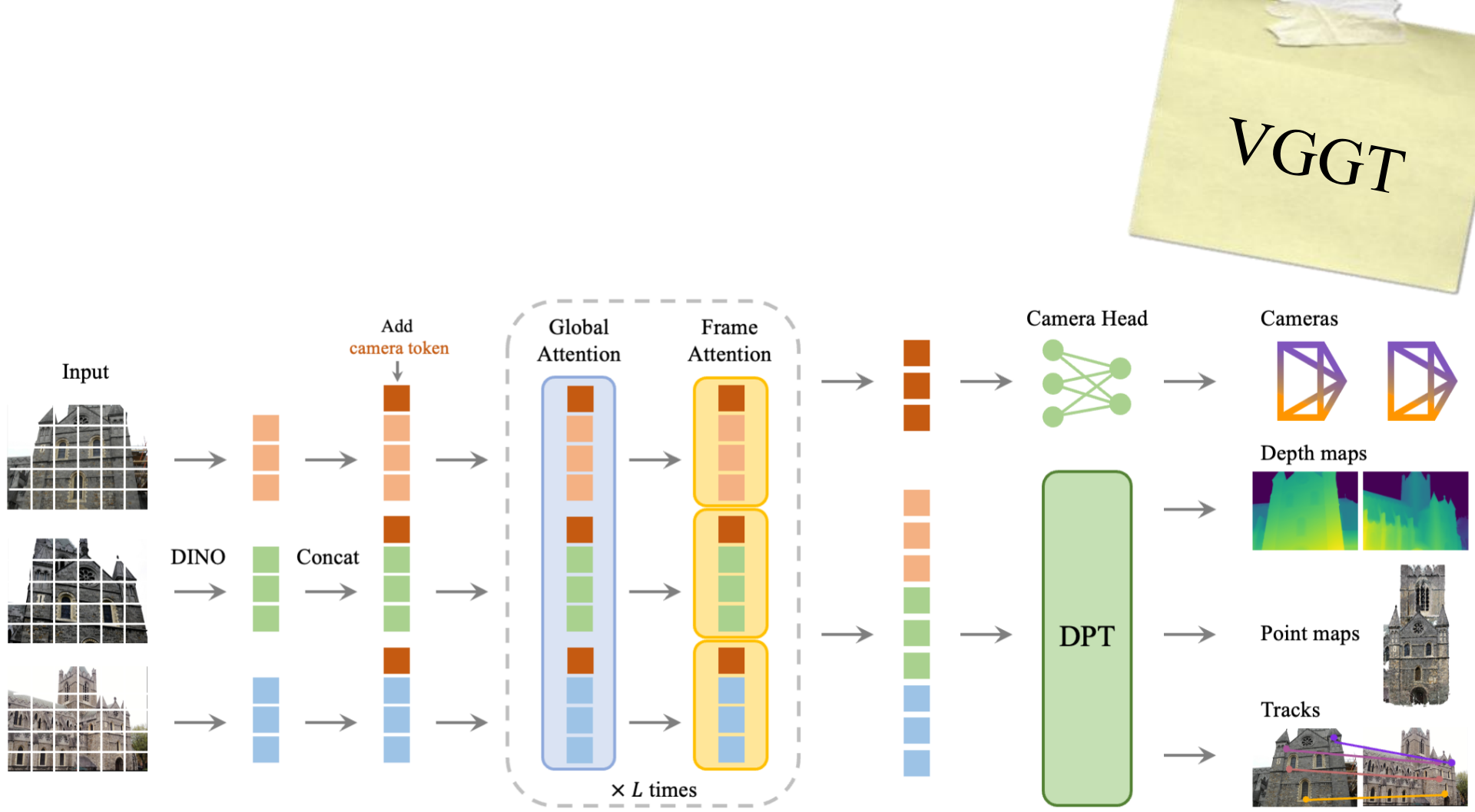

Input
Ideally they are identical.
$\mathbf{x} \approx \mathbf{x}'$
Reconstructed input
x
Encoder
g_ϕ
Bottleneck!
z
Decoder
f_θ
x′
An compressed low dimensional representation of the input.
3x3 conv, 64
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
$F(x)$
$H(x) = F(x$
64-d
3x3, 64
relu
3x3, 64
relu
112 x 112 x 128
56 x 56 x
x 1 x 1000
Input DEM
Predicted Mask
Conv 3x3, ReLU
MaxPool 2x2
Up-conv 2x2
Dropout, then conv 3x3, ReLU
Copy
Conv 1x1, sigmoid


Input
Ideally they are identical.
$\mathbf{x} \approx \mathbf{x}'$
Reconstructed input
x
Encoder
Bottleneck!
Training set
Discriminator
Real
Fake
Random noise
Generator
Fake image
Conv 3x3, ReLU
MaxPool 2x2
Up-conv 2x2
Dropout, then conv 3x3, ReLU
Copy
Conv 1x1, sigmoid


Input
Ideally they are identical.
$\mathbf{x} \approx \mathbf{x}'$
Reconstructed input
x
Encoder
Bottleneck!
4
Training
Random noise
Transformer (N = 1)
Output Probabilities
Softmax
Linear
Final Linear
Add & Norm
Feed Forward
Multi-Head Attention
Masked Multi-Head Attention
Nx
Encoder
Decoder
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)
Copy
Conv 1x1, sigmoid

结构工程


rs, we would like
ne authors for
ork.
Best Paper
VGGT: Visual Geometry Grounded Transformer
Jianyuan Wang[1,2] Minghao Chen[1,2] Nikita Karaev[1,2] Andrea Vedaldi[1,2]
Christian Rupprecht[1] David Novotny[2]
[1]Visual Geometry Group, University of Oxford [2]Meta AI
CVPR


VGGT
Input
Add
camera token
Global
Attention
Frame
Attention
Camera Head
Cameras
DINO
Concat
Depth maps
DPT
Point maps
Tracks
× L times


Microsoft
CVPR

机器学习和计算机视觉还是彼此相互依存、相互促进的关系，而不是对立的竞争关系

应用

人脸检测

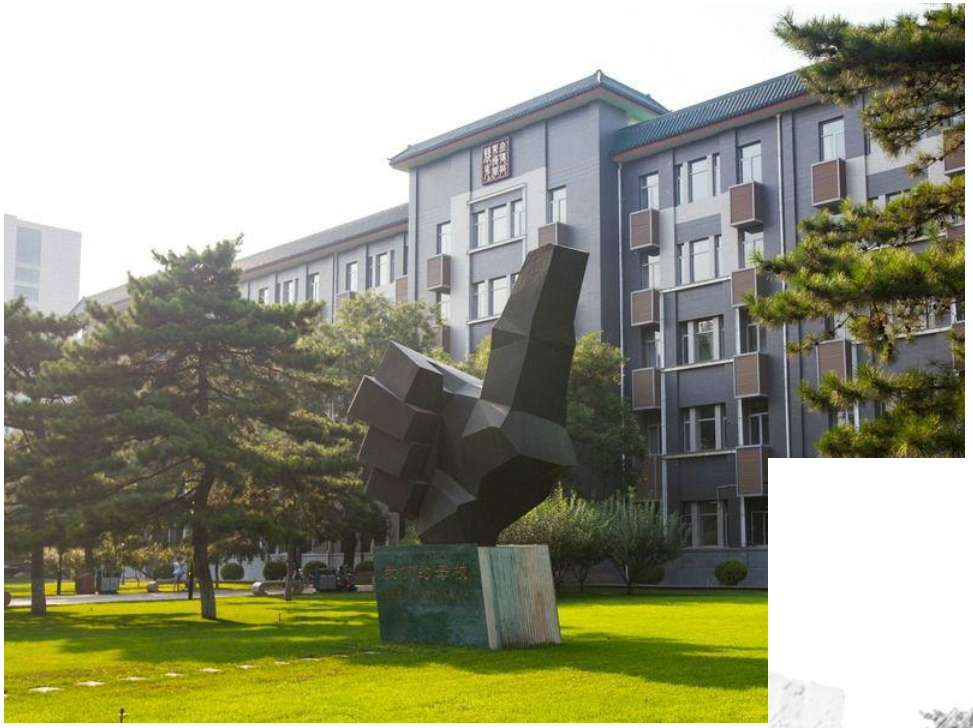

●●●○○ 中国联通
22:10
49%
返回
颜龄终极版，我也是拼...
天天P图：火遍全亚洲的美妆神器
测颜龄终极版已上线~美妆后再测小5岁哦
一秒测颜龄 晒出真实的你
35岁
博士后
分享炫耀
一键减龄
再试一次？去天天P图玩咯
立即打开

人脸检测


人脸识别
请通过，谢谢

表情识别

特效



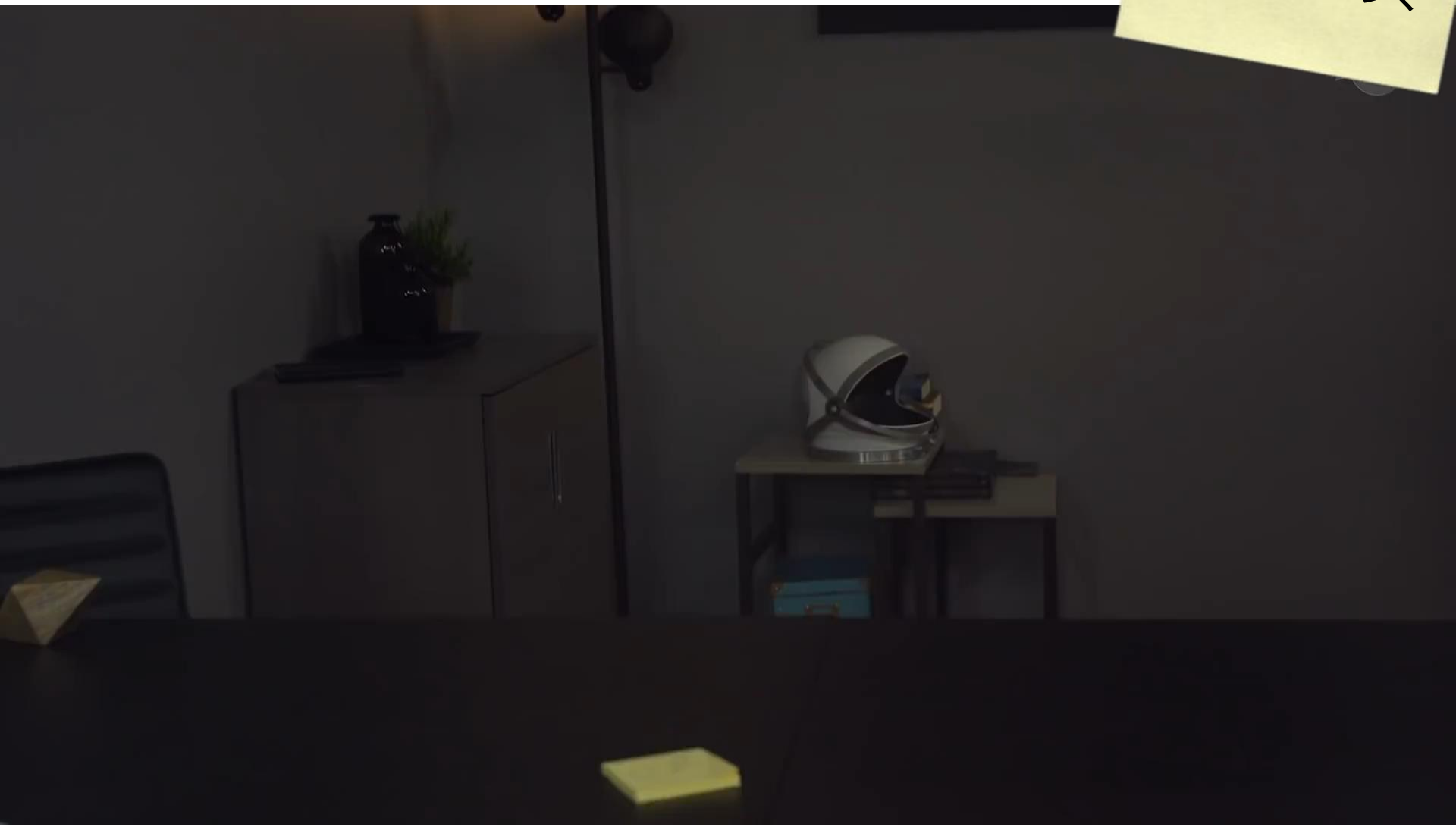
图像
风格迁移


生物识别

安防监控

安防监控


BBC
中文
alhua
中國各地都湧現裝有人工智能的監控系統
But you don't need to see my work to apply Artificial Intelligence everywhere.


增强现实

增强现实

人脸识别

ELECTRICITY MAY BE THE DRIVER. One day your car may speed along an electric super-highway, its speed and steering automatically controlled by electronic devices embedded in the road. Highways will be made safe—by electricity! No traffic jams . . . no collisions . . . no driver fatigue.

无人驾驶

Horizon Robotics

Horizon Journey™ 2 based Mono-Camera Vision Perception Solution for ADAS


地图绘制
金鸿运宾馆
霖祥宾馆
绿岛苑
绿岛苑-东区
聚福苑小区
鲁班大厦
朝阳路
G102
大黄庄教师楼
北京电力建设有限公司
图书馆
图书馆
金宝生活
传媒附中
万东国际文化创意产业园
新知楼
三间房南里
北京大黄庄医院
2号教学楼
北京第二外国语学院
定福庄南里
花园闸北里
21号楼
福慧嘉苑
门诊部
主楼
内蒙古大饭店
中国传媒大学逸夫综合教学楼
京通快速
传媒大学
三间房派出所
朝阳区三间房乡茶家坟小区
惠河北街
通惠河
惠河南街
如家快捷酒店
北花园小区1号院
北花园
乐栋300A座
北花园小区3号院
家有购物集团
北京众宜轩四合院酒店
万豪卡通
北京三间房动漫大厦


地图绘制
Google Street View
Explore the world at google.com/streetview
Google Street View
8BJX008

谷歌街景采集车

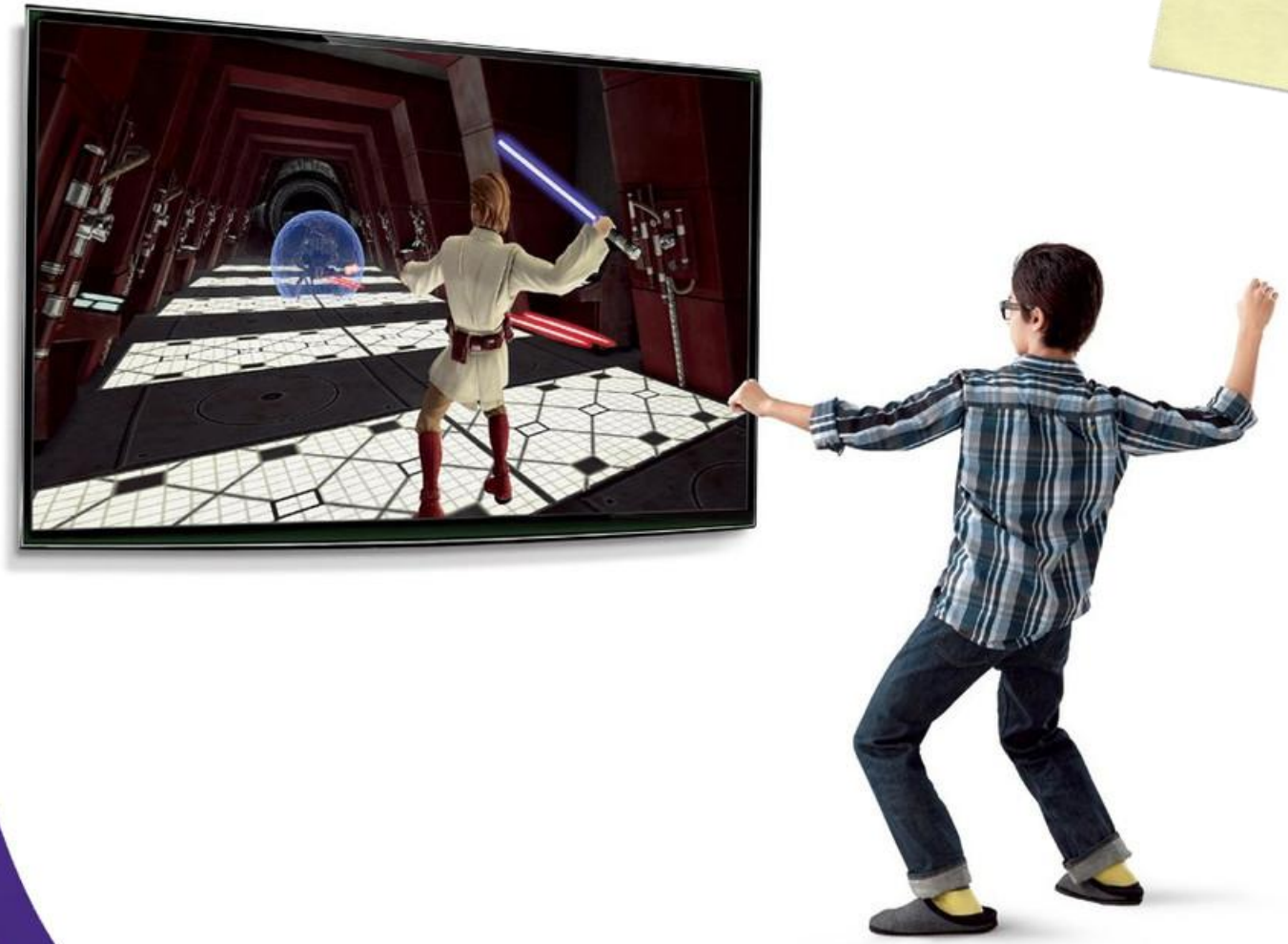

生产分拣


电影特效


电影特效

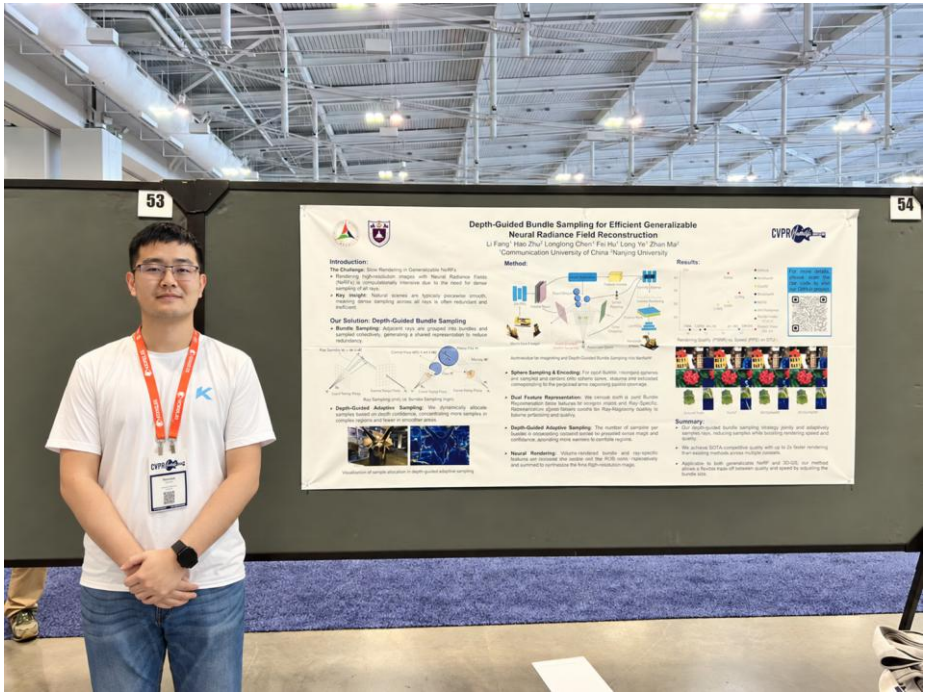
电影特效

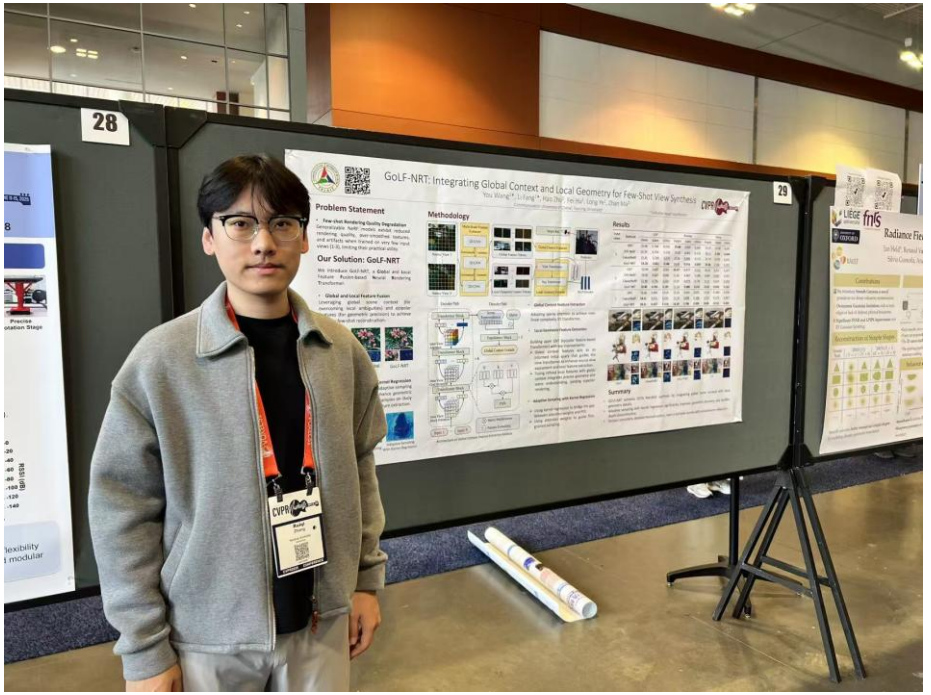
游戏



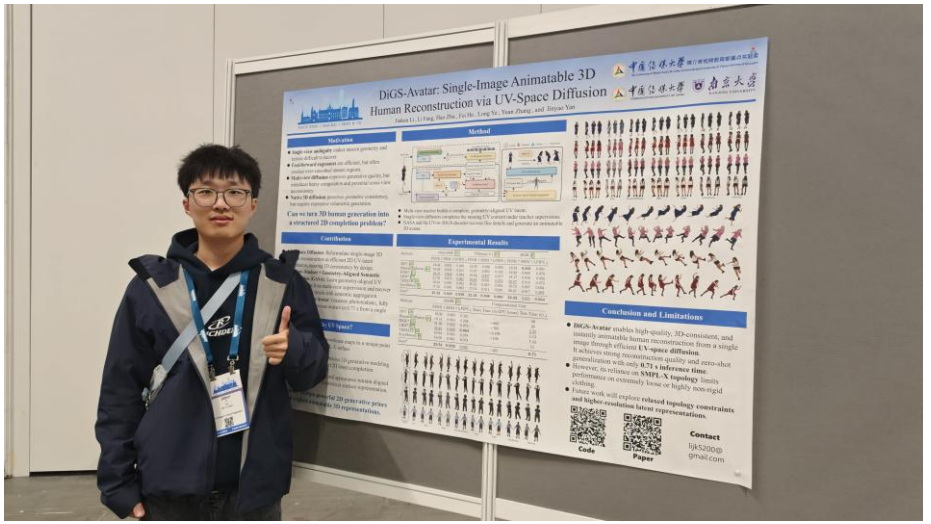

CVPR 2025，美国 纳什维尔


超过1.2万人出席！
CVPR 2025，美国 纳什维尔

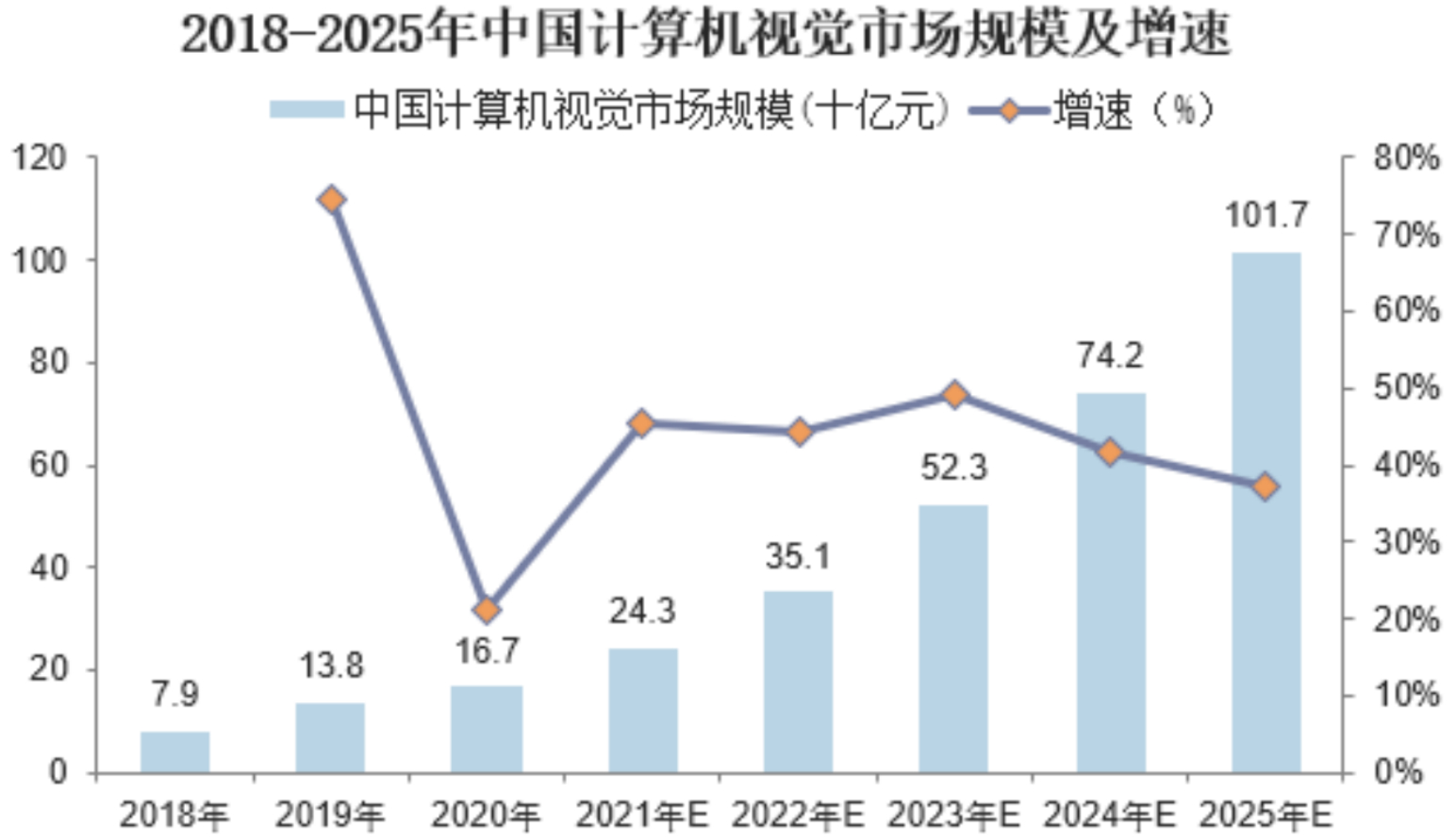




CVPR Nashville JUNE 11-15, 2025



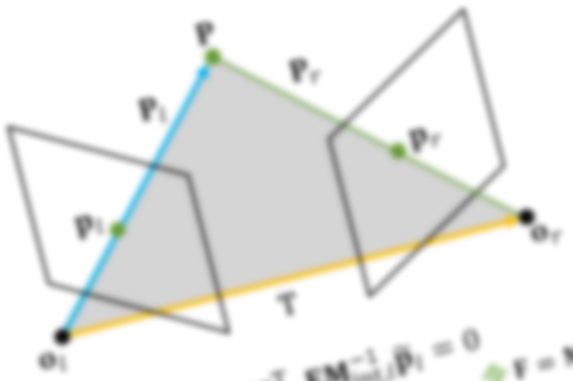

Google
Microsoft
Meta
ByteDance
字节跳动
Baidu 百度
科大讯飞
iFLYTEK
amazon
MEGVII 旷视
商汤
sensetime
Tencent 腾讯
dji
HUAWEI
OpenAI
intel
NVIDIA
TESLA



数据来源：华经产业研究院

视觉信息处理的基础分析

视觉信息处理的基础分析

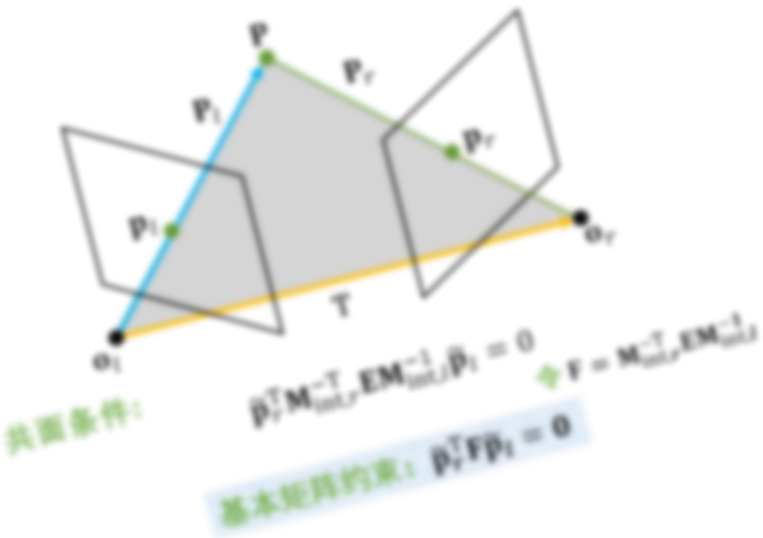


共面条件：$\mathbf{p}_r^T\mathbf{M}_{int,r}^{-T}\mathbf{E}\mathbf{M}_{int,l}^{-1}\mathbf{p}_l = 0$ ⇒ $\mathbf{F} = \mathbf{M}_{int,r}^{-T}\mathbf{E}\mathbf{M}_{int,l}^{-1}$

基本矩阵约束：$\mathbf{p}_r^T\mathbf{F}\mathbf{p}_l = \mathbf{0}$

视觉信息处理的基础分析

利用这些分析的算法



课题

图像形成

课题

图像形成

图像表示

课题

图像形成

图像表示

特征提取

课题

图像形成

图像表示

特征提取

立体视觉

课题

图像形成

图像表示

特征提取

立体视觉

运动分析

图像形成

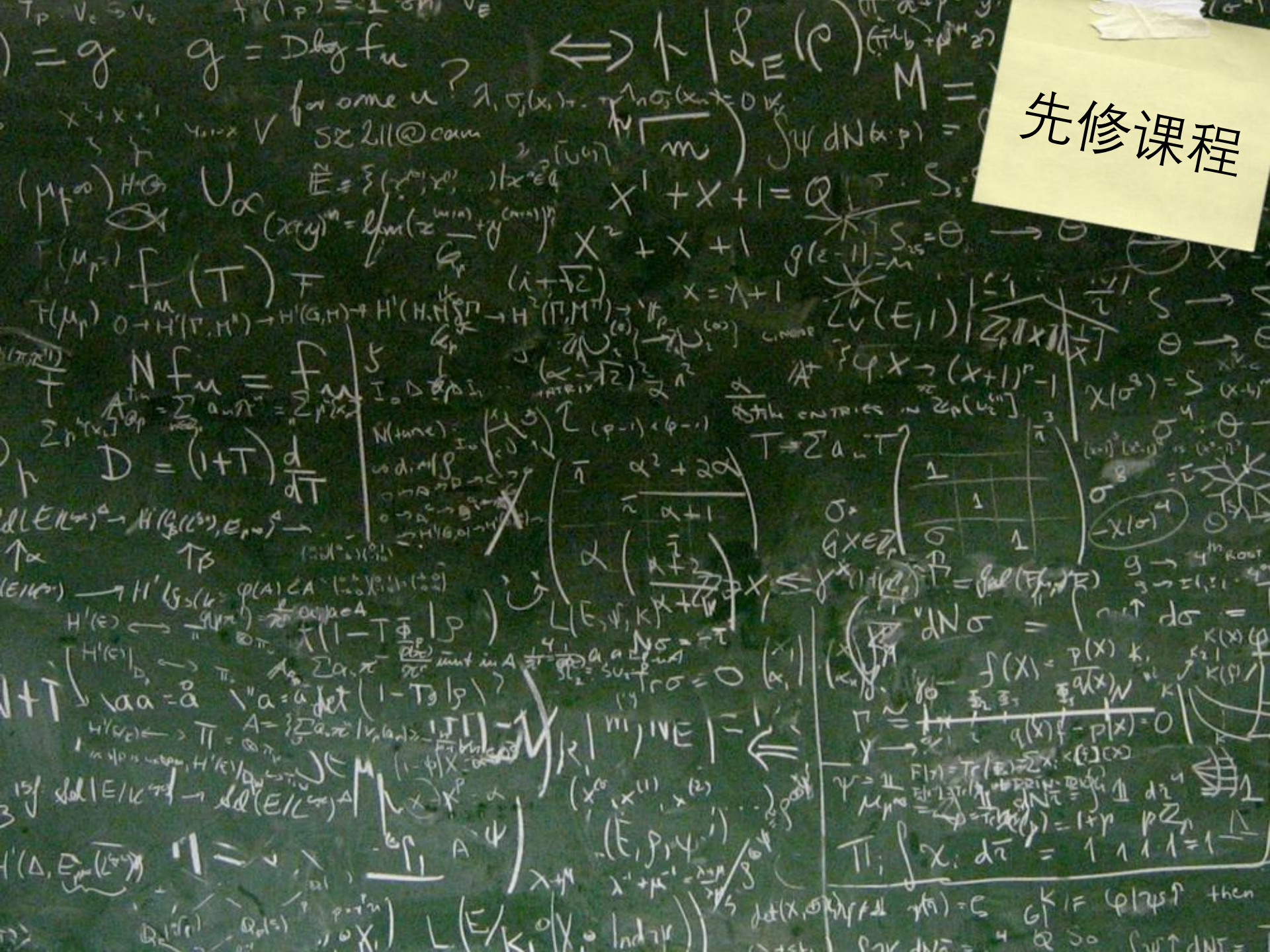
图像表示

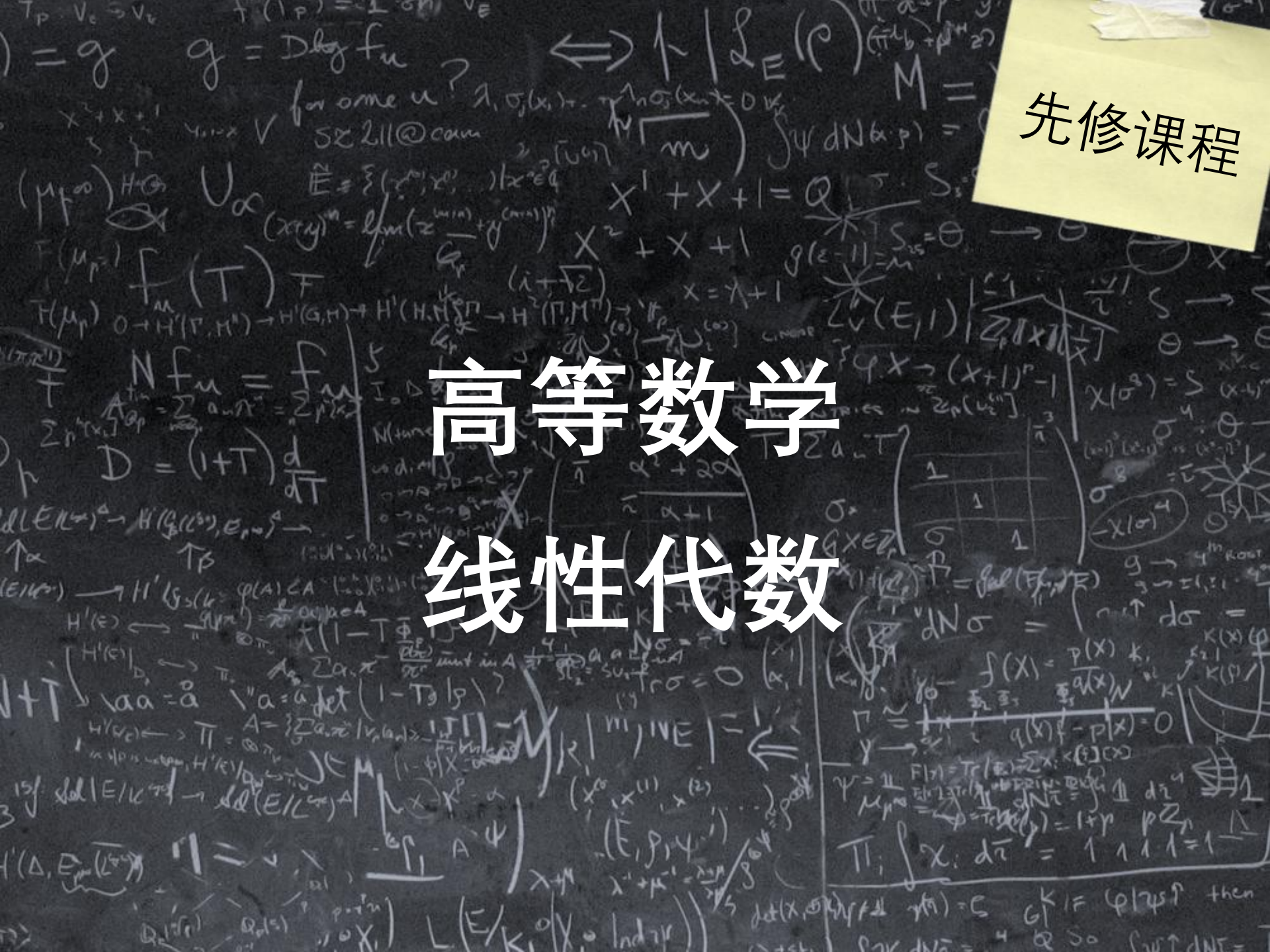
特征提取

立体视觉

运动分析

机器学习


先修课程


先修课程
高等数学
线性代数

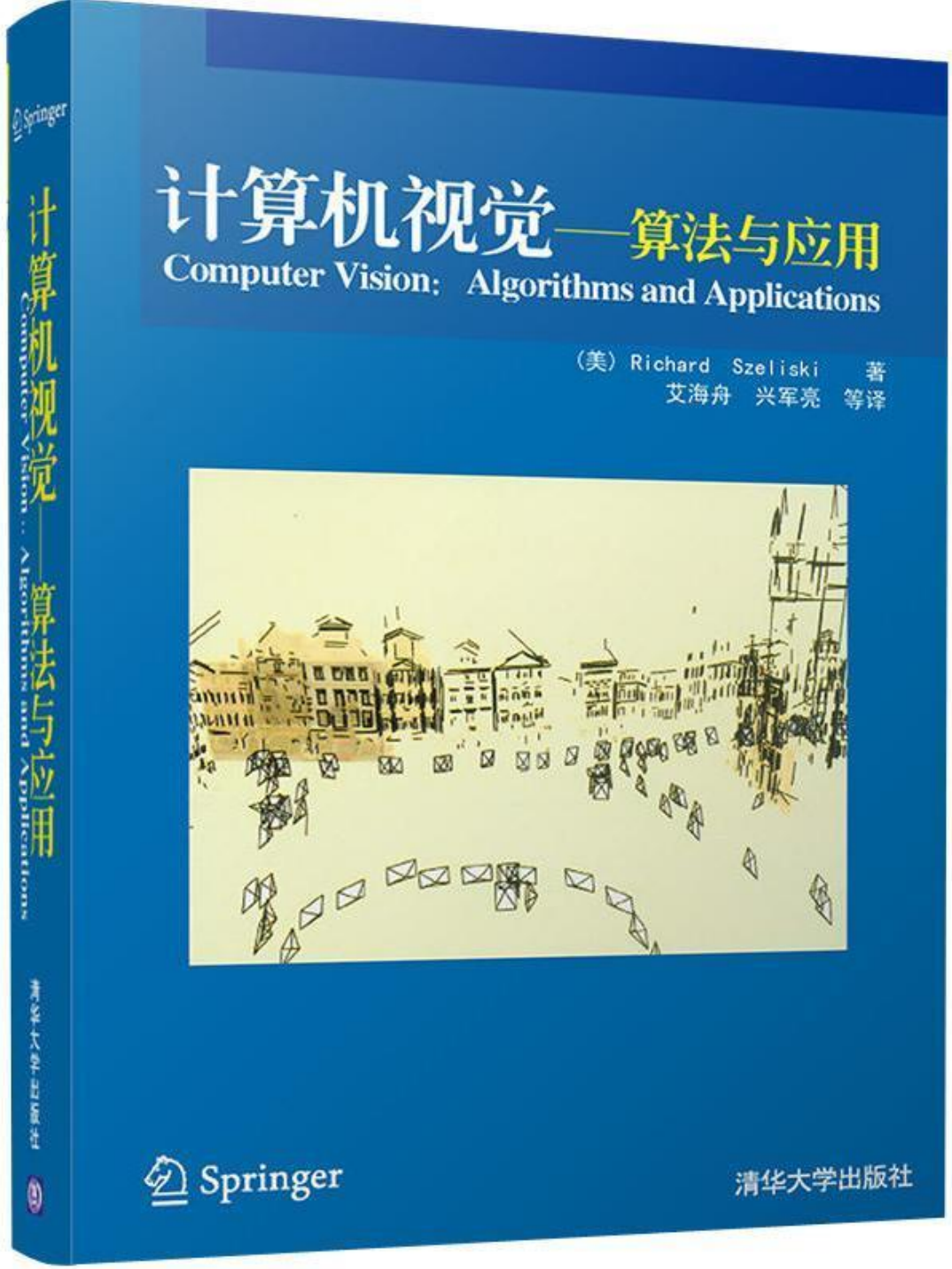
前方高能！


考核

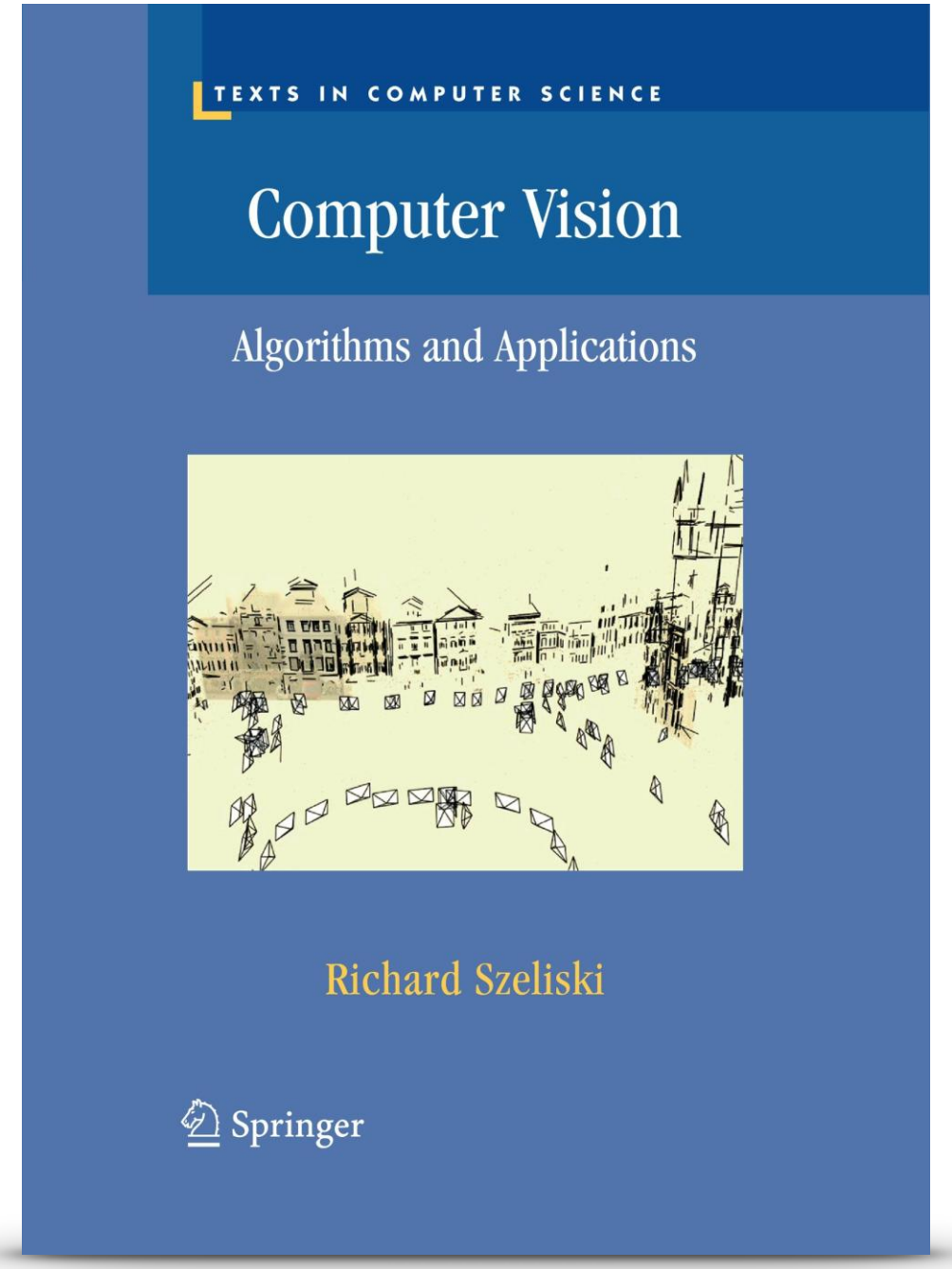
三次作业：40%

三次作业：40%
期末项目：60%


python
TM

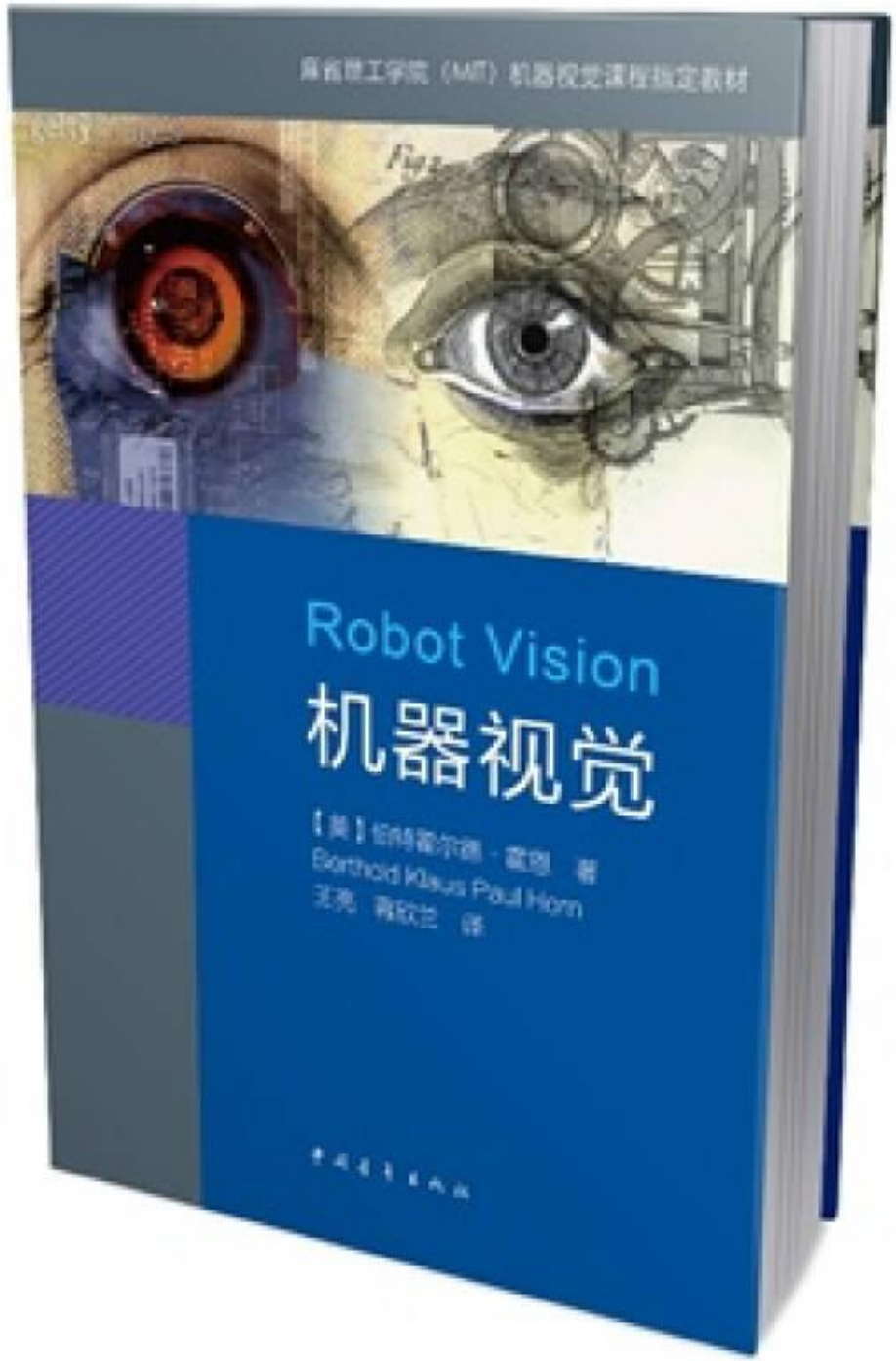

计算机视觉——算法与应用
Computer Vision: Algorithms and Applications
(美) Richard Szeliski 著
艾海舟 兴军亮 等译
Springer
清华大学出版社

教材



教材

在线免费提供
szeliski.org/Book



教材

研究生课程：计算机视觉

时间：2019-02-27 浏览： 8224

计算机视觉

Computer Vision

任课教师：方力

课程编号：2115530096

上课时间：1-9,11-17周，周三1-2节

上课地点：48教B906

课程描述：本课程介绍了视觉的基本概念，重点是计算机科学和工程。 特别是，该课程涵盖相机成像过程，图像表示，特征提取，立体视觉，运动分析，3D参数估计和应用等。

办公地址：中国传媒大学国重大楼 801C

联系电话：010-65783425

电子邮件：cuc_cv@163.com

先修课程：高等数学、线性代数

教材：Richard Szeliski著，艾海洲，兴军亮等译. 计算机视觉—算法与应用. 北京：清华大学出版社，2012.

参考资料：Berthold Horn著，王亮，蒋欣兰译. 机器视觉. 北京：中国青年出版社，2014.

课件下载：