深度学习平台与应用

第一讲：绪论

范琦

fanqi@nju.edu.cn

2024年9月4日

大　　纲

- **课程概览**
- **深度学习历史**
- **深度学习应用**
- **深度学习平台**
- **课程后续安排**

课程内容

南京大学 NANJING UNIVERSITY

- 绪论
- 线性分类器
- 正则化与优化
- 神经网络与反向传播
- 深度学习科研（Special）
- 图像分类
- 卷积神经网络
- 循环神经网络
- 注意力机制和Transformer
- 目标检测和分割
- 视频理解
- 网络可视化
- 深度自监督学习
- 多模态模型
- 3D视觉模型
- 深度生成模型

课程关系

南京大学 NANJING UNIVERSITY

- 前置课程：
  - 高等数学，微积分
  - 概率论与数理统计，线性代数
  - 最优化方法导论，人工智能导论，机器学习导论
- 后续课程：
  - 模式识别与计算机视觉
  - 自然语言处理

课程目标

- 了解深度学习基础知识
- 了解深度学习研究方向和领域
- 初步锻炼科研思维
- 培养论文写作能力
- 为后续实习/工作/深造奠定基础

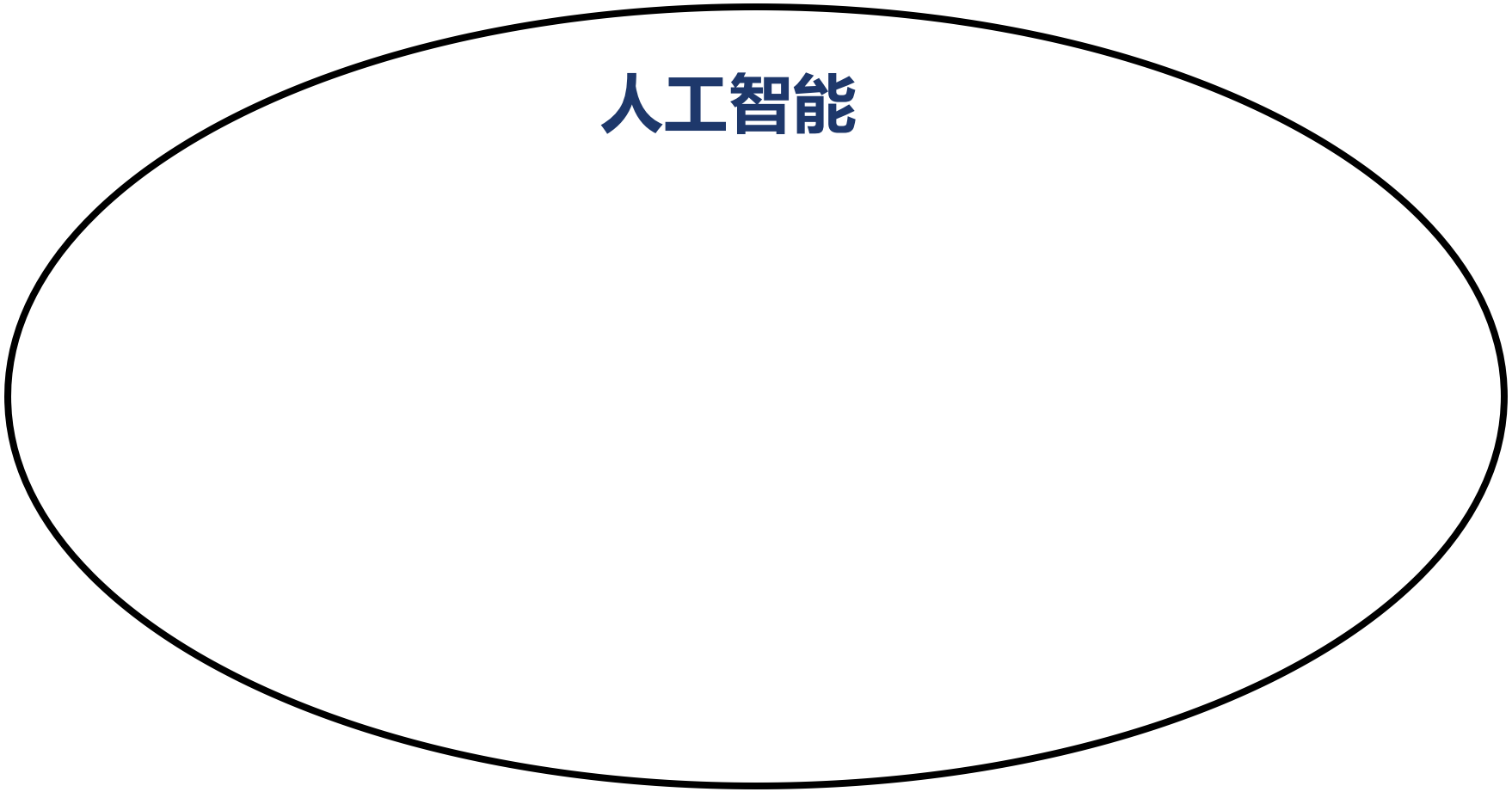
授课 & 答疑

- 线下授课
- 助教：
  - 段子轩（2021dwin@sina.com）
  - 鲁峰源（fengyuan_lu@qq.com）
- 答疑：面对面/邮件（**请并抄送给助教**）
- **课程考核：在第二节课公布**

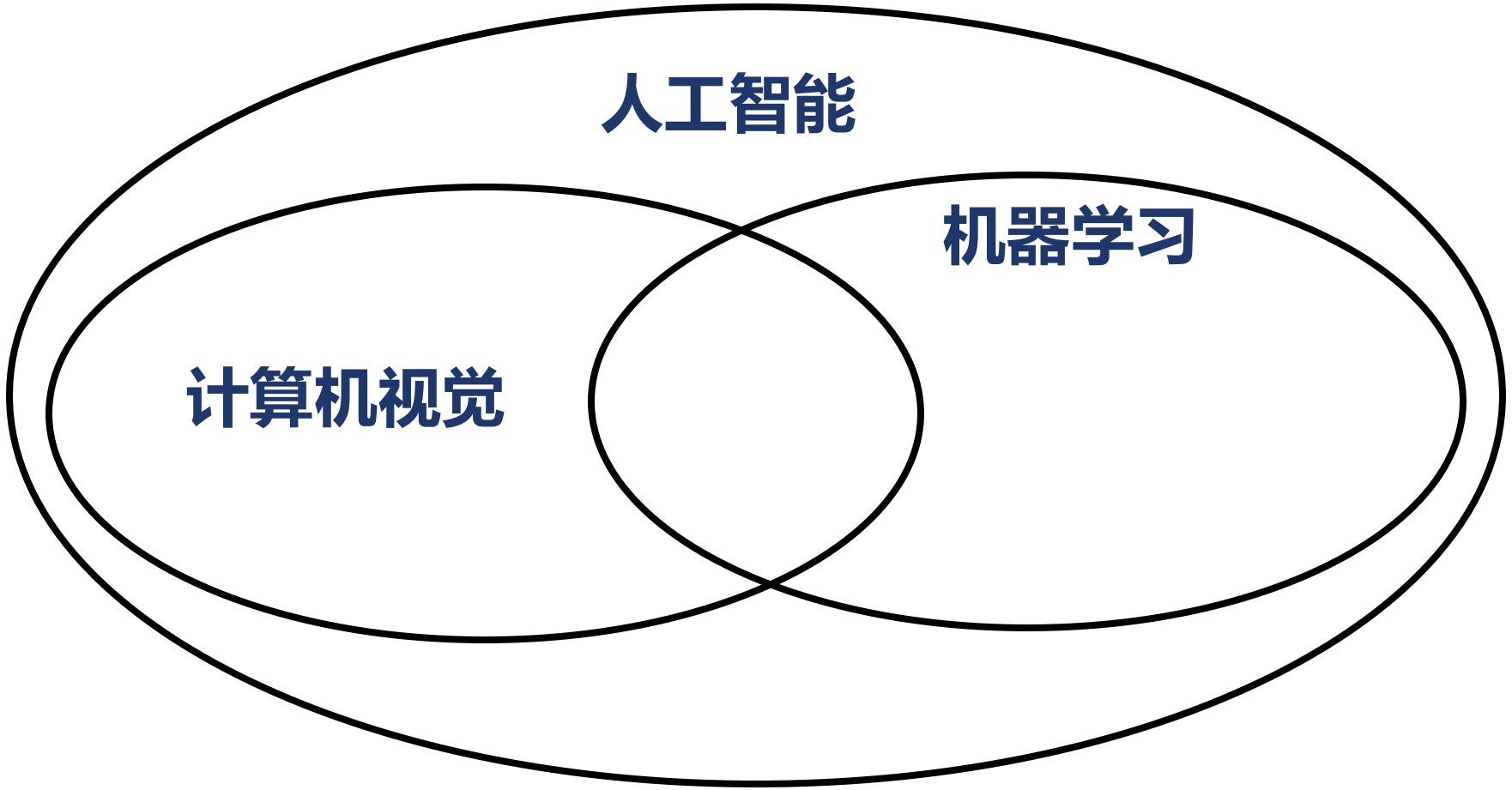
大　　纲

- 课程概览
- 深度学习历史
- 深度学习应用
- 深度学习平台
- 课程后续安排

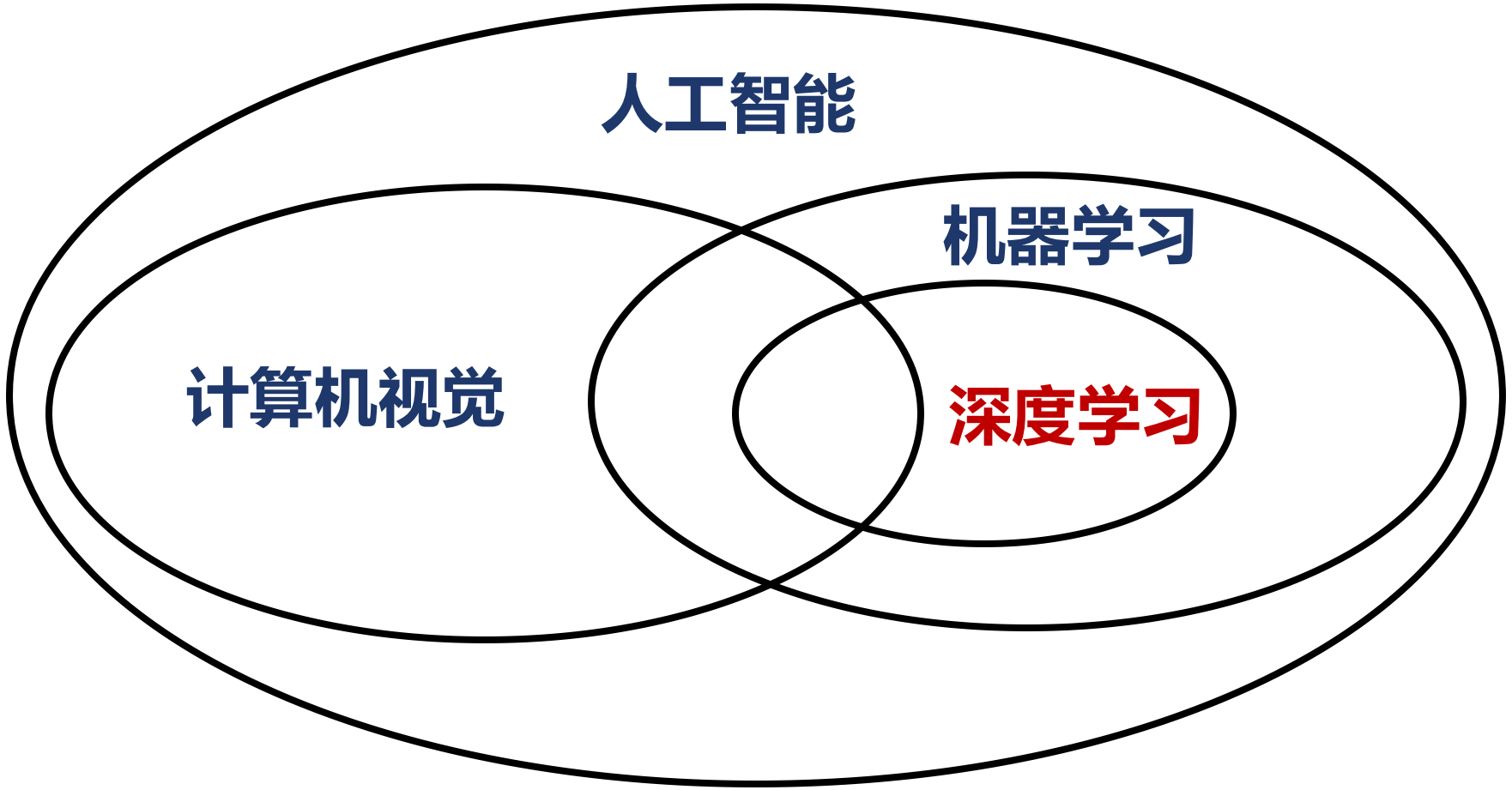
深度学习关系



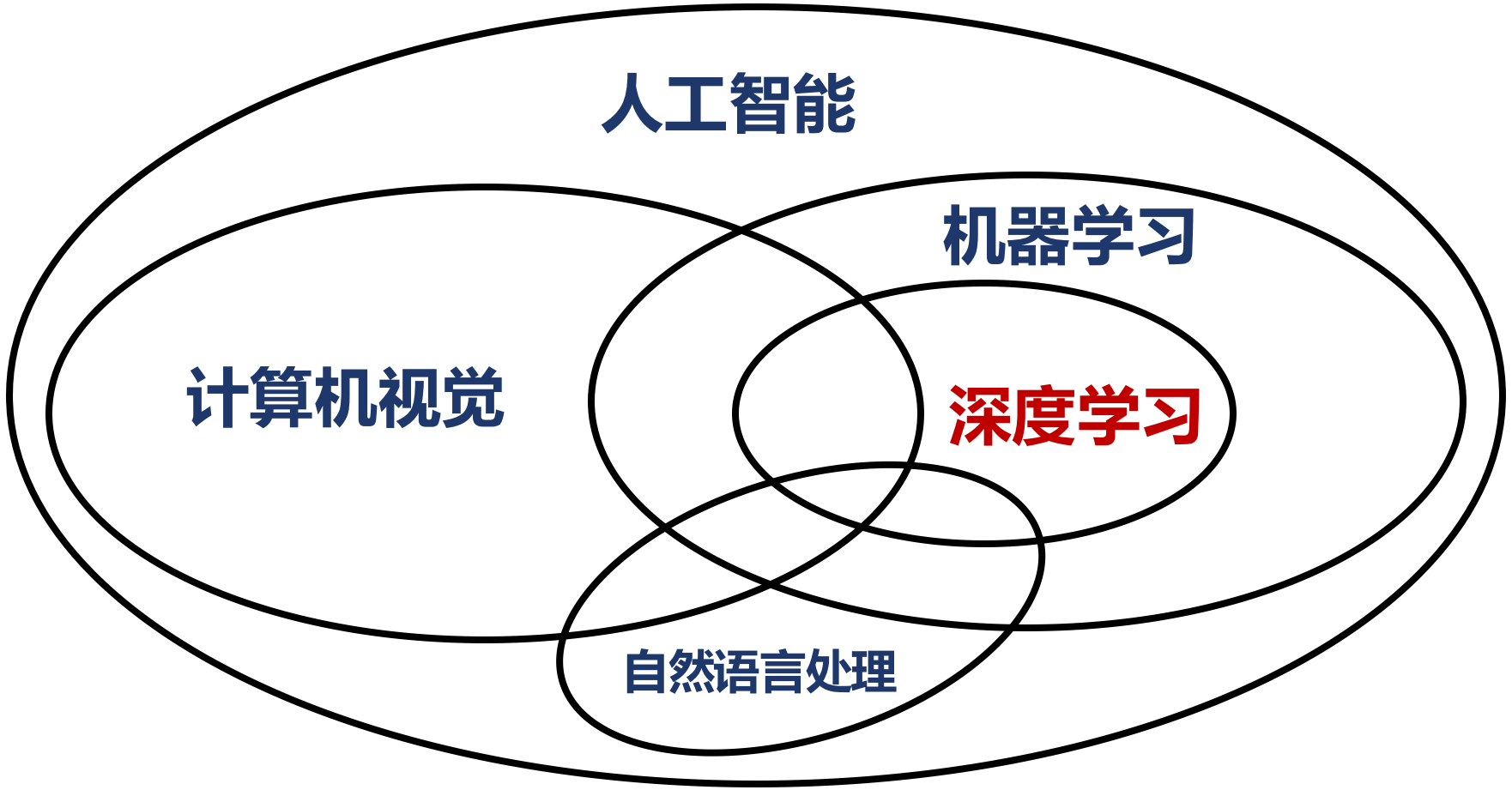
深度学习关系



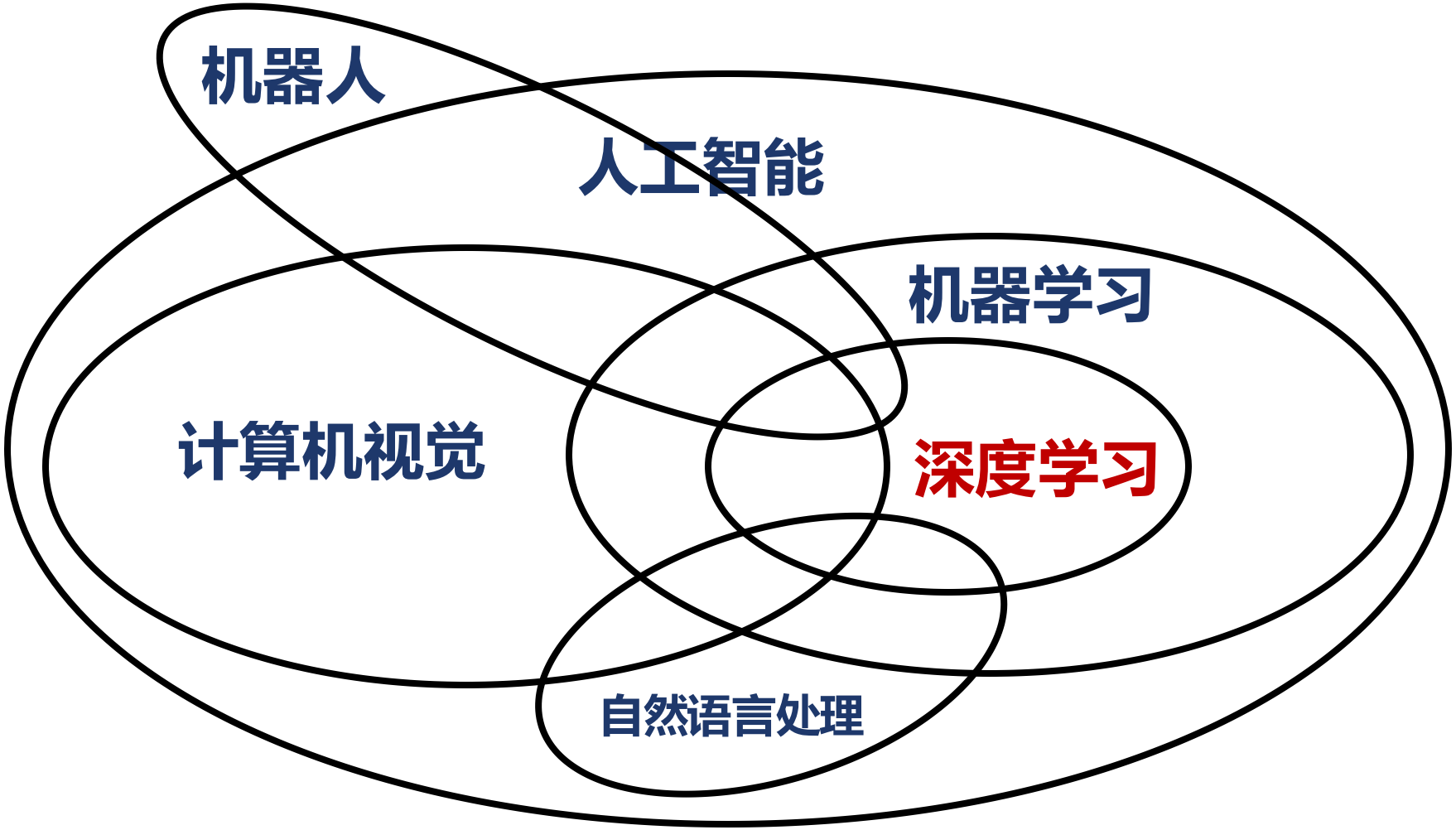
深度学习关系



深度学习关系

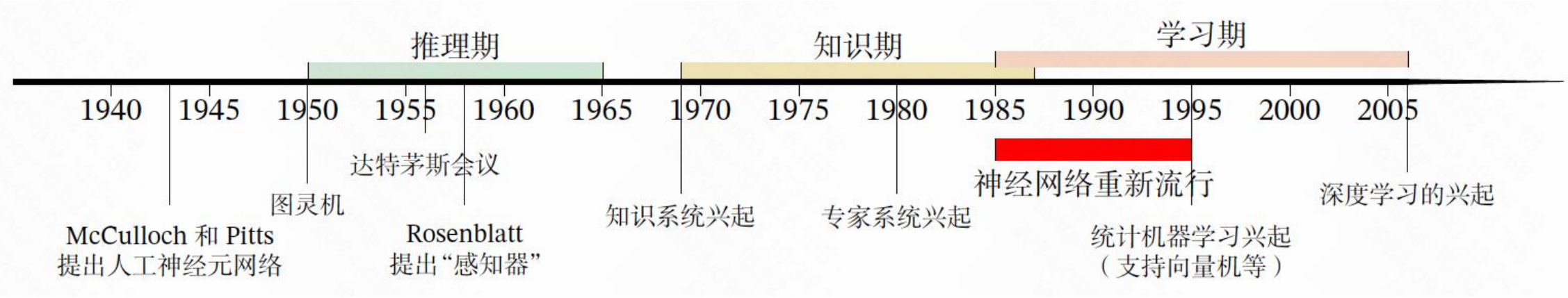


深度学习关系



深度学习关系

- 人工智能（Artificial Intelligence, AI）
  - 推理期：基于逻辑或者事实归纳出来一些规则
  - 知识期：领域专家构建的知识库+推理机
  - 学习期：让计算机从数据中自己学习



深度学习关系

- 机器学习（Machine Learning, ML）
  - 从有限的观测数据中学习出具有一般性的规律，并利用这些规律对未知数据进行预测
  - 浅层学习（传统机器学习）关注学习预测模型，**不涉及特征学习**，其特征主要靠**特征工程**来提取

深度学习关系

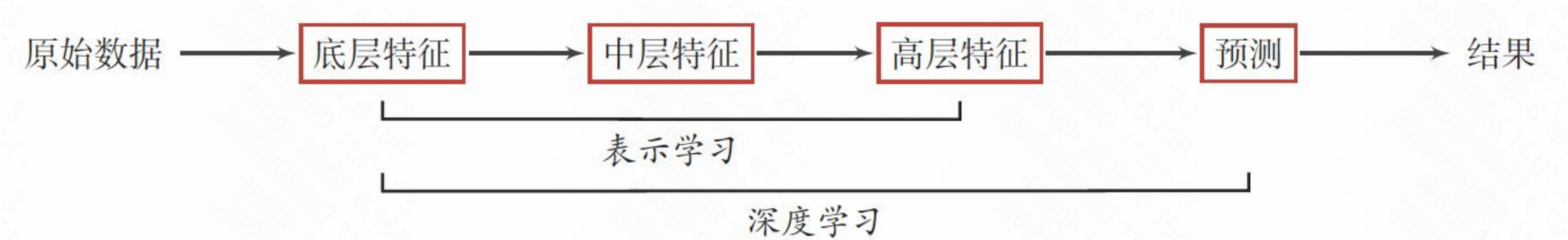
- 表示学习（Representation Learning）
  - 自动地学习出有效的特征，并提高最终机器学习模型的性能
  - 关键问题是解决语义鸿沟（Semantic Gap）
  - 语义鸿沟：输入数据的底层特征和高层语义信息之间的不一致性和差异性
  - 核心问题一：什么是一个好的表示
  - 核心问题二：如何学习到好的表示

深度学习关系

- 好的表示
  - 具有**很强的表示能力**，即同样大小的向量可以表示更多信息
  - 使后续的学习任务变得简单，即需要**包含更高层的语义信息**
  - **具有一般性**，是任务或领域独立的，可以迁移到其他任务上
- 学习好的表示
  - 关键是构建**具有一定深度的多层次特征表示**

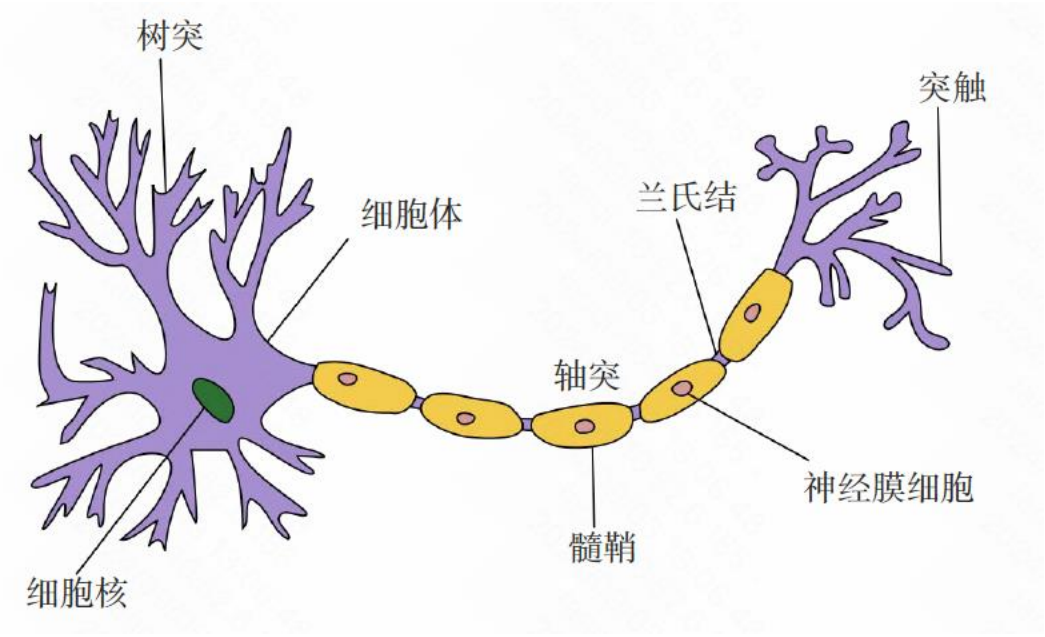
深度学习关系

- 深度学习（Deep Learning, DL）
  - 是机器学习的一个子问题，特征学习+预测模型
  - 构建并从数据中学习具有一定“深度”的模型
  - 关键问题是贡献度分配问题
  - 神经网络使用误差反向传播算法可以较好解决关键问题



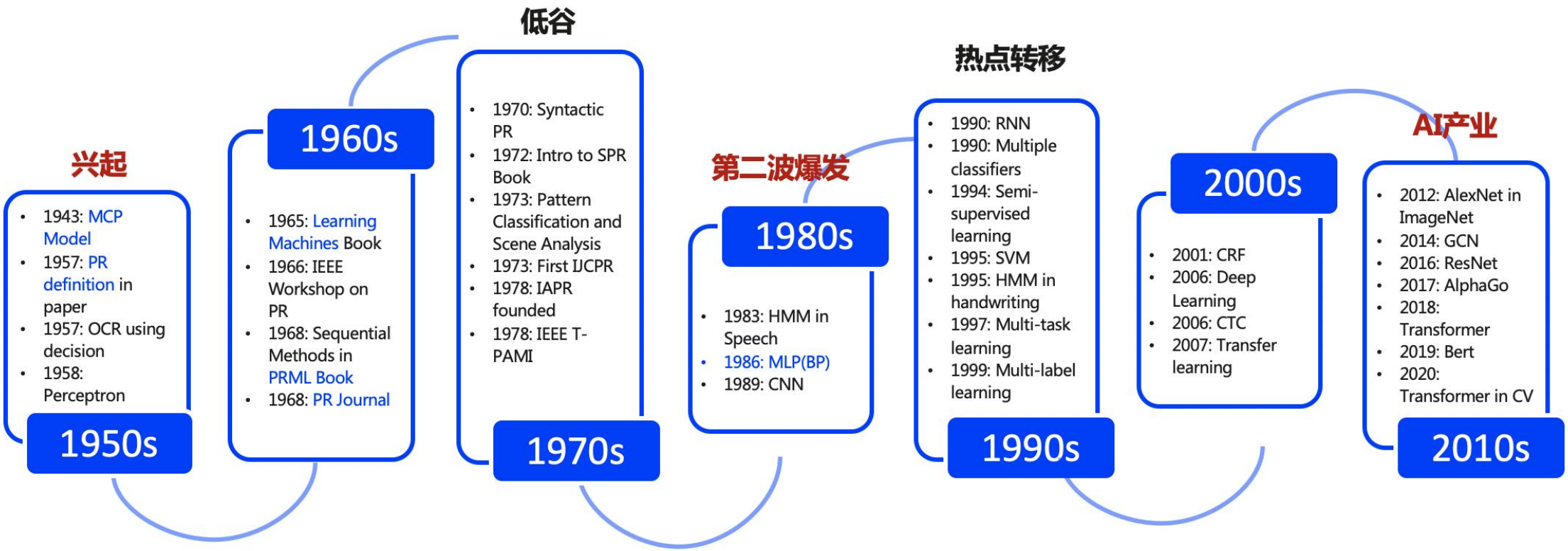
深度学习关系

- 人工神经网络
  - 模拟人脑神经网络而设计的一种计算模型
  - 由大量神经元通过极其丰富和完善的连接而构成的自适应非线性动态系统
  - 一个通用的函数逼近器
  - 反向传播算法



深度学习重要历史节点

深度学习发展历程



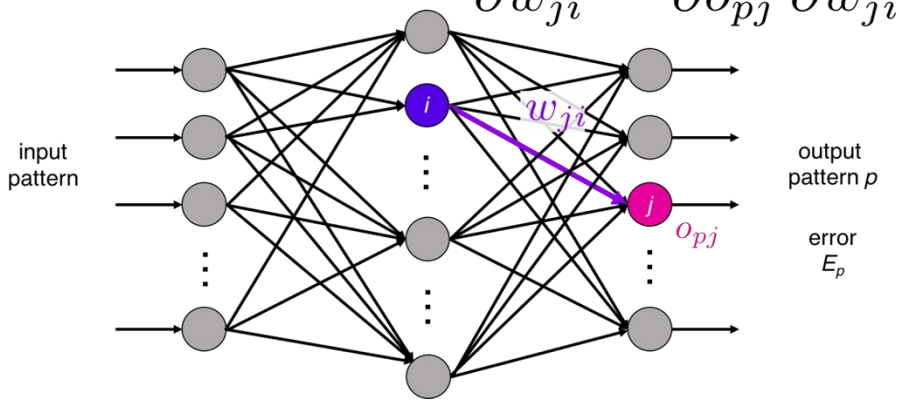
深度学习重要历史节点

- 深度学习发展历史阶段
  - 1943年~1969年：模型提出
  - 1969年~1983年：冰河期
  - 1983年~1995年：反向传播算法引起的复兴
  - 1995年~2006年：流行度降低
  - 2006年~至今：深度学习的崛起

深度学习重要历史节点

- 反向传播算法（Backprop: Rumelhart, Hinton, and Williams, 1986)
  - 反向传播计算梯度进行网络训练
  - 最成功的神经网络学习算法

$$\frac{\partial E_p}{\partial w_{ji}} = \frac{\partial E_p}{\partial o_{pj}} \frac{\partial o_{pj}}{\partial w_{ji}}$$



第1阶段：激励传播 [编辑]

每次迭代中的传播环节包含两步：

1. （前向传播阶段）将训练输入送入网络以获得预测结果；
2. （反向传播阶段）对预测结果同训练目标求差(损失函数)。

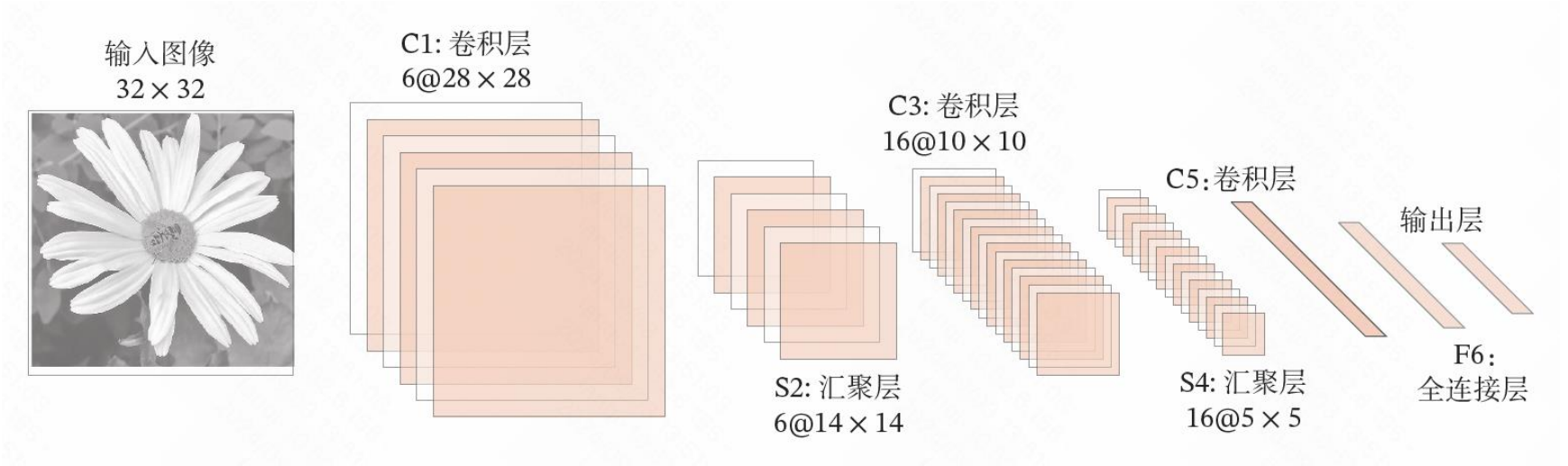
第2阶段：权重更新 [编辑]

对于每个突触上的权重，按照以下步骤进行更新：

1. 将输入激励和响应误差相乘，从而获得权重的梯度；
2. 将这个梯度乘上一个比例并取反后加到权重上。

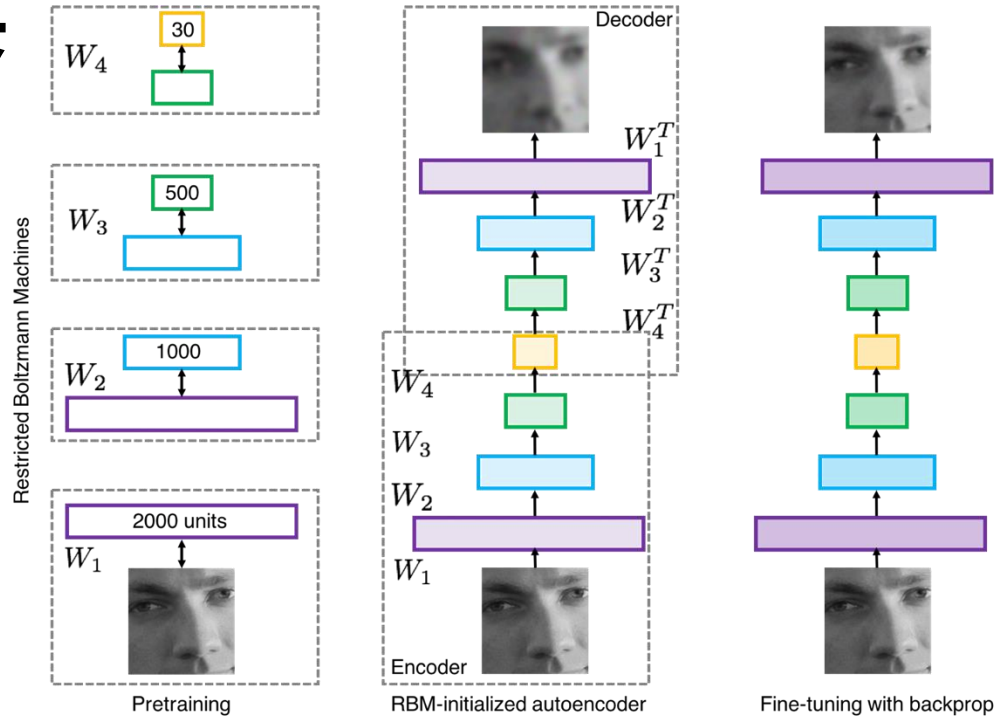
深度学习重要历史节点

- LeNet：Convolutional Networks: LeCun et al, 1998
  - 将反向传播算法引入卷积神经网络
  - 成功应用于手写字体识别系统，被银行系统广泛应用
  - 与现代神经网络非常相似



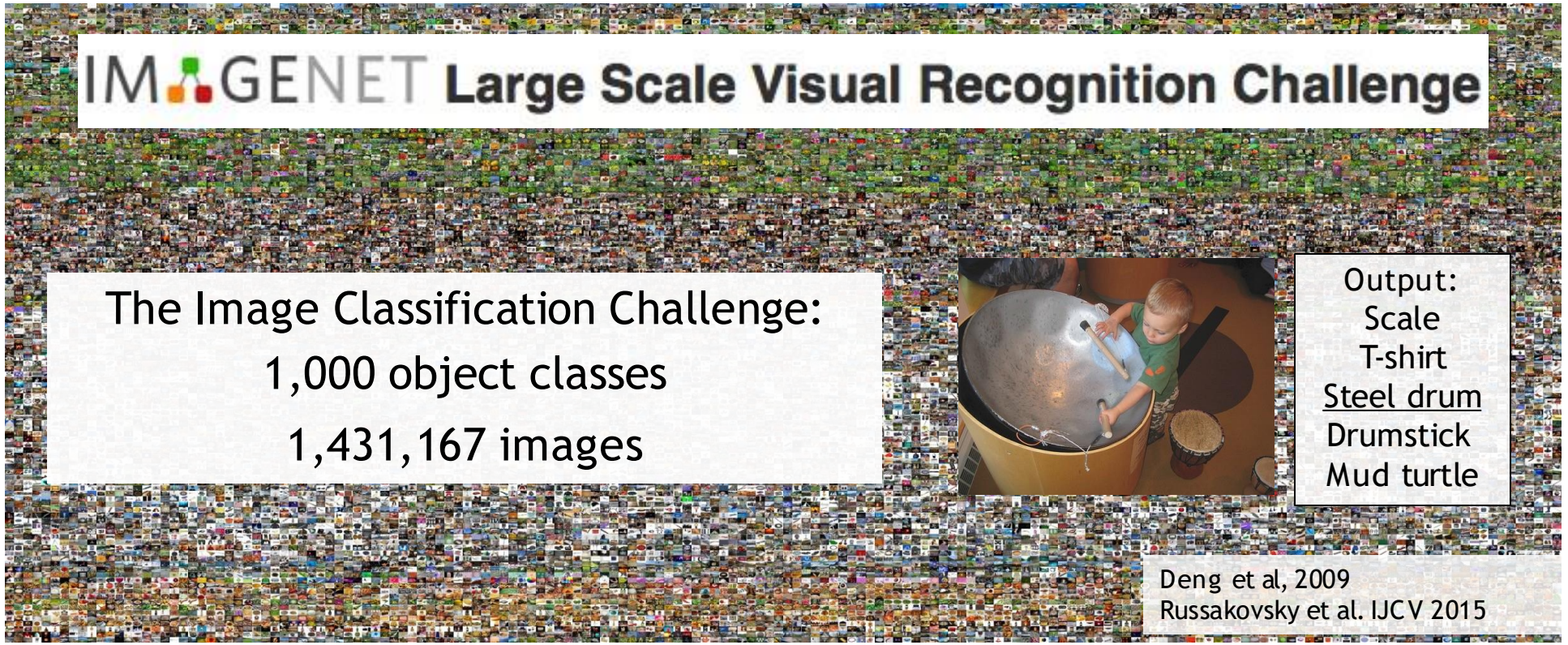
深度学习重要历史节点

- 深度信念网络（Hinton et al, 2006）
  - 提出“预训练 + 精调”方式来解决深度神经网络难以训练的问题
  - 逐层预训练学习一个深度信念网络
  - 将其权重作为一个多层前馈神经网络的初始化权重
  - 再用反向传播算法进行精调



深度学习重要历史节点

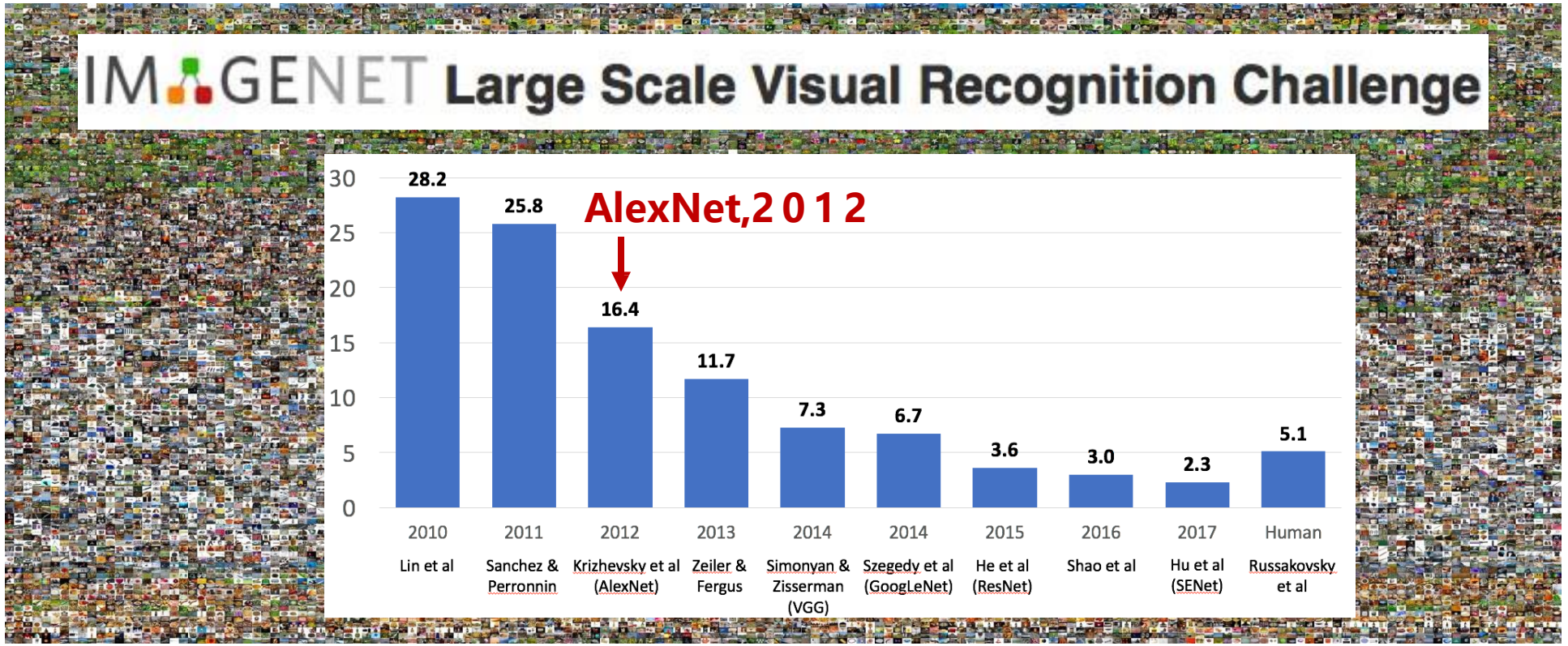
- 大规模数据集（ImageNet: Deng et al, 2009）
  - 1000个类别，1431167张图片



深度学习重要历史节点

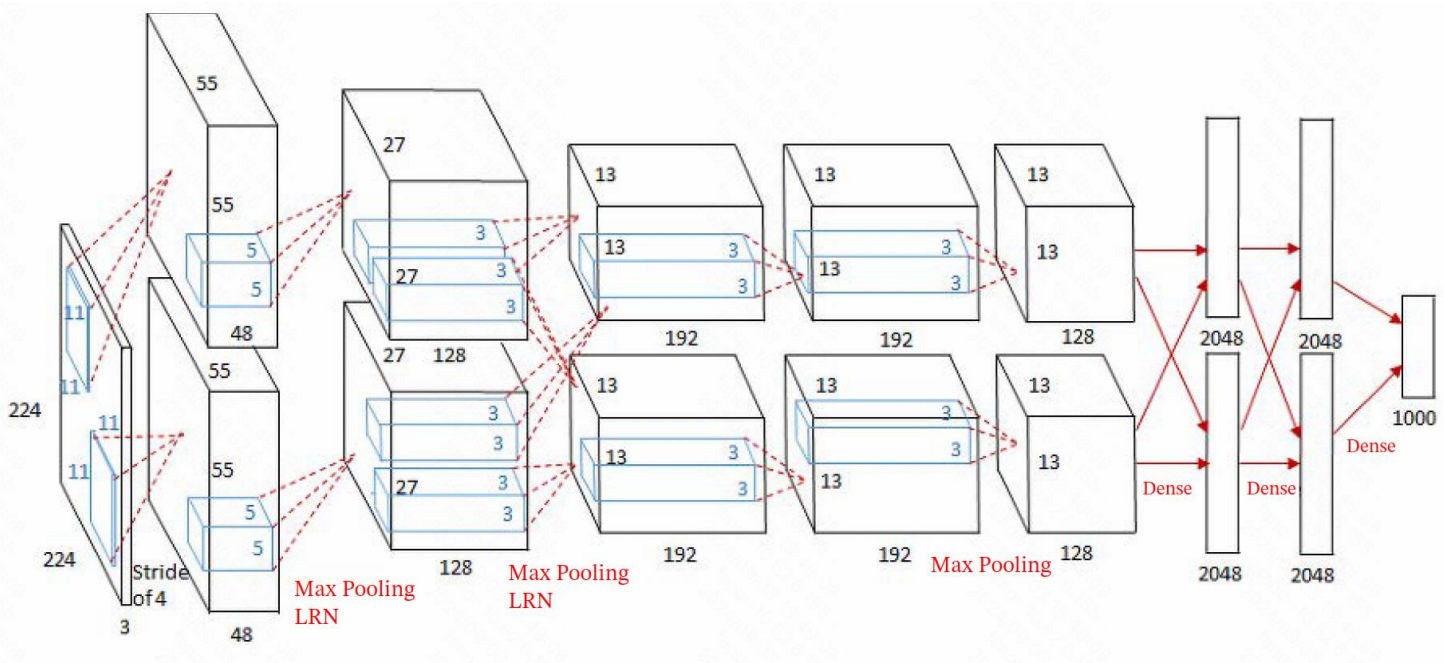
南京大学 NANJING UNIVERSITY

- 人们认识到数据对深度学习的重要性



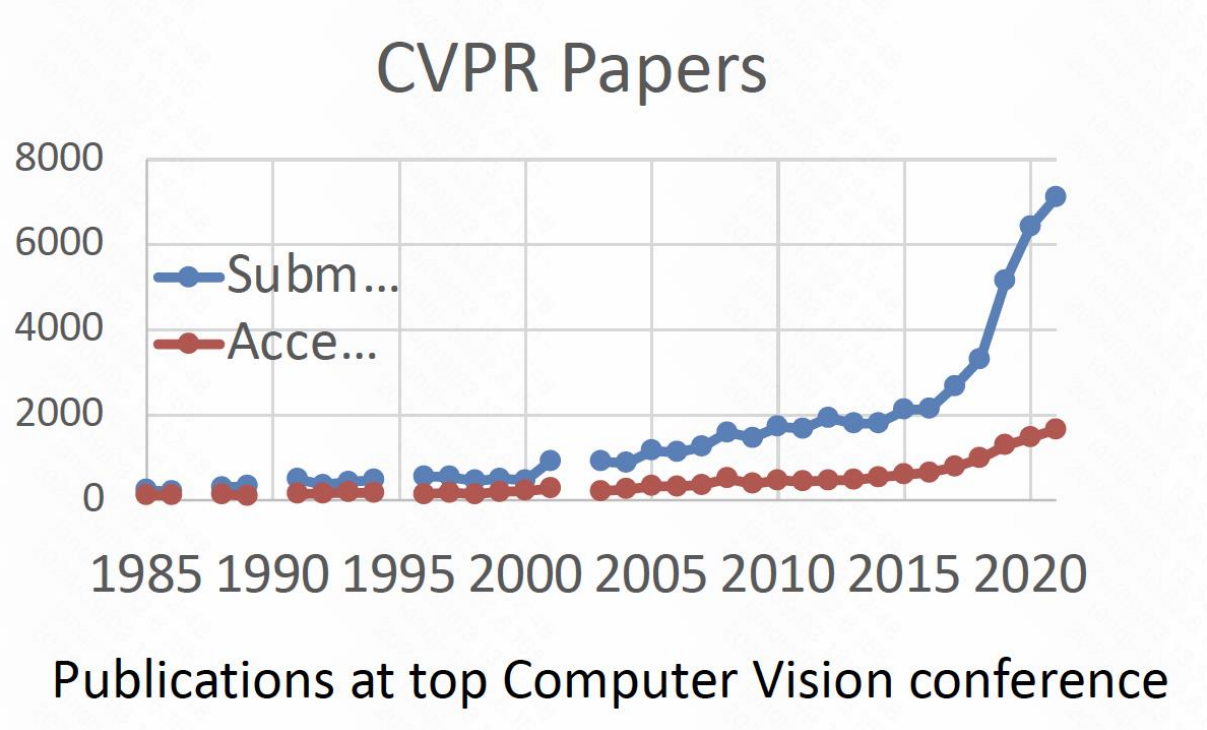
深度学习重要历史节点

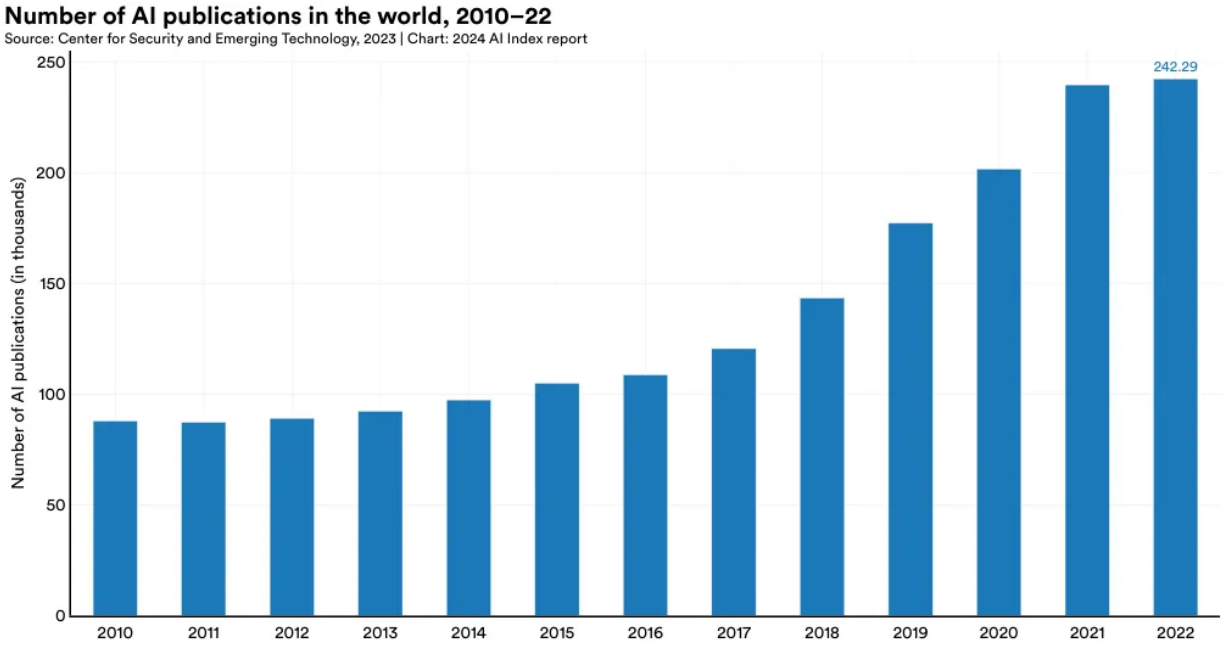
- 第一个现代深度卷积网络 AlexNet: Krizhevsky, et al, 2012
  - 端到端训练，使用GPU 进行并行训练
  - 使用 ReLU、Dropout、数据增强等技术



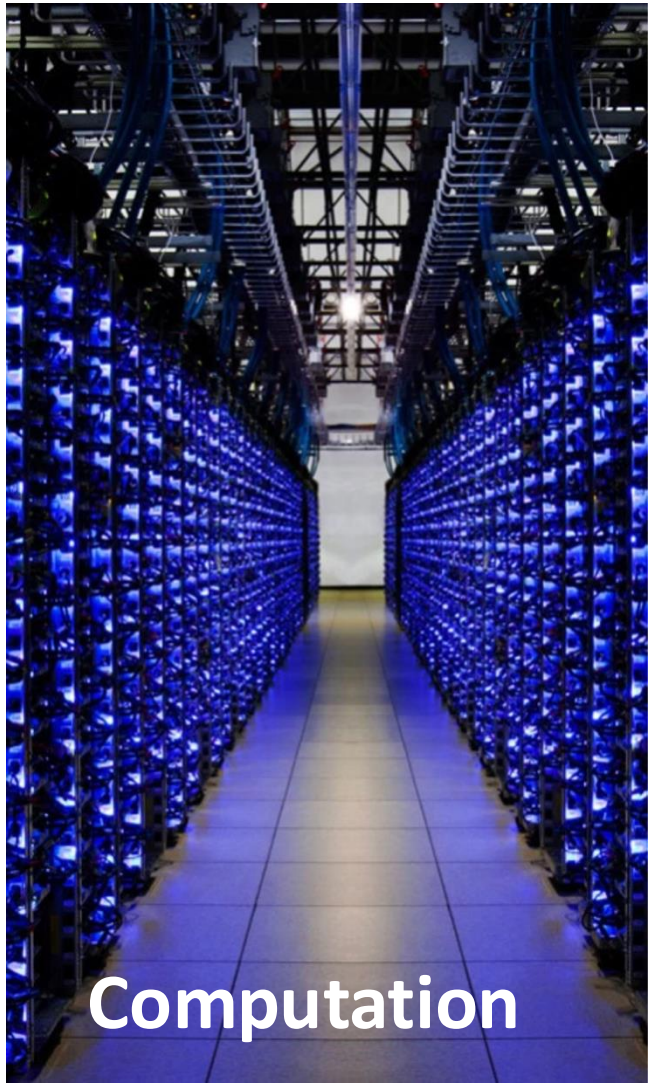
深度学习重要历史节点

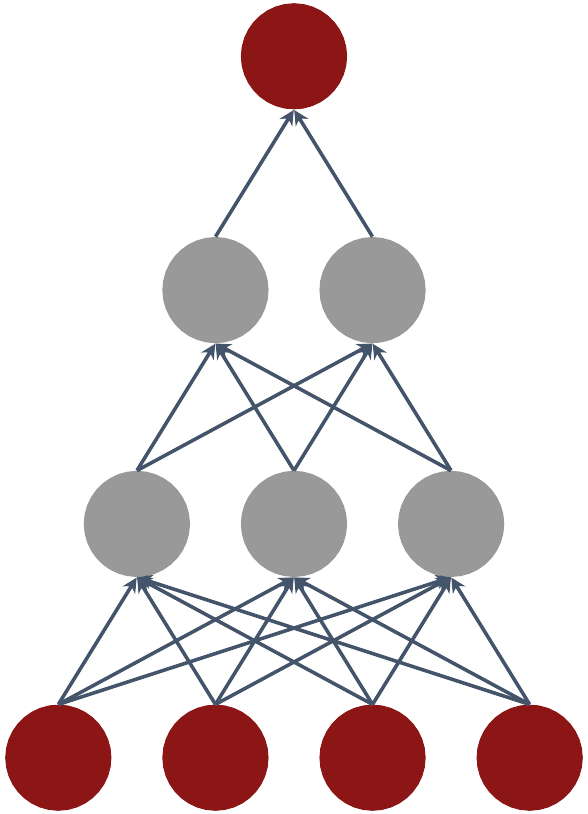
- **深度学习大爆发：2012年至今**



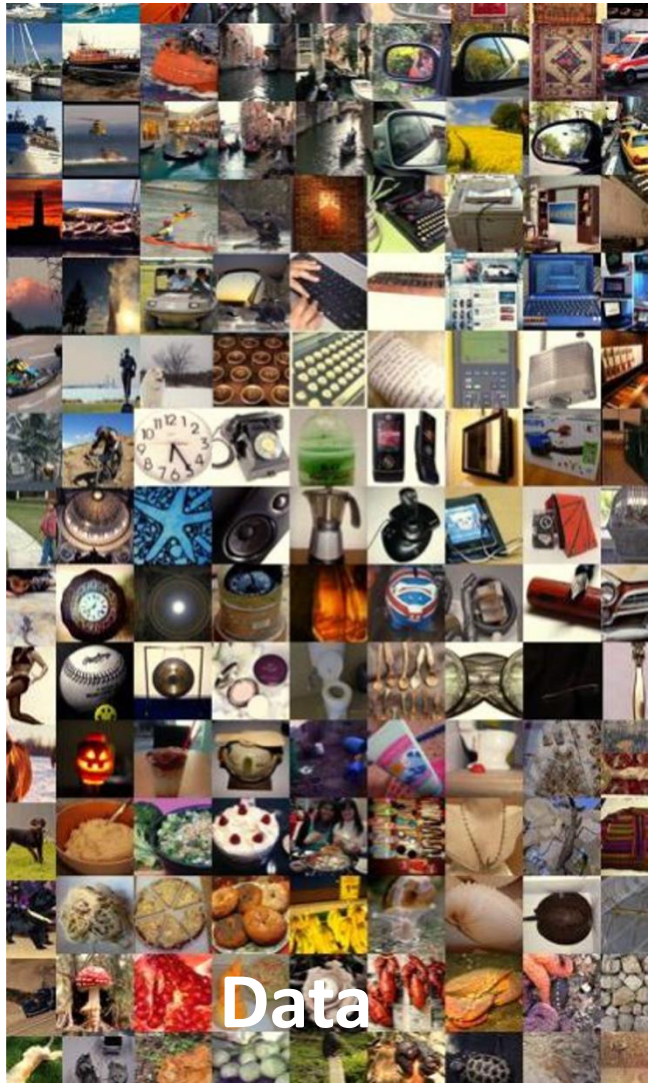


深度学习数据集





Algorithms



深度学习数据集

- Meta Pointer: A large collection organized by CV Datasets.
- Yet another Meta pointer
- Hugging face datasets: collection of generic datasets available on hugging face
- ImageNet: a large-scale image dataset for visual recognition organized by WordNet hierarchy
- SA-1B: dataset of a large number of images and segmentation masks to segment objects in those images
- COCO: large-scale object detection, segmentation, and captioning dataset
- Open Images: a dataset of ~9M images annotated with image-level labels, object bounding boxes, object segmentation masks, visual relationships, and localized narratives

深度学习数据集

- Cityscapes Dataset: This dataset focuses on semantic understanding of urban street scenes, with pixel-level annotations for various object classes such as cars, pedestrians, and roads
- DeepFashion: a large-scale clothes dataset containing over 800,000 diverse fashion images annotated with bounding boxes, clothing categories, and attributes
- Objaverse: a large-scale 3D asset database
- SUN Database: a benchmark for scene recognition and object detection with annotated scene categories and segmented objects
- Places Database: a scene-centric database with 205 scene categories and 2.5 millions of labelled images
- NYU Depth Dataset v2: a RGB-D dataset of segmented indoor scenes

深度学习数据集

- Flickr100M: 100 million creative commons Flickr images
- Labeled Faces in the Wild: a dataset of 13,000 labeled face photographs
- Human Pose Dataset: a benchmark for articulated human pose estimation
- YouTube Faces DB: a face video dataset for unconstrained face recognition in videos
- UCF101: an action recognition data set of realistic action videos with 101 action categories
- HMDB-51: a large human motion dataset of 51 action classes
- ActivityNet: A large-scale video dataset for human activity understanding
- Moments in Time: A dataset of one million 3-second videos
- Vision-language datasets: Visual Genome, Flickr30k, VQA v2, ADE20K, LAION

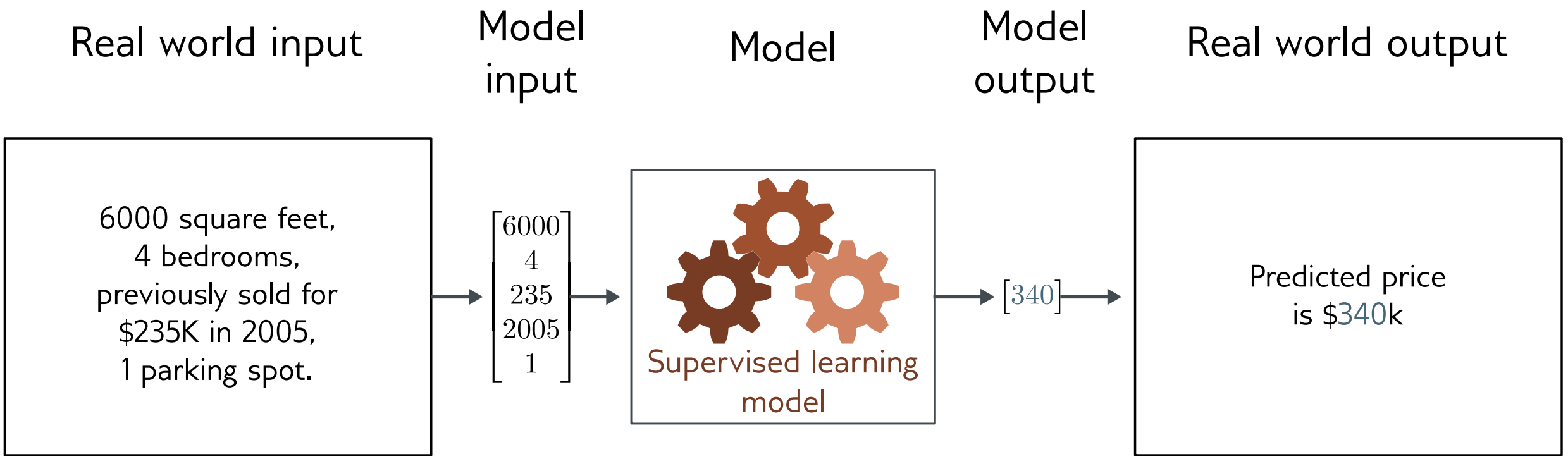
大　　纲

- 课程概览
- 深度学习历史
- 深度学习应用
- 深度学习平台
- 课程后续安排

深度学习应用

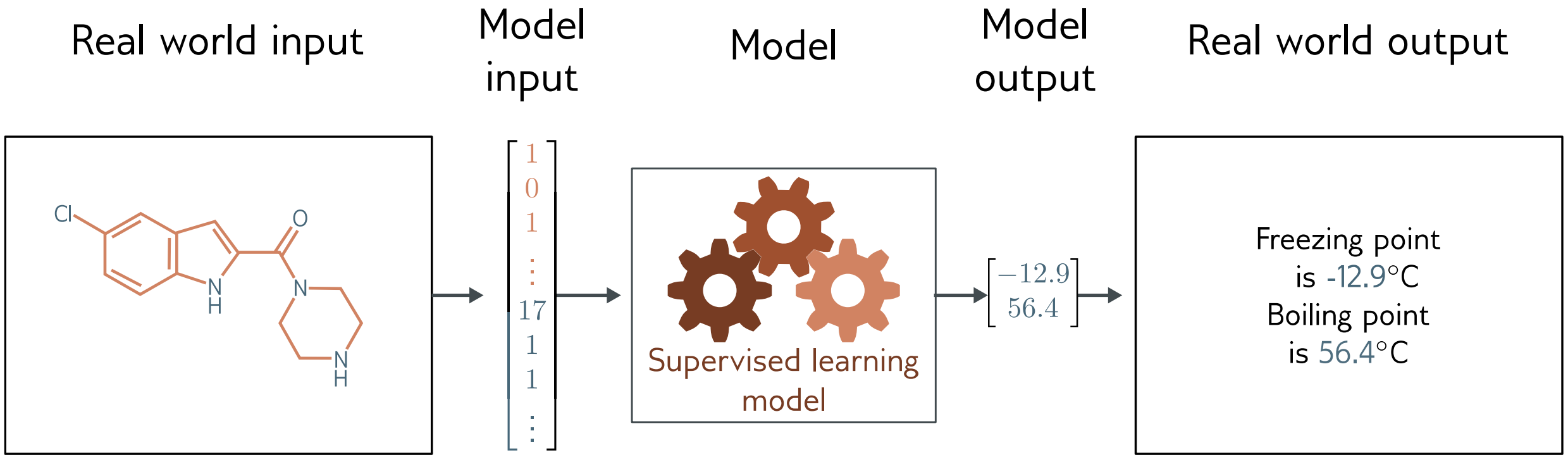
南京大学 NANJING UNIVERSITY

数值回归



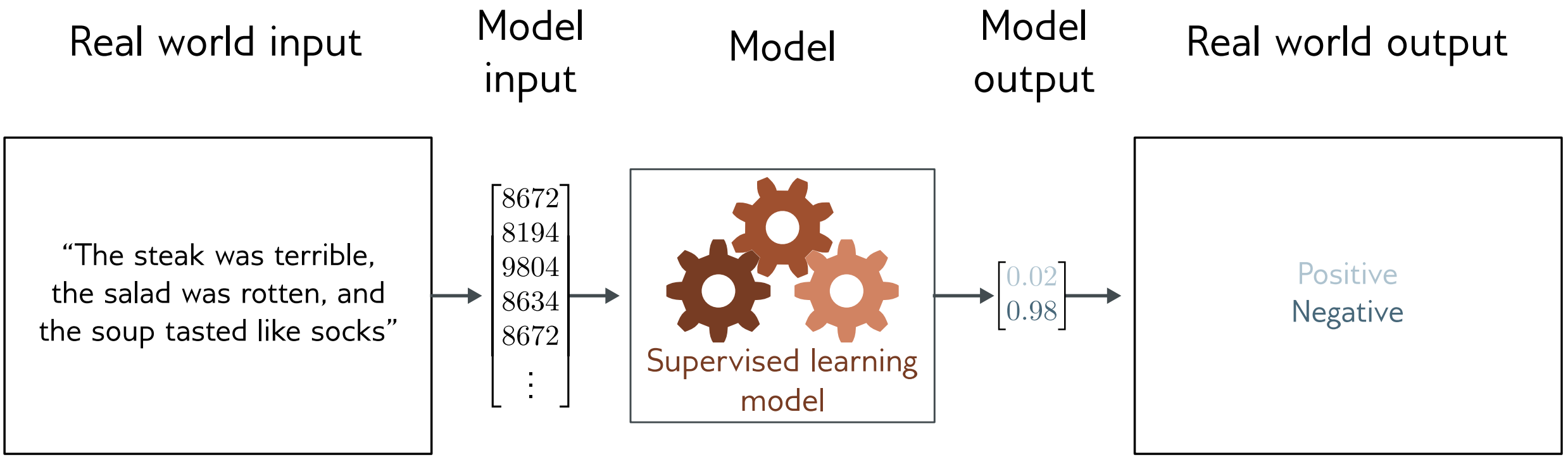
深度学习应用

图回归


Real world input
Model input
Model
Model output
Real world output
Cl
O
N
H
N
N
H
1
0
1
17
1
1
Supervised learning model
−12.9
56.4
Freezing point is -12.9°C
Boiling point is 56.4°C

深度学习应用

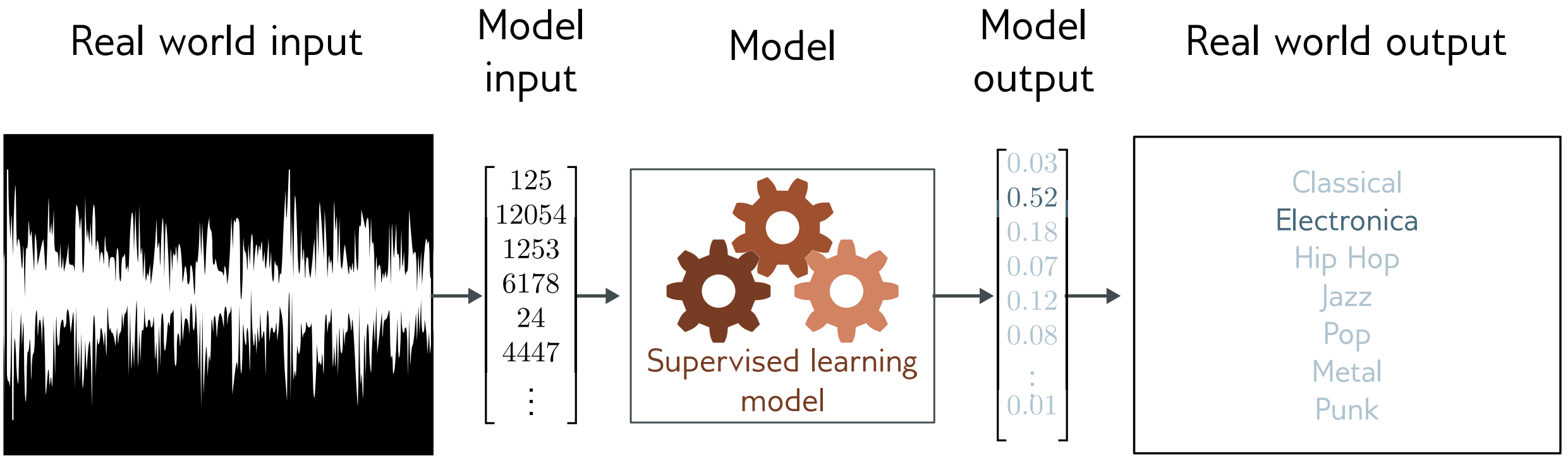
文本分类


Real world input
Model input
Model
Model output
Real world output
"The steak was terrible, the salad was rotten, and the soup tasted like socks"
8672
8194
9804
8634
8672
Supervised learning model
0.02
0.98
Positive
Negative

深度学习应用

南京大学 NANJING UNIVERSITY

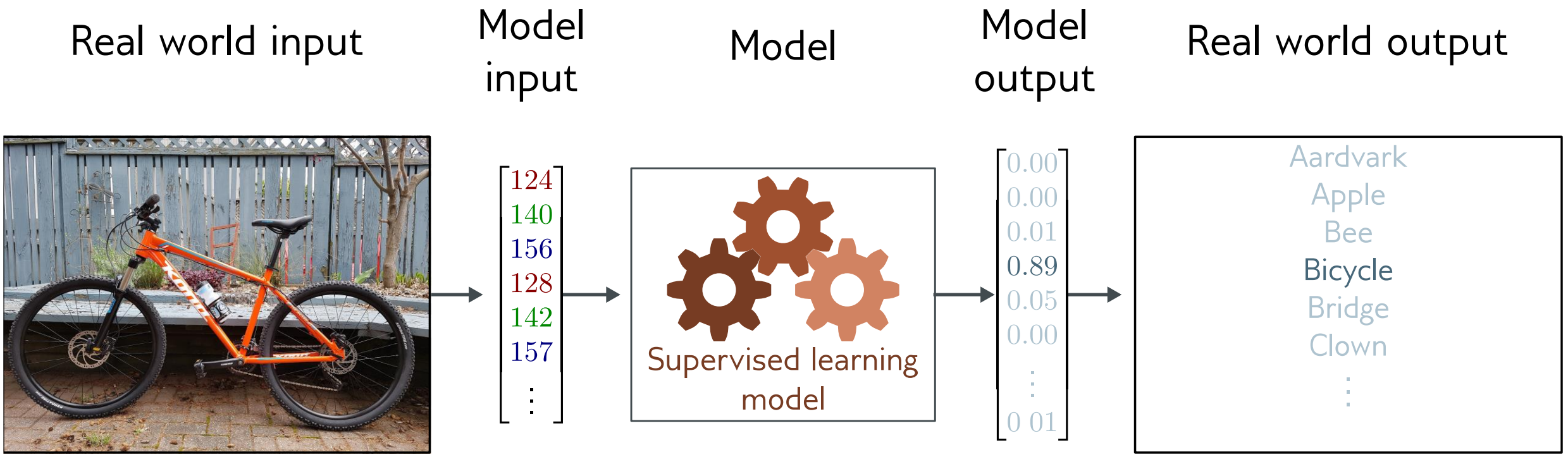
- **音乐分类**



深度学习应用

南京大学 NANJING UNIVERSITY

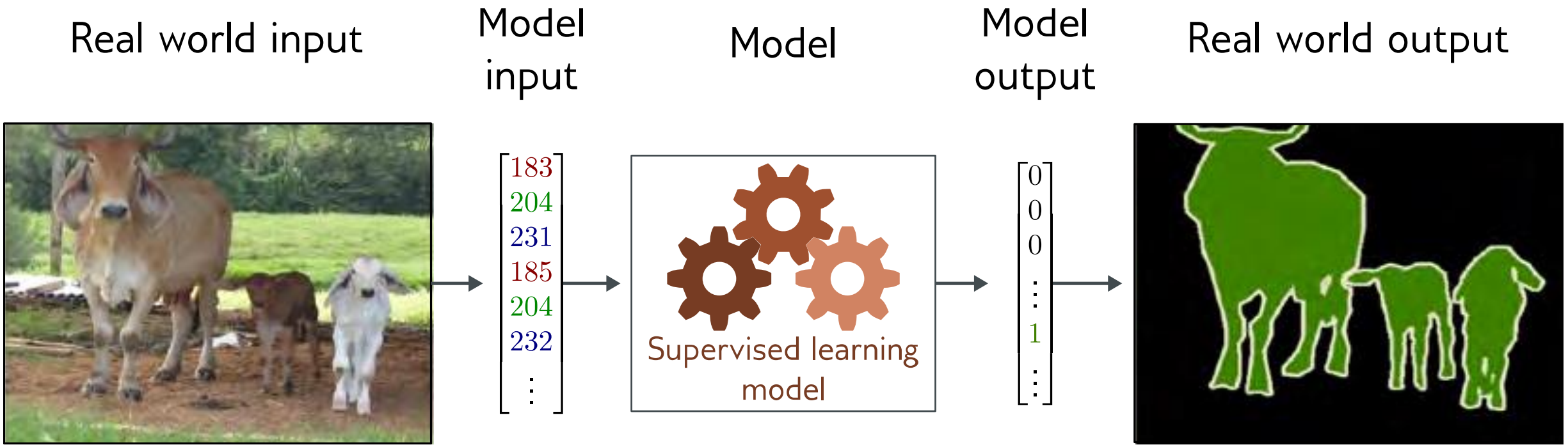
图像分类



深度学习应用

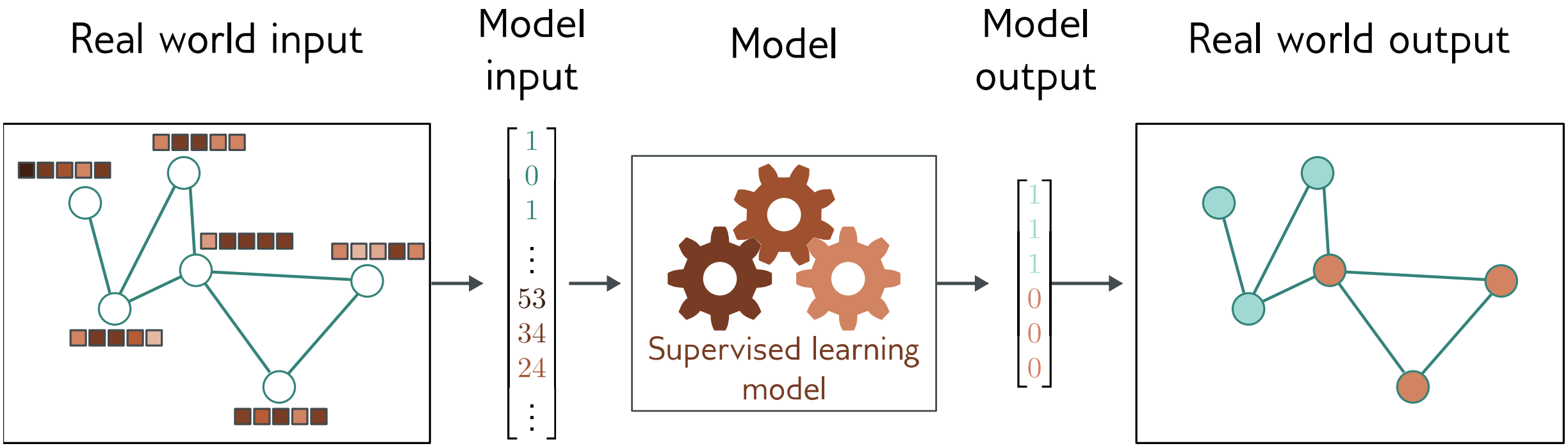
南京大学 NANJING UNIVERSITY

图像分割



深度学习应用

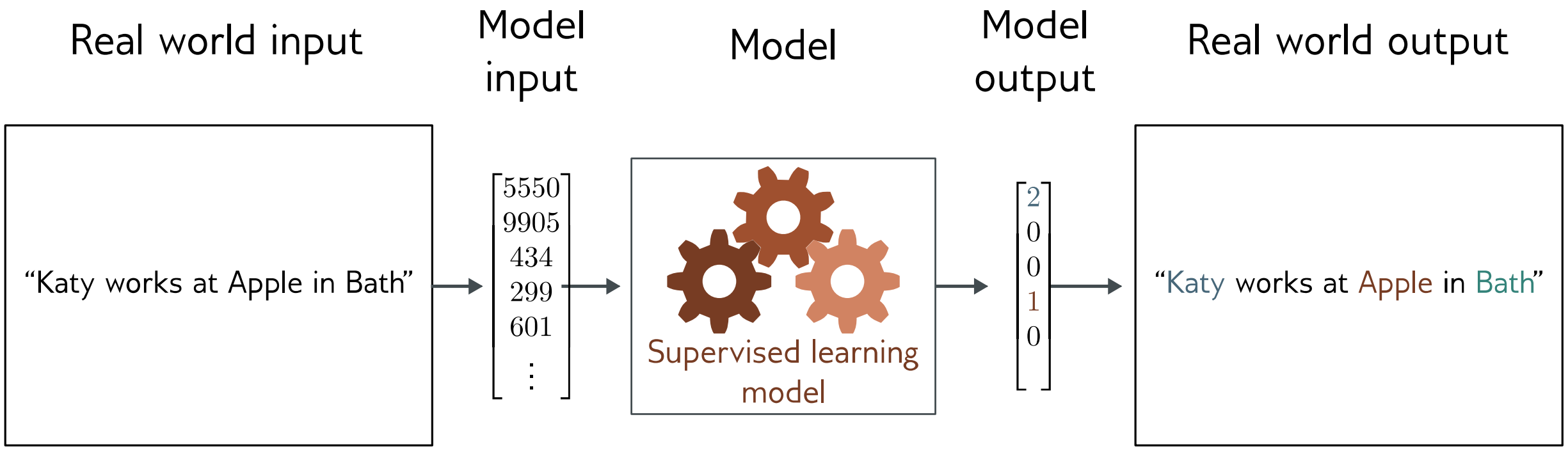
图节点分类



深度学习应用

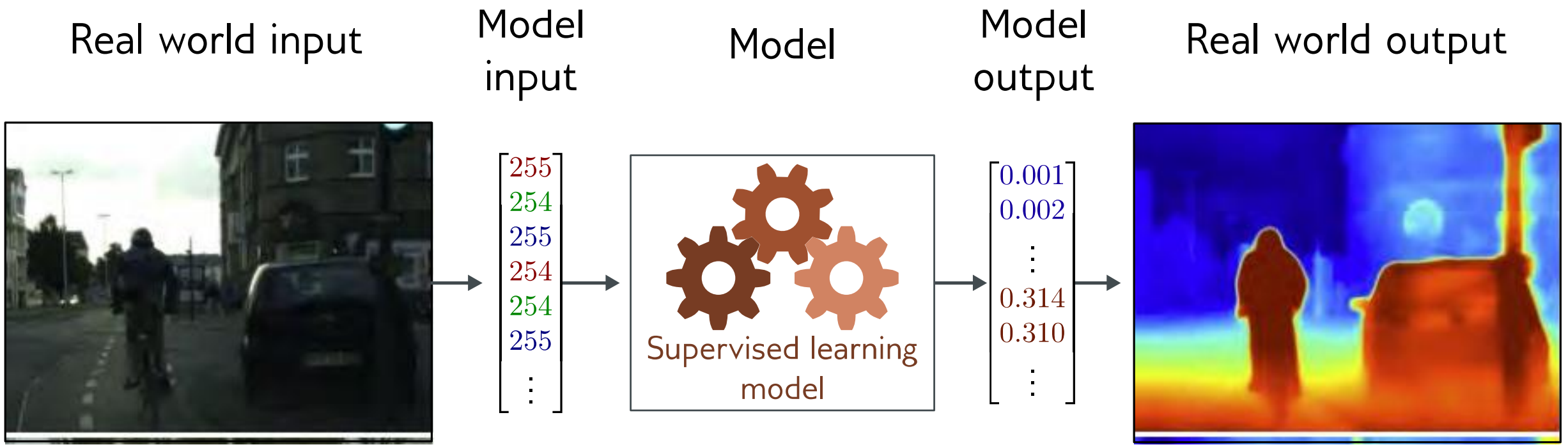
南京大学 NANJING UNIVERSITY

实体识别



深度学习应用

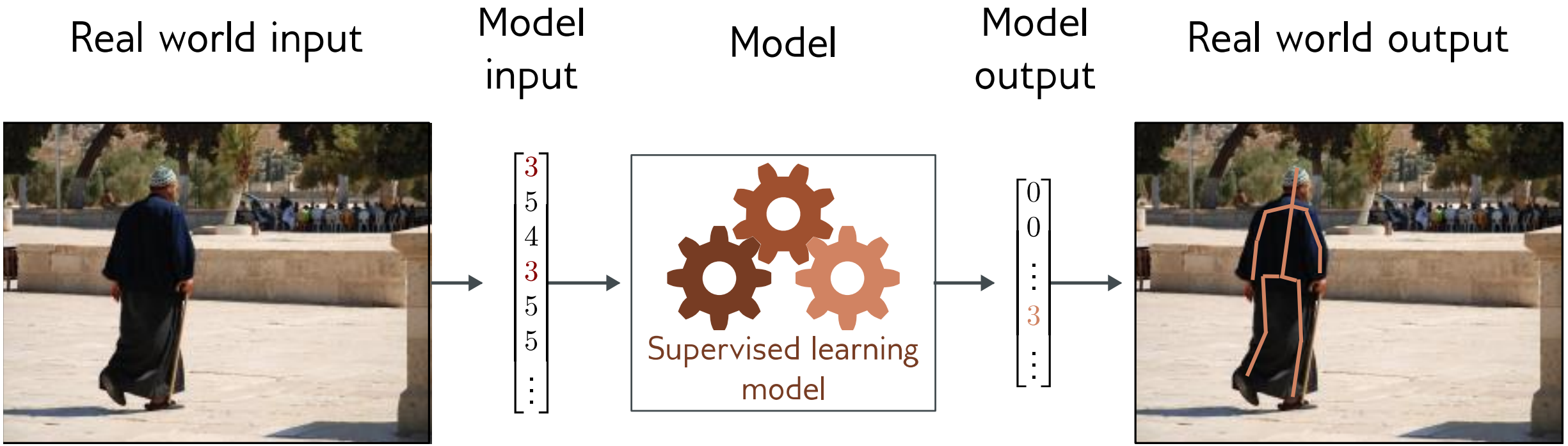
深度估计


Real world input
Model input
Model
Model output
Real world output
255
254
255
254
254
255
Supervised learning model
0.001
0.002
0.314
0.310

深度学习应用

南京大学 NANJING UNIVERSITY

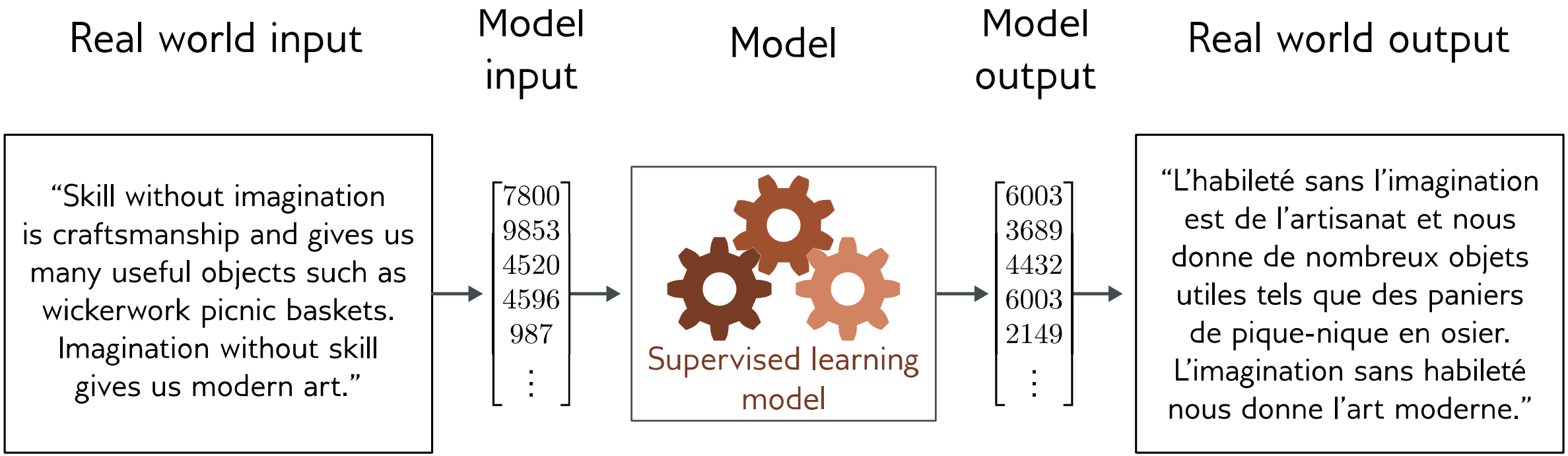
姿态估计



深度学习应用

南京大学 NANJING UNIVERSITY

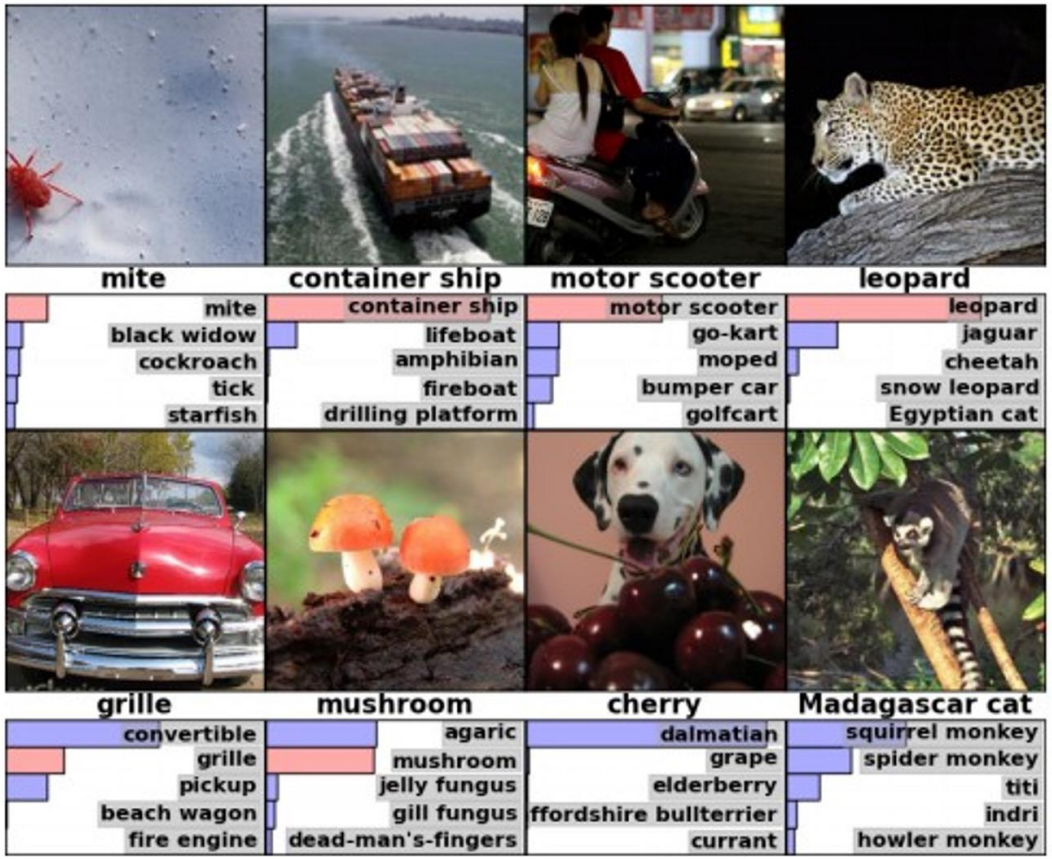
翻译

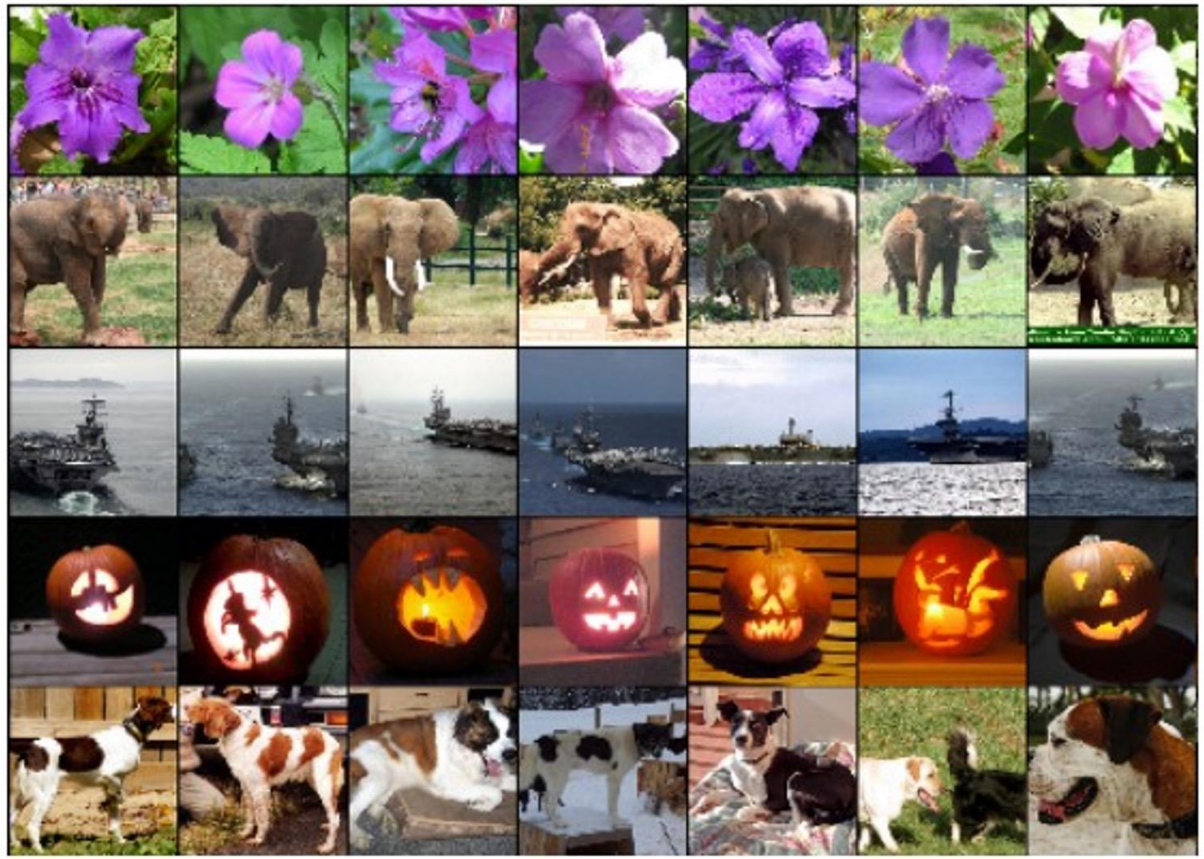


深度学习应用

南京大学 NANJING UNIVERSITY

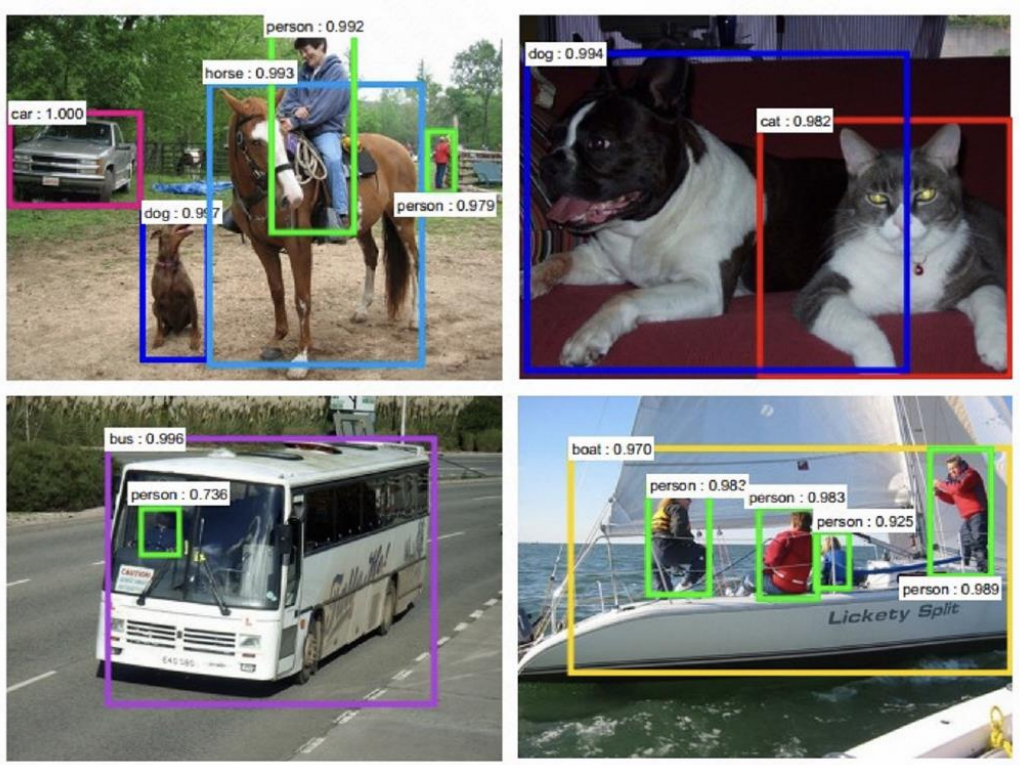
- 图像分类 & 图像检索

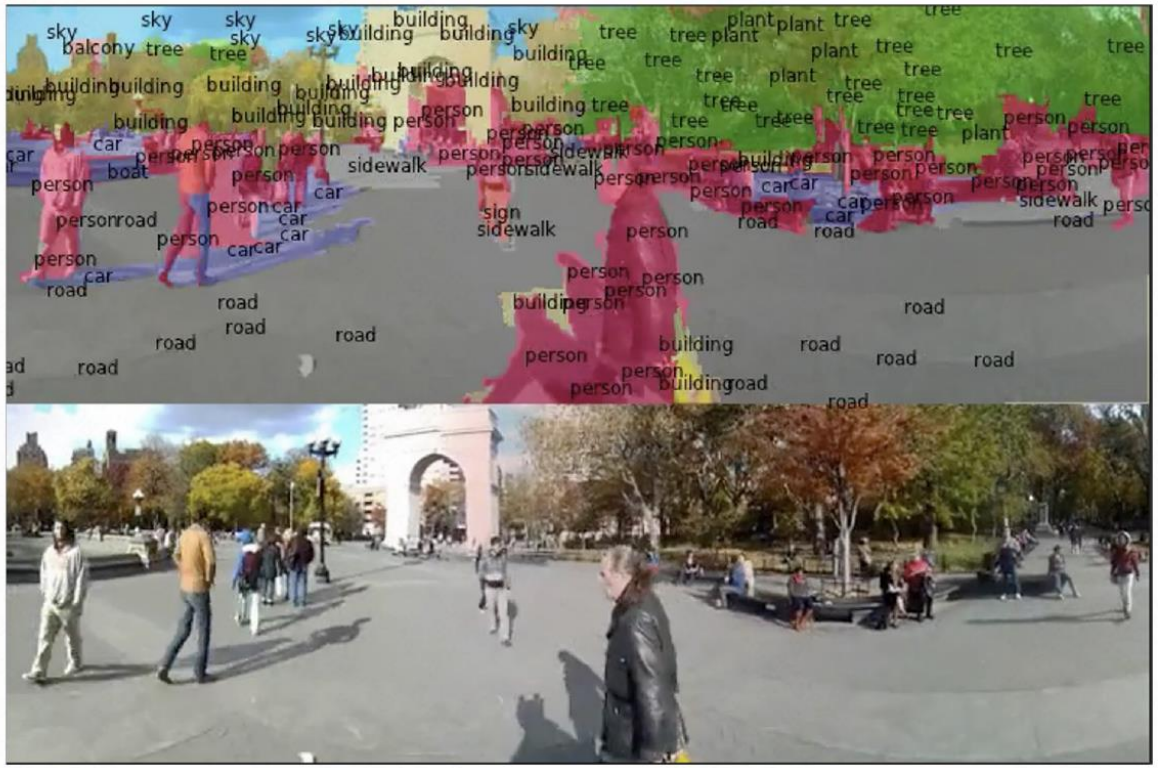




深度学习应用

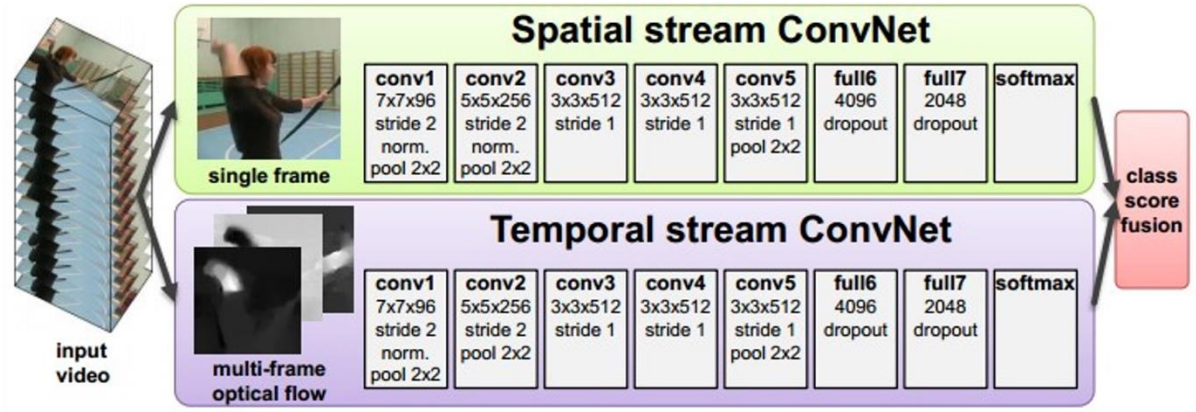
- 目标检测 & 图像分割

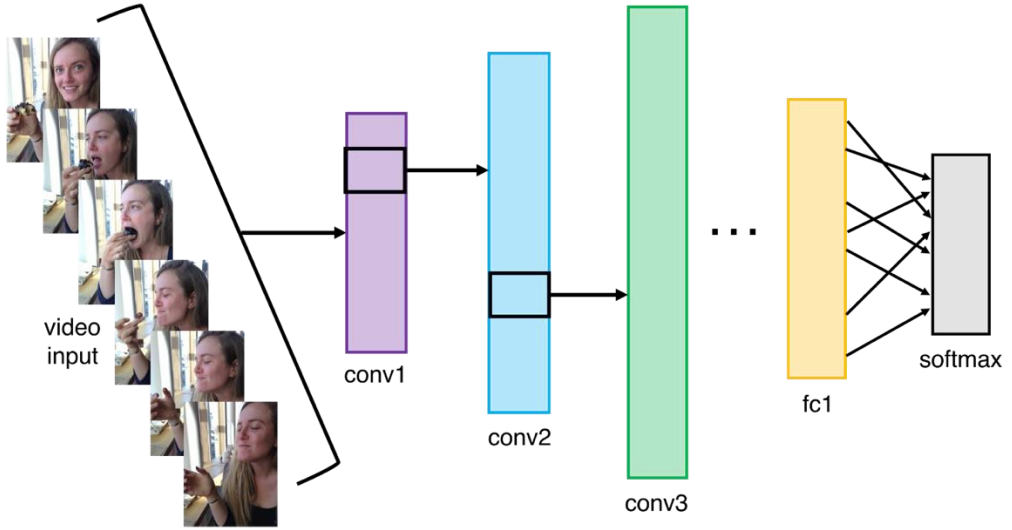


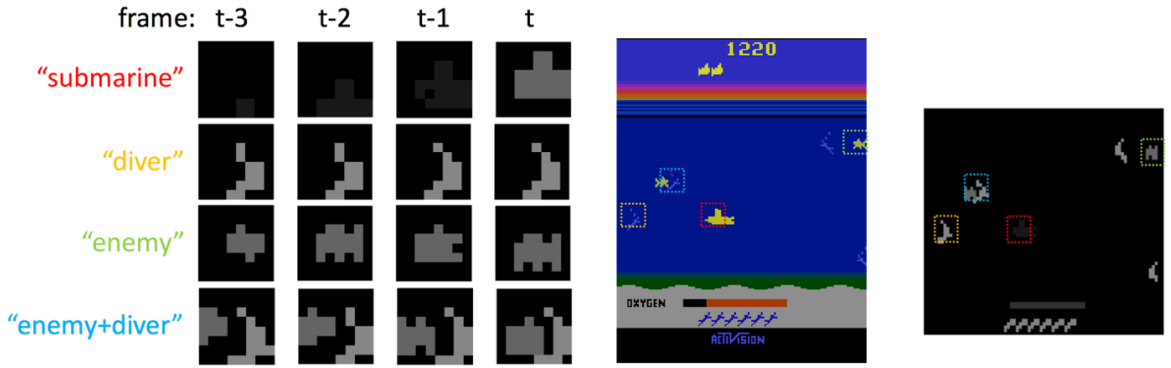


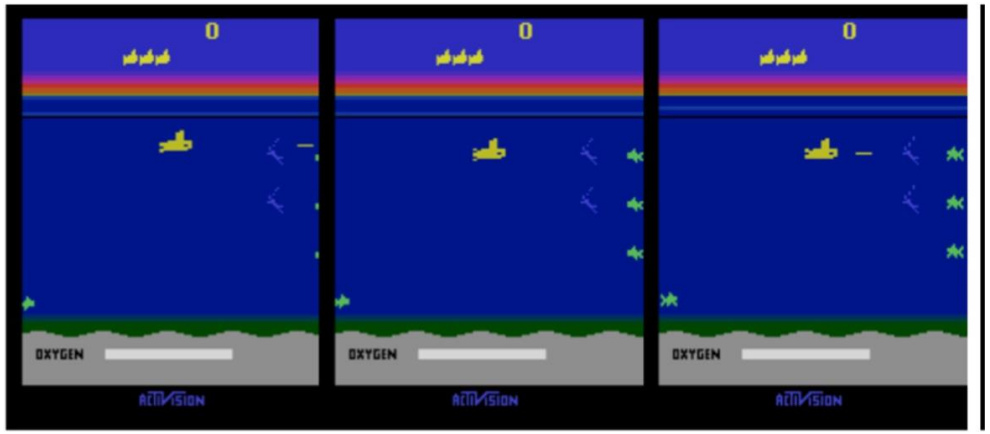
深度学习应用

- 视频分类 & 行为识别 & 强化学习



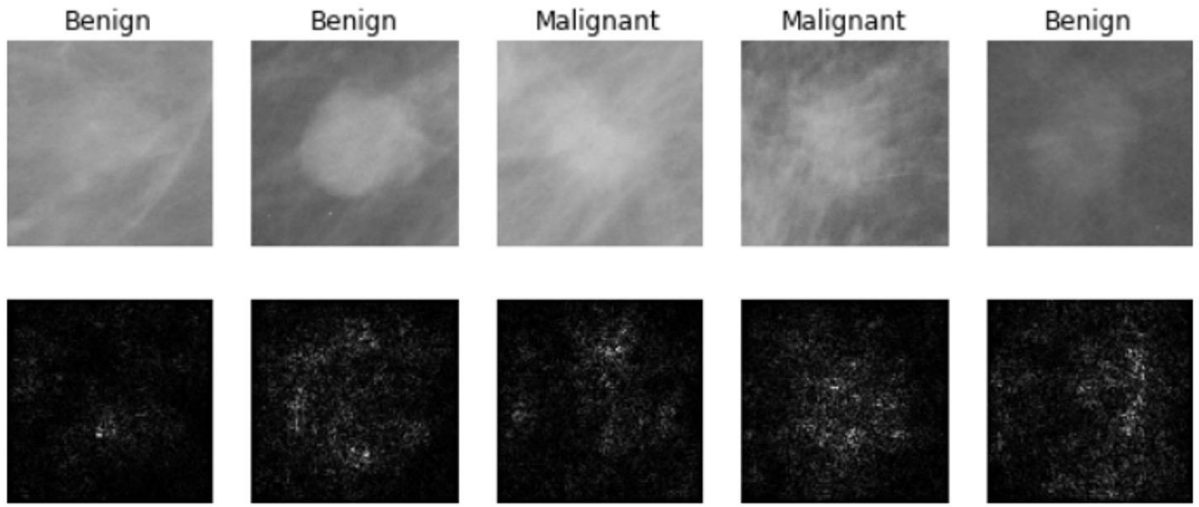






深度学习应用

- 医疗图像识别 & 星系识别 & 生物识别



南京大学 NANJING UNIVERSITY

图像描述

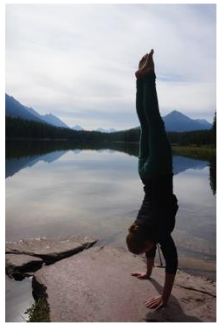
A white teddy bear sitting in the grass

A man in a baseball uniform throwing a ball

A woman is holding a cat in her hand

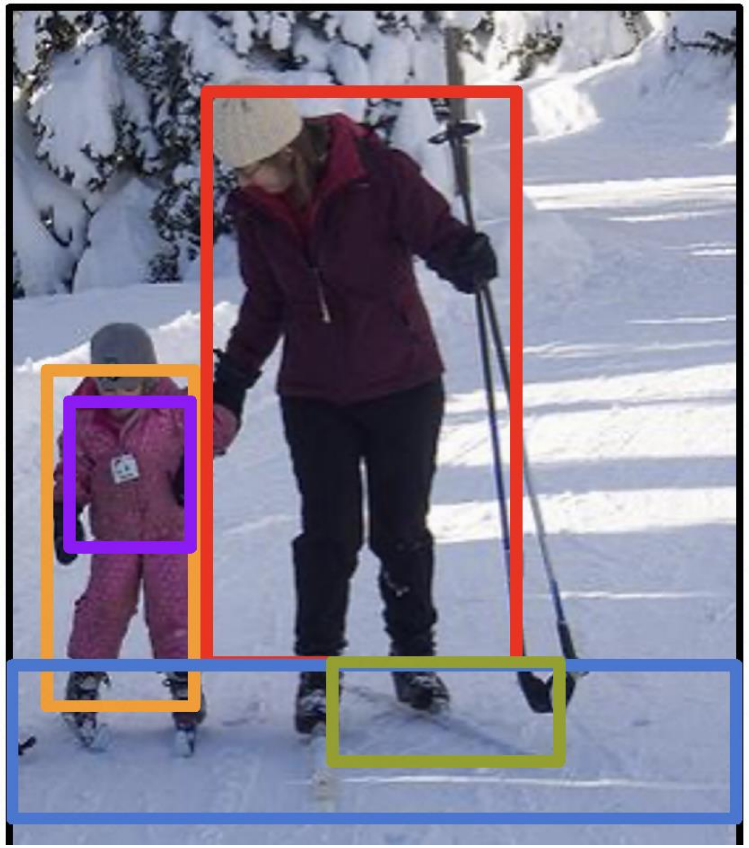
A man riding a wave on top of a surfboard

A cat sitting on a suitcase on the floor

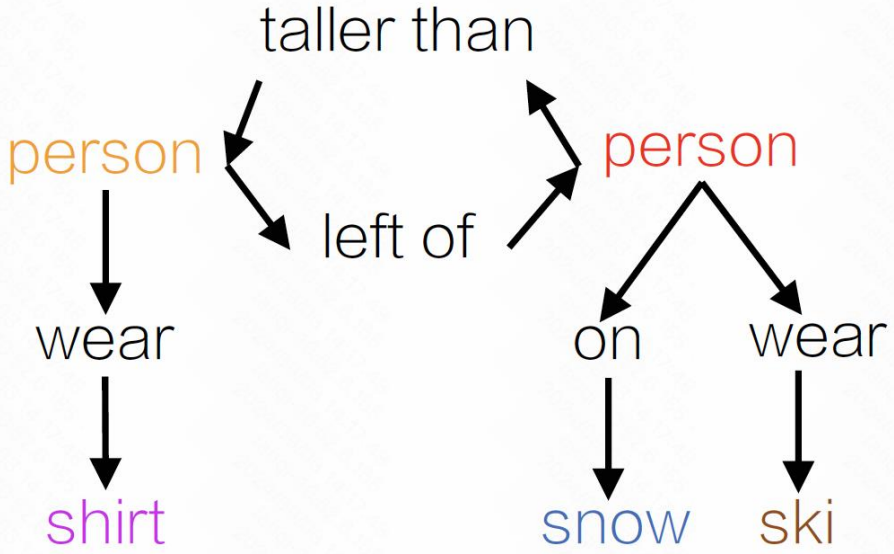


A woman standing on a beach holding a surfboard

视觉关系检测



Results:
spatial, comparative, asymmetrical,
verb, prepositional



深度学习应用

南京大学 NANJING UNIVERSITY

- 风格迁移

深度学习应用

南京大学 NANJING UNIVERSITY

- **图像生成**

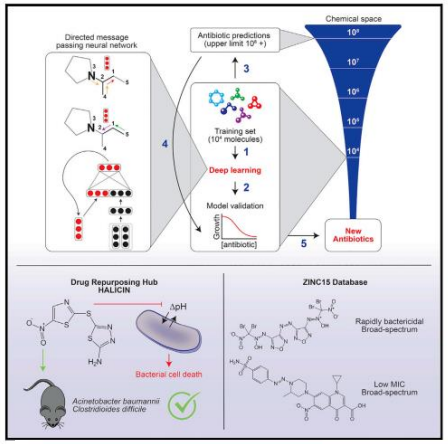
深度学习应用

AI4Science

Cell

A Deep Learning Approach to Antibiotic Discovery

Graphical Abstract



Authors

Jonathan M. Stokes, Kevin Yang, Kyle Swanson, ..., Tommi S. Jaakkola, Regina Barzilay, James J. Collins

Correspondence

regina@csail.mit.edu (R.B.), jimjc@mit.edu (J.J.C.)

In Brief

A trained deep neural network predicts antibiotic activity in molecules that are structurally different from known antibiotics, among which Halicin exhibits efficacy against broad-spectrum bacterial infections in mice.

nature biomedical engineering

ARTICLES

https://doi.org/10.1038/s41551-018-0195-0

Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning

Ryan Poplin[1,4], Avinash V. Varadarajan[1,4], Katy Blumer[1], Yun Liu[1], Michael V. McConnell[2,3], Greg S. Corrado[1], Lily Peng[1,4]* and Dale R. Webster[1,4]

Traditionally, medical discoveries are made by observing associations, making hypotheses from them and then designing and running experiments to test the hypotheses. However, with medical images, observing and quantifying associations can often be difficult because of the wide variety of features, patterns, colours, values and shapes that are present in real data. Here, we show that deep learning can extract new knowledge from retinal fundus images. Using deep-learning models trained on data from 284,335 patients and validated on two independent datasets of 12,026 and 999 patients, we predicted cardiovascular risk factors not previously thought to be present or quantifiable in retinal images, such as age (mean absolute error within 3.26 years), gender (area under the receiver operating characteristic curve (AUC) = 0.97), smoking status (AUC = 0.71), systolic blood pressure (mean absolute error within 11.23 mmHg) and major adverse cardiac events (AUC = 0.70). We also show that the trained deep-learning models used anatomical features, such as the optic disc or blood vessels, to generate each prediction.

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

Article | Open access | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

John Jumper ✉, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, ... Demis Hassabis ✉

+ Show authors

Nature **596**, 583–589 (2021) | Cite this article

1.42m Accesses | **12k** Citations | **3493** Altmetric | Metrics

Can you write me a report analyzing this chest X-ray?

Findings:

- Devices: None.
- Lungs: No pneumothorax. No substantial pleural effusion. Lungs appear clear.
- Cardiomediastinal: Normal heart size. Mediastinal contours within normal limits.
- Other: No acute skeletal abnormality.

Impression:

No active disease seen in chest.

Enter a question here

深度学习应用

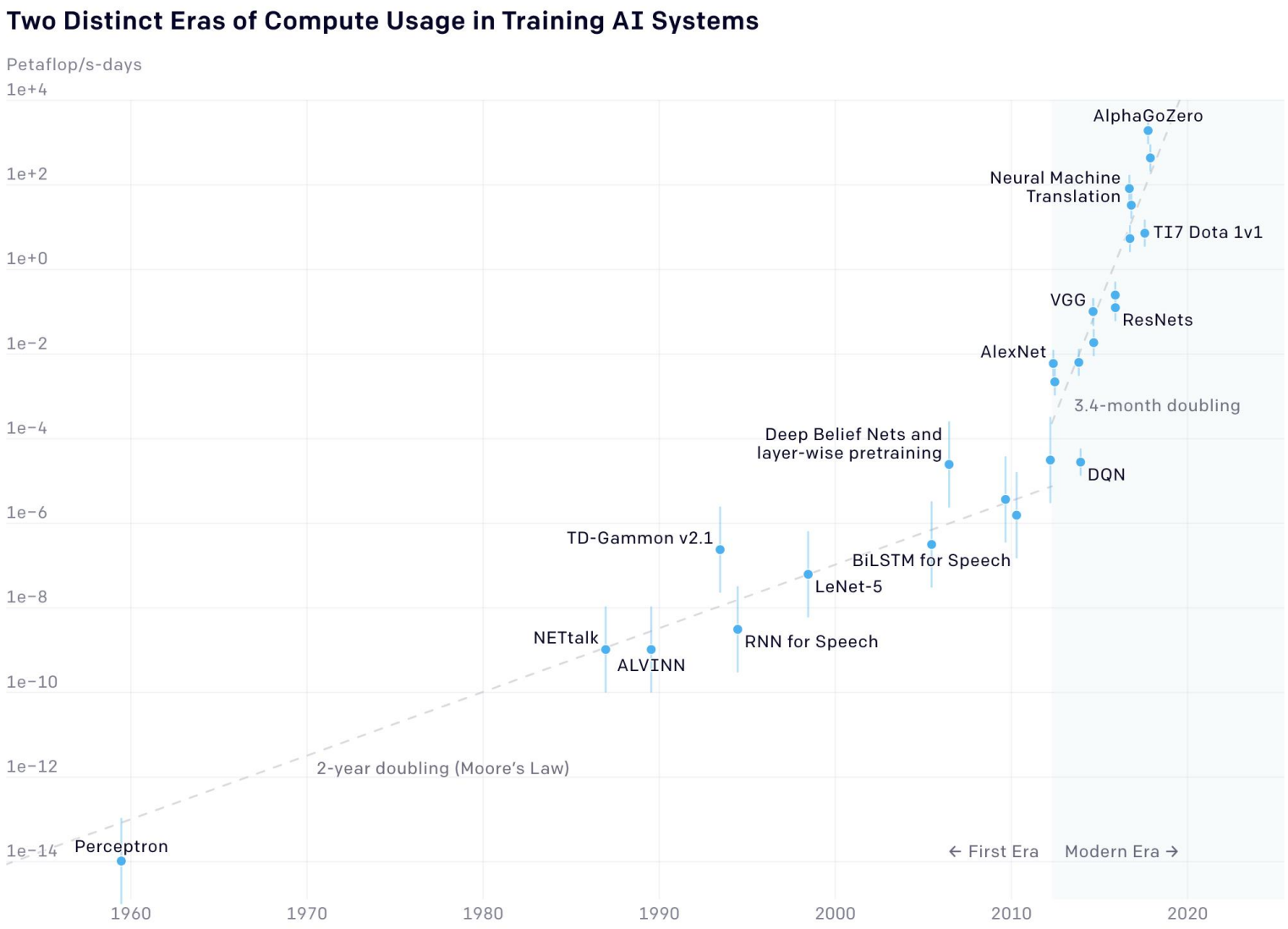
- 不同任务间的共性
  - 输入和输出之间具有非常复杂的关系（浅层学习难以处理）
  - 但输入和输出都遵循一定的规则（可以通过深度学习处理）

“A Kazakh man on a horse holding a bird of prey”

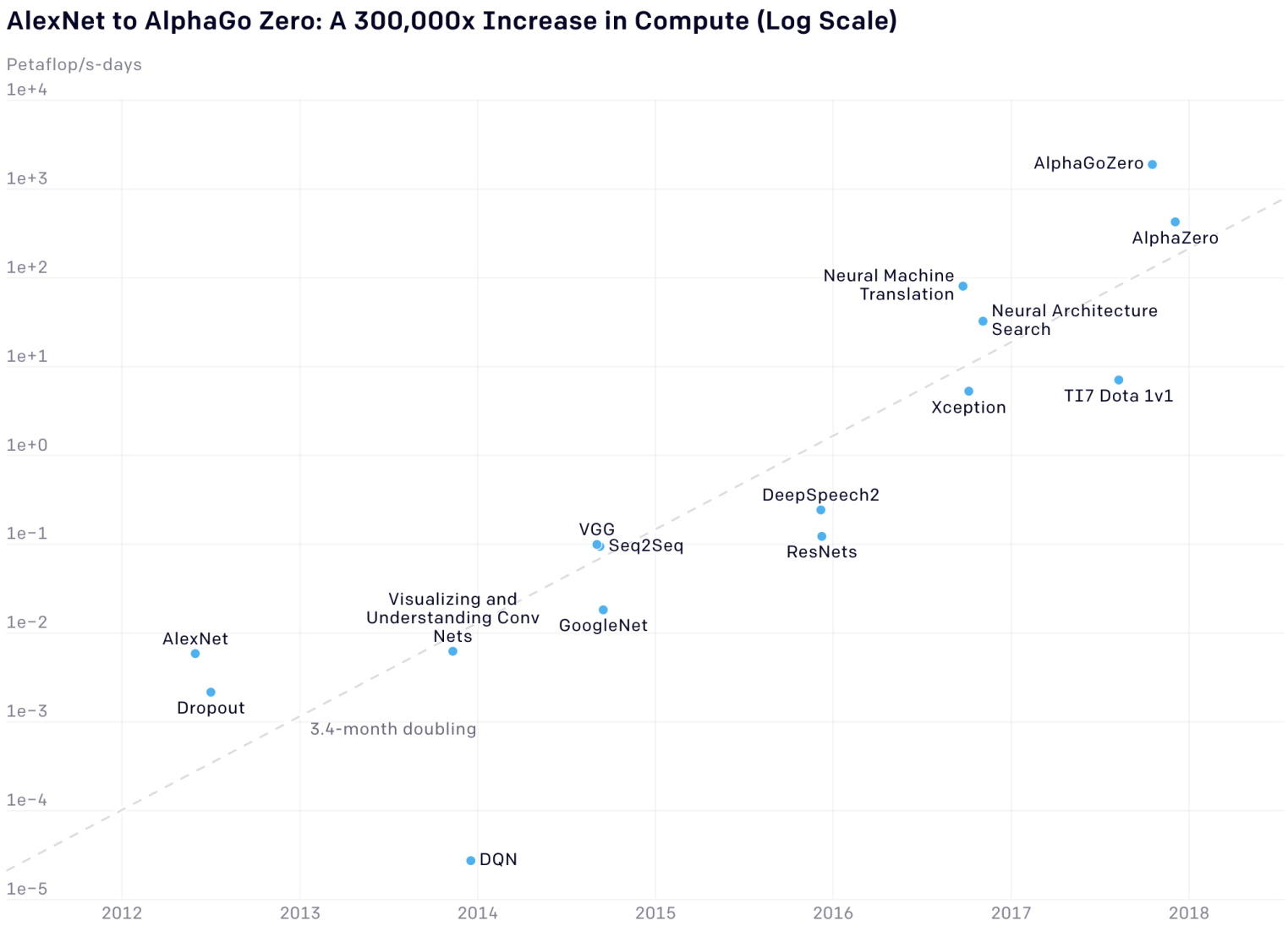
语言遵循语法规则

自然图像也有规则

深度学习算力需求


Two Distinct Eras of Compute Usage in Training AI Systems
Petaflop/s-days
1e+4
1e+2
1e+0
1e−2
1e−4
1e−6
1e−8
1e−10
1e−12
1e−14
1960
1970
1980
1990
2000
2010
2020
AlphaGoZero
Neural Machine Translation
TI7 Dota 1v1
VGG
ResNets
AlexNet
3.4-month doubling
Deep Belief Nets and layer-wise pretraining
DQN
TD-Gammon v2.1
BiLSTM for Speech
LeNet-5
NETtalk
ALVINN
RNN for Speech
2-year doubling (Moore's Law)
Perceptron
← First Era
Modern Era →

深度学习算力需求


AlexNet to AlphaGo Zero: A 300,000x Increase in Compute (Log Scale)
Petaflop/s-days
1e+4
1e+3
1e+2
1e+1
1e+0
1e−1
1e−2
1e−3
1e−4
1e−5
2012
2013
2014
2015
2016
2017
2018
AlphaGoZero
AlphaZero
Neural Machine Translation
Neural Architecture Search
TI7 Dota 1v1
Xception
DeepSpeech2
VGG
Seq2Seq
ResNets
Visualizing and Understanding Conv Nets
GoogleNet
AlexNet
Dropout
3.4-month doubling
DQN

大　　纲

- 课程概览
- 深度学习历史
- 深度学习应用
- 深度学习平台
- 课程后续安排

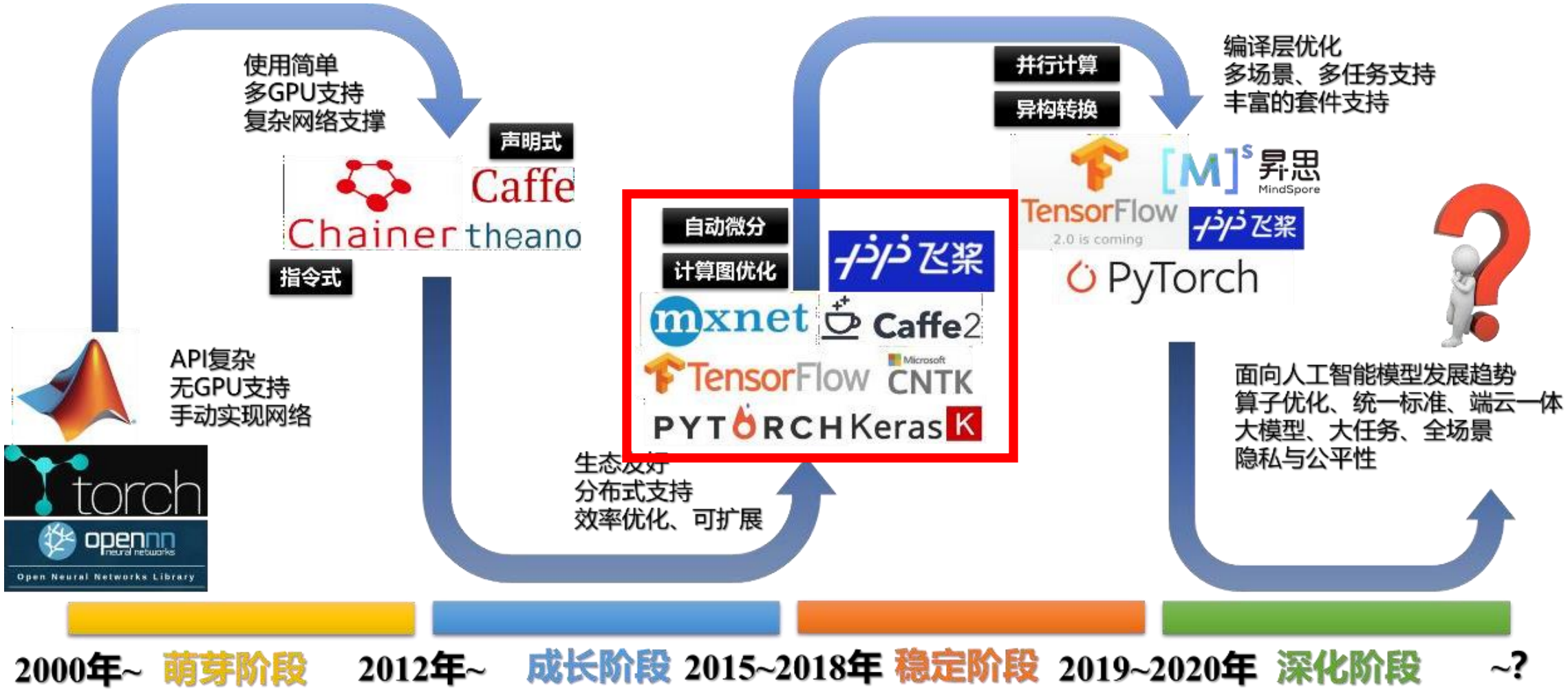
深度学习平台


使用简单
多GPU支持
复杂网络支撑
声明式
Caffe
Chainer
theano
指令式
API复杂
无GPU支持
手动实现网络
torch
opennn
neural networks
Open Neural Networks Library
自动微分
计算图优化
mxnet
飞桨
Caffe2
TensorFlow
Microsoft
CNTK
PYTORCH
Keras
生态友好
分布式支持
效率优化、可扩展
并行计算
异构转换
编译层优化
多场景、多任务支持
丰富的套件支持
TensorFlow
2.0 is coming
昇思
MindSpore
飞桨
PyTorch
面向人工智能模型发展趋势
算子优化、统一标准、端云一体
大模型、大任务、全场景
隐私与公平性
2000年~ 萌芽阶段
2012年~ 成长阶段
2015~2018年 稳定阶段
2019~2020年 深化阶段
~?

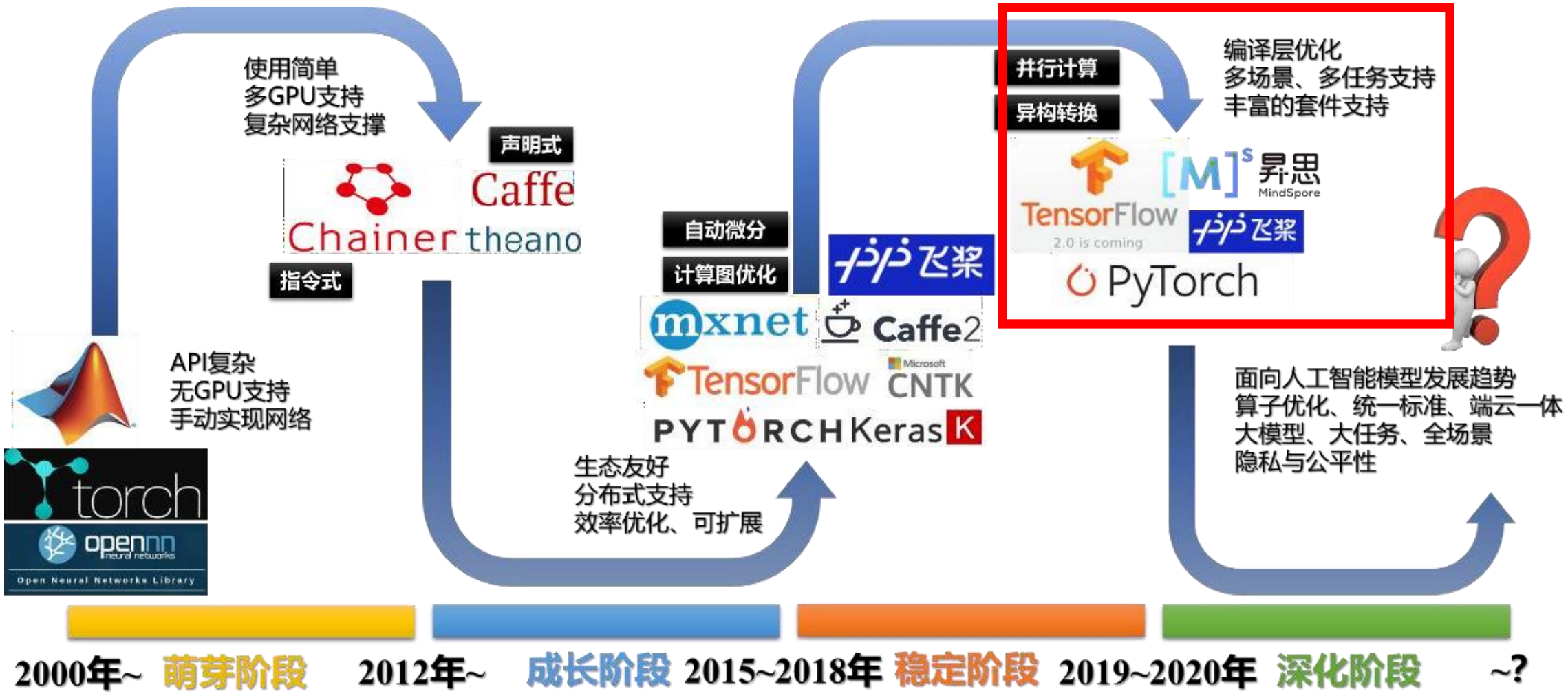
深度学习平台


使用简单
多GPU支持
复杂网络支撑
声明式
Caffe
Chainer
theano
指令式
API复杂
无GPU支持
手动实现网络
torch
opennn
neural networks
Open Neural Networks Library
自动微分
计算图优化
mxnet
Caffe2
TensorFlow
Microsoft CNTK
PYTORCH
Keras
生态友好
分布式支持
效率优化、可扩展
并行计算
异构转换
编译层优化
多场景、多任务支持
丰富的套件支持
TensorFlow
2.0 is coming
昇思
MindSpore
飞桨
PyTorch
面向人工智能模型发展趋势
算子优化、统一标准、端云一体
大模型、大任务、全场景
隐私与公平性
2000年~ 萌芽阶段
2012年~ 成长阶段
2015~2018年 稳定阶段
2019~2020年 深化阶段
~?

深度学习平台


使用简单
多GPU支持
复杂网络支撑
声明式
Caffe
Chainer
theano
指令式
API复杂
无GPU支持
手动实现网络
torch
opennn
neural networks
Open Neural Networks Library
自动微分
计算图优化
飞桨
mxnet
Caffe2
TensorFlow
Microsoft
CNTK
PYTORCH
Keras
生态友好
分布式支持
效率优化、可扩展
并行计算
异构转换
编译层优化
多场景、多任务支持
丰富的套件支持
TensorFlow
2.0 is coming
昇思
MindSpore
飞桨
PyTorch
面向人工智能模型发展趋势
算子优化、统一标准、端云一体
大模型、大任务、全场景
隐私与公平性
2000年~ 萌芽阶段
2012年~ 成长阶段
2015~2018年 稳定阶段
2019~2020年 深化阶段
~?

深度学习平台

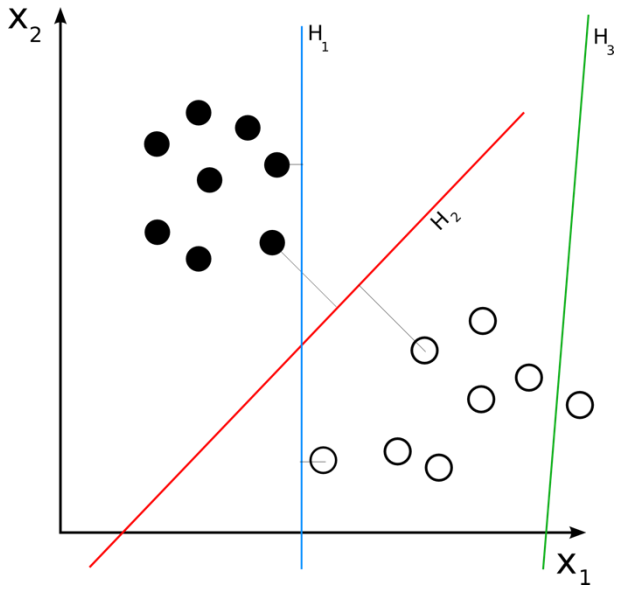

使用简单
多GPU支持
复杂网络支撑
声明式
Caffe
Chainer
theano
指令式
API复杂
无GPU支持
手动实现网络
torch
opennn
neural networks
Open Neural Networks Library
自动微分
计算图优化
mxnet
Caffe2
TensorFlow
Microsoft CNTK
PYTORCH
Keras
生态友好
分布式支持
效率优化、可扩展
并行计算
异构转换
编译层优化
多场景、多任务支持
丰富的套件支持
TensorFlow
2.0 is coming
昇思
MindSpore
飞桨
PyTorch
面向人工智能模型发展趋势
算子优化、统一标准、端云一体
大模型、大任务、全场景
隐私与公平性
2000年~ 萌芽阶段
2012年~ 成长阶段
2015~2018年 稳定阶段
2019~2020年 深化阶段
~?

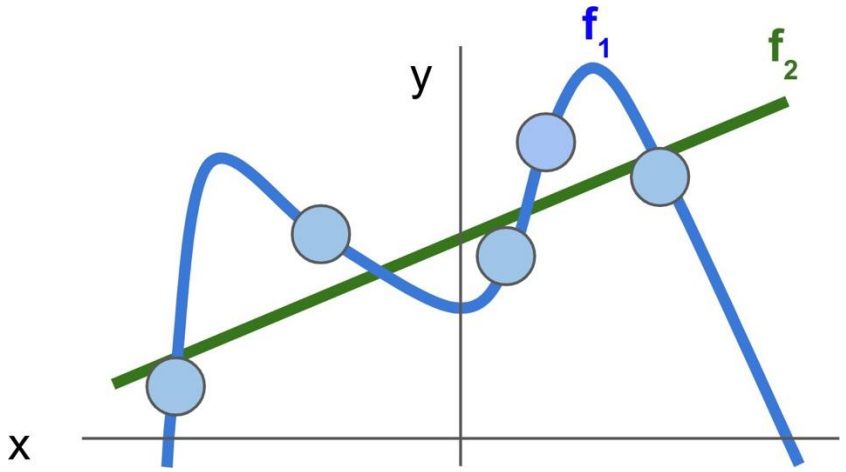
大　　纲

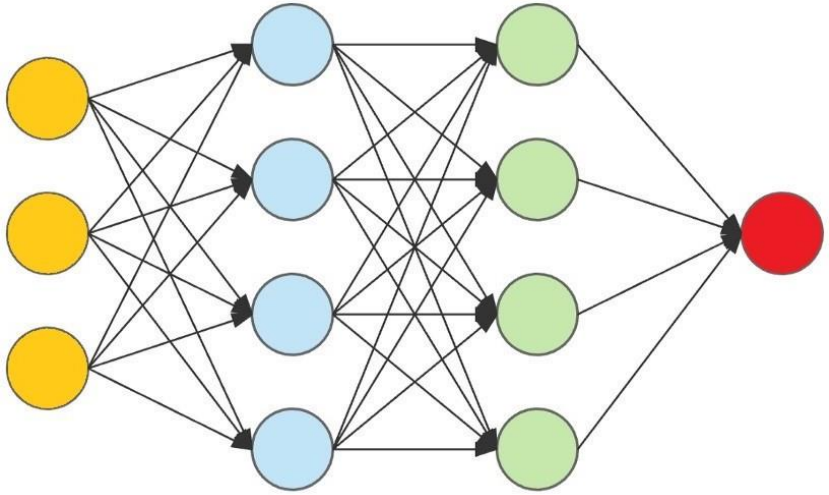
- 课程概览
- 深度学习历史
- 深度学习应用
- 深度学习平台
- 课程后续安排

课程安排

- 深度学习基础
  - 线性分类器、正则化与优化、神经网络





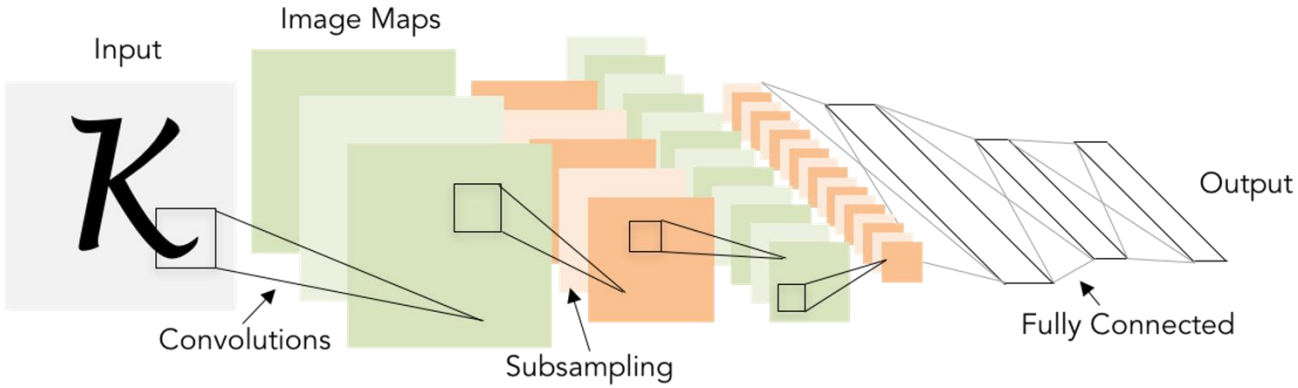


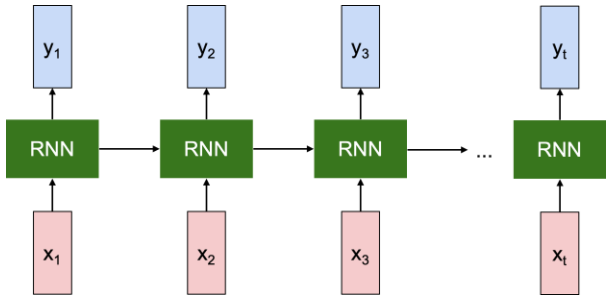
课程安排

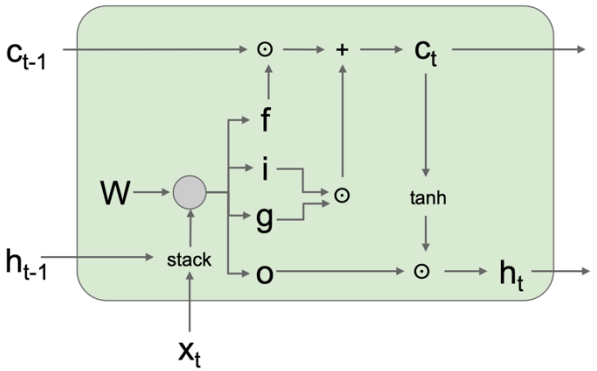
- 基础视觉任务
  - 图像理解（分类、检测、分割）、视频理解，网络可视化

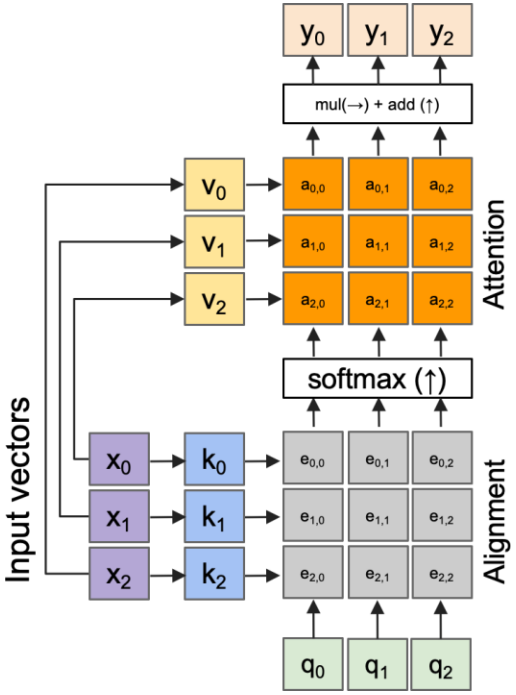
课程安排

- 基础神经网络
  - 卷积神经网络、循环神经网络、Transformer



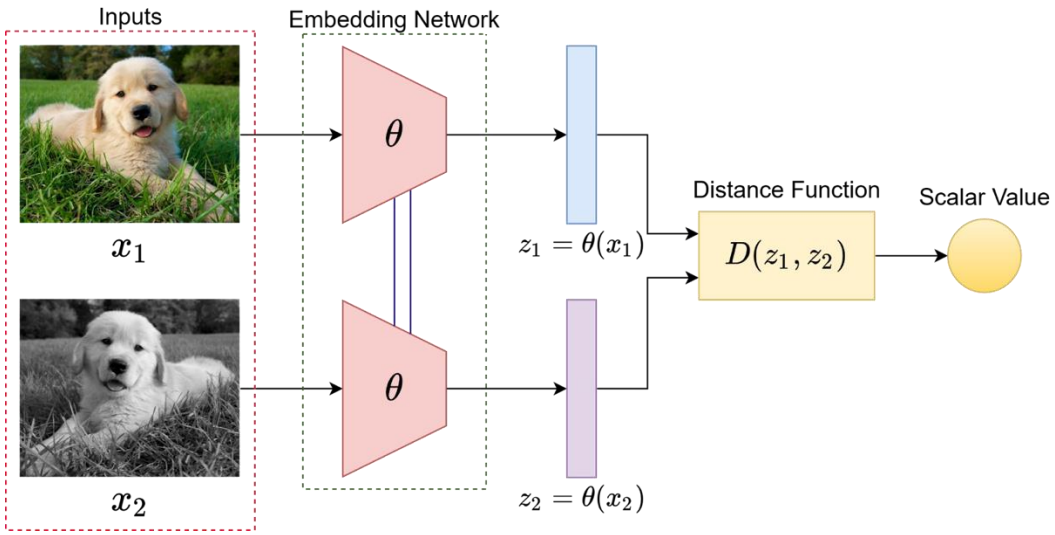


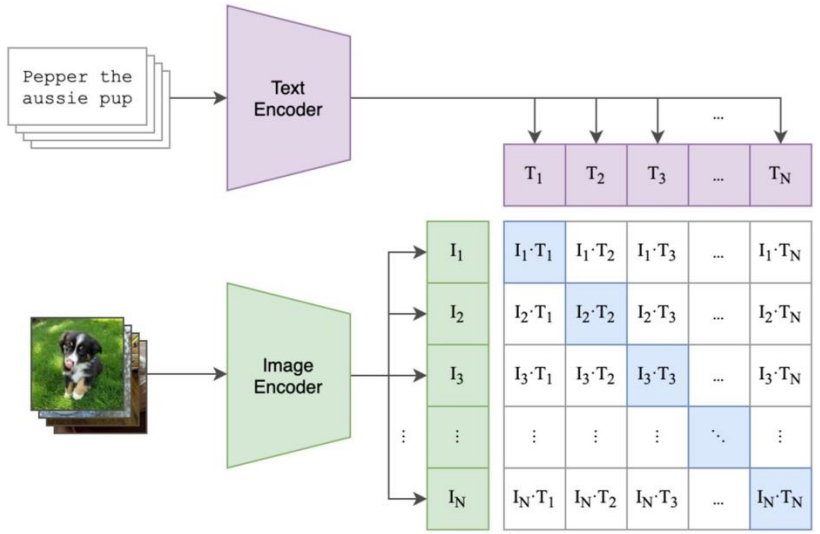




课程安排

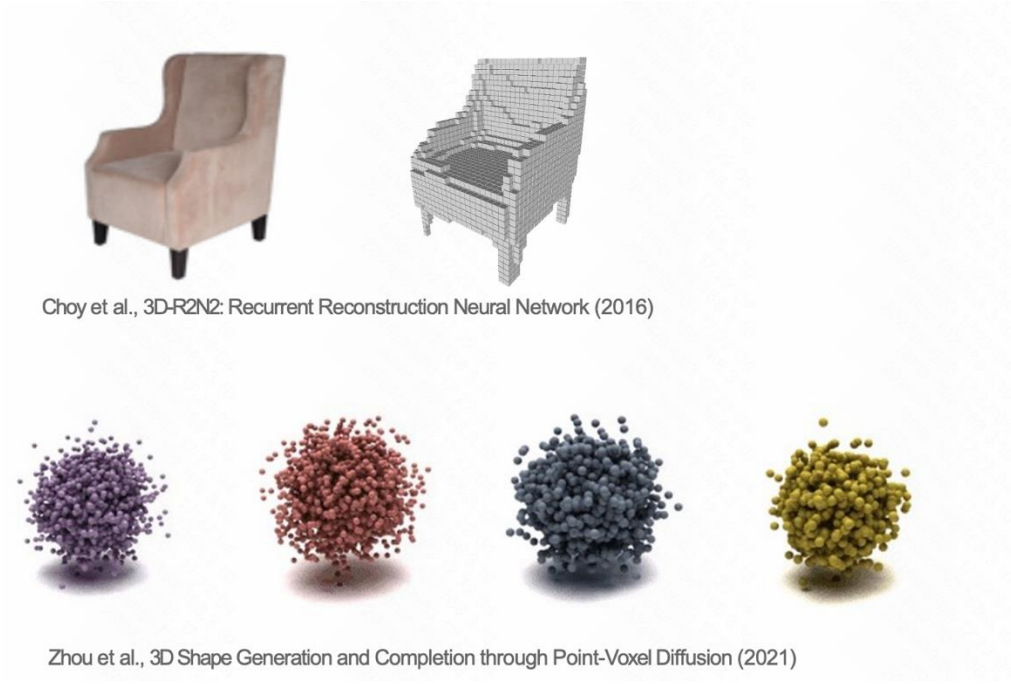
- 更多深度学习任务
  - 自监督学习、多模态模型





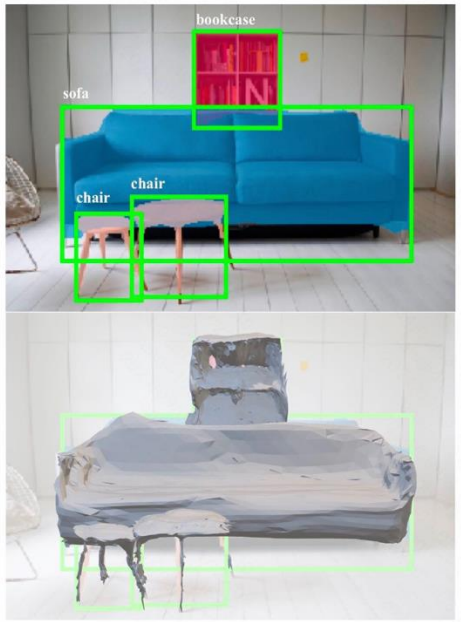
课程安排

- 更多深度学习任务
  - 生成模型、 3D模型



Choy et al., 3D-R2N2: Recurrent Reconstruction Neural Network (2016)

Zhou et al., 3D Shape Generation and Completion through Point-Voxel Diffusion (2021)



Gkioxari et al., "Mesh R-CNN", ICCV 2019