



河南大學

明德新民 止于至善

人工智能与图像处理

陈小潘

计算机与信息工程学院



目录

1. 智能技术发展背景

2. 机器学习

3. 感知器

4. 人工神经网络

5. 卷积神经网络

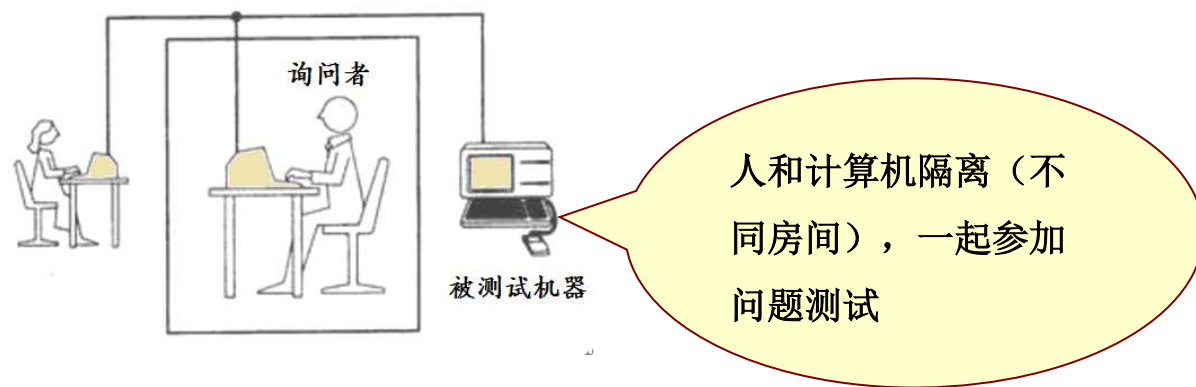
6. 深度学习

7. Tensorflow

8. Pytorch

1.人工智能发展背景

- 让机器具备智能是人们长期追求的目标。
- 上世纪50年代，数学家Alan Turing提出图灵测试，采用回答问题方式，辨别机器是否具有智能。



- 人类评估者会通过文本信息的方式跟受试者进行交流，并判断受试者究竟是人类还是机器。如果机器能骗过30%的人类评估者，那就可以认为它具有了智能。

1.人工智能发展背景

1956年夏，麦卡锡、明斯基等科学家在美国达特茅斯学院开会研讨“如何用机器模拟人的智能”，麦卡锡首次提出了“人工智能”（**Artificial Intelligence**）这一术语，用来概括会议的主题：“试图让机器精确模拟学习或其他智能行为”。由此，“人工智能”作为一个学科有了名字、有了明确使命，其创始人们雄心勃勃地宣告：任何智能的要素都能被精确定义并交给机器实现。

1.人工智能发展背景

1956 Dartmouth Conference: The Founding Fathers of AI



John McCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



Trenchard More

1.人工智能发展背景

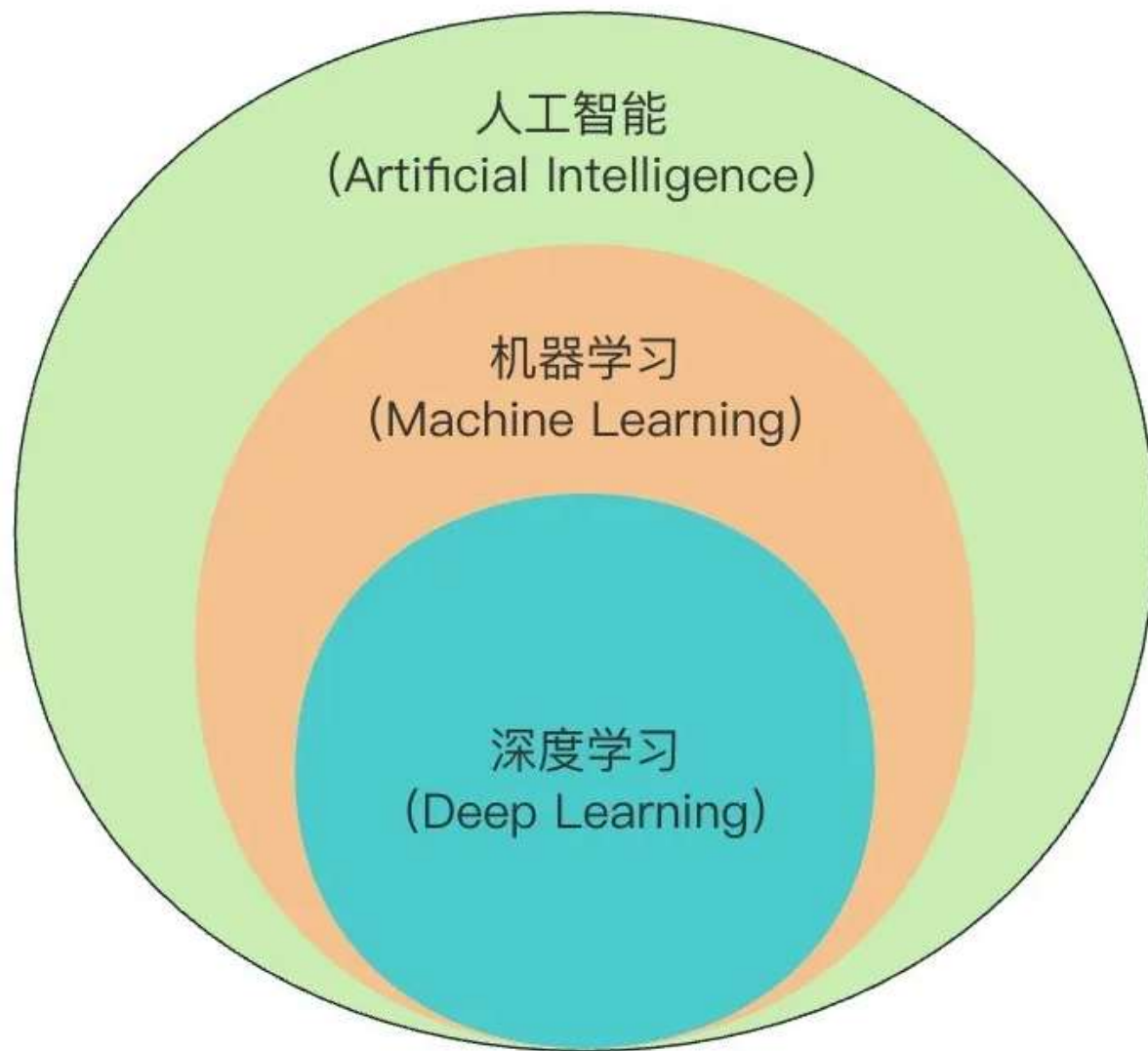
人工智能是研究开发能够模拟、延伸和扩展人类智能的理论、方法、技术及应用系统的一门新的技术科学，研究目的是促使智能机器**会听**（语音识别、机器翻译等）、**会看**（图像识别、文字识别等）、**会说**（语音合成、人机对话等）、**会思考**（人机对弈、定理证明等）、**会学习**（机器学习、知识表示等）、**会行动**（机器人、自动驾驶汽车等）。

人工智能作为引领新一轮科技革命和产业变革的战略性信息技术，现已成为**驱动科技创新、经济发展、社会进步和民生改善的重要力量**。

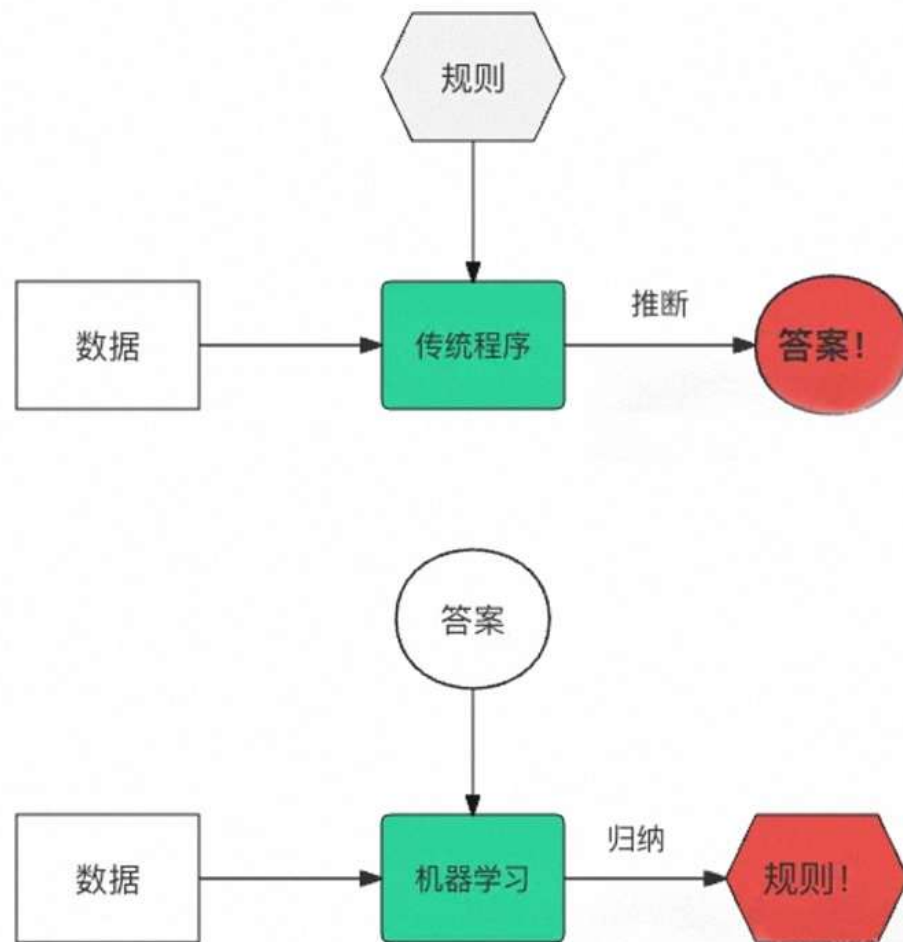
1.人工智能发展背景

- 要实现人工智能，必备基础：理解语言、学习、记忆、推理、决策
- 机器学习（Machine Learning，ML）：机器学习就是让机器（计算机）也能像人类一样，通过观察大量的数据和训练，发现事物规律（规则），获得某种分析问题、解决问题的能力。即让机器去学习、执行。
- 机器学习赋予计算机从数据中自动学习和提取知识的能力，无需明确的编程指令。它利用算法和模型，让计算机能够分析和处理大量数据，发现隐藏的模式、规律和关系。
- 机器学习涉及：概率论、统计学、逼近论、凸分析、计算复杂性理论等。

1.人工智能发展背景



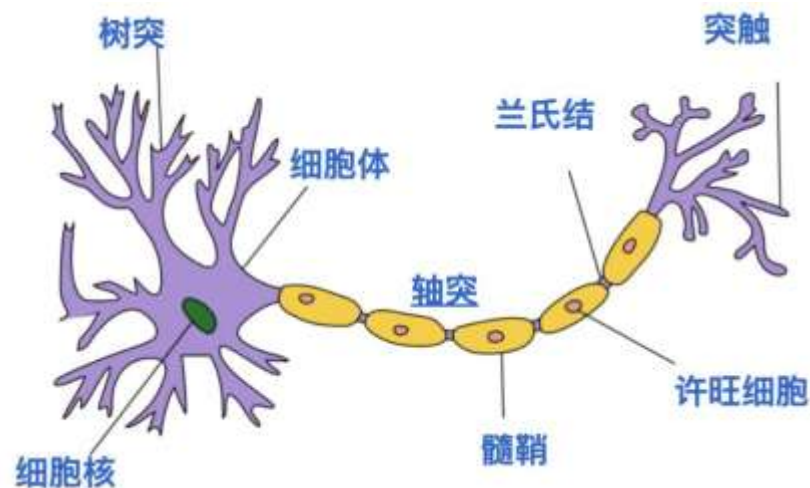
1.人工智能发展背景



机器学习和传统程序的区别

1.人工智能发展背景

- 人工神经网络（Artificial Neural Network, ANN），也简称神经网络，是众多机器学习算法中比较接近生物神经网络特性的**数学模型**。
- 人工神经网络是基于**生物学中神经网络**的基本原理，在理解和抽象了人脑结构和外界刺激响应机制后，以网络拓扑知识为理论基础，模拟人脑的神经系统对复杂信息的处理机制的一种数学模型。



1.人工智能发展背景

- 树突具有接受刺激并将冲动传入细胞体的功能
- 细胞体具有联络和整合输入信息并传出信息的作用
- 轴突的主要功能是将神经冲动由细胞体传至其他神经元
- 神经元通过树突接收来自其他神经元的信号，来自树突的信号在细胞体中积累，如果产生的信号强度超过一定的阈值，神经元就把信息传递给轴突。否则，信号会被神经元杀死，不再传播。

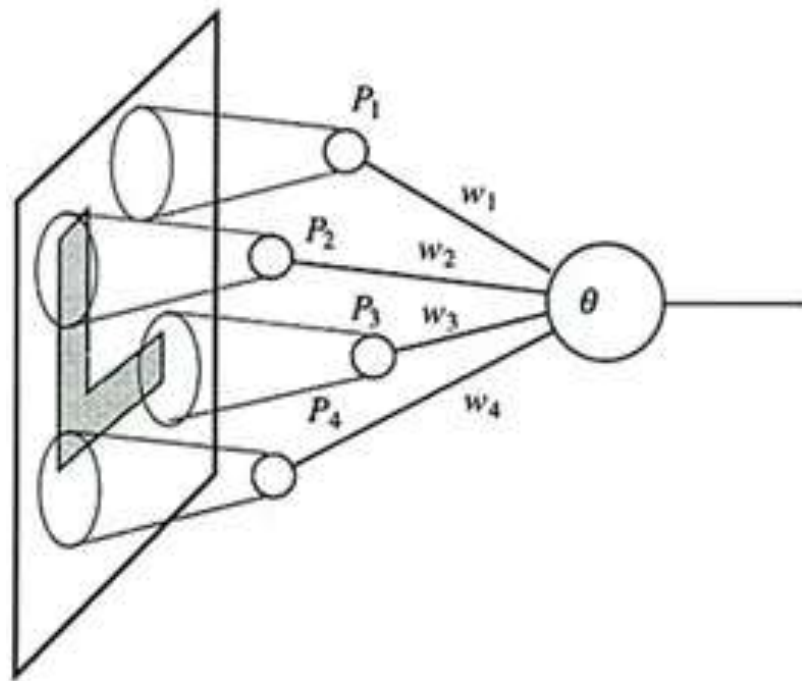
1.人工智能发展背景

人工神经网络是由大量的处理单元（神经元）互相连接而形成的复杂网络结构，是对人脑组织结构和运行机制的某种抽象、简化和模拟。以数学模型模拟神经元活动，是基于模仿大脑神经网络结构和功能而建立的一种信息处理系统。

人工神经网络有**多层和单层**之分，每一层包含若干神经元，各神经元之间用带可变权重的有向弧连接，网络通过对已知信息的反复学习训练，通过逐步调整改变神经元连接权重的方法，达到处理信息、模拟输入输出之间关系的目的。

1.人工智能发展背景

- 感知器，又称感知机，是最简单的神经网络，只有一层，一个神经元；一种最简单形式的前馈神经网络。
- 由美国计算机科学家Roseblatt 1957年提出的线性分类器的经典学习算法。



1.人工智能发展背景

感知器： 1957年，美国神经学家Frank Rosenblatt（罗森布拉特）在Cornell航空实验室中，成功在IBM 704机上完成了感知机的仿真。两年后，他又成功实现了能够识别一些英文字母、基于感知机的神经计算机——Mark1，并于1960年6月23日，展示与众。

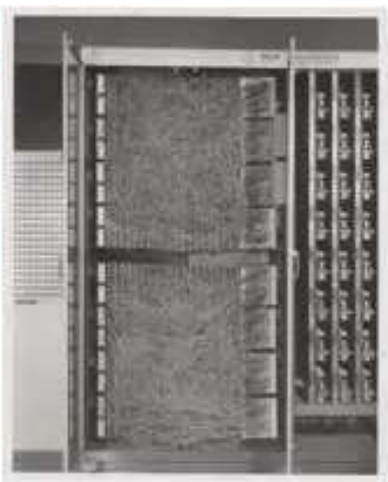


图 Rosenblatt和Mark1 感知机

1.人工智能发展背景

感知机最终被证明不能处理诸多的模式识别问题。1969年，Marvin Minsky和Seymour Papert在《Perceptrons》一书中，仔细分析了以感知机为代表的单层神经网络系统的功能及局限，证明感知机不能解决简单的异或（XOR）等线性不可分问题，但Rosenblatt和Minsky及Papert等人在当时已经了解到多层神经网络能够解决线性不可分的问题。

由于Rosenblatt等人没能够及时推广感知机学习算法到多层神经网络上，又由于《Perceptrons》在研究领域中的巨大影响，及人们对书中论点的误解，造成了人工神经领域发展的长年停滞及低潮，直到人们认识到多层感知机没有单层感知机固有的缺陷及反向传播算法在80年代的提出，才有所恢复。1987年，书中的错误得到了校正，并更名再版为《Perceptrons - Expanded Edition》。

1.人工智能发展背景

- 1986年，Geoffrey Hinton等发表的论文介绍了用于训练多层神经网络的反向传播算法。

nature

[Explore content](#) ▾

[Journal information](#) ▾

[Publish with us](#) ▾

[Subscribe](#)

[nature](#) > [letters](#) > [article](#)

Published: 09 October 1986

Learning representations by back-propagating errors

David E. Rumelhart, Geoffrey E. Hinton & Ronald J. Williams

[Nature](#) **323**, 533–536 (1986) | [Cite this article](#)

70k Accesses | **10556** Citations | **222** Altmetric | [Metrics](#)

1.人工智能发展背景

- 2000年以后，因为当时计算机的计算能力不能支持训练大规模的神经网络，人工神经网络又一次陷入低潮。
- 2018年，Hinton与Yann LeCun(杨立昆)和Yoshua Bengio(约书亚·本希奥)一起获得了2018年图灵奖，以表彰他们使深度神经网络成为计算的关键组成部分的概念和工程突破。
“深度学习三巨头”。辛顿称为“AI教父”、“神经网络之父”，杨立昆被称为“卷积网络之父”。

1.人工智能发展背景



图 早期的图灵三巨头

1.人工智能发展背景



图3：辛顿著名的“徒子徒孙”们

1.人工智能发展背景

- 2006 年，Hinton等发现多层前馈神经网络可以先通过逐层预训练，再用反向传播算法进行精调的方式进行有效学习。
- GPU实现大规模并行计算： 计算机计算能力的提高，可以训练大规模的人工神经网络。

1.人工智能发展背景

The **Mark I Perceptron** machine was the first implementation of the perceptron algorithm.

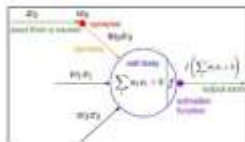
The machine was connected to a camera that used 20×20 cadmium sulfide photocells to produce a 400-pixel image.

recognized letters of the alphabet

$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

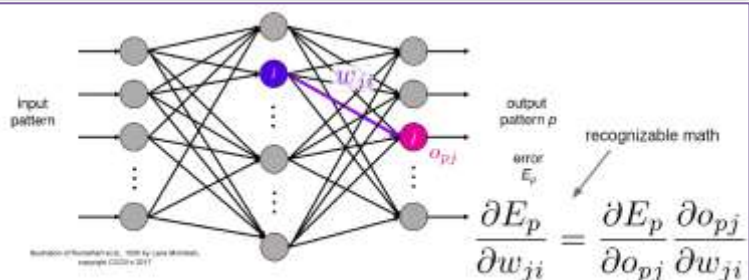
update rule:

$$w_i(t+1) = w_i(t) + \alpha(d_j - y_j(t))x_{ji}$$



Frank Rosenblatt, ~1957: Perceptron

Widrow and Hoff, ~1960: Adaline/Madaline



Rumelhart et al., 1986: First time back-propagation became popular

[Hinton and Salakhutdinov 2006]

Reinvigorated research in Deep Learning

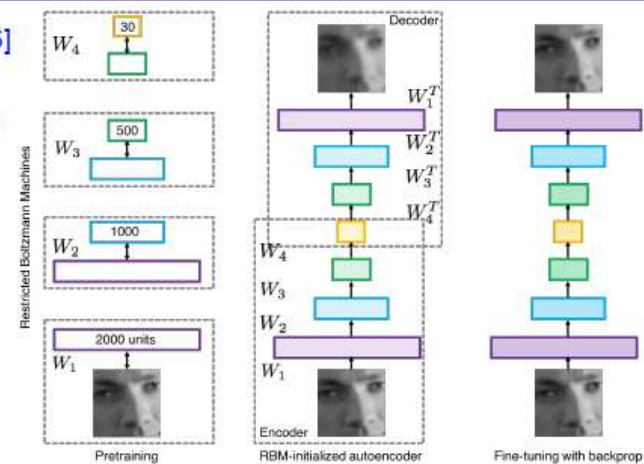


Illustration of Hinton and Salakhutdinov 2006 by Lane McIntosh, copyright CS231n 2017

ImageNet Classification with Deep Convolutional Neural Networks
[Krizhevsky, Sutskever, Hinton, 2012]

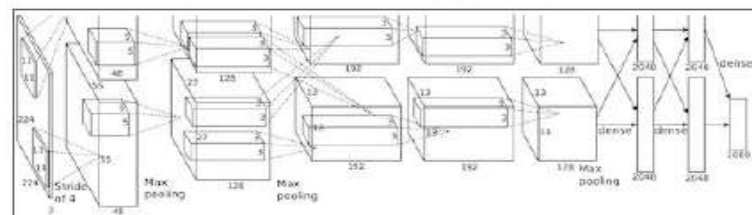
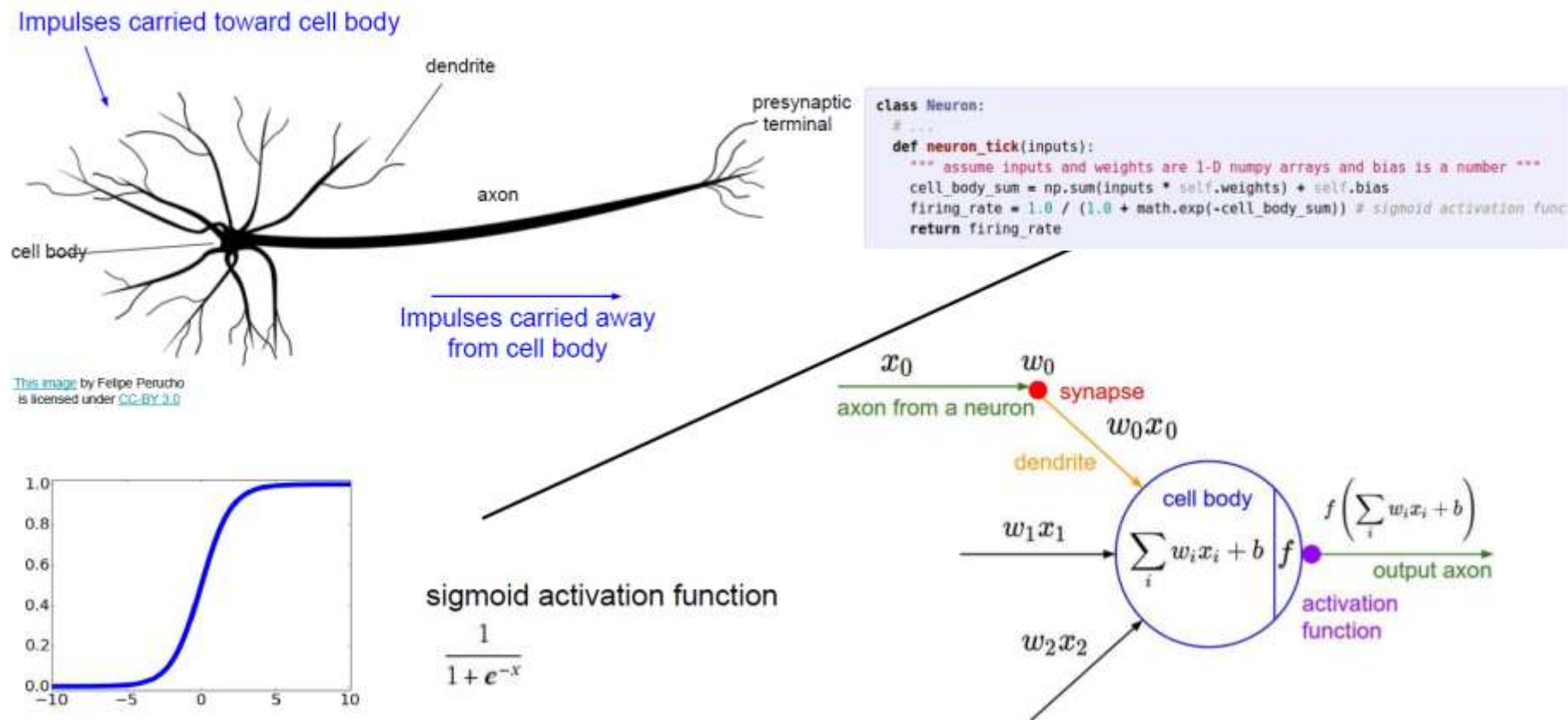


Figure copyright Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, 2012. Reproduced with permission.

“AlexNet”

<http://cs231n.stanford.edu/slides/2017>

1.人工智能发展背景



<http://cs231n.stanford.edu/slides/2017>

1.人工智能发展背景

计算机视觉革命：AlexNet用 GPU「看」世界

2012年，Krizhevsky等人在《ImageNet Classification with Deep Convolutional Neural Networks》中提出的AlexNet，以ReLU激活函数、Dropout正则化和GPU加速训练为核心，在ImageNet竞赛中将错误率降低10%，引发计算机视觉革命。这一突破让计算机仿佛装上“人工视网膜”，真正开始“看懂”图像。

1.人工智能发展背景

■ 激活函数是什么？

简单地说，激活函数就是加入到人工神经网络中的一个函数，目的在于帮助神经网络从数据中学习**复杂模式**。相比于人类大脑中基于神经元的模型，**激活函数是决定向下一个神经元传递何种信息的单元**，这也正是激活函数在人工神经网络中的作用。

激活函数接收前一个单元输出的信号，并将其转换成某种可以被下一个单元接收的形式。

1.人工智能发展背景

■ 为什么需要激活函数？

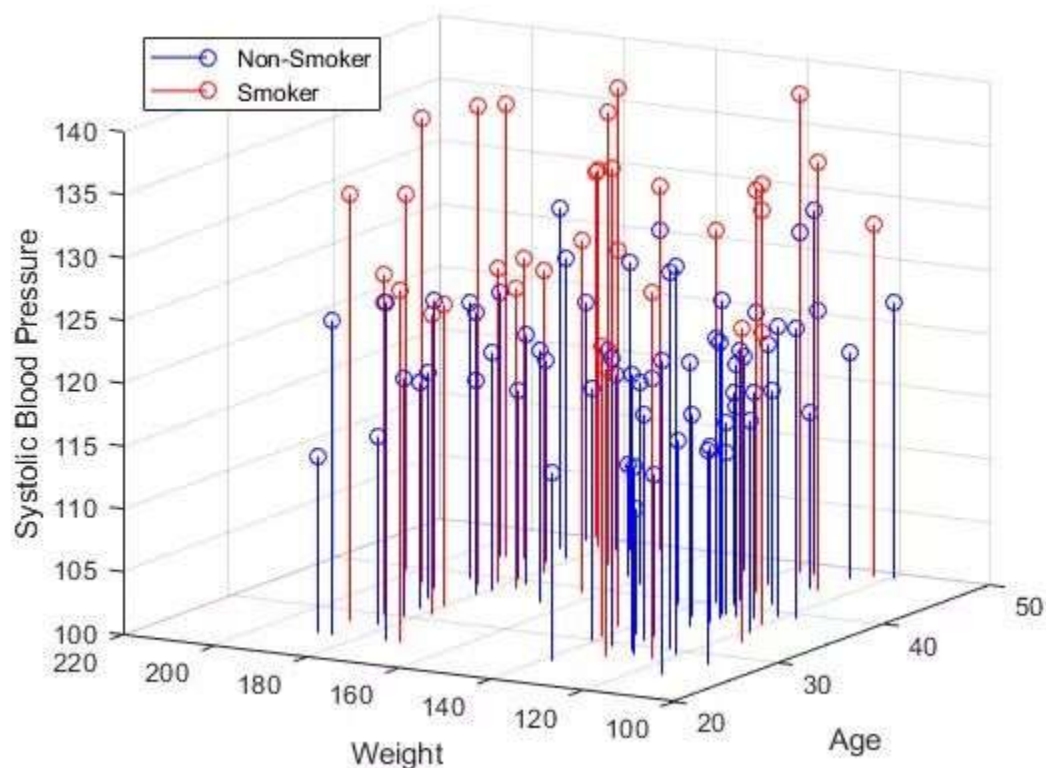
1. 除了生物学方面的相似性外，激活函数还有助于根据要求将神经元的输出值限定在一定的范围内。

因为激活函数的输入是 $W*x+b$ ，其中 W 是单元的权重， x 是输入值，然后加上偏置 b 。如果输出值不被限定在某个范围内，它可能会变得非常大，特别是在具有数百万个参数的深层神经网络中，从而导致计算量过大。例如，有一些激活函数（如 softmax）对于不同的输入值（0 或 1）会输出特定的值。

1.人工智能发展背景

■ 为什么需要激活函数？

2. 激活函数最重要的特点是它具有在神经网络中加入非线性的能力。



1.人工智能发展背景

■ 为什么需要激活函数？

2. 激活函数最重要的特点是它具有在神经网络中加入非线性的能力。

假定上图所示的情况对应的线性函数有三个属性（体重、收缩压和年龄），它使我们得到一条贯穿三维空间的直线，但它永远也无法学会一种模式来将一个人准确地区分为吸烟者和不吸烟者（以手头上现有的分类问题为例）。原因很简单，定义这个分类的模式根本不是线性的。

1.人工智能发展背景

2024 年诺贝尔**物理学奖**：授予美国新泽西州普林斯顿大学John J. Hopfield（约翰·霍普菲尔德）和加拿大多伦多大学的Geoffrey E. Hinton（杰弗里·辛顿），表彰他们「**在使用人工神经网络实现机器学习方面奠基性发现和发明**」。

瑞典皇家科学院当天发布的新闻公报说，长期以来，机器学习对于研究非常重要，包括对大量数据的排序和分析。约翰·霍普菲尔德和杰弗里·辛顿使用物理学工具构建方法，为当今强大的机器学习奠定了基础。基于人工神经网络的机器学习目前正在彻底改变科学、工程和日常生活。

1.人工智能发展背景



The image is a blue rectangular graphic announcing the Nobel Prize in Physics 2024. At the top left is a gold Nobel medal. To its right, the text reads "NOBELPRISET I FYSIK 2024" and "THE NOBEL PRIZE IN PHYSICS 2024". At the top right is the logo of the Kungl. Vetenskaps Akademi (The Royal Swedish Academy of Sciences). Below the medal and text are two portraits of the winners. The left portrait is of John J. Hopfield, a man with white hair wearing a suit and tie. Below his portrait is his name "John J. Hopfield" and his affiliation "Princeton University, NJ, USA". The right portrait is of Geoffrey E. Hinton, an older man with grey hair wearing a dark sweater. Below his portrait is his name "Geoffrey E. Hinton" and his affiliation "University of Toronto, Canada". At the bottom of the graphic, the award citation is written in Swedish and English. The Swedish text is "för grundläggande upptäckter och uppfinningar som möjliggör maskininlärning med artificiella neuronnätverk" and the English text is "for foundational discoveries and inventions that enable machine learning with artificial neural networks".

NOBELPRISET I FYSIK 2024
THE NOBEL PRIZE IN PHYSICS 2024

KUNGL. VETENSKAPS AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

John J. Hopfield
Princeton University, NJ, USA

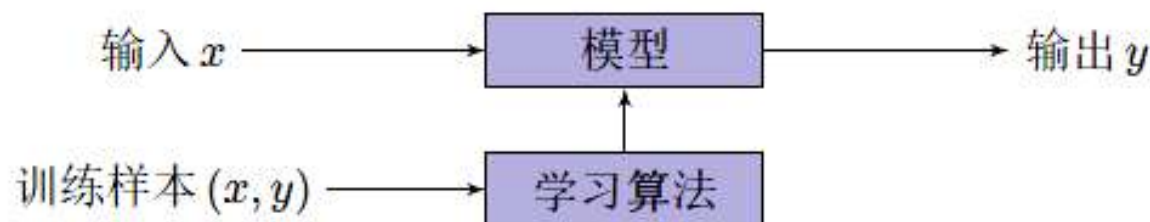
Geoffrey E. Hinton
University of Toronto, Canada

"för grundläggande upptäckter och uppfinningar som möjliggör maskininlärning med artificiella neuronnätverk"
"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

2. 机器学习

- **机器学习：** 给定一些训练样本 (x_i, y_i) ，（其中 x_i 是输入， y_i 是需要预测的目标），让计算机自动寻找一个决策函数 $f(\cdot)$ 来建立 x 和 y 之间的关系。

$$\hat{y} = f(\phi(x), \theta), \quad \text{或者} \quad \hat{y} = f(x, \theta).$$



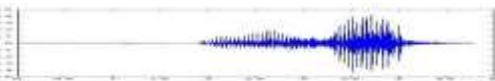
机器学习系统示例

机器学习在给定数据集的情况下，可以自动学习模式并使用学习结果来预测或做出决策。它依靠数学和统计学算法来构建模型，这些模型可以使计算机在没有明确编程的情况下自主学习。

2.机器学习

机器学习≈寻找一个函数

- 语音识别

$$f(\text{  }) = \text{“How are you”}$$

- 图片识别 $f(\text{  }) = \text{“Cat”}$

- 对话系统 $f(\text{“Hi”}) = \text{“Hello”}$
(你说的话) (系统的回答)

2. 机器学习

(a) 机器学习算法的类型

- 有监督学习 (**Supervised Learning**)
- 无监督学习 (**Unsupervised Learning**)
- 增强学习 (**Reinforcement Learning**)
- 自监督学习 (**Self-Supervised Learning, SSL**)
- 迁移学习 (**Transfer Learning**)

2. 机器学习

■ 有监督学习

- ✓ 利用一组已知输入 x 和输出 y 的数据对模型进行学习，在学习过程中尽量使输出标记 \hat{y} 和真实标记 y 相接近。 y 称之为标签。
- ✓ 应用场景：图像分类、语音识别、垃圾邮件检测等。
- ✓ 有监督学习分为回归和分类
- ✓ 常见算法：线性回归、逻辑回归、支持向量机（SVM）、决策树、随机森林、神经网络等。

2. 机器学习

■ 有监督学习

- ✓ 回归问题（Regression）：如果输出 y 是连续值（实数或连续整数）， $f(x)$ 的输出也是连续值，对于所有已知或未知的 (x, y) ，使得 f 输出和 y 尽可能地一致。
- ✓ 例如：输入是可能是种种跟预测PM2.5有关的指数，包括今天的PM2.5的数值、平均温度、平均的臭氧浓度等等，输出是明天中午的PM2.5的数值，这就是一个回归问题。
- ✓ 损失函数通常定义为平方误差

$$L(y, f(x, \theta)) = \| y - f(x, \theta) \|^2$$

2. 机器学习

■ 有监督学习

分类（classification）

- ✓ 分类任务要让机器做选择题。人类先准备好一些选项，这些选项称为类别（class），现在要找的函数的输出就是从设定好的选项里面选择一个当作输出，该任务称为分类。
- ✓ 举个例子，每个人都有邮箱账户，邮箱账户里面有一个函数，该函数可以检测一封邮件是否为垃圾邮件。分类不一定只有两个选项，也可以有多个选项。
- ✓ AlphaGo下棋，下一步棋局的位置。

2. 机器学习

■ 无监督学习

- ✓ 用来学习的数据不包含输出目标（或者说数据不带有标签或标记），需要学习算法自动学习到一些有价值的信息。一个典型的无监督学习问题就是聚类（Clustering）。无监督学习主要用于数据的探索和归类。
- ✓ 应用场景：聚类分析、降维、异常检测等。
- ✓ 常见算法：K均值聚类（K-Means）、层次聚类、主成分分析（PCA）、独立成分分析（ICA）等。

2. 机器学习

■ 半监督学习 (Semi-Supervised Learning)

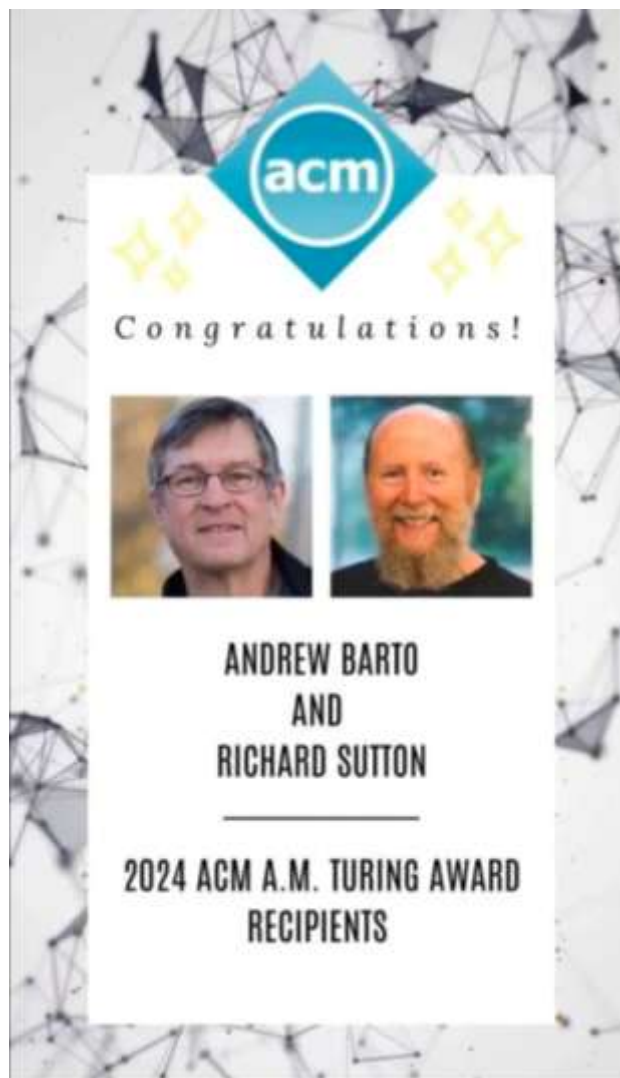
- ✓ 半监督学习结合了监督学习和无监督学习的特点，使用少量的标签数据和大量的无标签数据来训练模型。这样可以在减少标注成本的同时，仍然获得较高的模型性能。标注样本的成本非常高（人力、物力），尤其是高质量样本（对医学影像的标注更是需要专业的临床医生）。
- ✓ 应用场景：图像分类（当大部分图像没有标签时）、语音识别等。
- ✓ 常见算法：图半监督学习、生成对抗网络（GAN）等。

2. 机器学习

■ 增强学习（Reinforcement Learning）

- ✓ 也称**强化学习**，强化学习是通过让智能体（Agent）在与环境的交互中学习，根据反馈（奖励或惩罚）来优化其行为策略。**目标是通过试错过程最大化累积奖励。**
- ✓ 萨顿和巴托受行为主义心理学的启发，于**20 世纪 80 年代**开发了强化学习（**RL**）。**强化学习与监督学习和无监督学习并称为三大基本机器学习范式。**
强化学习教人工智能代理通过尝试和犯错，做出能实现最优结果的决策，这与人类的学习方式类似。
- ✓ 应用场景：游戏AI、机器人控制、自动驾驶等。
- ✓ 常见算法：Q学习、深度Q网络（DQN）、策略梯度方法、Actor-Critic方法等。

2. 机器学习



自1980年代起，**Andrew Barto** 和 **Richard Sutton**在一系列论文中提出了强化学习的主要思想，还构建了强化学习的数学基础，并开发了强化学习的重要算法。两人合著的

《**Reinforcement Learning: An Introduction**》一直是强化学习领域最经典的教材之一。

2024年，ACM图灵奖将素有“计算机界诺贝尔奖”之称的殊荣授予强化学习领域奠基人**Andrew Barto** 和 **Richard Sutton**（安德鲁·巴托与理查德·萨顿），标志着人工智能技术发展史迎来又一里程碑。从AlphaGo重塑围棋博弈范式，到特斯拉无人驾驶革新交通系统，强化学习作为智能系统的“决策中枢”，正以颠覆性力量推动医疗诊断、智能制造、金融风控等领域的范式变革。**Richard Sutton** 更是有「强化学习之父」的美誉。

2. 机器学习

■ 自监督学习（Self-Supervised Learning, SSL）

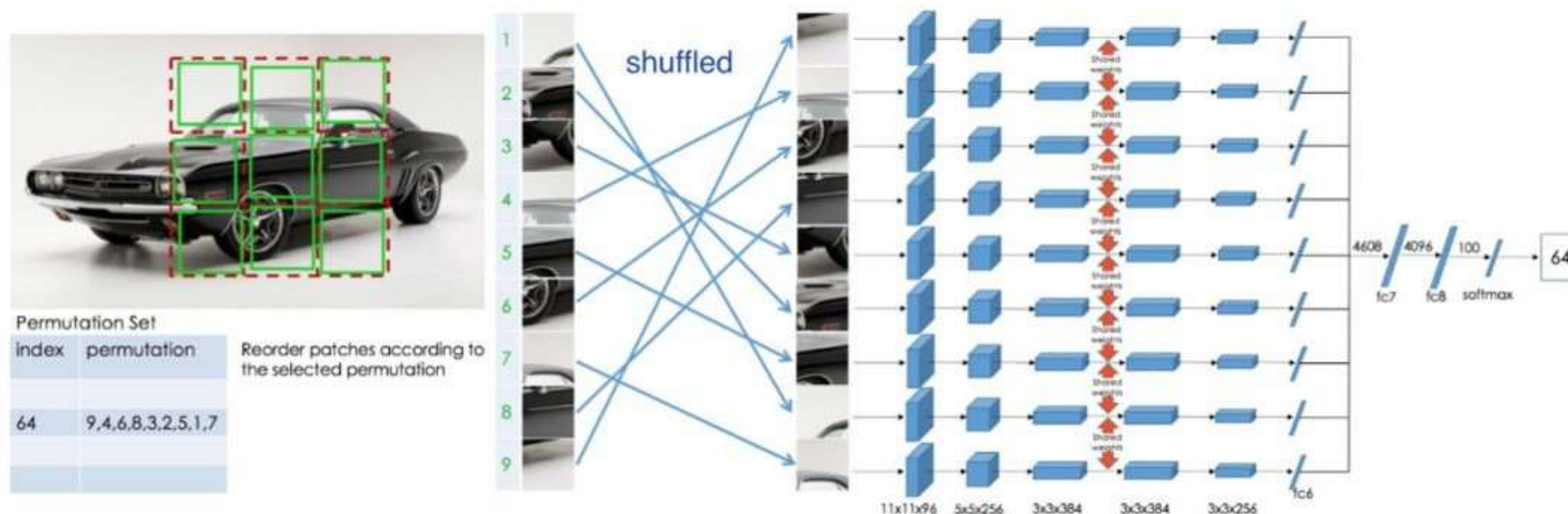
- ✓ 无标注的学习方式
- ✓ Yann LeCun最初在2019年4月在Facebook（后改名为 Meta）上的一篇帖子上提出了“自监督学习”这个词。
- ✓ 自监督学习是一种通过数据本身生成标签进行训练的方法，旨在通过预测数据的一部分来学习其表示。它不需要手动标注的数据，而是通过设计某种预测任务，从数据中自我学习。
- ✓ 应用场景：自然语言处理（如BERT模型）、图像处理等。
- ✓ 常见算法：对比学习、自动编码器等。

2. 机器学习

■ 自监督学习常见任务

a、图像重组（Jigsaw Puzzles）

将图像分为不同的patch，比如九宫格，然后让网络预测不同patch的相对位置信息。
这个过程中可以提高模型的局部特征提取能力以及全局空间信息提取能力。

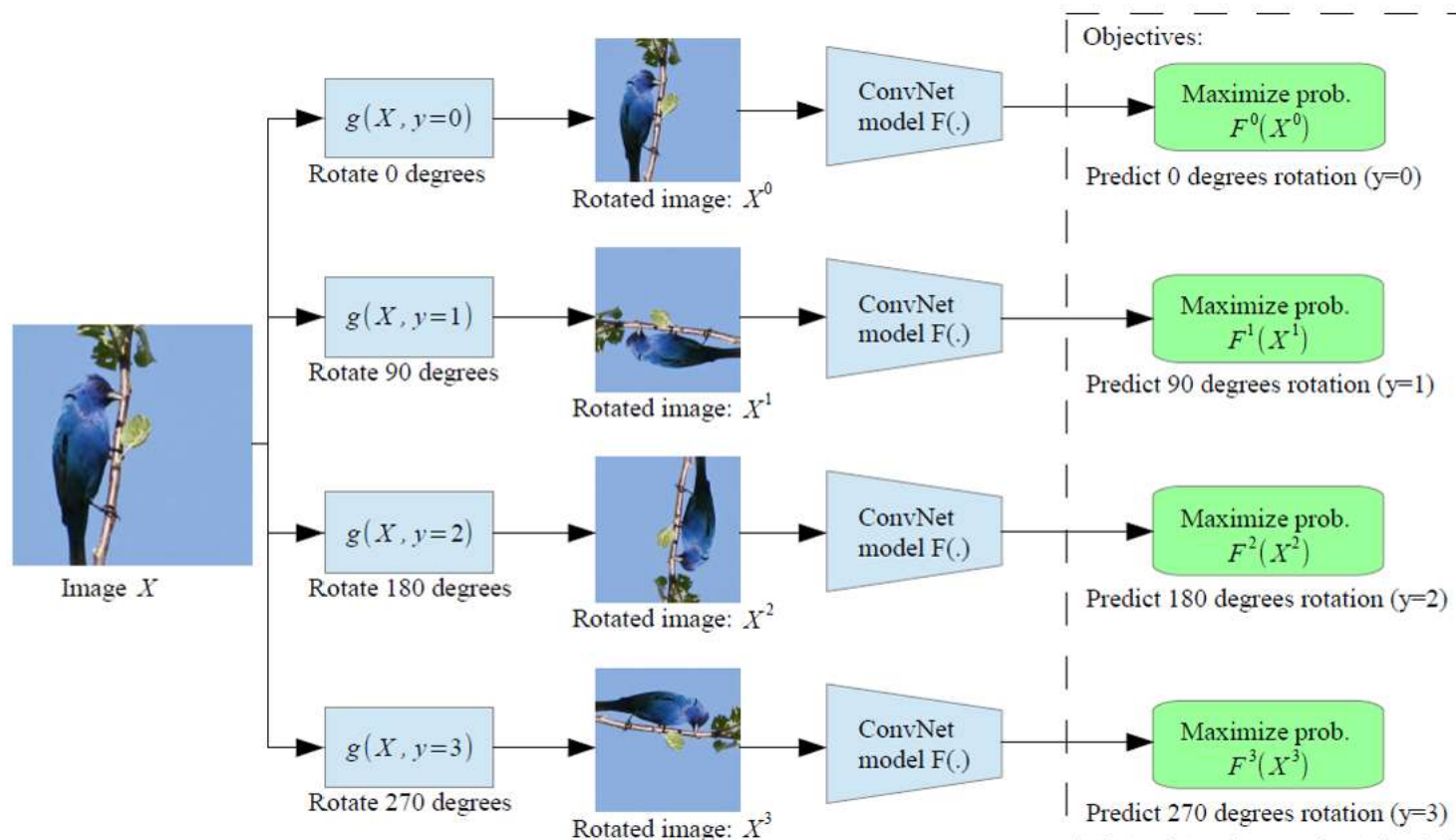


2. 机器学习

■ 自监督学习常见任务

b、图像旋转角度预测（Image Colorization）

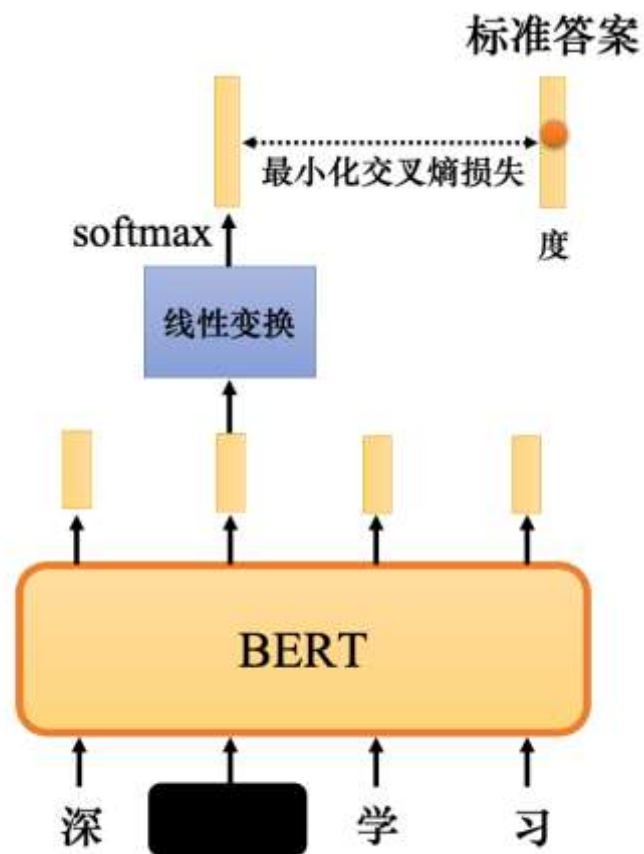
将训练集中的图像进行随机旋转，然后通过旋转角回归任务来训练网络。



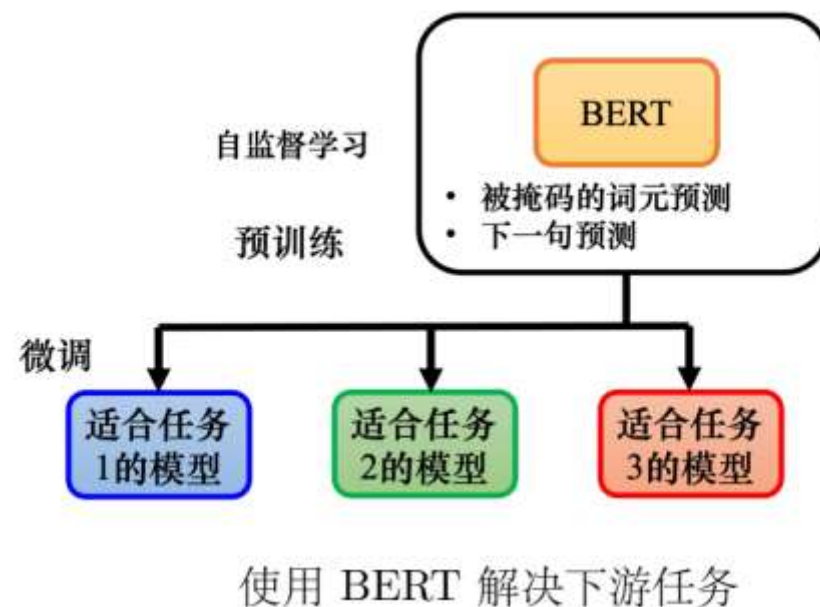
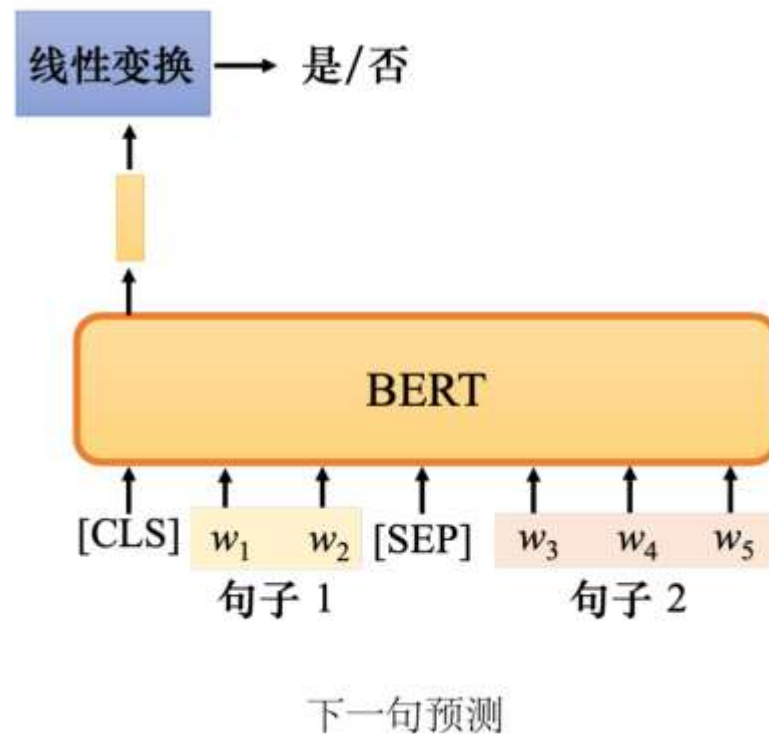
2. 机器学习

■ 自监督学习常见任务

BERT



BERT 的训练过程



2. 机器学习

■ 迁移学习（Transfer Learning）：

- ✓ 迁移学习利用在一个任务中学到的知识，应用到另一个相关任务中。这种方法尤其适用于数据不足或相似领域的任务。
- ✓ 应用场景：跨领域文本分类、图像识别等。
- ✓ 常见算法：模型微调、特征提取等。

2. 机器学习

(b) 如何衡量决策函数的好坏

- **建立一些准则：** 机器学习算法中定义一个损失函数
- 在所有的训练样本上来评价决策函数的风险：

$$R(\theta) = \frac{1}{N} \sum_{i=1}^N L(y^{(i)}, f(x^{(i)}, \theta))$$

- 风险函数 R 根据已知训练样本（经验数据）计算得到，称经验风险
- **学习的目标：** 找到一个参数使得经验风险 R 最小

$$\theta^* = \operatorname{argmin}_{\theta} R(\theta)$$

2. 机器学习

(b) 如何衡量决策函数的好坏

■ 过拟合与欠拟合现象

- ✓ 过拟合：在训练数据上表现良好，在未知数据上表现差；
- ✓ 欠拟合：在训练数据和未知数据上表现都很差；

■ 目前机器学习应用时多数出现的是过拟合现象。

2. 机器学习

(c) 机器学习中概念

■ 数据

是指所有能在计算机程序处理的对象的总称，可以是数字、字母和符号等。在不同的任务中，表现形式不一样，比如图像、视频、声音、文字、传感器数据等。

■ 特征

从数据中抽取的一些可以表征它们特性的数值称特征， 这些数值型特征一般可以表示为向量形式，也称为特征向量。

2. 机器学习

- **机器学习特征提取**：指从原始数据中提取出**有用的特征**，以便用于机器学习算法的训练和模型的建立。这些特征可以是数值、文字、图像甚至声音等。通过特征提取，可以帮助机器学习算法更好地理解数据，实现更好的数据分析和分类。
- **特征提取的目标**：找到一组**最能代表数据的特征集合**，这些特征可以使机器学习算法从数据中学到更有意义的信息。
- **特征提取的过程**：包括数据预处理、特征选择、特征提取和特征表示等步骤。在这些过程中，需要考虑到数据的种类、数量、质量和分布等因素，以及机器学习算法的适用性和需求等因素。

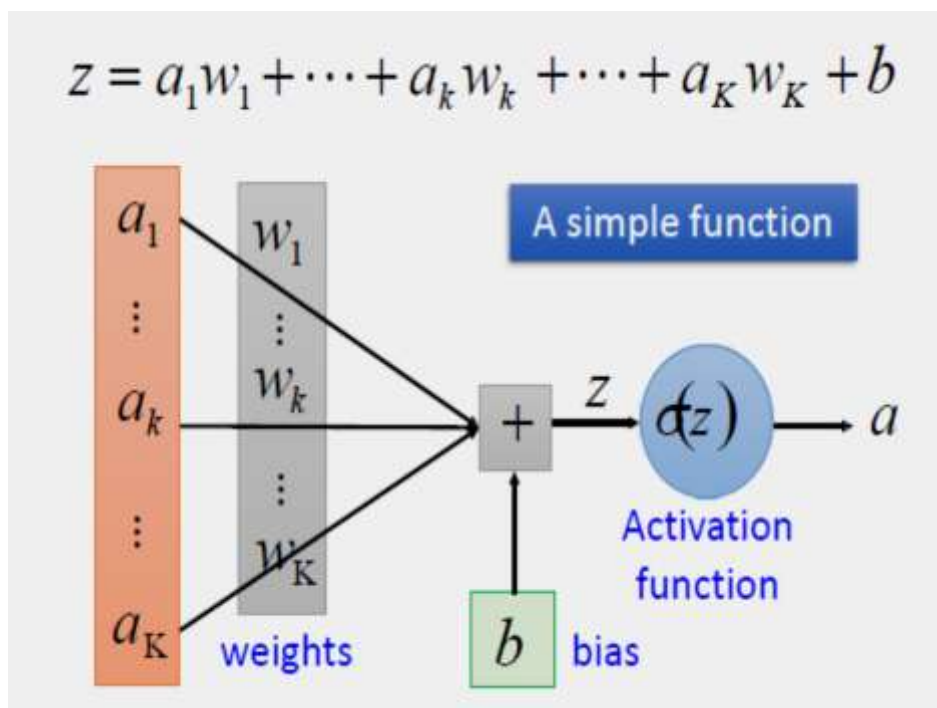
2. 机器学习

■ 样本

- ✓ 样本是按照一定的抽样规则从全部数据中取出的一部分数据，是实际观测得到的数据。
- ✓ 在有监督学习中，需要提供一组有输出目标的样本用来学习模型以及检验模型的好坏。

2. 机器学习

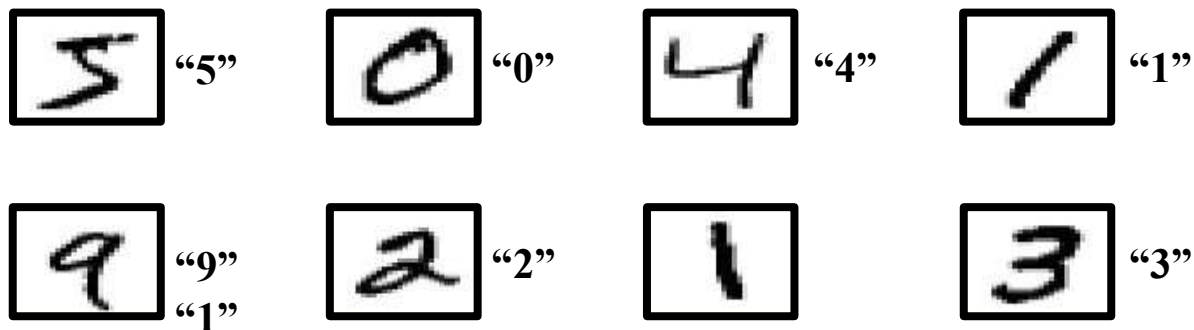
步骤1：定义一组函数



2. 机器学习

步骤2：训练数据

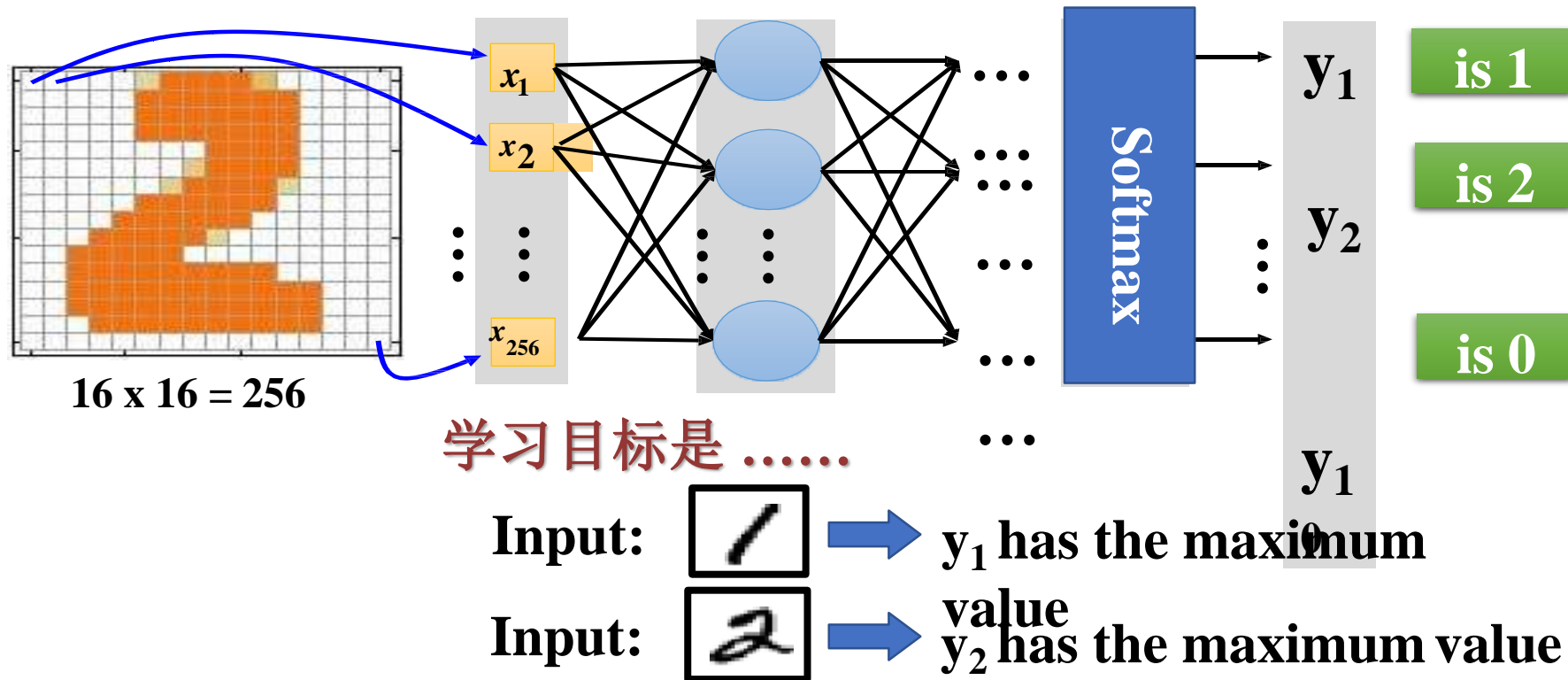
- 准备训练数据:图片和它们的标签



学习目标是在训练数据上定义的

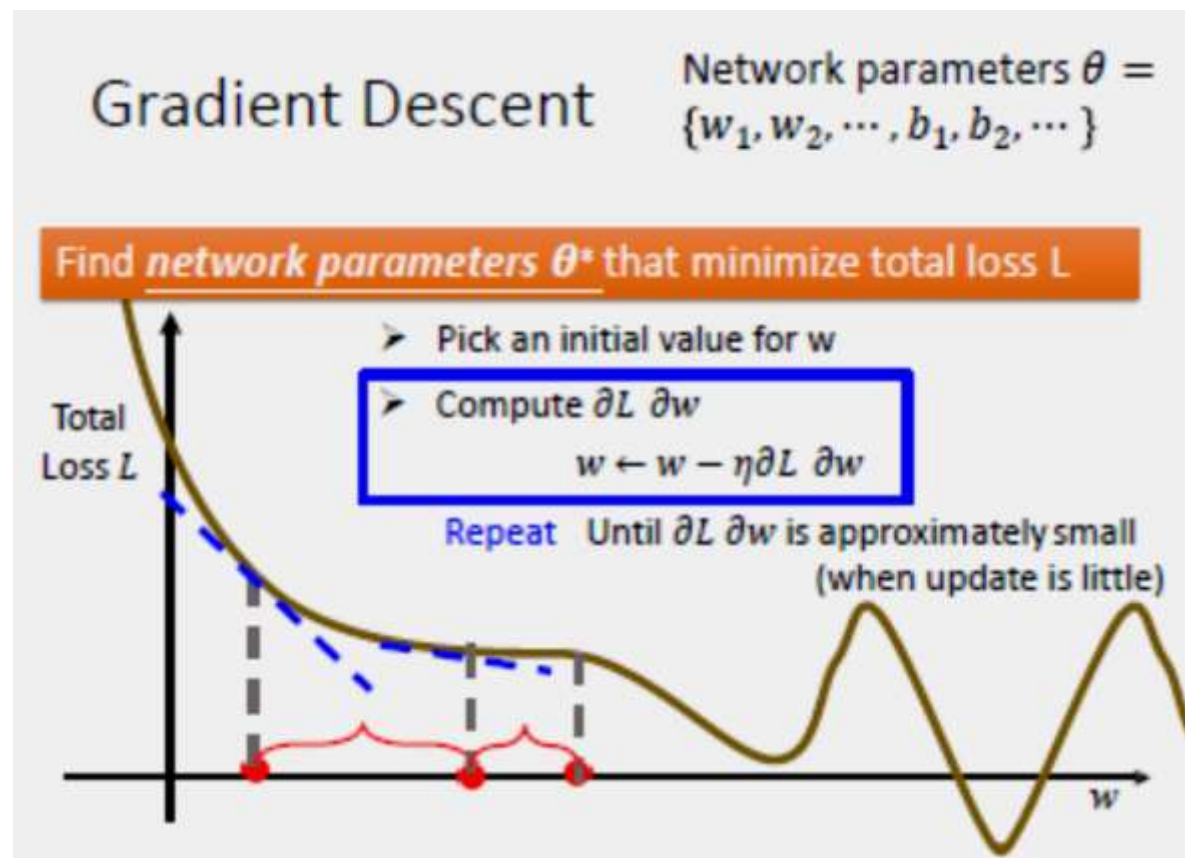
2. 机器学习

优化目标

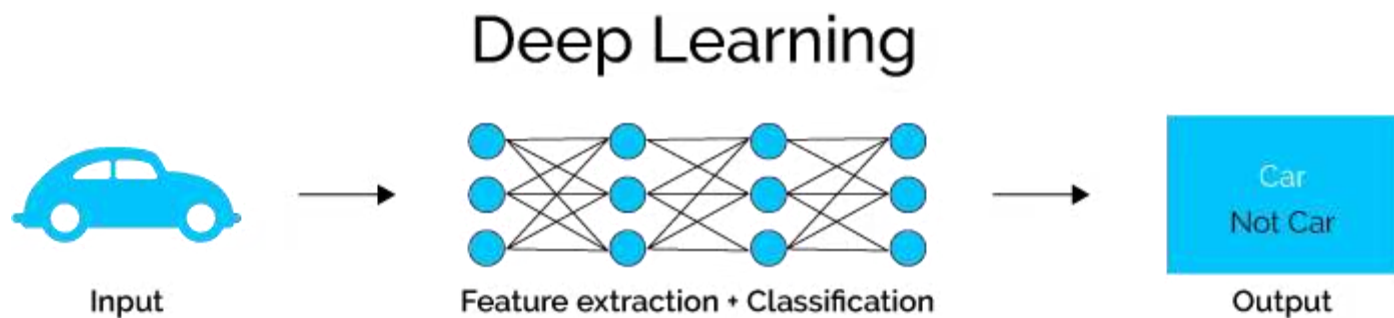
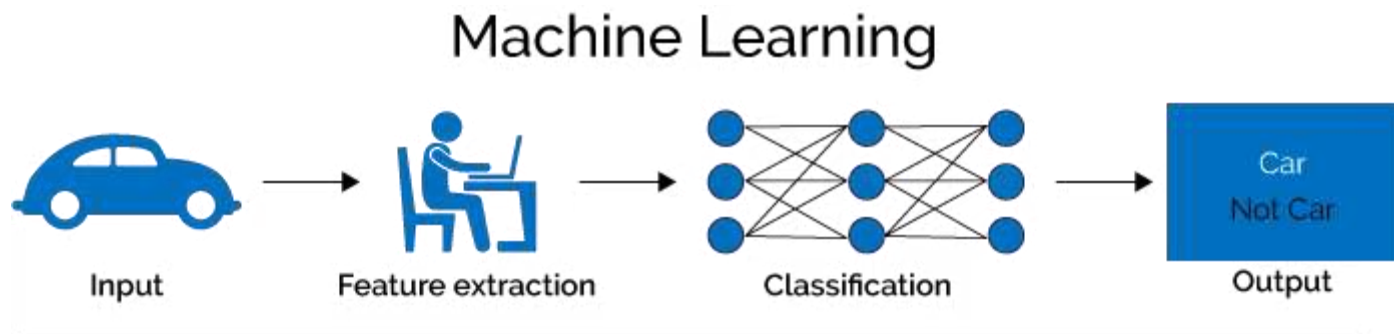


2. 机器学习

Step 3: 选择最优函数



2. 机器学习



3. 深度学习

- 深度学习是一种机器学习的方法（范式，paradigm），它利用由多个层次构成的神经网络来模拟和解决复杂问题。
- 它的核心思想是通过层次化的学习和特征提取，逐步提高模型对数据的表征能力和预测能力。
- 在深度学习中，核心技术包括反向传播算法、卷积神经网络、循环神经网络、生成对抗网络等。深度学习具有广泛的应用场景，如图像识别、语音识别、自然语言处理、人工智能等领域。

3. 深度学习

- 深度学习通常由多个神经网络层组成（至少有三层：输入层、隐藏层和输出层），它们通常将更简单的模型组合在一起，将数据从一层传递到另一层来构建更复杂的模型。通过训练大量数据自动得出模型，不需要人工特征提取环节。
- 深度学习算法试图从数据中自动学习高级功能，这是深度学习的一个非常独特的部分。

3. 深度学习

深度学习特征提取

- 深度学习特征提取是指利用深度神经网络模型对原始数据进行自动学习提取特征的过程，使得原始数据能够更好地表示为一组**高层次、抽象**的特征，从而能够更好地应用于分类、识别、检测等任务中。
- 深度学习特征提取在计算机视觉、自然语言处理、语音识别等领域被广泛使用。通过使用深度神经网络对数据进行特征提取，可以大大提高模型的准确度和鲁棒性。
- 深度学习可以通过堆叠多层神经网络来学习输入数据的**复杂特征**。
- 在深度学习中，每一层神经网络都可以看作是对输入数据的一种**映射**，而每一层的输出都可以作为下一层的输入。通过不断地堆叠多层神经网络并对权重进行训练，可以逐渐提取出更复杂的特征，从而达到更精确的分类和预测。

4. 卷积神经网络

目前，已有近40种神经网络模型，根据连接的拓扑结构，神经网络模型可以分为：

✓ **前向网络：**网络中各个神经元接受前一级输入，并输出到下一级，网络中没有反馈，实现信号从输入空间到输出空间的变换，信息处理能力来自于简单非线性函数的多次复合。网络结构简单，易于实现。

✓ **反馈网络：**网络内神经元间有反馈，信息处理是状态的变换，可以用动力学系统理论处理。系统的稳定性与联想记忆功能有密切关系。

5. 卷积神经网络

- 卷积神经网络（Convolutional Neural Networks, CNN）是一种前馈神经网络
- CNN受生物学上感受野（Receptive Field）的机制而提出，感受野主要是指听觉系统、本体感觉系统和视觉系统中神经元的一些性质。比如在视觉神经系统中，一个神经元的感受野是指视网膜上的特定区域，只有这个区域内的刺激才能够激活该神经元。

5. 卷积神经网络

为什么要引入CNN

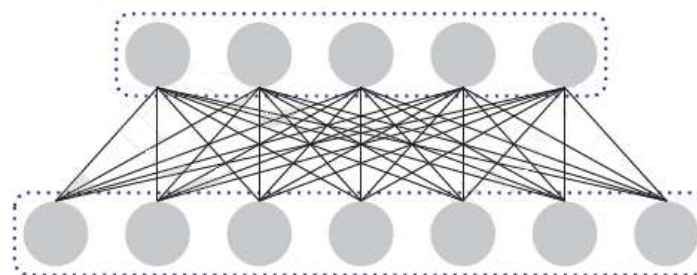
- 在人工的全连接神经网络中，每相邻两层之间的每个神经元之间都是有边相连的。
- 当输入层的特征维度变得很高时，这时全连接网络需要训练的参数就会增大很多，计算速度就会变得很慢
- 例如一张黑白的 28×28 的手写数字图片，输入层的神经元就有784(28×28)个

5. 卷积神经网络

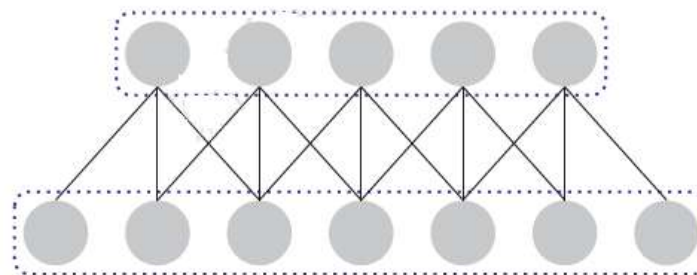
在卷积神经网络（Convolutional Neural Network, CNN）中，卷积层的神经元只与前一层的**部分**神经元节点相连，即它的神经元间的连接是非全连接的，且同一层中某些神经元之间的连接的权重 w 和偏移 b 是**共享的（即相同的）**，这样大量地减少了需要训练参数的数量。

5.卷积神经网络

- 卷积层：用卷积来代替全连接
 - ✓ 全连接前馈神经网络中，如果第 l 层有 n^l 个神经元，第 $l-1$ 层有 $n^{(l-1)}$ 个神经元，连接边有 $n^l \times n^{(l-1)}$ 个，也就是权重矩阵有 $n^l \times n^{(l-1)}$ 个参数。当 m 和 n 都很大时，权重矩阵的参数非常多，训练的效率会非常低。
 - ✓ 如果采用卷积来代替全连接：参数大大减少



(a) 全连接层



(b) 卷积层

5.卷积神经网络

卷积神经网络CNN的结构一般包含以下层：

- 输入层：用于数据的输入
- 卷积层：使用卷积核进行特征提取和特征映射
- 激励层：由于卷积也是一种线性运算，因此需要增加非线性映射
- 池化层：进行下采样，对特征图稀疏处理，减少数据运算量。
- 全连接层：通常在CNN的尾部进行重新拟合，减少特征信息的损失
- 输出层：用于输出结果

5.Tensorflow

- TensorFlow是一个非常强大且流行的开源机器学习框架，可以让开发者轻松构建和部署各种类型的机器学习模型，包括但不限于深度学习模型。
- TensorFlow工作流程由三个不同的部分定义，即数据预处理、构建模型和训练模型，从而进行预测。
- TensorFlow可用于开发自然语言处理、图像识别、手写识别以及基于计算的不同模拟（例如偏微分方程）等各种任务模型。

5.Tensorflow

软件架构

Caffe

(UC Berkeley)



Caffe2

(Facebook)

Paddle

(Baidu)

Torch

(NYU / Facebook)



PyTorch

(Facebook)

CNTK

(Microsoft)

Theano

(U Montreal)



TensorFlow

(Google)

MXNet

(Amazon)

Developed by U Washington, CMU, MIT, Hong Kong U, etc but main framework of choice at AWS

6. PyTorch

- PyTorch是一个基于Torch的Python机器学习包，它是一个基于编程语言Lua的开源机器学习包
- 主要特点：
 - ✓ 具有强大的GPU加速度的张量计算（如NumPy）
 - ✓ 自动微分建立和训练神经网络
- 与TensorFlow不同
 - ✓ TensorFlow在运行模型之前必须先定义整个计算图
 - ✓ PyTorch允许动态定义图。

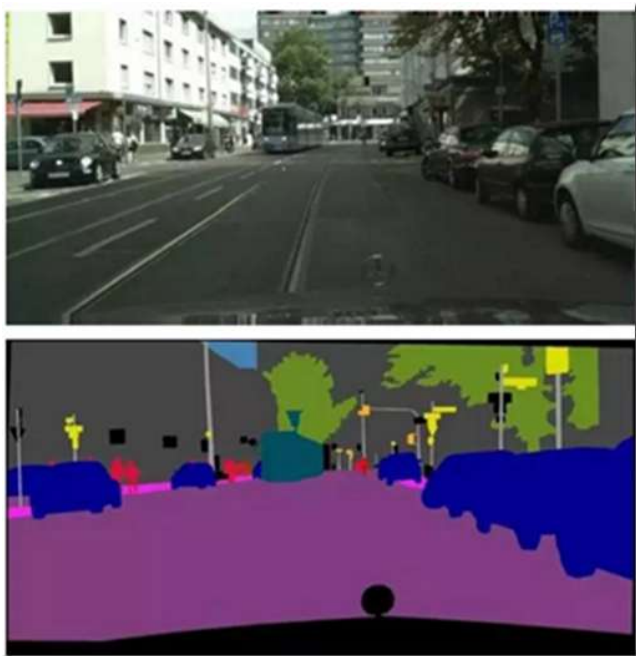
语义分割



语义分割概念

- 语义分割（semantic segmentation）是计算机视觉中十分重要的领域。
- 图像语义分割：根据图像的语义（蕴含的含义）来进行分割，即指定图像中每个像素的类别，即实现像素级别的分类。
- 查文献的关键词： pixel-wise segmentation

语义分割实例



predict →



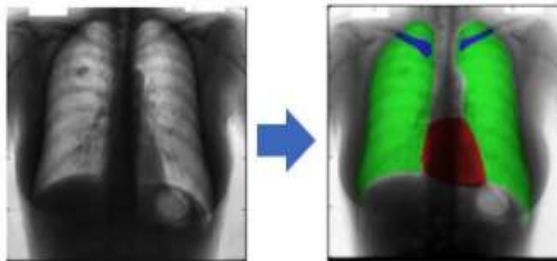
Person
Bicycle
Background

- 图像语义涉及理解：每幅图像中都蕴含一些信息
- 每幅图像就是一个“故事场景” (An image is a story)

语义分割的应用



自动驾驶



医学影像诊断



目标对象分割



面部分割，用于表情分析



精准农业

自动驾驶动态分割实例



图像语义分割与目标检测的区别

分割精度区别

- 语义分割： 像素级分割，将每个像素指定类别
- 目标检测： 只需给定包含目标的最小区域即可，窗口级



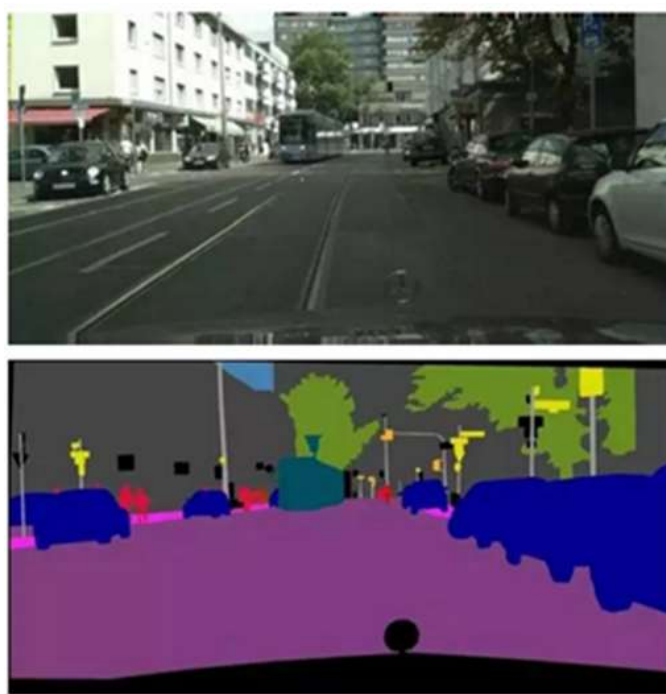
窗口级目标检测
(detection)



像素级分割
(segmentation)

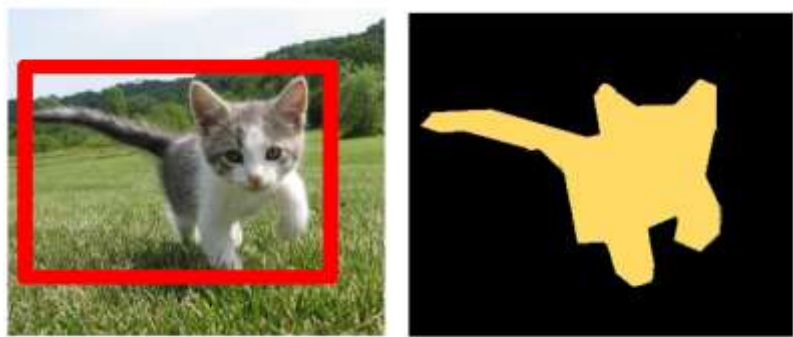
图像语义的本质

- 图像语义，就是图像内容的含义。图分割目标是将一张RGB图像（ $\text{height} \times \text{width} \times 3$ ）或是灰度图（ $\text{height} \times \text{width} \times 1$ ）作为输入，输出的是分割图。
- 输出图像的每一个像素的类别标签（ $\text{height} \times \text{width} \times 1$ ）



语义分割类型

(1) 单实例与多实例分割



单目标实例及分割结果



多目标实例及分割结果

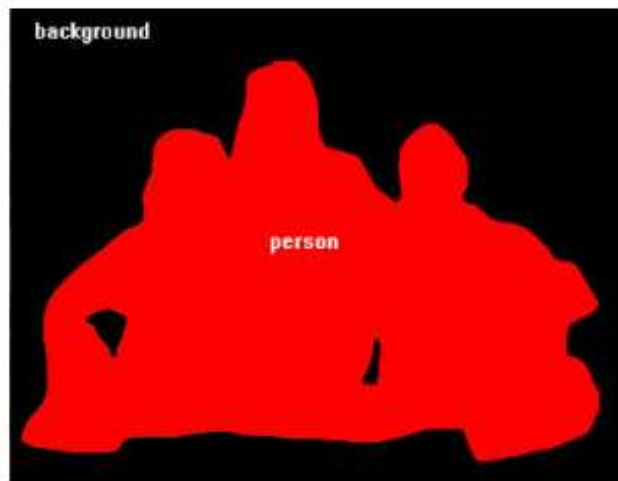
语义分割类型

(2) 语义分割与实例分割的区别

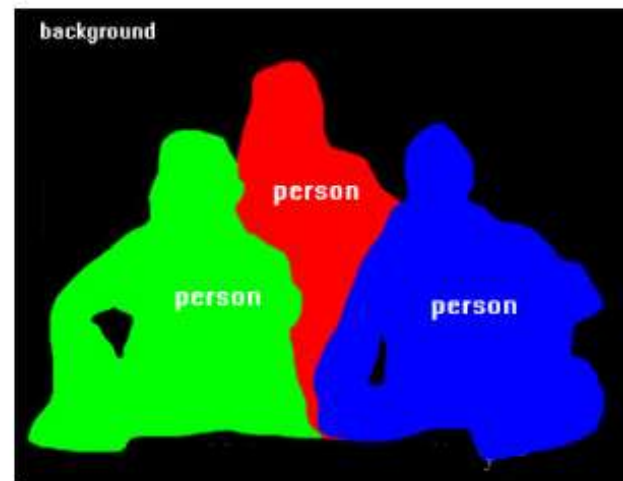
- 实例分割：对图像中的每个像素划分出对应的实例类别
- 实例分割关键：区别开不同的实例



输入图像



语义分割结果



实例分割结果

语义分割类型

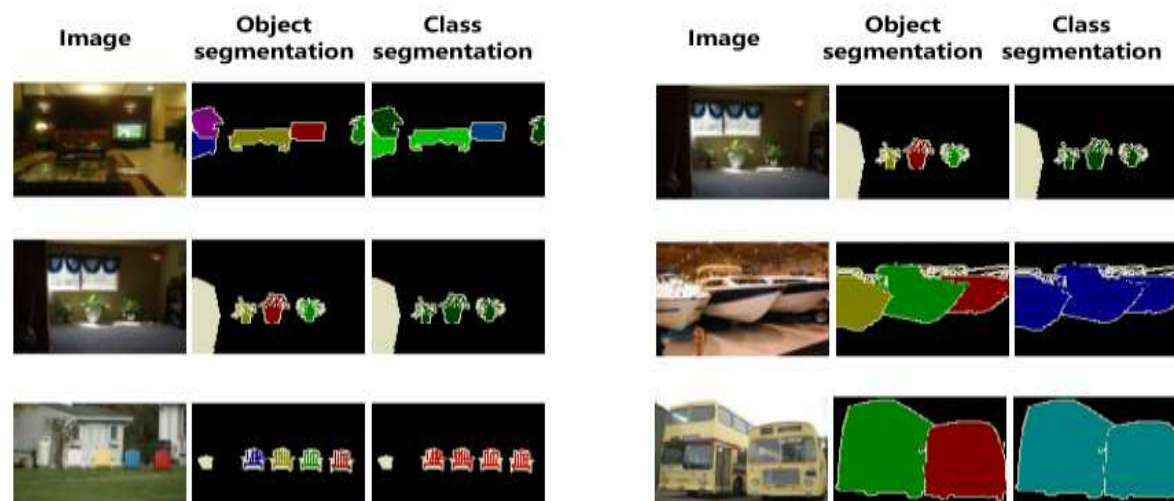
语义分割数据集

PASCAL VOC2011 Example Segmentations

Below are training examples for the segmentation taster, each consisting of:

- **the training image**
- **the object segmentation**
pixel indices correspond to the first, second, third object etc.
- **the class segmentation**
pixel indices correspond to classes in alphabetical order (1=aeroplane, 2=bicycle, 3=bird, 4=boat, 5=bottle, 6=bus, 7=car, 8=cat plant, 17=sheep, 18=sofa, 19=train, 20=tv/monitor)

For both types of segmentation image, index 0 corresponds to background and index 255 corresponds to 'void' or unlabelled.



语义分割类型

- 在早期、图像的语义分割采用传统的机器学习方法，如随机森林分类与决策树分类等方法。
- 2012年，AlexNet出现，Krizhevsky、Alex等采取滑动窗口的方式，利用周围像素对中心像素进行分类，即取每个像素为中心的小图像块输入卷积神经网络来预测该像素的语义标签。
- 2014年，全卷积神经网络(Fully Convolutional Networks, FCN)出现，它在AlexNet结构的基础上[去掉末端全连接层](#)，引入了端到端（End-to-End）的全卷积网络，大大地提高了分割精度。
- 2015年，基于编码器-解码器结构的分割技术得到发展，U-Net网络被提出，可以用于获取上下文和位置信息，采用数据增强、加权损失等策略；同年，SegNet将最大池化指数转移至解码器中，改善了分割分辨率。

语义分割类型

- 2016年，DeepLab V1结合深度卷积神经网络和条件随机场模型，改善了分割结果。
- 2018年，DeepLab V2被提出，基于空洞空间金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)结构，以不同采样率的空间卷积进行采样，可以捕捉更多的图像上下文信息。
- 2019年，DANet引入双注意力机制(Dual Attention, DA)获取上下文关系，即对通道以及空间利用注意力策略进行学习。
- 2020年，上下文感知的分割策略出现，用于语义分割的动态规划路径选择网络，可根据物体尺寸分布情况，动态生成网络拓扑、传输特征，网络计算成本较低。
- 2021年，设计基于Transformer的语义分割网络，它将语义分割视为序列到序列的预测任务，通过 Transformer的每一层建模全局上下文，增强了语义分割功能。

基于FCN的语义分割

- 2014年，加州大学伯克利分校的Long等人发表于CVPR的论文“Fully Convolutional Networks for Semantic Segmentation”提出了FCN分割方法，在不带有全连接层的情况下能进行密集预测。论文获得并荣获Best paper honorable mention。
- FCN能够将传统的分类网络转换为适用于分割任务的网络结构。与经典的CNN在卷积层之后使用全连接层得到固定长度的特征向量进行分类不同，FCN可以接受任意尺寸的输入图像，并通过转置卷积层对最后一个卷积层的feature map进行上采样，使其恢复到输入图像相同的尺寸。这样，FCN可以对每个像素都产生一个预测，同时保留了原始输入图像中的空间信息，最后在上采样的特征图上进行逐像素分类，从而解决了语义级别的图像分割问题。
- FCN为语义分割技术的发展奠定了基础。

基于FCN的语义分割

FCN的思想很直观，即直接进行像素级别端到端的语义分割，它可以基于主流的深度卷积神经网络模型（CNN）来实现。在FCN中，传统的全连接层fc6和fc7均是由卷积层实现，而最后的fc8层则被替代为一个21通道（channel）的1x1卷积层，作为网络的最终输出。之所以有21个通道是因为PASCAL VOC 的数据中包含21个类别（20个object类别和一个「background」类别）。

基于FCN的语义分割

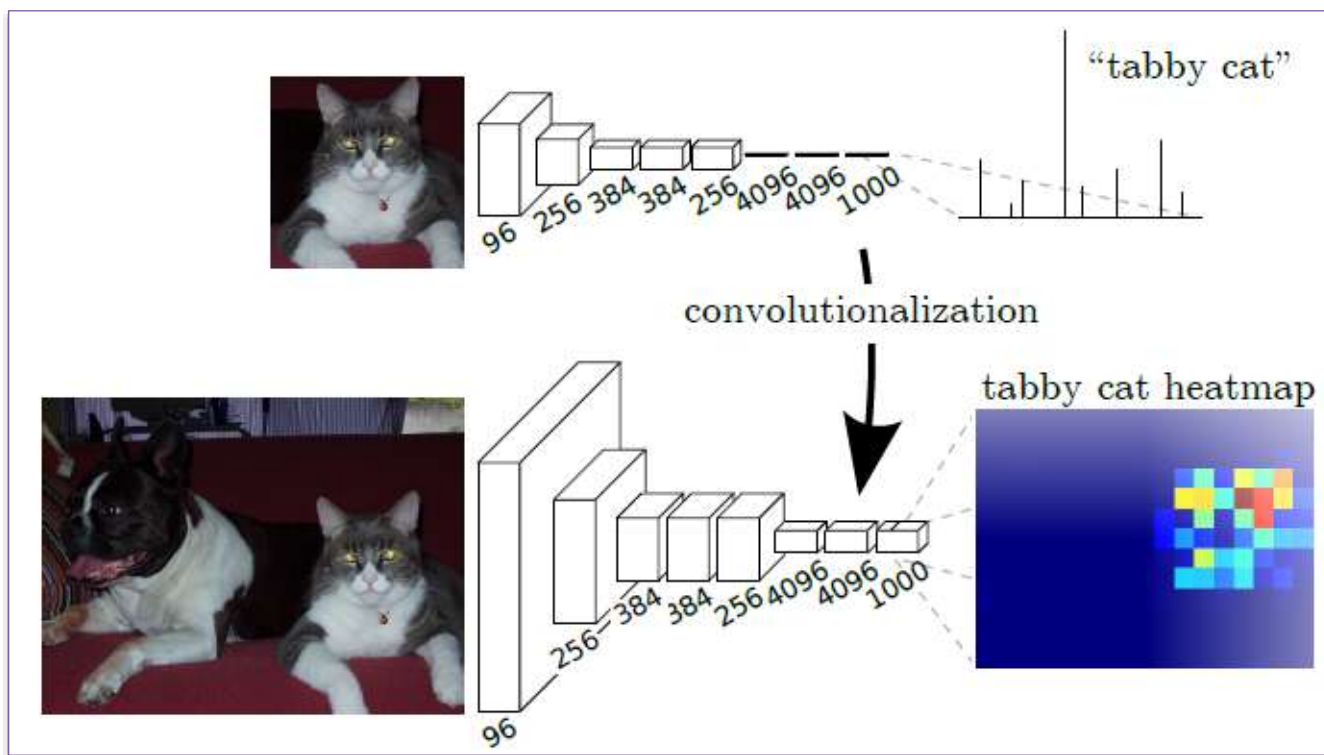


图 全连接层替换为卷积层

基于FCN的语义分割

- FCN将现有的CNN网络（如 AlexNet）作为网络的编码模块，用转置卷积层（transposed-convolution）作为解码模块，将低分辨率特征图上采样至全分辨率（与输入图像分辨率相同），从而得到分割图。
- 由于像素级（pixel-wise）语义分割需要输出整幅图像每个像素的分类，要求网络输出的特征图应该是二维的。
- 在AlexNet基础上需要替换掉全连接层，改换为卷积层，即将最后三个全连接层替换为卷积层。

基于FCN的语义分割

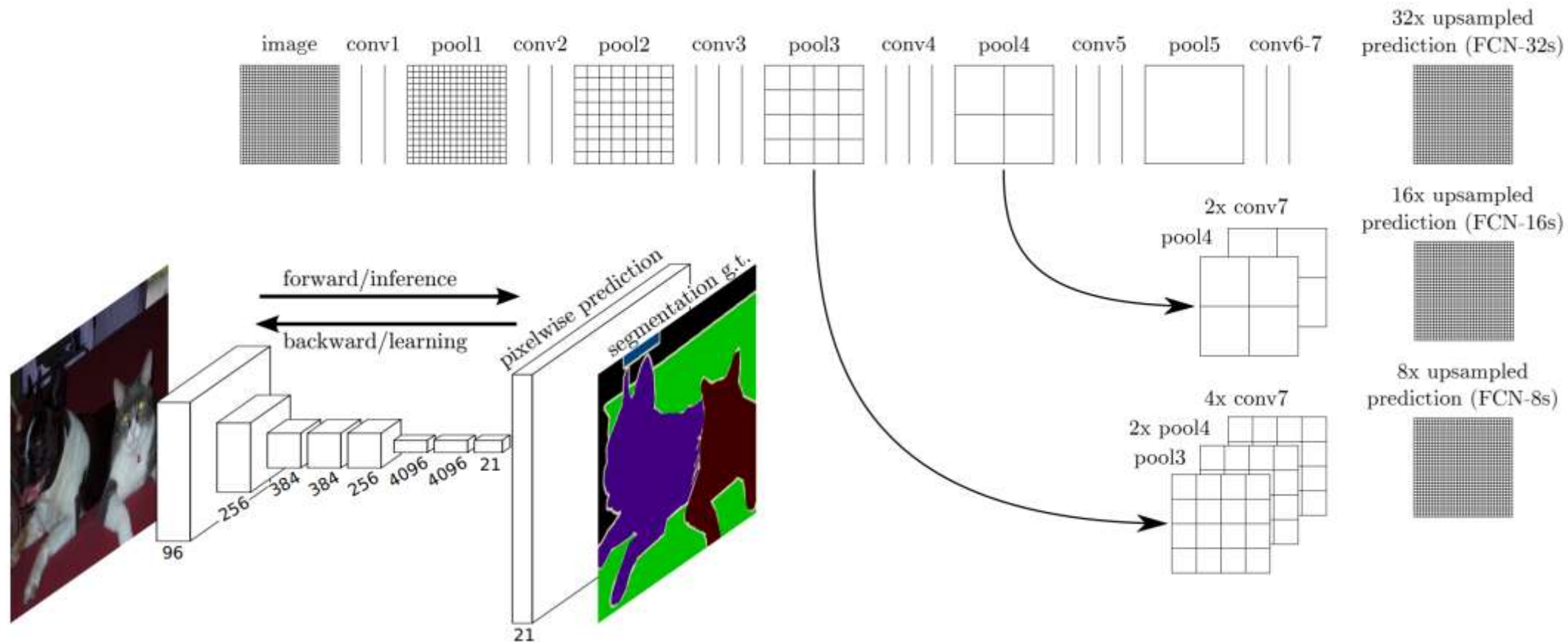


Figure 1. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation.

基于FCN的语义分割

FCN的层次卷积结构(1)

- 对原图像进行卷积conv1、pool1后原图像缩小为1/2
- 对图像进行第二次conv2、pool2后图像缩小为1/4
- 进行第三次卷积操作conv3、pool3缩小为原图像的1/8，此时保留pool3的featureMap
- 再对图像进行第四次卷积操作conv4、pool4，缩小为原图像的1/16，保留pool4的featureMap
- 最后对图像进行第五次卷积操作conv5、pool5，缩小为原图像的1/32

基于FCN的语义分割

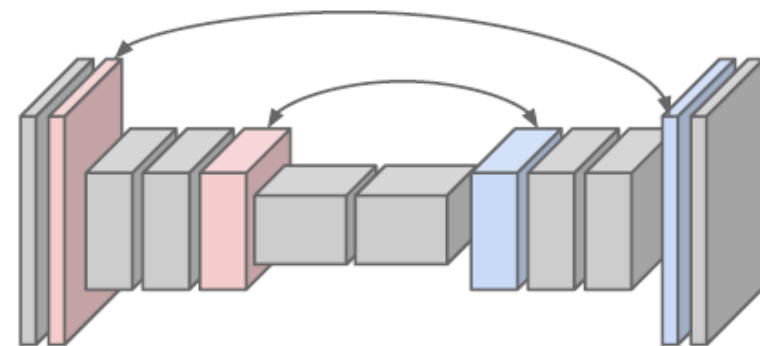
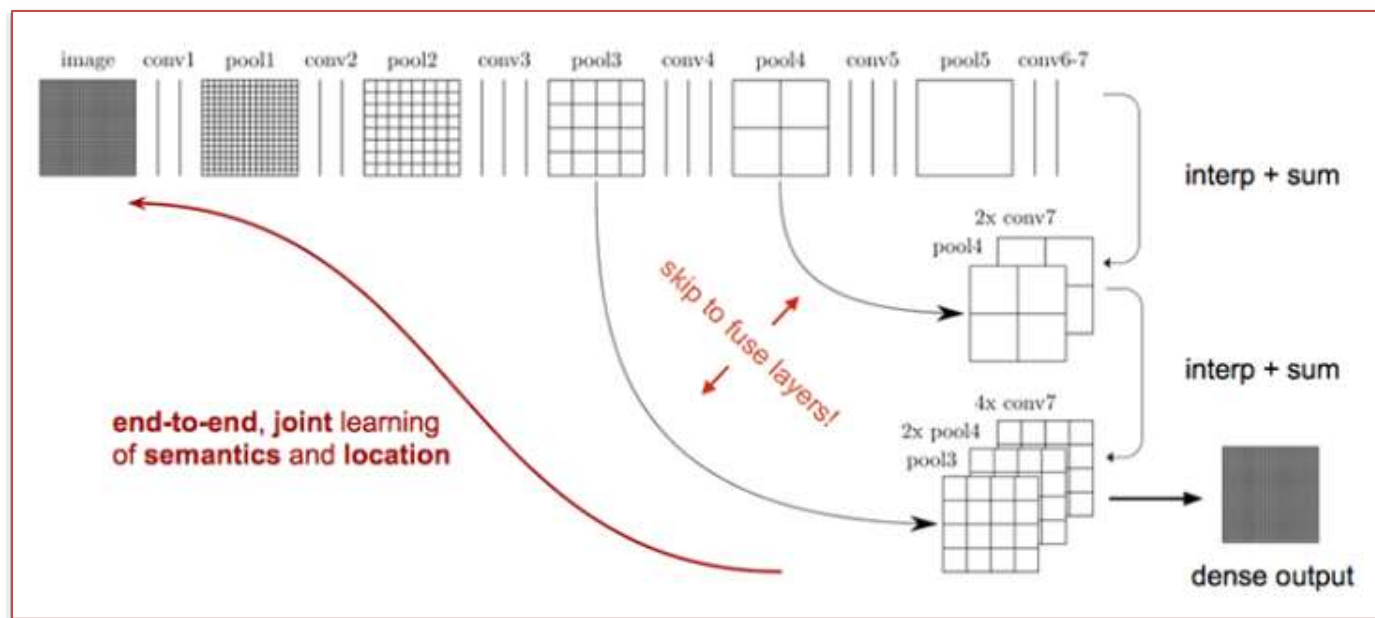
FCN的层次卷积结构(2)

- 将全连接换成卷积操作conv6、conv7，图像featureMap数量改变，图像大小依然为原图的1/32，此时featureMap成为热图heatMap
- 根据精度要求，需要跨层连接（跳跃结构）：
深层的特征对应全局的语义信息，浅层的特征对应着目标的位置信息，语义分割二者都需要考虑，所以引入了跳跃结构。

基于FCN的语义分割

FCN设计中的跨层连接（跳跃结构）

- 限于精度问题，由于不能够很好地还原图像当中的特征，向前迭代，借助低层细节特征还原
- 跨层融合粗细语义，能学习得到较好的语义特征，得到满意的空間预测结果



基于FCN的语义分割

应用跳跃结构得到三种不同的模型：

- 直接对pool5进行32x上采样后，将得到的特征图送入Softmax分类器，得到密集预测结果 → FCN-32s
- 对pool5进行2x上采样得到与pool4相同尺寸的上采样特征，并与pool4逐点相加得到特征图，对此特征图进行16x上采样，将得到的特征图送给Softmax分类器，得到密集预测结果 → FCN-16s
- 对pool5进行2x上采样得到与pool4相同尺寸的上采样特征，对此特征进行2x上采样得到与pool3相同尺寸的上采样特征，并与pool3逐点相加得到特征图，对此特征图进行8x上采样，将得到的特征图送给Softmax分类器，得到密集预测结果 → FCN-8s

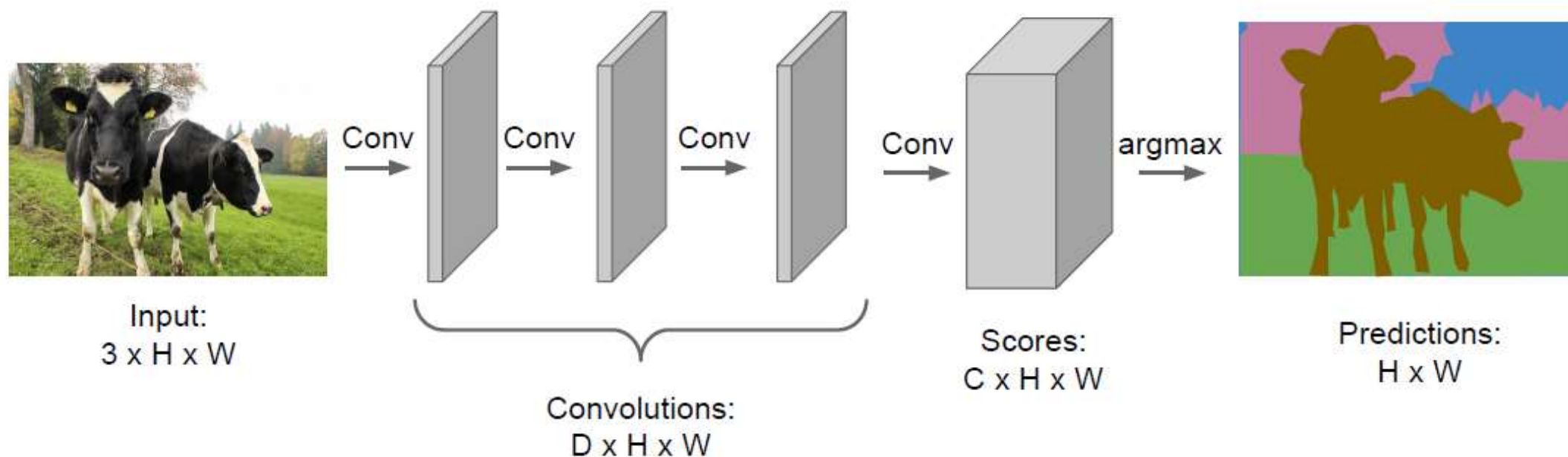
基于FCN的语义分割

将全连接层转换为卷积层的路线

- 全连接层和卷积层两类层的神经元都是计算点积，它们的函数形式相同。因此，**两者相互转化是可能的**。把全连接层转换为卷积层，实际上是用一个和输入的图像一样大小的卷积核去做这个操作
- 假设一个输出为4096的全连接层，输入数据的尺寸是 $7*7*512$ ，这个全连接层可以被等效地看作一个 $\text{kernel}=7$ ， $P=0$ ， $S=1$ ， $\text{num_out}=4096$ 的卷积层。即将滤波器的尺寸设置为和输入数据体尺寸一致，结果就和全连接层作用一样。

基于FCN的语义分割

- 设计卷积层用于像素的类别的预测
- 如果卷积特征尺度与输入图像一致，复杂度很高



基于FCN的语义分割

FCN方法优点

- 利用已经训练好的supervised pre-training的网络，无需从头训练，只需要微调fine-tuning即可，训练效率高
- 采用skip layer（跨层）的方法，在浅层处减小upsampling的步长，得到浅层精细特征fine-layer和高层得到的coarse-layer做融合，然后再upsampling可以得到准确的预测

基于FCN的语义分割

FCN语义分割方法缺点

- 得到的结果不够精细。进行8倍上采样虽然比32倍的效果好了很多，但是上采样的结果还是比较模糊和平滑，对图像中的细节不敏感。
- 对像素进行分类，没有充分考虑像素与像素之间的关系，忽略了像素类别的空域一致性规则。

基于FCN的语义分割

- 为什么说如果一个神经网络里面只有卷积层，那么输入的图像大小是可以任意的。
但是如果神经网络里不仅仅只有卷积层，还有全连接层，那么输入的图像的大小必须是固定的？
- ✓ 卷积层的参数和输入大小无关，它仅仅是一个卷积核在图像上滑动，不管输入图像多大都没关系。图像进行卷积时，因为每一个卷积核中的权值都是共享的，因此无论输入图像的尺寸多大，都可以按步长滑动做卷积，不同之处在于经过卷积运算，不同大小的输入图片所提取出的卷积特征的大小是不同的。

基于FCN的语义分割

- 为什么说如果一个神经网络里面只有卷积层，那么输入的图像大小是可以任意的。但是如果神经网络里不仅仅只有卷积层，还有全连接层，那么输入的图像的大小必须是固定的？
- ✓ 全连接层的参数与输入图像大小有关，全连接层的计算其实相当于输入的特征图数据矩阵和全连接层权值矩阵进行内积，在配置一个网络时，全连接层的参数维度是固定的，所以两个矩阵要能够进行内积，则输入的特征图的数据矩阵维数也需要固定。

深度学习与图像增强

深度学习与低光图像增强的背景

- 在数字图像处理领域，**低光图像增强（Low-Light Image Enhancement）**是一个常见的问题。由于各种原因，如不良的天气条件、室内照明不足或摄像机的限制，拍摄的图像可能会显得过暗。这种情况下，图像的细节很难辨认，这对于许多应用，如视频监控、医学成像或摄影，都是不理想的。
- 传统的图像增强方法，如直方图均衡化、伽马校正等，虽然可以提高图像的亮度，但往往会**牺牲图像的细节和对比度**。近年来，深度学习，特别是卷积神经网络（CNN），已经在许多图像处理任务中取得了显著的进展，包括低光图像增强。

深度学习与图像增强

发表于Pattern Recognition（PR）2017上的论文**LLNet: A deep autoencoder approach to natural low-light image enhancement**，提出了LLNET，是第一个真正意义上将深度学习应用在图像增强上的网络。

LLNet基于CNN，使用深度自编码的方法提取图像特征。它的主要目标是从低光图像中恢复细节并提高对比度。具体结构包括多个卷积层、激活函数和池化层，这些层共同工作，从输入的低光图像中提取特征，并重构增强后的图像。通过encoder部分，网络学习到图像的特征，并将其压缩到一个更小的维度。然后，通过decoder部分，它试图从这些特征中重建增强的图像。

深度学习与图像增强

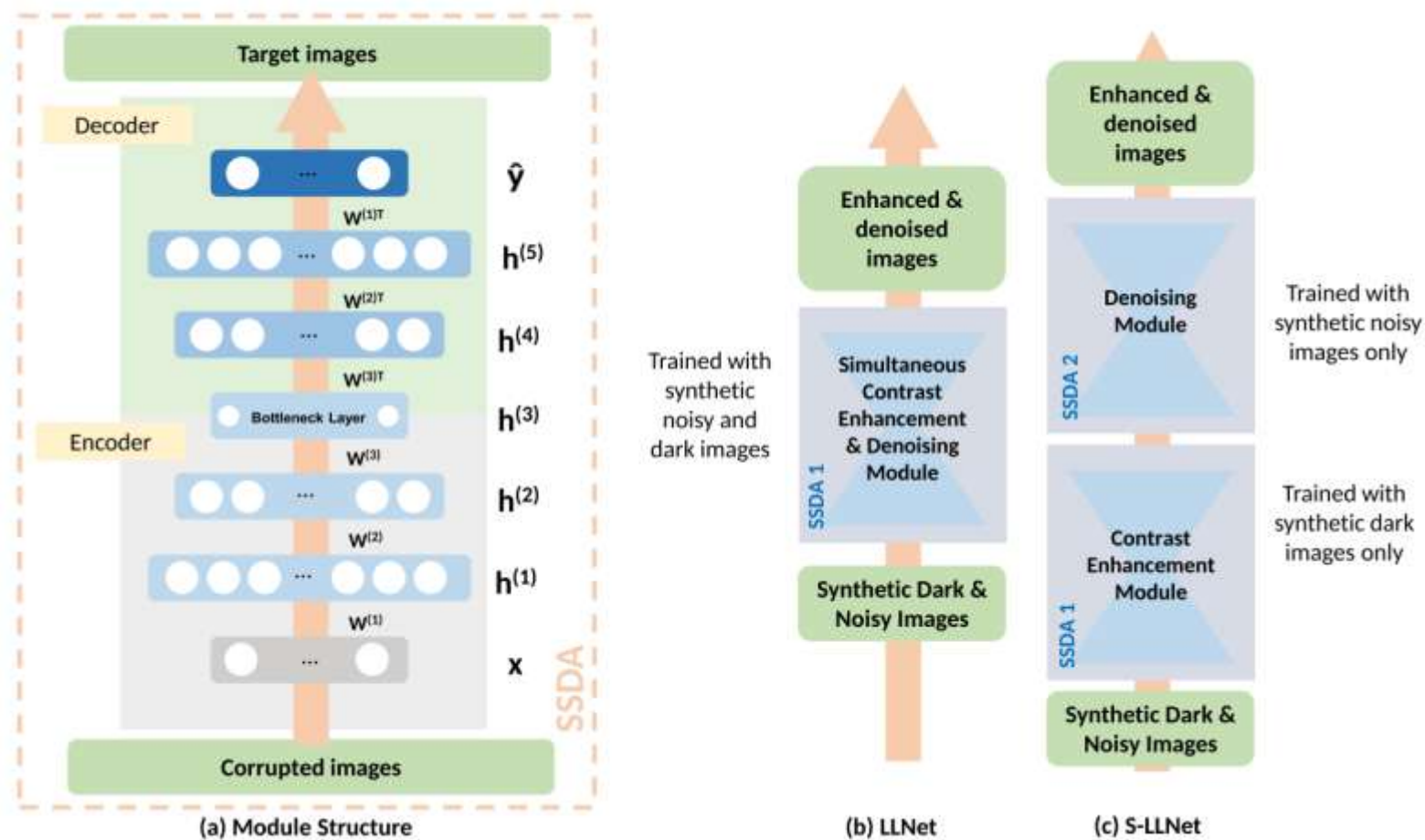


图 LLNet网络结构

深度学习与图像增强

论文中提出了两种网络

- LLNet是将亮度低并且含有噪声的图片输入编解码器中进行训练。
- S-LLNET则用亮度低的图片和含有噪声的图片分别训练SSDA模块并将两个模块连接起来。

When training both LLNet and S-LLNet, each DA is trained by error back-propagation to minimize the sparsity regularized reconstruction loss as described in Xie et al. [25]:

$$\begin{aligned}\mathcal{L}_{\text{DA}}(\mathcal{D}; \theta) = & \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \|y_i - \hat{y}(x_i)\|_2^2 + \beta \sum_{j=1}^K \text{KL}(\hat{\rho}_j \| \rho) \\ & + \frac{\lambda}{2} (\|W\|_F^2 + \|W'\|_F^2)\end{aligned}\tag{3}$$

整个SSDA的损失函数如下所示：

$$\mathcal{L}_{SSDA}(\mathcal{D}; \theta) = \frac{1}{N} \sum_{i=1}^N \|y_i - \hat{y}(x_i)\|_2^2 + \frac{\lambda}{L} \sum_{l=1}^{2L} \|W^{(l)}\|_F^2$$

where L is the number of stacked DAs and $W^{(l)}$ denotes weights for the l th layer in the stacked deep network. The sparsity inducing term is not needed for this step because the sparsity was already incorporated in the pre-trained DAs.

■ 模型评估

✓ PSNR

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (6)$$

The PSNR, in decibels (dB), is defined as:

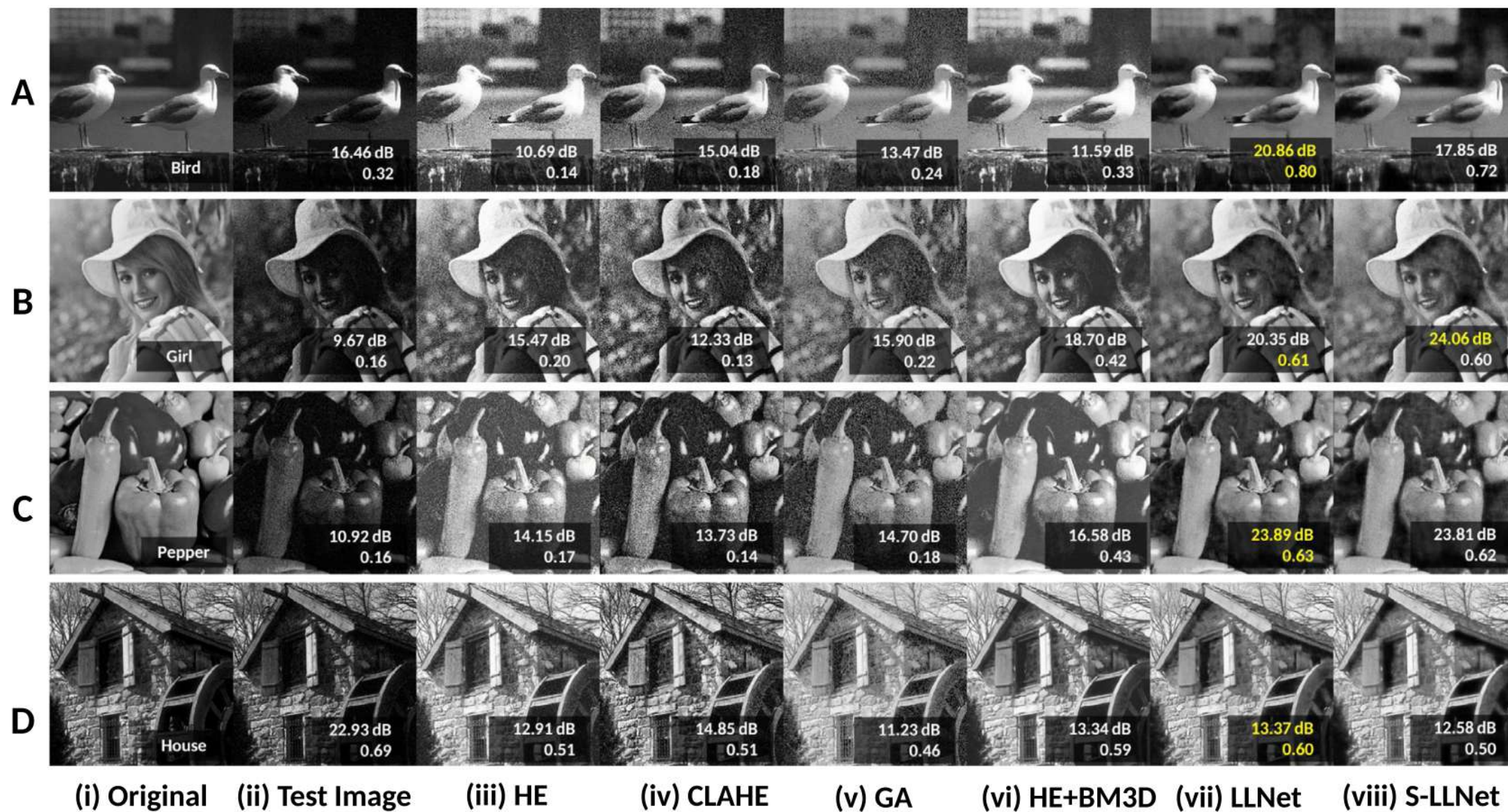
$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\max(I)^2}{\text{MSE}} \right) \quad (7)$$

Here, $\max(I)$ is the maximum possible pixel value of the image I .

✓ SSIM: 用来评价两幅图片的相似相似度

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

深度学习与图像增强



深度学习与图像增强

LLNet与其他同类技术相比表现出色。它不仅在增强低光图像的准确性方面表现出众，而且由于其深度学习的本质，还能够许多复杂的情况下提供更自然的增强效果。

与传统的图像增强技术相比，LLNet有以下优势：

- **自适应性**：不需要手动调整参数，LLNet可以自动识别并修复图像中的问题。
- **高效性**：尽管深度学习模型通常需要较大的计算资源，但预训练的模型可以在短时间内产生出色的结果。

人工智能 v.s. 传统图像处理技术



人工智能 v.s. 传统图像处理技术

- **传统的图像处理技术**主要依赖于数学和物理的理论和方法，如傅里叶变换、小波变换、边缘检测、滤波、形态学、特征提取等，它们通常需要人为地设定参数和规则，以适应不同的图像场景和需求。然而，随着图像数据的规模和复杂度的增加，传统的图像处理技术面临着很多挑战和局限性，如计算效率低、泛化能力差、难以处理高层语义信息等。
- 人工智能技术，尤其是深度学习技术，逐渐成为了图像处理领域的新兴力量，它能够利用大量的数据和强大的计算能力，自动地学习图像的特征和规律，从而实现更高效、更准确、更智能的图像处理任务，如图像分类、目标检测、人脸识别、图像生成、图像风格转换等。

人工智能技术的优势

(1) 数据驱动

人工智能技术能够从大量的数据中自动地学习图像的特征和规律，而不需要人为地设定参数和规则。这使得人工智能技术能够适应各种复杂和多变的图像场景和需求，具有很强的泛化能力和适应能力。

(2) 高度并行化，计算效率高

人工智能技术能够利用强大的计算能力，如GPU、TPU等，来加速图像处理的过程，实现实时或近实时的图像处理效果。这使得人工智能技术能够应对大规模和高速的图像处理任务，提高图像处理的效率和性能。

人工智能技术的优势

(3) 智能化

人工智能技术能够处理**高层语义信息**，如图像的内容、风格、情感等，从而实现更高级和更复杂的图像处理任务，如图像生成、图像风格转换、图像修复、图像增强等。这使得人工智能技术能够创造出更丰富和更有趣的图像处理效果，提升图像处理的质量和价值。

人工智能技术的劣势

■ 数据依赖

需要大量的数据来训练和优化模型，而不是利用先验的知识和理论。这使得在数据量不足或数据质量不高的情况下，容易出现过拟合或欠拟合的问题，导致模型的泛化能力下降或出现错误和偏差。

■ 计算资源密集

需要大量的计算能力来训练和运行模型，而不是利用简单的算法和公式。这使得人工智能技术在计算资源有限或计算成本高昂的情况下，难以实现或部署，限制了其普及和应用的范围和场景。

人工智能技术的劣势

■ 可解释性低

人工智能方法的内部机制和原理往往是黑盒的，难以理解和解释，而不是利用清晰的逻辑和推理。这使得人工智能技术在出现错误或异常的情况下，难以找到原因和解决方案，也难以获得用户和社会的信任和认可。

玄学？

传统图像处理技术的优势

■ 理论基础强

主要依赖于数学和物理的理论和方法，如傅里叶变换、小波变换、边缘检测、滤波、形态学、特征提取等，它们通常有着清晰的定义和公式，能够对图像进行精确和稳定的分析和处理，而不需要大量的数据和计算能力。

■ 计算简单

主要利用简单的算法和公式来实现图像处理的功能，如灰度化、二值化、直方图均衡化、阈值分割、颜色空间转换等，它们通常不需要复杂的网络结构和优化方法，能够在普通的CPU上快速地运行和部署，降低了图像处理的计算成本和资源消耗。

传统图像处理技术的优势

■ 可解释性高

内部机制和原理往往是白盒的，容易理解和解释，能够利用清晰的逻辑和推理来说明图像处理的过程和结果，而不是依赖于复杂的网络结构和优化方法。这使得传统图像处理技术在出现错误或异常的情况下，容易找到原因和解决方案，也容易获得用户和社会的信任和认可。

传统图像处理技术的劣势

■ 参数调节

主要依赖于人为地设定参数和规则，以适应不同的图像场景和需求，如滤波器的大小、阈值的选择、特征的提取等，它们通常需要**经验和专业知识**，而不是自动地学习和优化。这使得传统图像处理技术在图像数据的规模和复杂度增加的情况下，容易出现参数的不合适或不一致的问题，**导致图像处理的效果和性能下降或出现错误和偏差。**

■ 泛化能力差

传统方法主要利用先验的知识和理论，以处理特定的图像问题和任务，如边缘检测、特征提取、图像压缩等，它们通常难以处理高层语义信息，如图像的内容、风格、情感等，而不是利用大量的数据和计算能力。这使得传统图像处理技术在图像场景和需求变化的情况下，难以适应和调整，**限制了其应用的范围和场景。**

传统图像处理技术的劣势

■ 创新性低

传统技术主要实现基本和常见的图像处理功能，如图像的分析、处理和增强等，它们通常难以实现更高级和更复杂的图像处理任务，如图像生成、图像风格转换、图像修复、图像增强等，而不是利用智能化的技术和方法。这使得传统图像处理技术在图像处理的质量和方面，难以创造出更丰富和更有趣的图像处理效果，降低了其竞争力和吸引力。

两种技术结合和互补

■ 数据预处理

在使用人工智能技术，尤其是深度学习技术，来处理图像时，通常需要对图像数据进行一些预处理，以提高数据的质量和一致性，减少数据的噪声和冗余，增加数据的多样性和平衡性等。

这些预处理的步骤，往往需要借助传统图像处理技术，如图像的裁剪、缩放、旋转、翻转、平移、色彩变换、直方图均衡化、噪声去除、锐化、模糊等，来实现。这些预处理的技术，可以有效地提升人工智能技术的图像处理的效果和性能。

两种技术结合和互补

■ 特征提取

深度学习技术来处理图像时，通常需要从图像中提取有用的特征，以表示图像的内容和属性，供后续的图像处理任务使用，如图像分类、目标检测、人脸识别等。这些特征的提取，往往需要借助传统图像处理技术，如边缘检测、角点检测、SIFT、SURF、HOG等，来实现。这些特征的提取，可以有效地提高人工智能技术的图像处理的准确度和鲁棒性。

两种技术的结合和互补

■ 特征提取

Wavelet Approximation-Aware Residual Network for Single Image Deraining, TPAMI 2023

基于小波近似感知残差网络的单图像去雨

作者: Wei-Yen Hsu; Wei-Chi Chang

雨痕会影响图像的视觉质量，并干扰人类视觉和计算机视觉算法。对于许多户外计算机视觉应用来说，恢复被雨水降级的图像是重要的。与利用时间冗余和降雨动态的视频去雨方法不同，单图像去雨方法只能访问邻近像素的空间信息以及雨和背景场景的视觉特征。这使得从单个图像中去除雨水比从由多帧组成的视频中去除雨水更加困难。

两种技术结合和互补

■ 特征提取

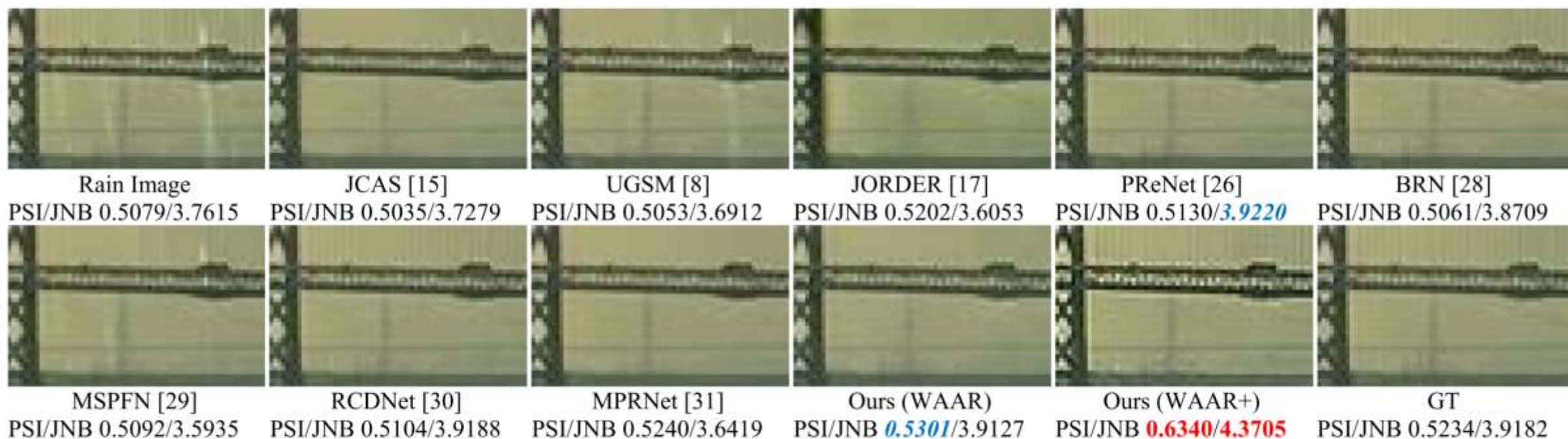
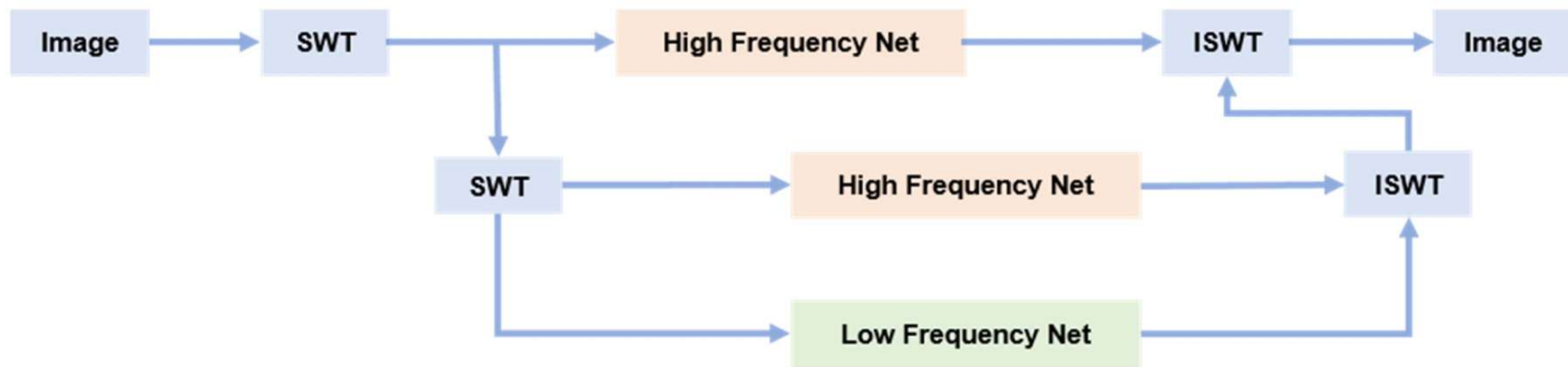


Fig. 1. Performance comparison of the images produced by the state-of-the-art methods for deraining. Our WAAR/WAAR+ can produce structure-preserving and edge-sharp images than the baseline methods.

两种技术的结合和互补

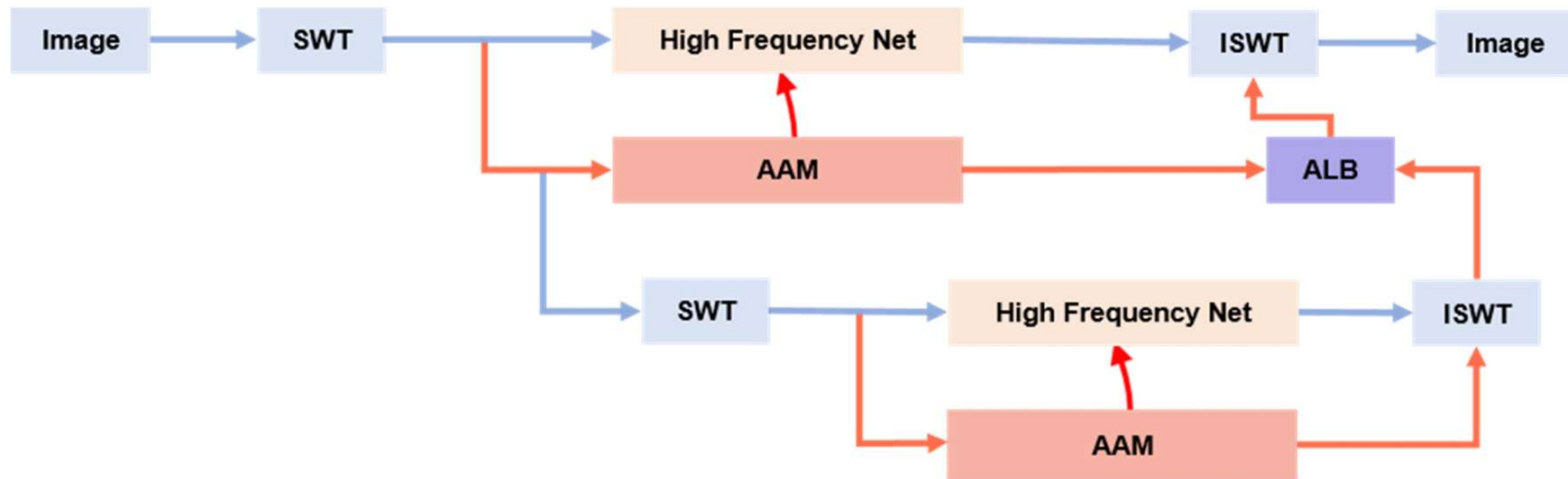
- 大多数现有的基于CNNs的深度去雨方法通常学习从有雨图像到干净无雨图像的直接映射，并且它们的架构越来越复杂。
- 由于雨水与物体边缘和背景混合的限制，很难分离雨水和物体/背景，图像的边缘细节在重建过程中无法有效恢复。
- 为了解决这个问题，提出了一种小波近似感知的残差网络（WAAR）。在该网络中能够在每一层次上分别有效去除低频结构和高频细节中的雨水，尤其是在每一层次的低频子图像中。经过小波变换后，提出了近似感知（AAM）和近似层级混合（ALB）机制，以进一步帮助每个级别的低频网络递归恢复低频子图像的结构和纹理，而高频网络可以通过块连接有效地消除雨痕，并通过调整超参数实现不同程度的边缘细节增强。
- 引入了块连接来丰富高频网络中的高频细节，有助于揭示高频与低频特征之间的潜在相互依赖关系。

两种技术的结合和互补



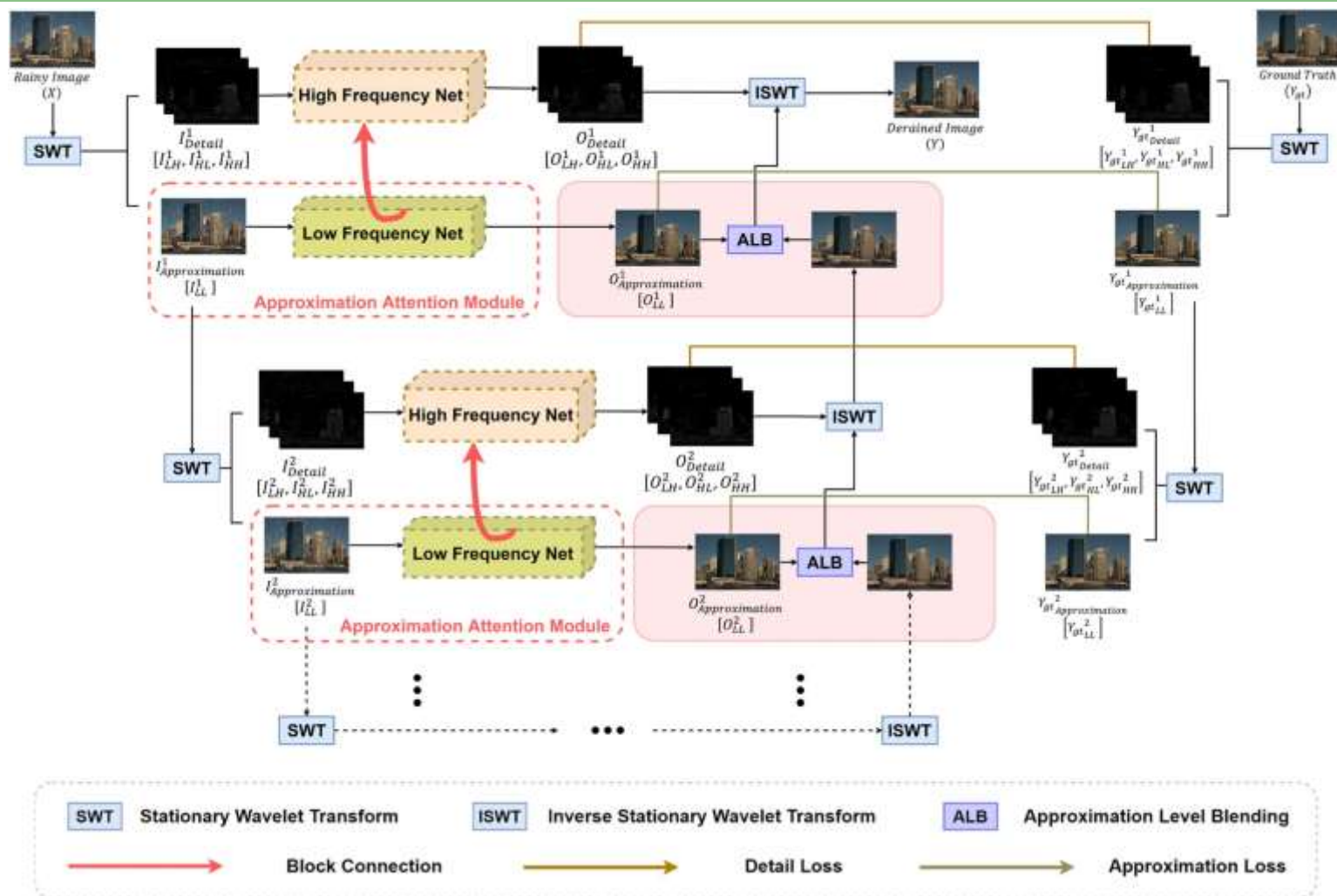
(a) Traditional wavelet transform-based rain removal network

两种技术的结合和互补



(b) Proposed WAAR/WAAR+ rain removal network

两种技术的结合和互补



Architecture of the proposed WAAR network

两种技术结合和互补

Efficient Frequency-Domain Image Deraining with Contrastive Regularization, ECCV 2024, 北航

作者: **Ning Gao, Xingyu Jiang, Xiuhui Zhang, and Yue Deng**

源码: **<https://github.com/deng-ai-lab/FADformer>**

动机: 大多数当前的单图像去雨（SID）方法都是基于Transformer，通过全局建模实现高质量的重建。然而，它们的架构仅从空间域构建长距离特征，这在保持有效性的同时带来了显著的计算负担。此外，这些方法要么在训练中忽视了负样本信息，要么未能充分利用负样本中存在的雨迹模式。

两种技术结合和互补

Efficient Frequency-Domain Image Deraining with Contrastive Regularization, ECCV 2024, 北航

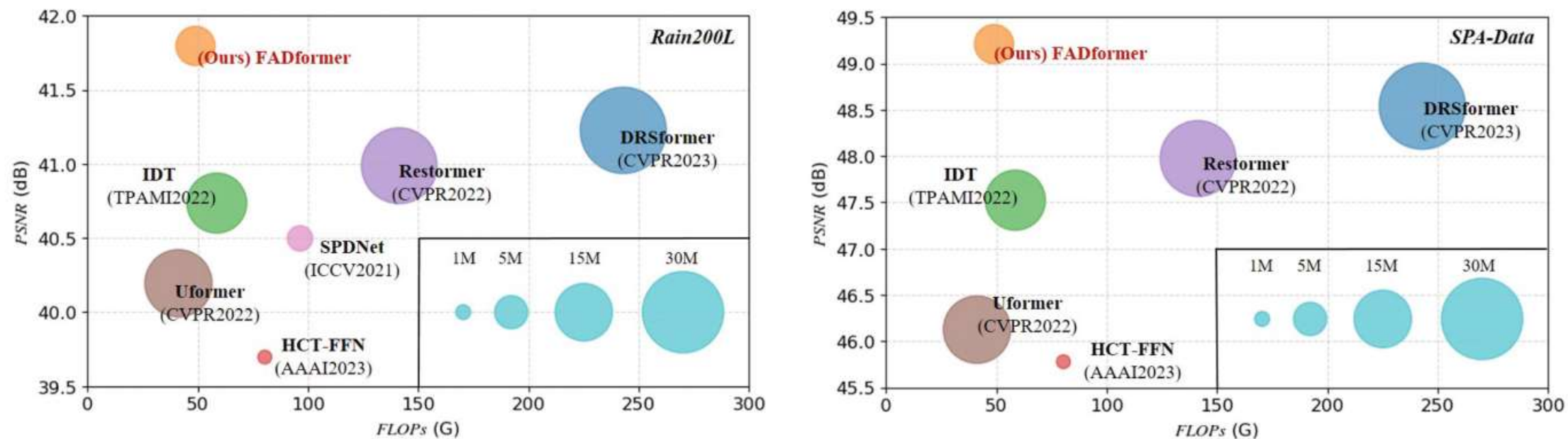


Fig. 1. Comparison results on the **Rain200L** [50] dataset and **SPA-Data** [45] dataset. The circle size reflects the number of model parameters. The proposed method achieves a better trade-off between model complexity and performance.

两种技术结合和互补

Efficient Frequency-Domain Image Deraining with Contrastive Regularization, ECCV 2024, 北航
创新点

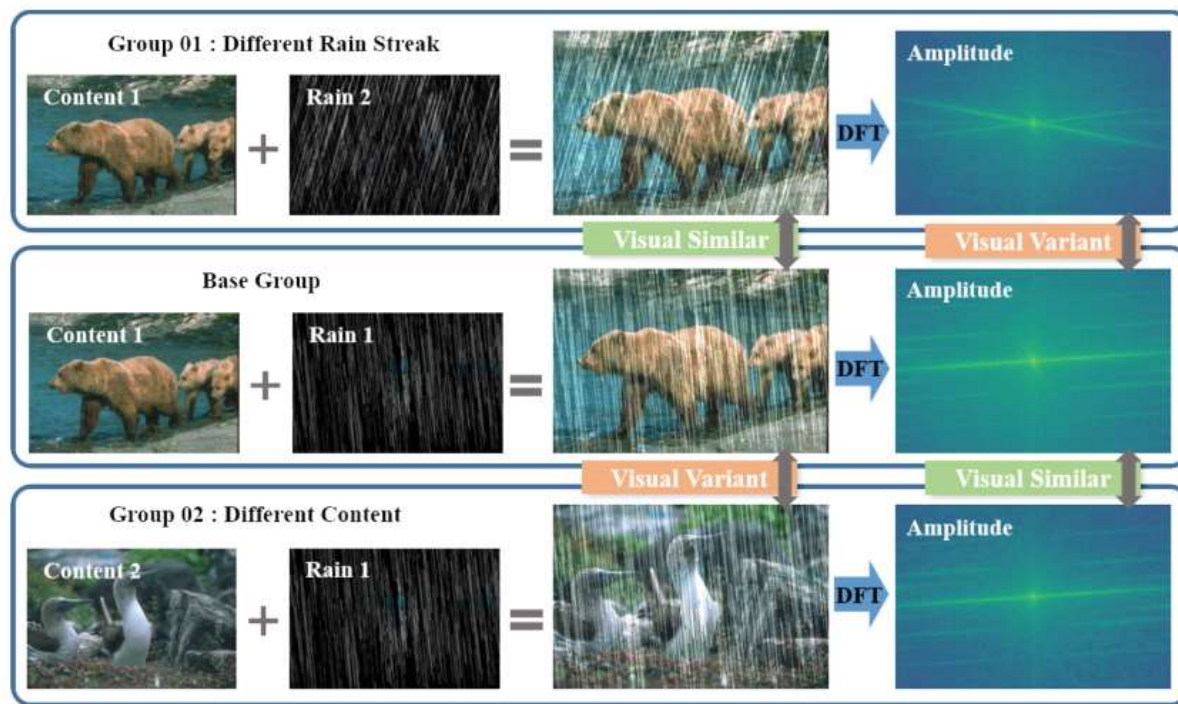
- 提出了一种高效的频率感知去雨框架（**FADformer**）：该框架通过在频域中捕获特征来实现高效的雨迹去除，这与仅依赖于空间域特征的方法相比，提供了一种新的视角和解决方案。
- 构建了融合傅里叶卷积混合器（**FFCM**）：FFCM能够在空间和频域中进行卷积操作，这使得它在保持效率的同时具备了局部-全局特征捕获的能力，与自注意力机制相比，具有更高的计算效率。
- 引入了先验门控前馈网络（**PGFN**）：PGFN通过引入残差通道先验（RCP）信息，以门控方式增强了局部细节和结构特征的恢复能力，这种结合先验知识的方法在去雨任务中显示出了其有效性。
- 设计了频域对比正则化（**FCR**）：作者提出了一种新的频域对比正则化方法，该方法利用频域特征作为对比空间，有效地利用了负样本信息，并显著提高了去雨性能，这一方法在多个数据集和任务中都显示出了其普遍的有效性。

两种技术结合和互补

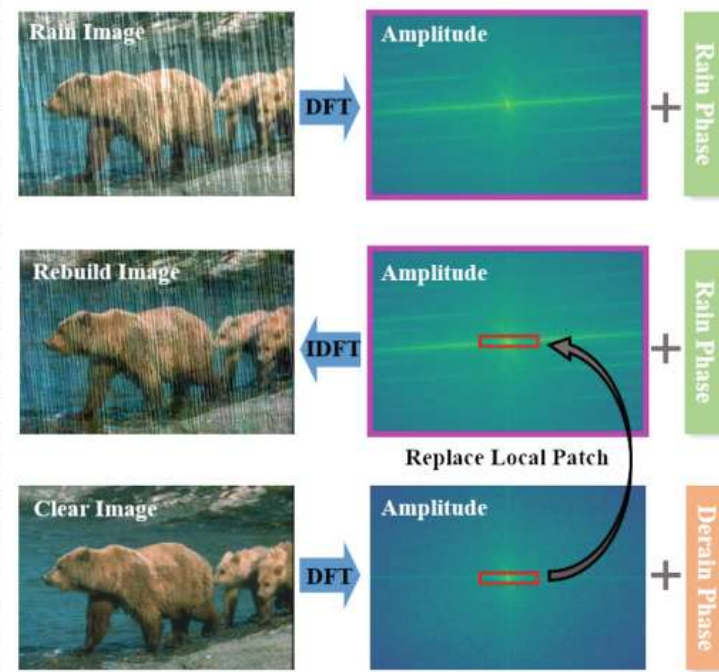
Efficient Frequency-Domain Image Deraining with Contrastive Regularization, ECCV 2024, 北航

目标:

- (1) 建立一个既高效又擅长全局-局部建模的去雨网络,
- (2) 创建一个利用负样本增强去雨性能的对比正则化。



(a) The Significance and Difference of Rain Streak Patterns in Frequency Domain



(b) The Global Influence from Frequency Domain

Fig. 2. Motivation. (a) By independently modifying the background content and foreground degraded rain streaks in rainy images, we observe that images with identical rain streaks but different content are relatively similar in the frequency domain, whereas images with different rain streaks but the same content exhibit significant differences. This highlights the Discrete Fourier Transform's (DFT) sensitivity to rain degradation patterns. (b) Replacing local patches in the amplitude components of rain images after DFT has a global impact on the spatial domain after IDFT. This demonstrates the global modeling capacity of frequency domain feature processing.

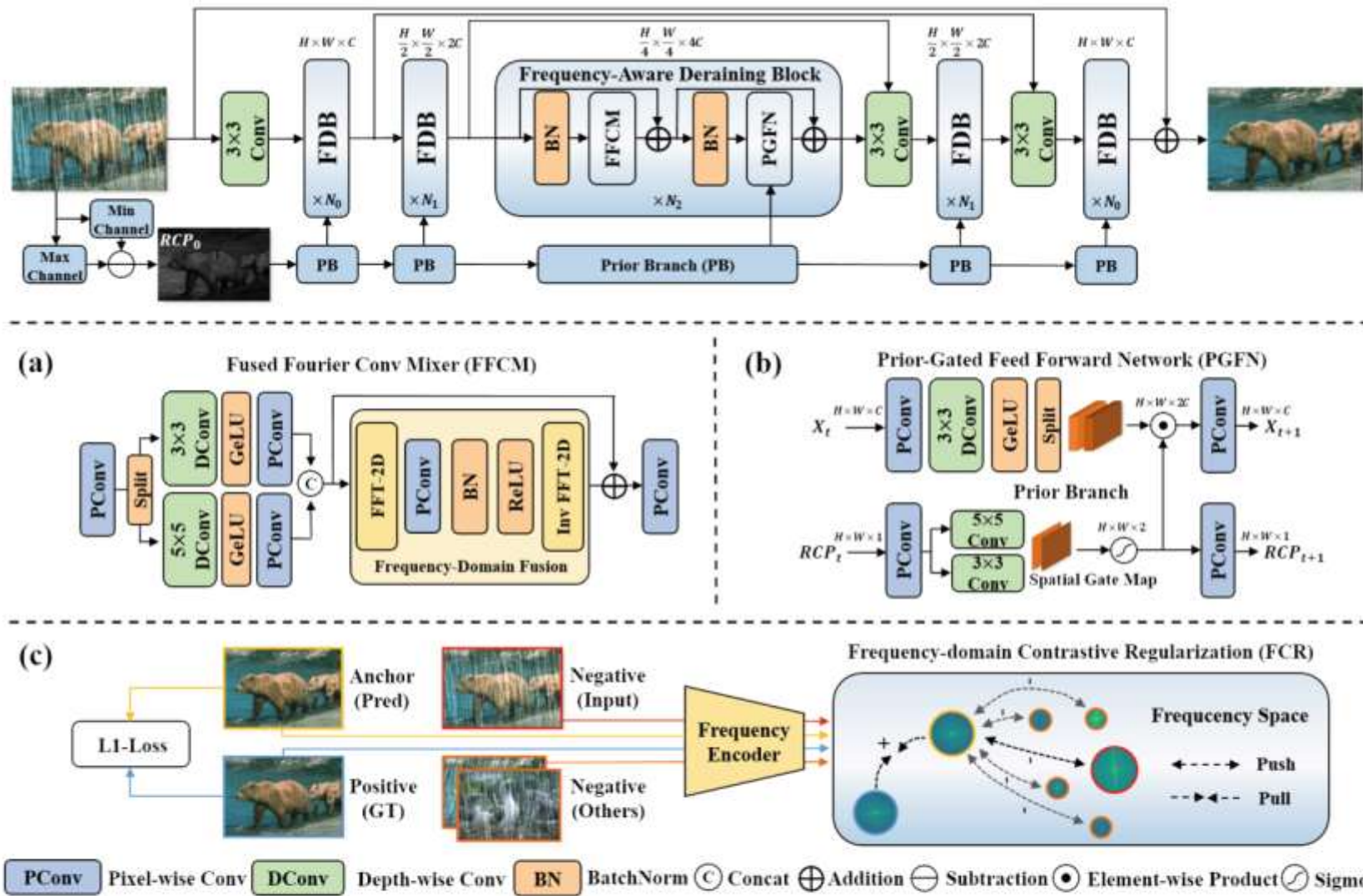


Fig. 3. Illustration of our FADformer framework containing Fused Fourier Convolution Mixer (a, FFCM) and Prior-Gated Feed Forward Network (b, PGFN) with Frequency-domain Contrastive Regularization (c, FCR) for single image deraining.

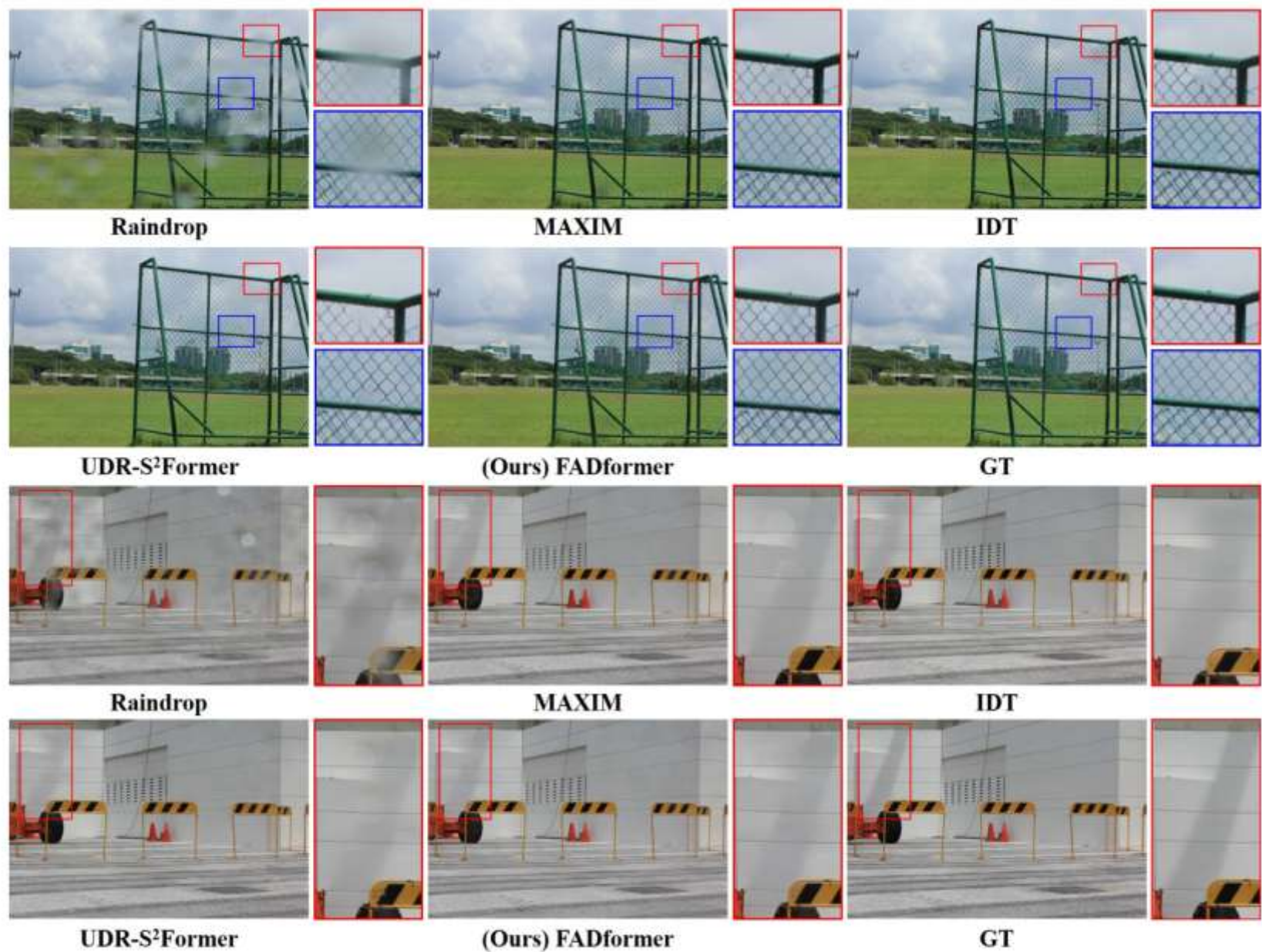


Fig. 1: Visual comparison of Raindrop [18] results.

两种技术结合和互补

■ 模型优化

通常需要对模型进行一些优化，以降低模型的复杂度和计算量，提高模型的速度和效率，适应不同的计算资源和部署场景，如移动设备、边缘计算等。**这些优化方法往往需要借助传统图像处理技术**，如图像的压缩、量化、低秩近似、稀疏编码等，来实现。这些优化方法可以有效地提升人工智能技术的图像处理的可扩展性和可移植性。

THE END
Thank You.

