ISC 2019 第七届互联网安全大会

人工智能及其安全的STEAM教育

黄永洪

码有引力公司创始人、

重庆邮电大学高级工程师

小鹅助理

扫码添加小鹅助理，与数万科技圈人士

分享重量级活动PPT、干货培训课程、高端会议免费

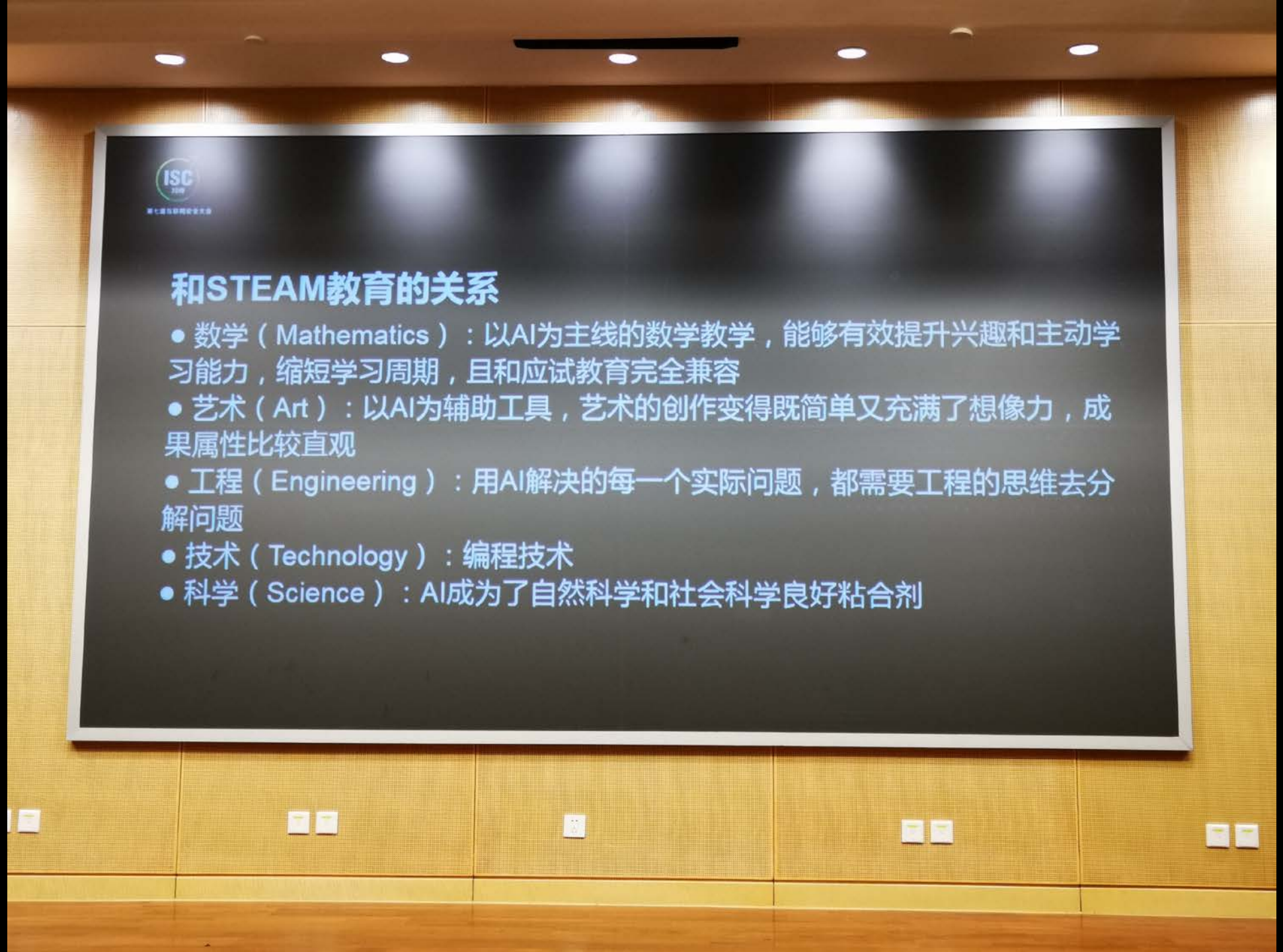
门票

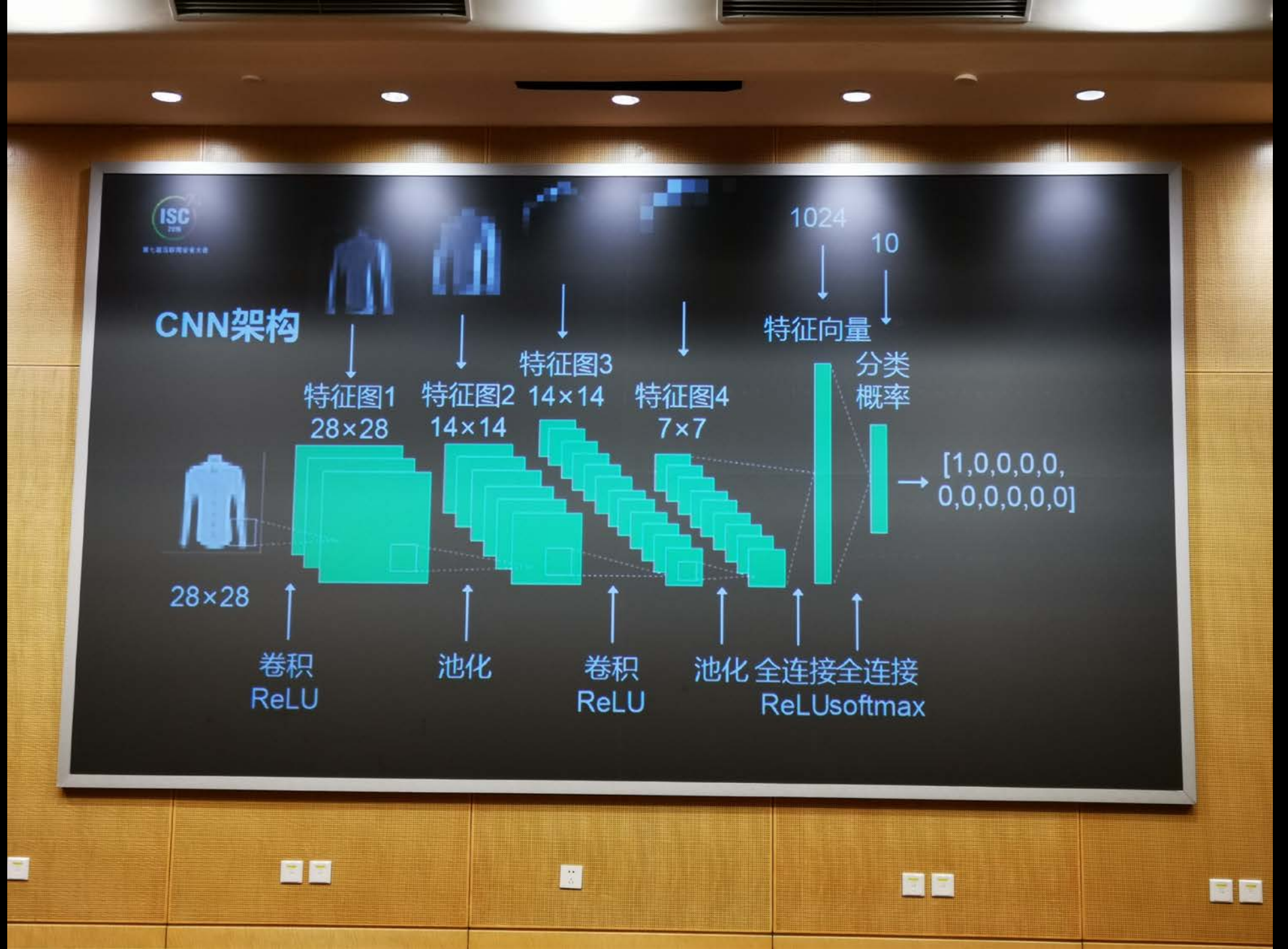

ISC
黄永洪
码有引力公司创始人
重庆邮电大学 高级工程师


ISC
人工智能及其安全的
STEAM教育
重庆邮电大学　黄永洪

人工智能及其安全的几个典型问题

- 神经网络的黑箱，导致了人工智能系统的安全比传统信息系统的安全更加具有不确定性和挑战性，打开黑箱，则是人机协作共同提高的过程
- 逃逸攻击的穷举性与主动防御，具有和免杀相同的特点，而GAN为暴力攻击和对抗训练提供无穷无尽的素材
- GAN数据的非现实性，使得真假辨别成为一件极其困难的事情，也让AI和艺术走得更近
- 从利用L-BFGS到GAN的逃逸，无一不通过距离产生，从L1、L2距离，到KL、JS散度，再到Wasserstein距离，是从具体到抽象的距离。存在更抽象的距离吗？


ISC
和STEAM教育的关系
● 数学（Mathematics）：以AI为主线的数学教学，能够有效提升兴趣和主动学习能力，缩短学习周期，且和应试教育完全兼容
● 艺术（Art）：以AI为辅助工具，艺术的创作变得既简单又充满了想像力，成果属性比较直观
● 工程（Engineering）：用AI解决的每一个实际问题，都需要工程的思维去分解问题
● 技术（Technology）：编程技术
● 科学（Science）：AI成为了自然科学和社会科学良好粘合剂


CNN架构
特征图1
28×28
特征图2
14×14
特征图3
14×14
特征图4
7×7
1024
10
特征向量
分类
概率
[1,0,0,0,0,
0,0,0,0,0,0]
28×28
卷积
ReLU
池化
卷积
ReLU
池化
全连接
ReLU
全连接
softmax

CNN – 卷积

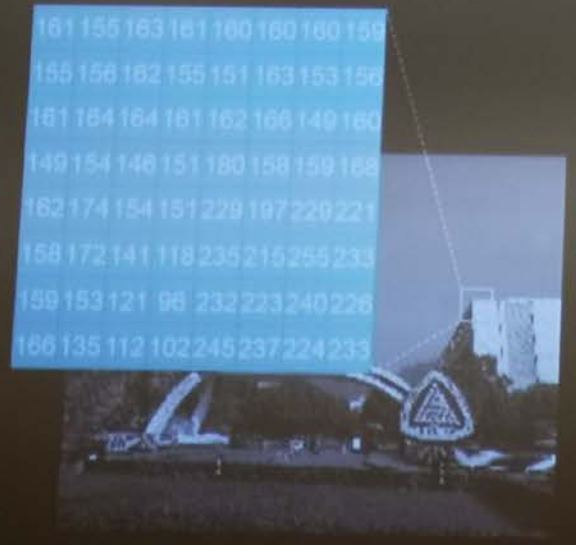
Filter 1

1	1	1
0	0	0
-1	-1	-1

*

=

10	10	-3	-8	3	4
24	22	-9	-20	-30	-13
-1	10	-47	-88	-178	-172
-22	20	-17	-79	-208	-216
57	109	86	26	-40	-42
58	82	35	-16	-1	9



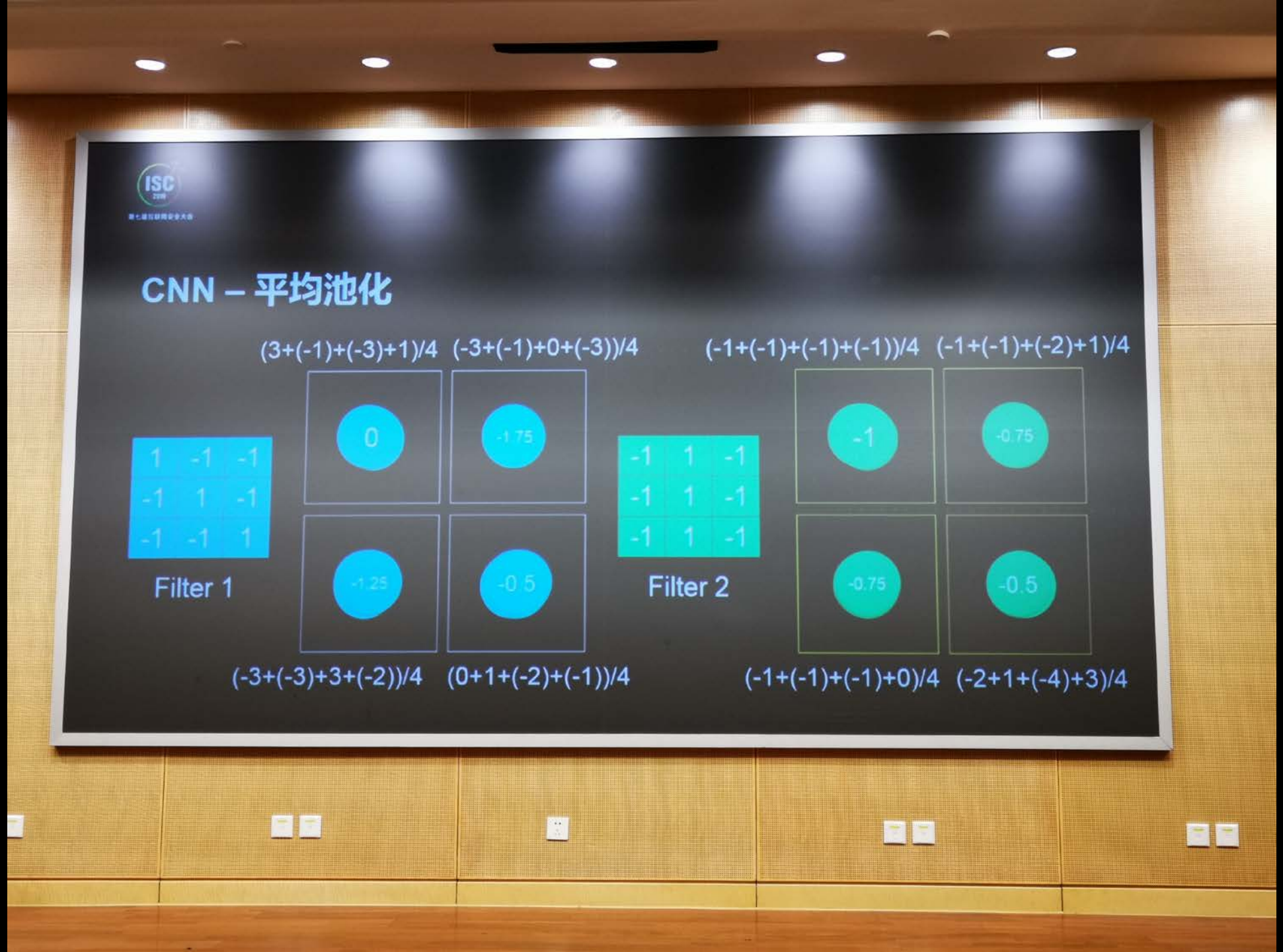
Filter 2

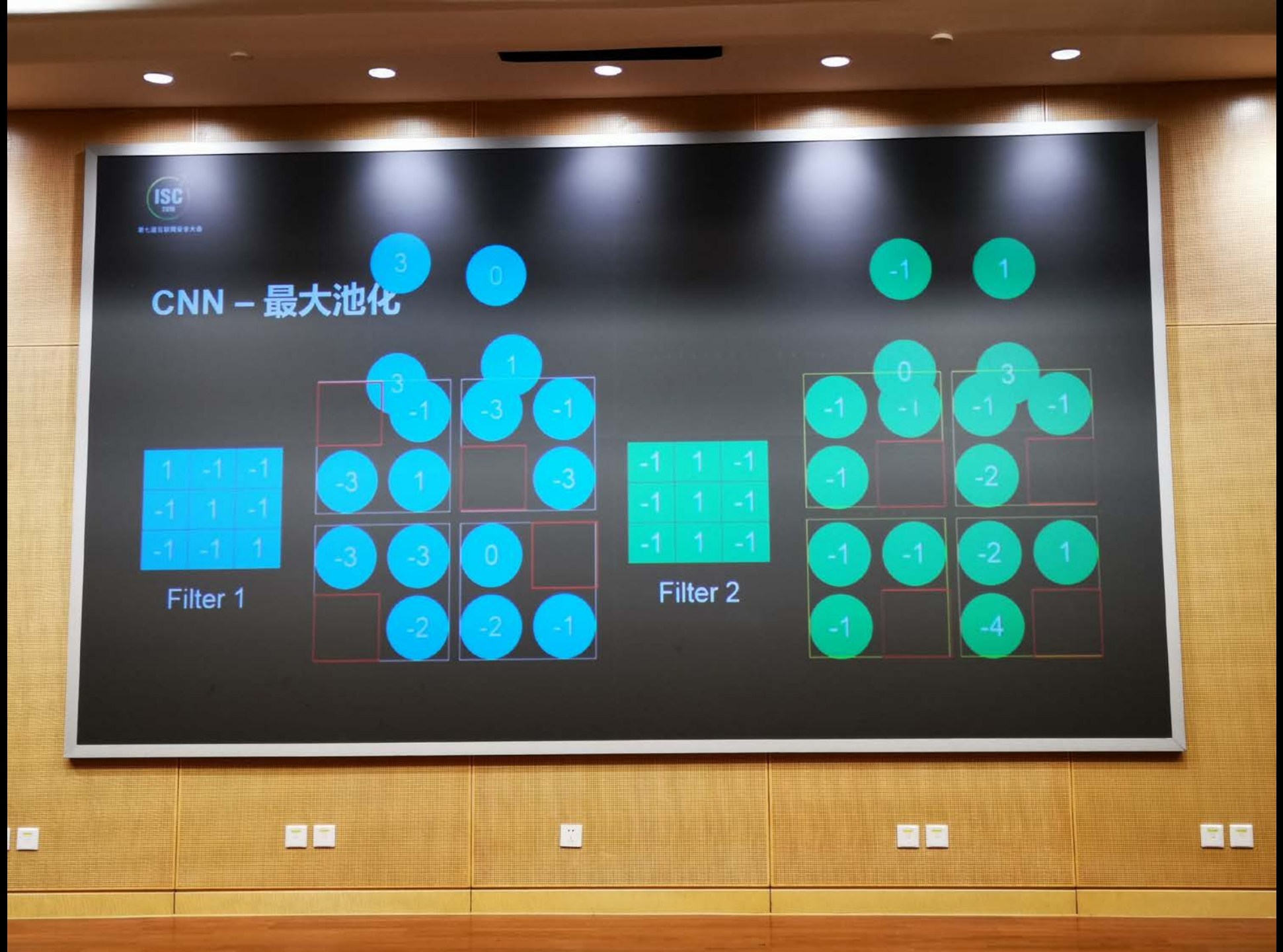
1	0	-1
1	0	-1
1	0	-1

*

=

-12	-2	16	-12	11	14
-7	7	-21	-20	32	3
8	29	-107	-58	34	-28
28	80	-203	-150	1	-52
63	134	-280	-270	-28	-45
109	144	-338	-359	-7	-17


CNN – 平均池化
(3+(-1)+(-3)+1)/4
(-3+(-1)+0+(-3))/4
(-1+(-1)+(-1)+(-1))/4
(-1+(-1)+(-2)+1)/4
1 -1 -1
-1 1 -1
-1 -1 1
Filter 1
0
-1.75
-1.25
-0.5
-1 1 -1
-1 1 -1
-1 1 -1
Filter 2
-1
-0.75
-0.75
-0.5
(-3+(-3)+3+(-2))/4
(0+1+(-2)+(-1))/4
(-1+(-1)+(-1)+0)/4
(-2+1+(-4)+3)/4


CNN – 最大池化
Filter 1
Filter 2


ISC
扁平化
-1
1
0
3
扁平化
3
0
3
1
-1
1
0
3

以人工智能为主线学数学

小学1~3年级
四则混合运算（卷积运算）
分数
平均数（平均池化）

小学4~6年级：

$\left(1+\frac{1}{n}\right)^n$（e的初始理解）

$E=\frac{1}{2}\sum_k(y_k-t_k)^2$（梯度下降）

初中1~3年级：

π、e、无理数

$f(x)=\begin{cases}0 & x\le 0,\\ 1 & x>0\end{cases}$（阶跃）

$f(x)=\begin{cases}0 & x\le 0,\\ x & x>0\end{cases}$（ReLU激活函数）

高中1~3年级：

$f(x)=\frac{1}{1+e^{-x}}$（sigmoid激活函数）

$(1\quad 2)\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix}=(9\quad 12\quad 15)$（矩阵运算）

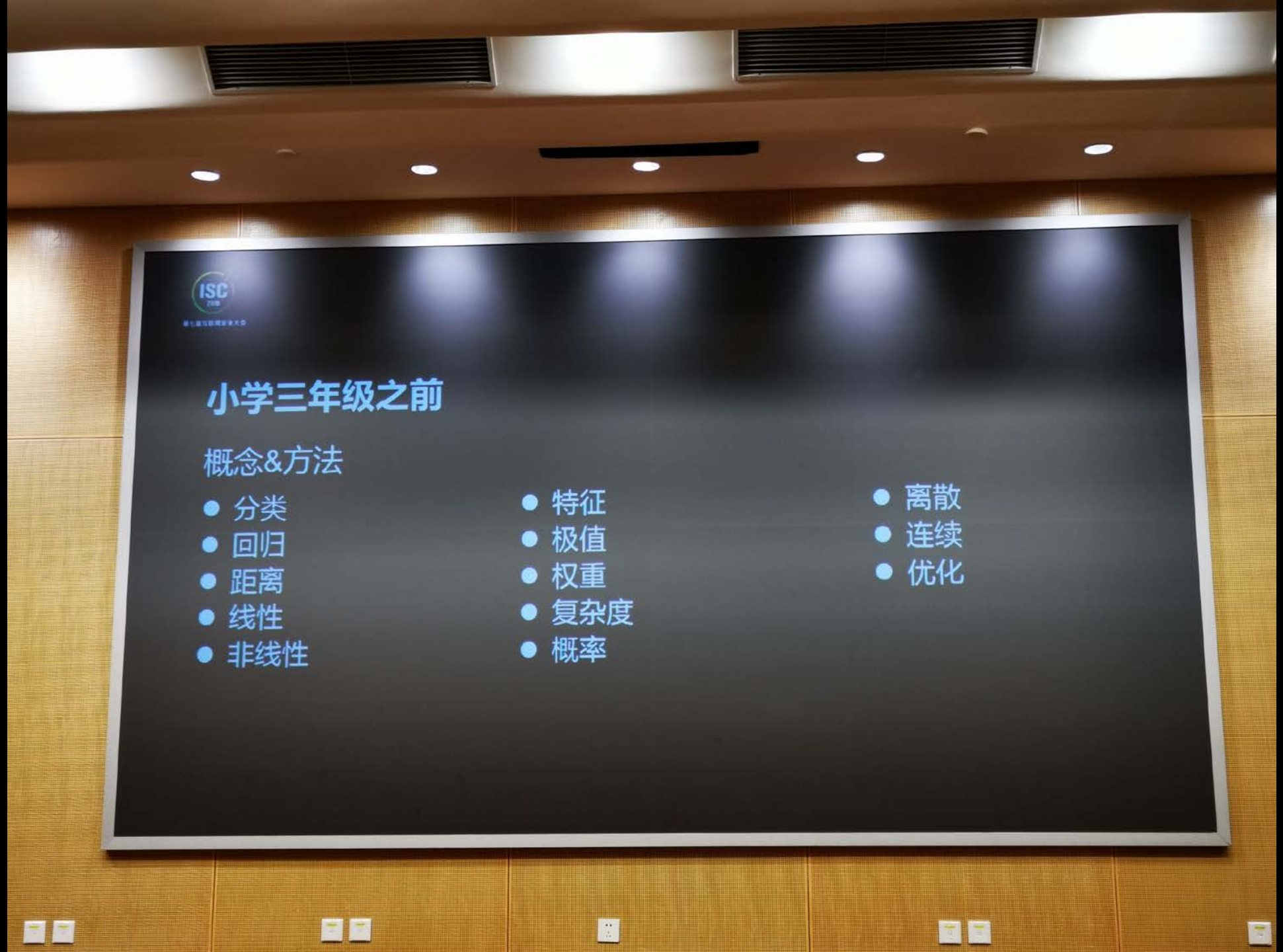
$\frac{df(x)}{dx}=\lim_{n\to 0}\frac{f(hx_1)-f(x)}{h}$（求导）

$y_k=\frac{e^{a_k}}{\sum_{k=1}^{n}e^{a_k}}$（softmax）

$e=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$（e的定义）

$E=-t_k\sum\log y_k$（梯度下降）

$D_{KL}(p||q)=\sum_{i=1}^{N}p(x_i)\cdot\log\frac{p(x_i)}{q(x_i)}$（GAN）


ISC
小学三年级之前
概念&方法
分类
回归
距离
线性
非线性
特征
极值
权重
复杂度
概率
离散
连续
优化

和编程的结合

人工智能的编程语言

- 程序设计，和使用计算机要解决的问题紧密相关。不系统学习操作系统、编译、网络，以及各种协议，能解决的问题有限。
- 小中阶段，以人工智能为载体，主要是理解程序的结构，选择，循环以及数据的表达和简单的输入输出。算法是很容易变为程序语言的。
- 习惯数学、算法与程序的相互转换

人工智能+数学+程序

卷积

池化

$$y(t) = \int_{-\infty}^{\infty} x(p)h(t-p)dp = x(t) * h(t)$$

$$S = \beta down(C) + b$$

```
def _convolve(img,fil):
    fil_heigh = fil.shape[0]                    #获取卷积核(滤波)的高度
    fil_width = fil.shape[1]                    #获取卷积核(滤波)的宽度

    conv_heigh = img.shape[0] - fil.shape[0] + 1    #确定卷积结果
的大小
    conv_width = img.shape[1] - fil.shape[1] + 1
    conv = np.zeros((conv_heigh,conv_width),dtype = 'uint8')

    for i in range(conv_heigh):
        for j in range(conv_width):    #逐点相乘并求和得到每一个点
            conv[i][j] = wise_element_sum(img[i:i + fil_heigh,j:j +
fil_width ],fil)
    return conv


def max_pooling_forward(x,pool_param):
    out=None
    N, C, H, W = x.shape
    HH, WW, stride = pool_param['pool_height'],
pool_param['pool_width'], pool_param['stride']
    H_out = (H-HH)/stride+1
    W_out = (W-WW)/stride+1
    out = np.zeros((N,C,H_out,W_out))
    for i in xrange(H_out):
        for j in xrange(W_out):
            x_mask=x[:,:,i*stride : i*stride+HH, j*stride : j*stride+WW]
            out[ :, :,i,j] = np.max(x_mask, axis=(2,3)) #中值池化的话，
就是np.median

    cache=(x,pool_param)
    return out,cache
```

和应试教育的兼容性

小升初考题：

将连续的奇数1, 3, 5, 7, 9, 11、、、按5个一行排成如下的数表：

（1）十字框中的五个数的平均数与中间数有什么关系？

（2）若将十字框上下左右平移，可框住另外的五个数，这五个数的和能等于2011吗？能等于2015吗？能等于2045吗？若能，请求出这五个数；若不能，请说明理由。

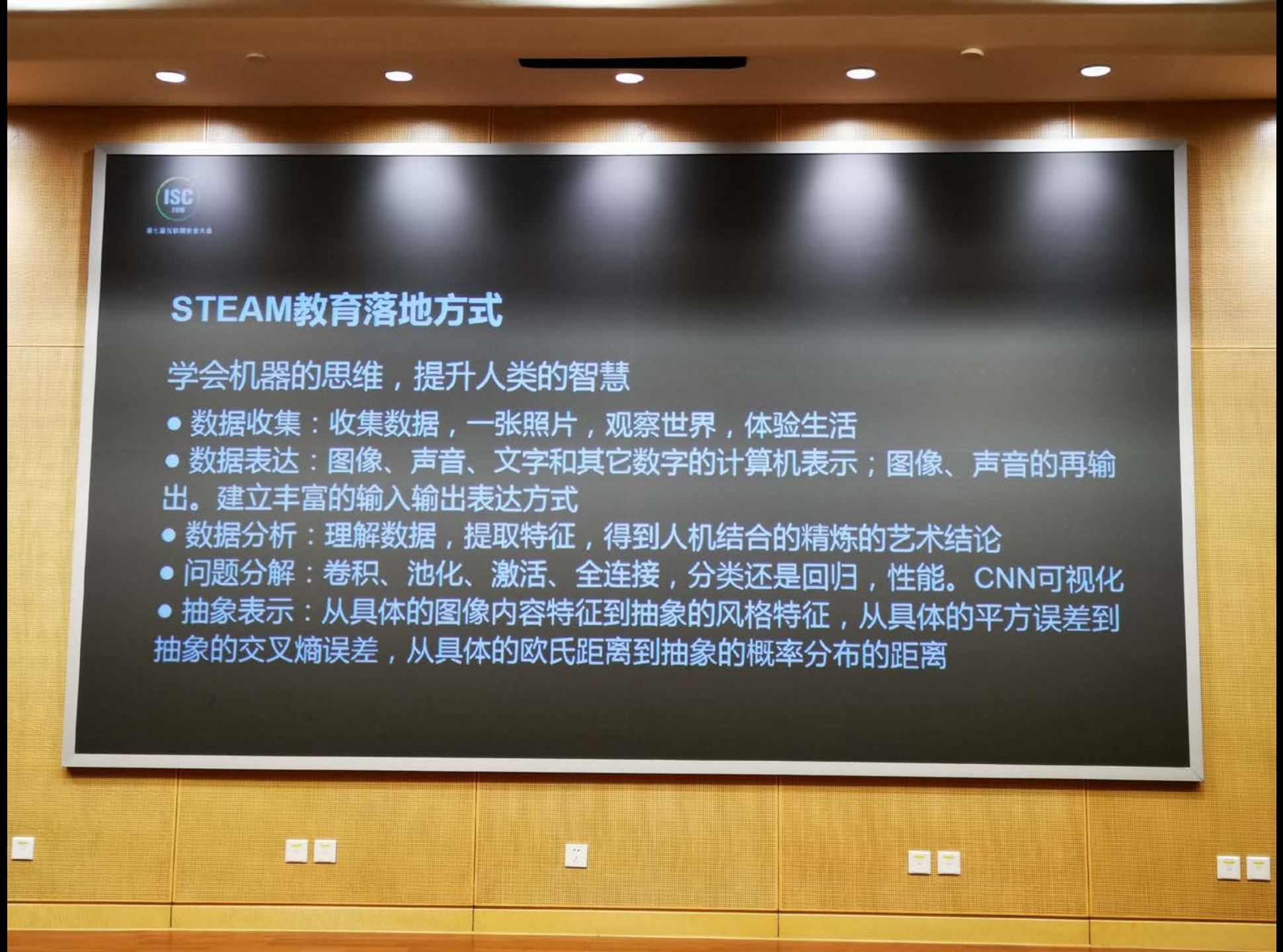
1	3	5	7	9
11	13	15	17	19
21	23	25	27	29
31	33	35	37	39
…	…	…	…	…

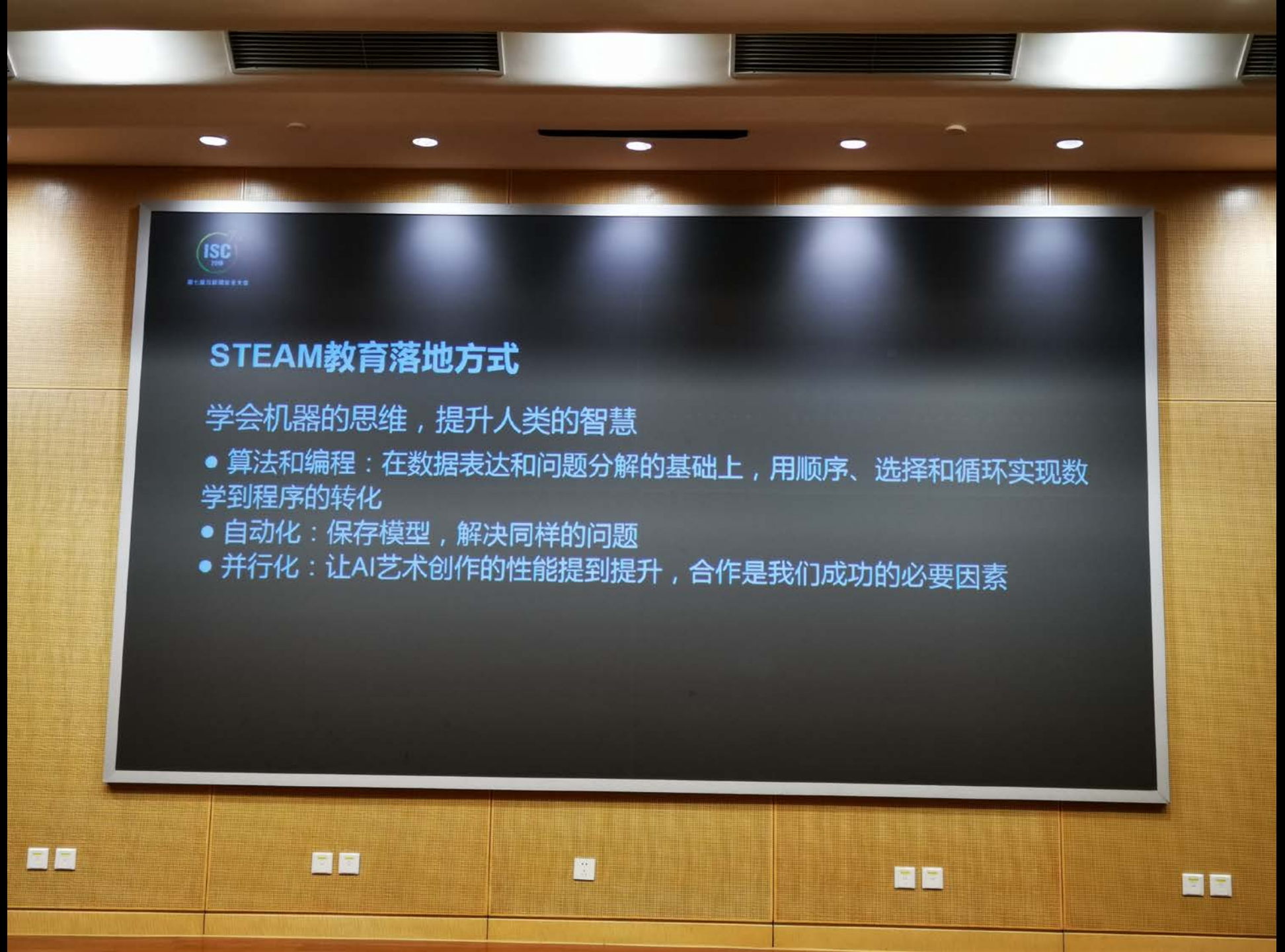
中考题：

已知x_1、x_2、y_1、y_2属于实数，满足：

$$\begin{cases} x_1^2 + y_1^2 = 1 \\ x_2^2 + y_2^2 = 1 \\ x_1x_2 + y_1y_2 = \dfrac{1}{2} \end{cases}$$

求$\dfrac{|x_1 + y_1 - 1|}{\sqrt{2}} + \dfrac{|x_2 + y_2 - 1|}{\sqrt{2}}$的最大值。


ISC
2019
第七届互联网安全大会
STEAM教育落地方式
学会机器的思维，提升人类的智慧
● 数据收集：收集数据，一张照片，观察世界，体验生活
● 数据表达：图像、声音、文字和其它数字的计算机表示；图像、声音的再输出。建立丰富的输入输出表达方式
● 数据分析：理解数据，提取特征，得到人机结合的精炼的艺术结论
● 问题分解：卷积、池化、激活、全连接，分类还是回归，性能。CNN可视化
● 抽象表示：从具体的图像内容特征到抽象的风格特征，从具体的平方误差到抽象的交叉熵误差，从具体的欧氏距离到抽象的概率分布的距离


ISC
STEAM教育落地方式
学会机器的思维，提升人类的智慧
• 算法和编程：在数据表达和问题分解的基础上，用顺序、选择和循环实现数学到程序的转化
• 自动化：保存模型，解决同样的问题
• 并行化：让AI艺术创作的性能提到提升，合作是我们成功的必要因素

小鹅助理

谢谢！

扫码添加小鹅助理，与数万科技圈人士

分享重量级活动PPT、干货培训课程、高端会议免费门票