

企业/机构致谢

- 本课程由【IBM】授权提供，选自IBM AI启蒙季系列课程。在此，特别致谢IBM对“云支教”助学计划的支持，以及对乡村儿童教育发展所做出的贡献。

什么是数字化

云支教全国标准课程1.0版



CONTENTS



01

二进制



02

数字化模型



03

数据处理

二进制

1997年

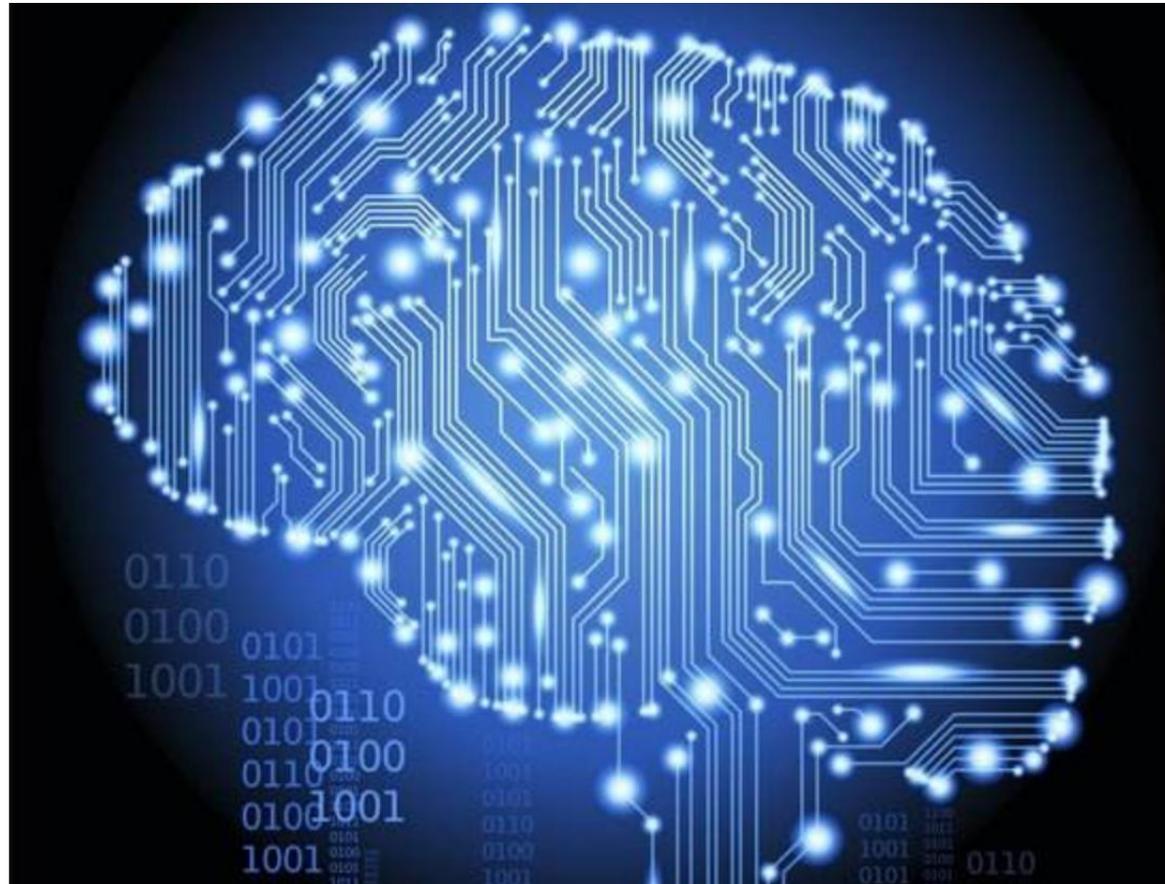
IBM 深蓝计算机
获得
国际象棋冠军



1997年5月11日，一台名为“深蓝”的超级电脑将棋盘上的一个兵走到C4位置时，人类有史以来最伟大的国际象棋名家卡斯帕罗夫不得不沮丧地承认自己输了。世纪末的一场人机大战终于以计算机的微弱优势取胜。

什么叫数字化

数字化就是将复杂多变的信息转变为可以度量的数字、数据，再以它们建立起数字化模型，转变为一系列二进制代码，引入计算机内部，进行统一处理，这就是数字化的基本过程。



十进制

$$\begin{array}{r} 198 \\ + \quad 8 \\ \hline 206 \end{array}$$

The diagram shows the addition of 198 and 8. A horizontal line is drawn under the 8. Below the line, the digit '1' is placed under the '9' and another '1' is placed under the '8'. The result, 206, is written below the line.

0, 1, 2, 3, ..., 8, 9

10, 11, 12, 13, ..., 18, 19

20, 21, 22, 23, ..., 28, 29

30, 31, 32, ...

...

50, 51, ...

⋮

100, 101, 102, ...

⋮

1000, 1001, ...

⋮

100000, 100001, 100002, ...

⋮

二进制

把信息用1和0表示，然后用计算机识别和处理

二进制

0000 0000->0

0000 0001->1

0000 0010->2

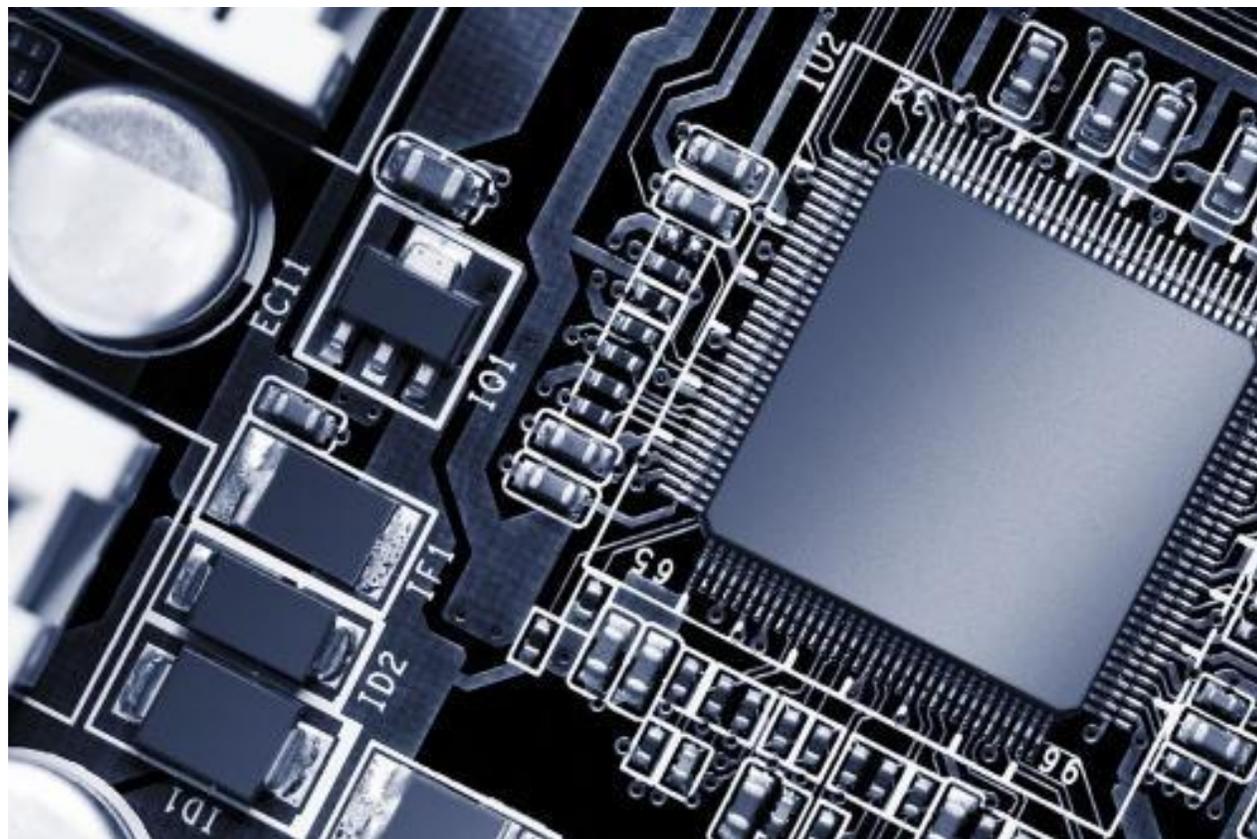
0000 0011->3

0000 0100->4

0000 0101->5

...

思考题：
想想十进制的8在二进制中如何表示？



一个字节是什么？

- 1. 字节是二进制数据的单位
- 2. 一个字节通常8位长
- 3. 字节通常简写为“B”，而位通常简写为小写“b”
- 4. 一个位就代表一个0或1（即二进制），每8个位（bit，简写为b）组成一个字节（Byte，简写为B），是最小一级的信息单位
- 5. 计算机存储器的大小通常用字节来表示

B: Byte, “字节”
b: bit, “位”, 又名“比特”



思考题：用1个字节最多可以表示多少个二进制数据？

启发：从0000 0000 到1111 1111

哪些可以被数字化?

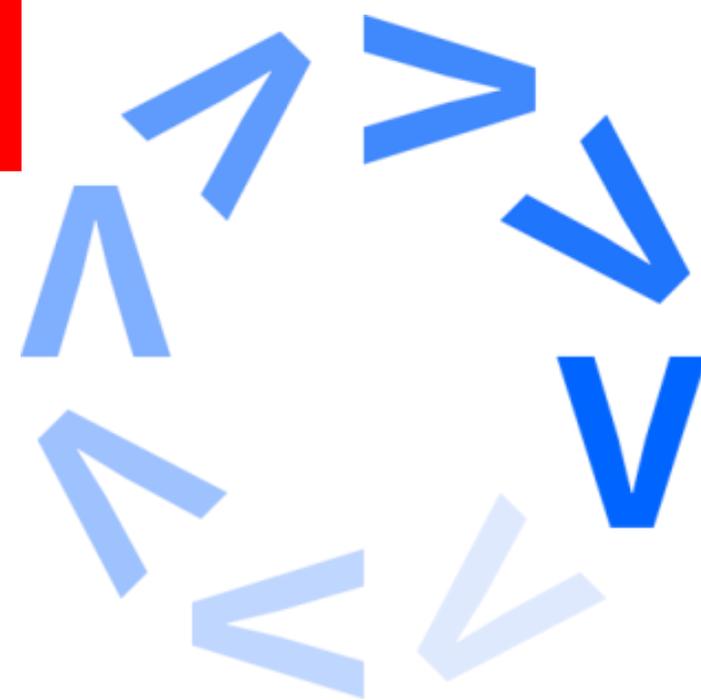
文字

图像

声音

视频

空间结构



文本的数字化

找一找

0100 1001
0100 0010
0100 1101

ASCII 字符代码表 一

高四位 低四位		ASCII非打印控制字符										ASCII 打印字符											
		0000					0001					0010	0011		0100	0101		0110		0111			
		0					1					2	3		4	5		6		7			
十进制	字符	ctrl	代码	字符解释	十进制	字符	ctrl	代码	字符解释	十进制	字符	十进制	字符	十进制	字符	十进制	字符	十进制	字符	ctrl			
0000	0	0	BLANK NULL	^@	NUL	空	16	▶	^P	DLE	数据链路转意	32		48	0	64	@	80	P	96	`	112	p
0001	1	1	☺	^A	SOH	头标开始	17	◀	^Q	DC1	设备控制 1	33	!	49	1	65	A	81	Q	97	a	113	q
0010	2	2	☹	^B	STX	正文开始	18	↕	^R	DC2	设备控制 2	34	"	50	2	66	B	82	R	98	b	114	r
0011	3	3	♥	^C	ETX	正文结束	19	!!	^S	DC3	设备控制 3	35	#	51	3	67	C	83	S	99	c	115	s
0100	4	4	♦	^D	EOF	传输结束	20	¶	^T	DC4	设备控制 4	36	\$	52	4	68	D	84	T	100	d	116	t
0101	5	5	♣	^E	ENQ	查询	21	§	^U	NAK	反确认	37	%	53	5	69	E	85	U	101	e	117	u
0110	6	6	♠	^F	ACK	确认	22	■	^V	SYN	同步空闲	38	&	54	6	70	F	86	V	102	f	118	v
0111	7	7	●	^G	BEL	震铃	23	↑↓	^W	ETB	传输块结束	39	'	55	7	71	G	87	w	103	g	119	w
1000	8	8	◻	^H	BS	退格	24	↑	^X	CAN	取消	40	(56	8	72	H	88	X	104	h	120	x
1001	9	9	○	^I	TAB	水平制表符	25	↓	^Y	EM	媒体结束	41)	57	9	73	I	89	Y	105	i	121	y
1010	A	10	◻	^J	LF	换行/新行	26	→	^Z	SUB	替换	42	*	58	:	74	J	90	Z	106	j	122	z
1011	B	11	♂	^K	VT	竖直制表符	27	←	^[ESC	转意	43	+	59	;	75	K	91	[107	k	123	{
1100	C	12	♀	^L	FF	换页/新页	28	└	^\ FS	文件分隔符	44	,	60	<	76	L	92	\	108	l	124		
1101	D	13	🎵	^M	CR	回车	29	↔	^] GS	组分隔符	45	-	61	=	77	M	93]	109	m	125	}	
1110	E	14	🎵	^N	SO	移出	30	▲	^_	RS	记录分隔符	46	.	62	>	78	N	94	^	110	n	126	~
1111	F	15	⚙	^O	SI	移入	31	▼	^-	US	单元分隔符	47	/	63	?	79	O	95	_	111	o	127	Δ

注：表中的ASCII字符可以用:ALT + “小键盘上的数字键”输入

文本的数字化

汉字及对应的 unicode 码														GB 码范围			
啊:21834	阿:38463	埃:22467	挨:25384	哎:21710	唉:21769	哀:21696	皑:30353	癌:30284	蔼:34108	矮:30702	艾:33406	碍:30861	爱:29233	隘:38552	鞍:38797	B0A1-	B0B0
氨:27688	安:23433	俺:20474	按:25353	暗:26263	岸:23736	腋:33018	案:26696	肮:32942	昂:26114	盎:30414	凹:20985	敖:25942	熬:29100	翱:32753	袄:34948	B0B1-	B0C0
傲:20658	奥:22885	懊:25034	澳:28595	芭:33453	捌:25420	扒:25170	叭:21485	吧:21543	芭:31494	八:20843	疤:30116	巴:24052	拔:25300	跋:36299	靶:38774	B0C1-	B0D0
把:25226	耙:32793	坝:22365	霸:38712	罢:32610	爸:29240	白:30333	柏:26575	百:30334	摆:25670	佰:20336	败:36133	拜:25308	裨:31255	斑:26001	班:29677	B0D1-	B0E0
搬:25644	扳:25203	般:33324	颁:39041	板:26495	版:29256	扮:25198	拌:25292	伴:20276	辮:29923	半:21322	办:21150	绊:32458	邦:37030	帮:24110	梆:26758	B0E1-	B0F0
榜:27036	膀:33152	绑:32465	棒:26834	磅:30917	蚌:34444	镑:38225	傍:20621	谤:35876	苞:33502	胞:32990	包:21253	褒:35090	剥:21093			B0F1-	B0FE
薄:34180	雹:38649	保:20445	堡:22561	饱:39281	宝:23453	抱:25265	报:25253	暴:26292	豹:35961	鲍:40077	爆:29190	杯:26479	碑:30865	悲:24754	卑:21329	B1A1-	B1B0
北:21271	辈:36744	背:32972	贝:36125	狈:38049	倍:20493	狈:29384	备:22791	惫:24811	焙:28953	被:34987	奔:22868	笨:33519	本:26412	笨:31528	崩:23849	B1B1-	B1C0
緹:32503	甬:29997	泵:27893	躅:36454	迸:36856	逼:36924	鼻:40763	比:27604	鄙:37145	笔:31508	彼:24444	碧:30887	蓖:34006	蔽:34109	毕:27605	毙:27609	B1C1-	B1D0
恣:27606	币:24065	庇:24199	痹:30201	闭:38381	敝:25949	弊:24330	必:24517	辟:36767	壁:22721	臂:33218	避:36991	陛:38491	鞭:38829	边:36793	编:32534	B1D1-	B1E0
贬:36140	扁:25153	便:20415	变:21464	下:21342	辨:36776	辩:36777	辩:36779	遍:36941	标:26631	彪:24426	膘:33176	表:34920	鳖:40150	憋:24971	别:21035	B1E1-	B1F0
瘴:30250	彬:24428	斌:25996	濒:28626	滨:28392	宾:23486	摈:25672	兵:20853	冰:20912	柄:26564	丙:19993	乘:31177	饼:39292	炳:28851			B1F1-	B1FE
病:30149	并:24182	玻:29627	菠:33760	播:25773	拨:25320	钵:38069	波:27874	博:21338	勃:21187	搏:25615	铂:38082	箔:31636	伯:20271	帛:24091	舶:33334	B2A1-	B2B0
脖:33046	膊:33162	渤:28196	泊:27850	驳:39539	捕:25429	卜:21340	哺:21754	补:34917	埠:22496	不:19981	布:24067	步:27493	簿:31807	部:37096	怖:24598	B2B1-	B2C0
擦:25830	猜:29468	裁:35009	材:26448	才:25165	财:36130	睬:30572	睬:36393	采:37319	彩:24425	菜:33756	蔡:34081	餐:39184	参:21442	蚕:34453	残:27531	B2C1-	B2D0
惭:24813	惨:24808	灿:28799	苍:33485	舱:33329	仓:20179	沧:27815	藏:34255	操:25805	糙:31961	槽:27133	曹:26361	草:33609	厕:21397	策:31574	侧:20391	B2D1-	B2E0
册:20876	测:27979	层:23618	蹭:36461	插:25554	叉:21449	茬:33580	茶:33590	查:26597	楂:30900	搽:25661	察:23519	岔:23700	差:24046	说:35815	拆:25286	B2E1-	B2F0
柴:26612	豺:35962	搀:25600	掺:25530	蝉:34633	馋:39307	谗:35863	缠:32544	铲:38130	产:20135	阐:38416	颤:39076	昌:26124	猖:29462			B2F1-	B2FE
场:22330	尝:23581	常:24120	长:38271	偿:20607	肠:32928	厂:21378	敞:25950	畅:30021	唱:21809	倡:20513	超:36229	抄:25220	钞:38046	朝:26397	嘲:22066	B3A1-	B3B0
潮:28526	巢:24034	吵:21557	炒:28818	车:36710	扯:25199	撤:25764	掣:25507	彻:24443	澈:28552	郴:37108	臣:33251	辰:36784	尘:23576	晨:26216	忱:24561	B3B1-	B3C0
沉:27785	陈:38472	趁:36225	衬:34924	撑:25745	称:31216	城:22478	橙:27225	成:25104	呈:21576	乘:20056	程:31243	惩:24809	澄:28548	诚:35802	承:25215	B3C1-	B3D0
逞:36894	骋:39563	秤:31204	吃:21507	痴:30196	持:25345	匙:21273	池:27744	迟:36831	弛:24347	驰:39536	耻:32827	齿:40831	侈:20360	尺:23610	赤:36196	B3D1-	B3E0
翅:32709	斥:26021	炽:28861	充:20805	冲:20914	虫:34411	崇:23815	宠:23456	抽:25277	酬:37228	畴:30068	踌:36364	稠:31264	愁:24833	筹:31609	仇:20167	B3E1-	B3F0
绸:32504	瞅:30597	丑:19985	臭:33261	初:21021	出:20986	橱:27249	厨:21416	蹒:36487	锄:38148	雏:38607	滁:28353	除:38500	楚:26970			B3F1-	B3FE
础:30784	储:20648	蠢:30679	搐:25616	触:35302	处:22788	揣:25571	川:24029	穿:31359	椽:26941	传:20256	船:33337	喘:21912	串:20018	疮:30126	窗:31383	B4A1-	B4B0
幢:24162	床:24202	闯:38383	创:21019	吹:21561	炊:28810	捶:25462	锤:38180	垂:22402	春:26149	椿:26943	醇:37255	唇:21767	淳:28147	纯:32431	蠢:34850	B4B1-	B4C0
戳:25139	绰:32496	疵:30133	茨:33576	磁:30913	雌:38604	辞:36766	慈:24904	瓷:29943	词:35789	此:27492	刺:21050	赐:36176	次:27425	聪:32874	葱:33905	B4C1-	B4D0
囱:22257	匆:21254	从:20174	丛:19995	凑:20945	粗:31895	醋:37259	簇:31751	促:20419	蹕:36479	篡:31713	窜:31388	摧:25703	崔:23828	催:20652	脆:33030	B4D1-	B4E0
瘁:30209	粹:31929	淬:28140	翠:32736	村:26449	存:23384	寸:23544	磋:30923	撮:25774	搓:25619	措:25514	挫:25387	错:38169	搭:25645	达:36798	答:31572	B4E1-	B4F0

图片的数字化

二值图：只用两种状态来表示图像——

0 / 1

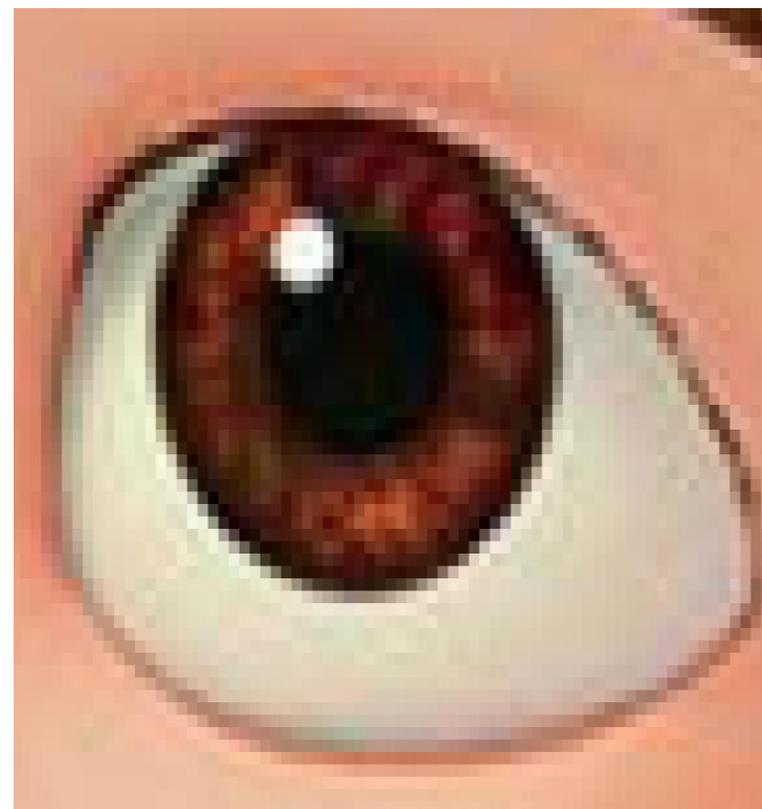
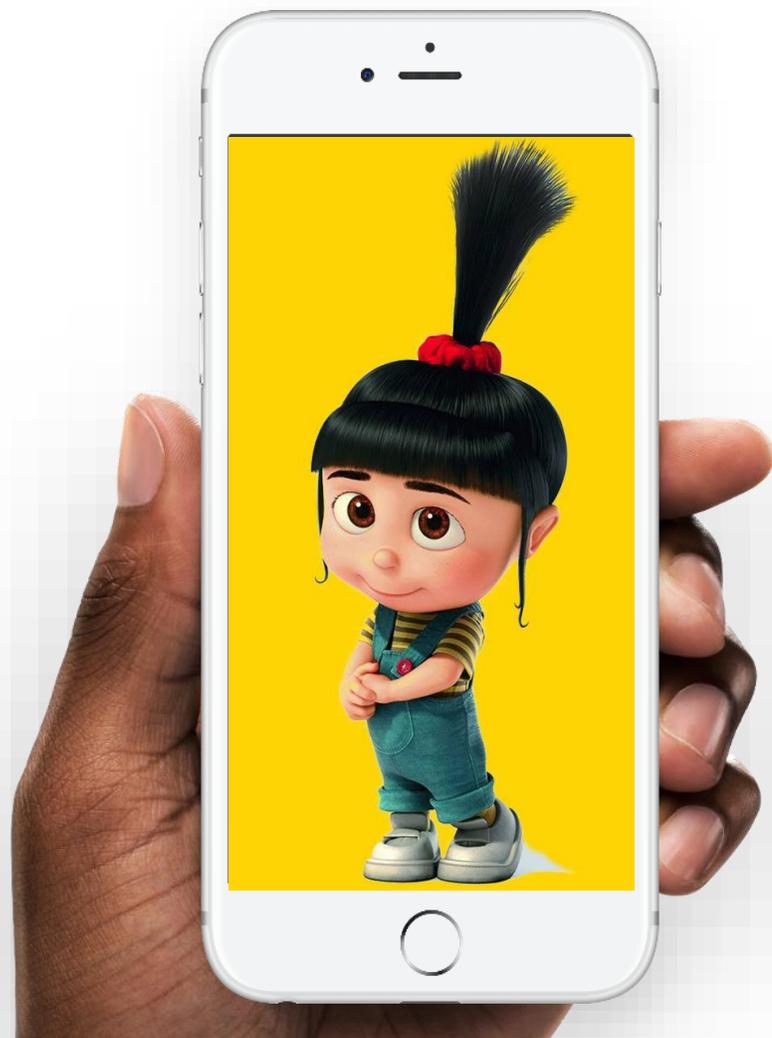
黑 / 白

红 / 蓝

...

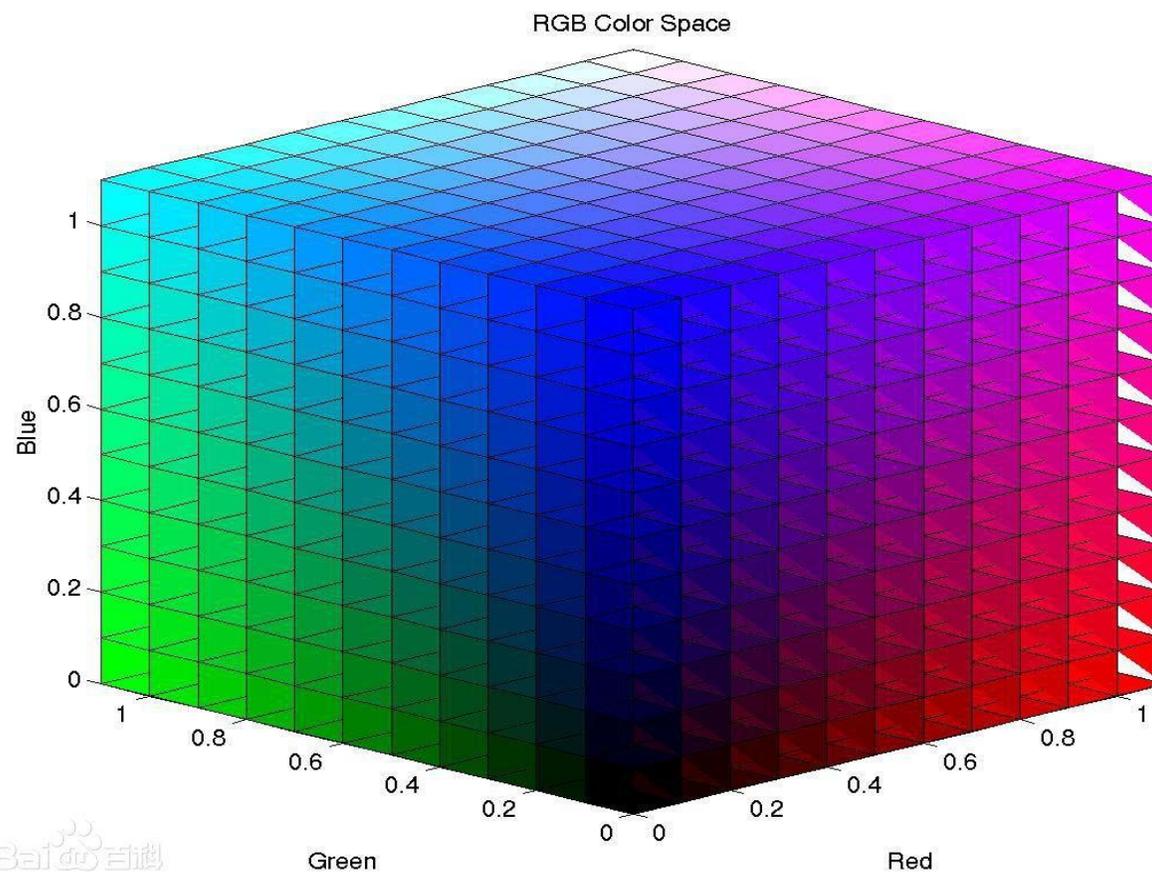


数字化世界



图像的数字化 之 三原色

三原色：R 红 G 绿 B 蓝
RGB 三原色以任意亮度组合，可以生成肉眼可见的所有颜色



三原色的组合可以形成上千万种颜色

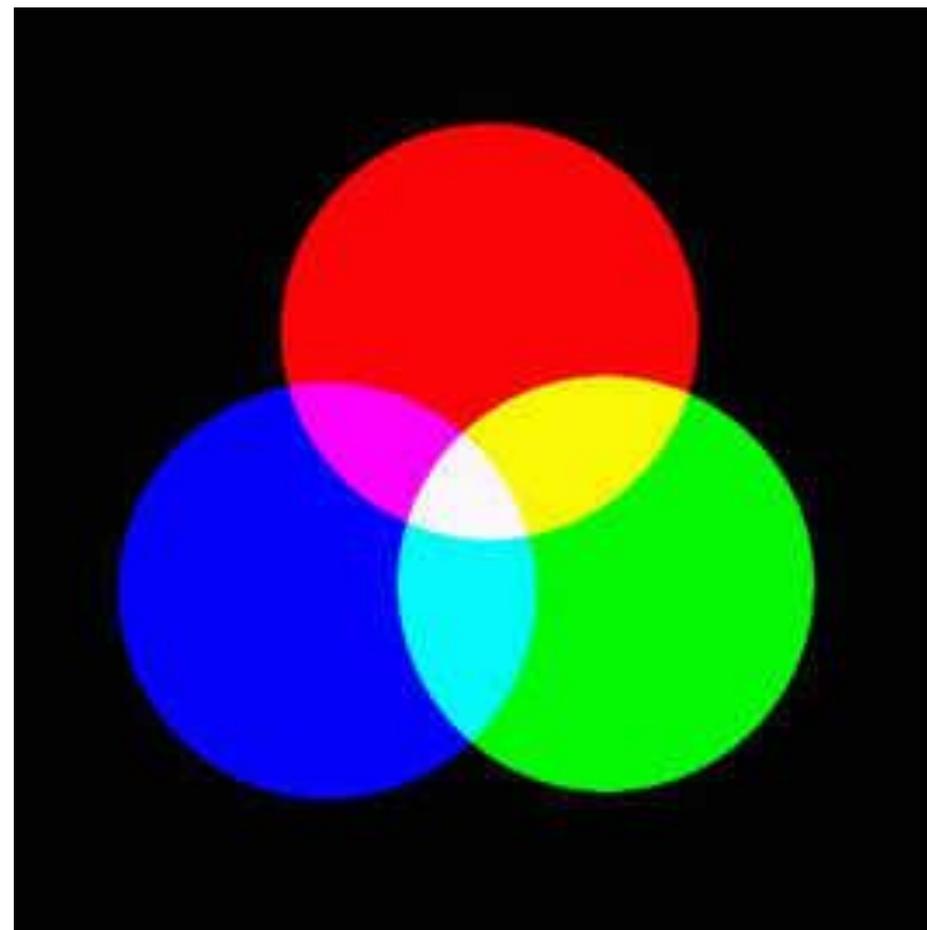
红色 Red、绿色 Green 和蓝色 Blue，每个都由一个字节（8位比特）来表示

0000 0000 -> 0
0000 0001 -> 1
...
1111 1111 -> 255

} 一个字节可以表示256个数字

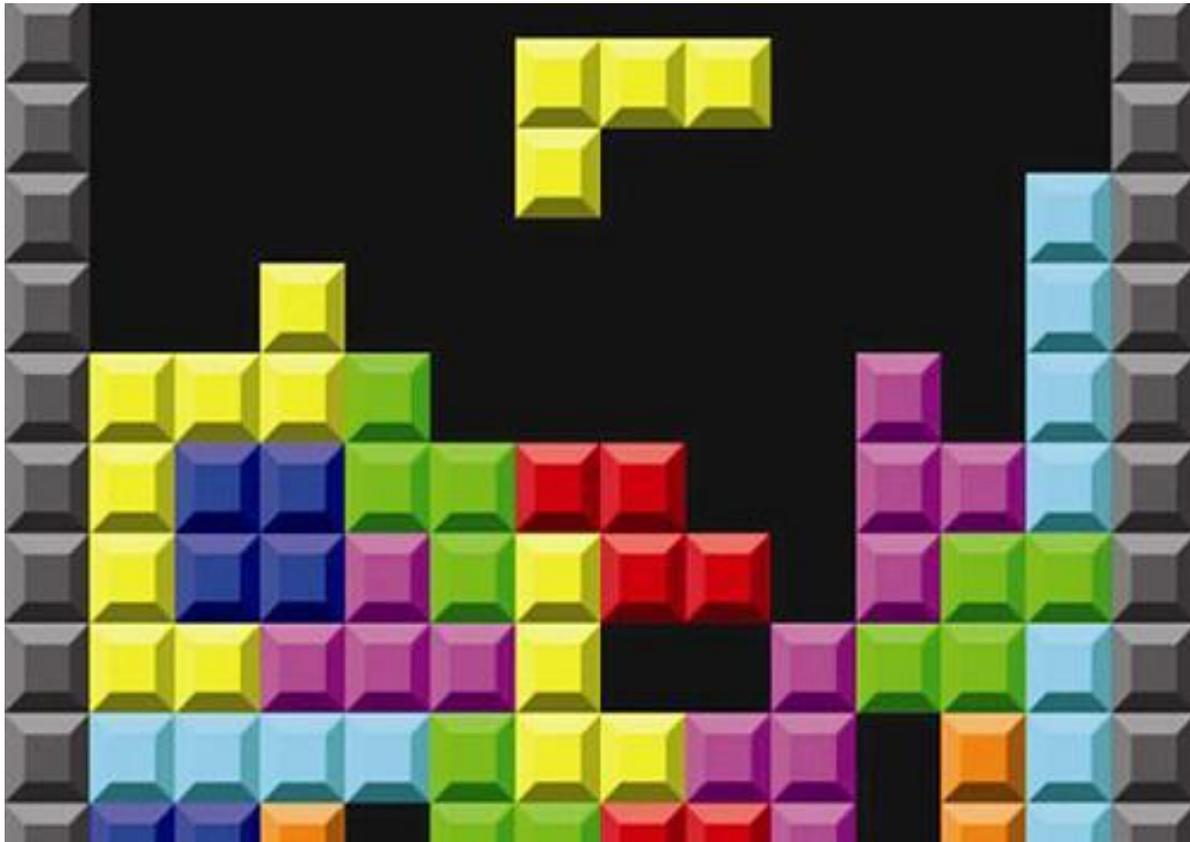
红 * 绿 * 蓝
 $256 * 256 * 256 = 16, 777, 216$ 种颜色

11111111,00000000,00000000;
11111111,00011000,00000000;
11111111,00000000,00000000;
11111111,11111111,00000000...



三维的“物体”呢？

从拼图到积木

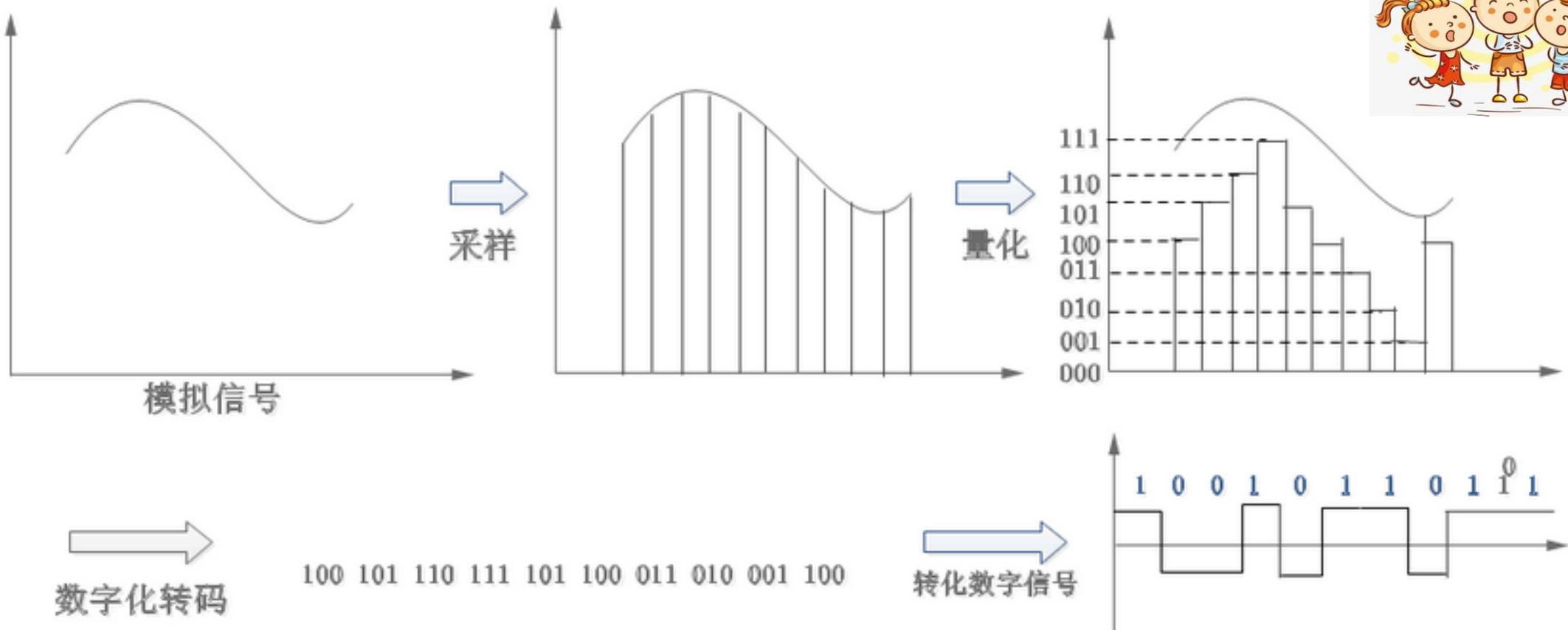


3D 数字化的例子：服装3D数字化平台

https://v.youku.com/v_show/id_XNDE2MDUwNzU4OA==.html?spm=a2hbt.13141534.0.13141534



声音的数字化



数字化的两个部分：数据和程序

信息（比如图片）被数字化以后就变成了数据
计算机对数据进行处理的过程（美图）就是程序

数据和程序都是0和1

数据可以被程序加工，变成另外的数据。



我们和周围的世界已经被数字化.....

2 MILLION BLOG POSTS are written.
Enough posts to fill Time Magazine for 770 years.



THE WORLD'S INFORMATION IS DOUBLING EVERY TWO YEARS, WITH A COLOSSAL **1.8 zettabytes** to be created & replicated in **2011**.
New information being created in 2011 also includes replicated information such as shared documents or duplicated DVDs.

294 BILLION emails are sent.



It would take 2 years to process that many pieces of mail in the US.

A PETABYTE IS A LOT OF DATA

- 1** PETABYTE = **20 MILLION** FOUR-DRAWER FILING CABINETS FILLED WITH TEXT
- 1** PETABYTE = **13.3 YEARS** OF HD-TV VIDEO
- 1.5** PETABYTES = **SIZE OF THE 10 BILLION** PHOTOS ON **FACEBOOK**
- 20** PETABYTES = **THE AMOUNT OF DATA PROCESSED BY GOOGLE PER DAY**

In terms of sheer volume, **1.8 ZB** of data is equivalent to:

- EVERY PERSON IN THE UNITED STATES TWEETING **3 TWEETS PER MINUTE** or **200 BILLION HD MOVIES** EACH **120 MINUTES** LONG
- 4,320** TWEETS PER DAY PER PERSON
- 1 = 10 billion**
- IT WOULD TAKE ONE PERSON 47 MILLION YEARS** OF 24/7 VIEWING TO WATCH EVERY MOVIE FOR **26,976** YEARS NON-STOP

In one day, enough information is consumed by internet traffic to fill **168 MILLION DVDS.**



1 = 1 MILLION

- 1 KB 千字节 = 1024 Byte
- 1 MB 兆字节 = 1024 KB = 1,048,576 Byte
- 1 G 吉字节 = 1,073,741,824 Byte
- 1 T 太字节 = 1,099,511,627,776 Byte
- 1 P 拍字节 = 1,125,899,906,842,624 Byte
- 1 E 艾字节 = 1.152921504606847e+18 Byte

我们看到了数字化的一个过程

早上买早点
写程序
接收短信
健身
读书



工程师的故事



今天你学到了什么？

文本转化中的ascii码

调色板的组合

彩色图像怎么数字化



什么可以被数字化

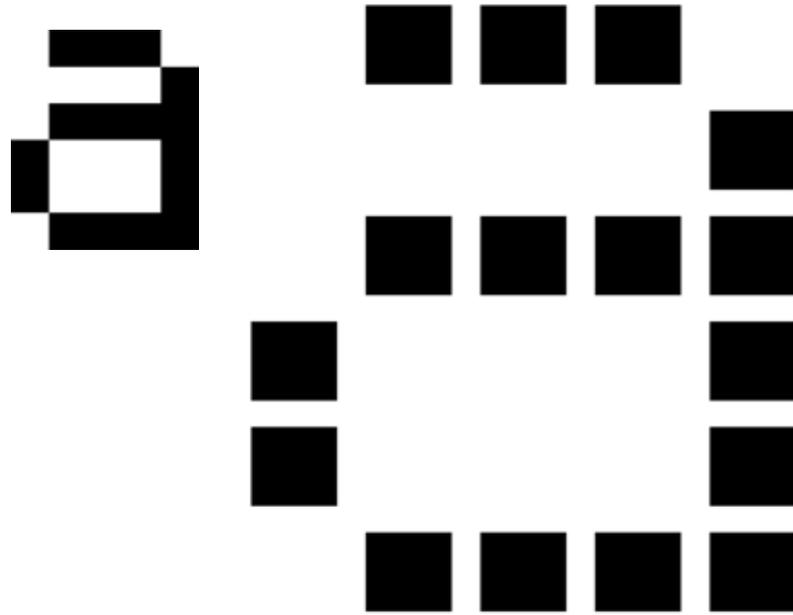
二进制中的位和字节

二进制和十进制的关系

.....



课后作业 -- 破解密电码



计算机显示屏由一系列小点（像素pixels）组成

每张黑白照片里，每一个像素要么是黑色要么是白色
放大字母“a”，可以看到哪些点是黑色哪些点是白色

破解密电码 时间：5分钟

白	黑	黑	黑	白	1, 3, 1
白	白	白	白	黑	4, 1
白	黑	黑	黑	黑	1, 4
黑	白	白	白	黑	0, 1, 3, 1
黑	白	白	白	黑	0, 1, 3, 1
白	黑	黑	黑	黑	1, 4

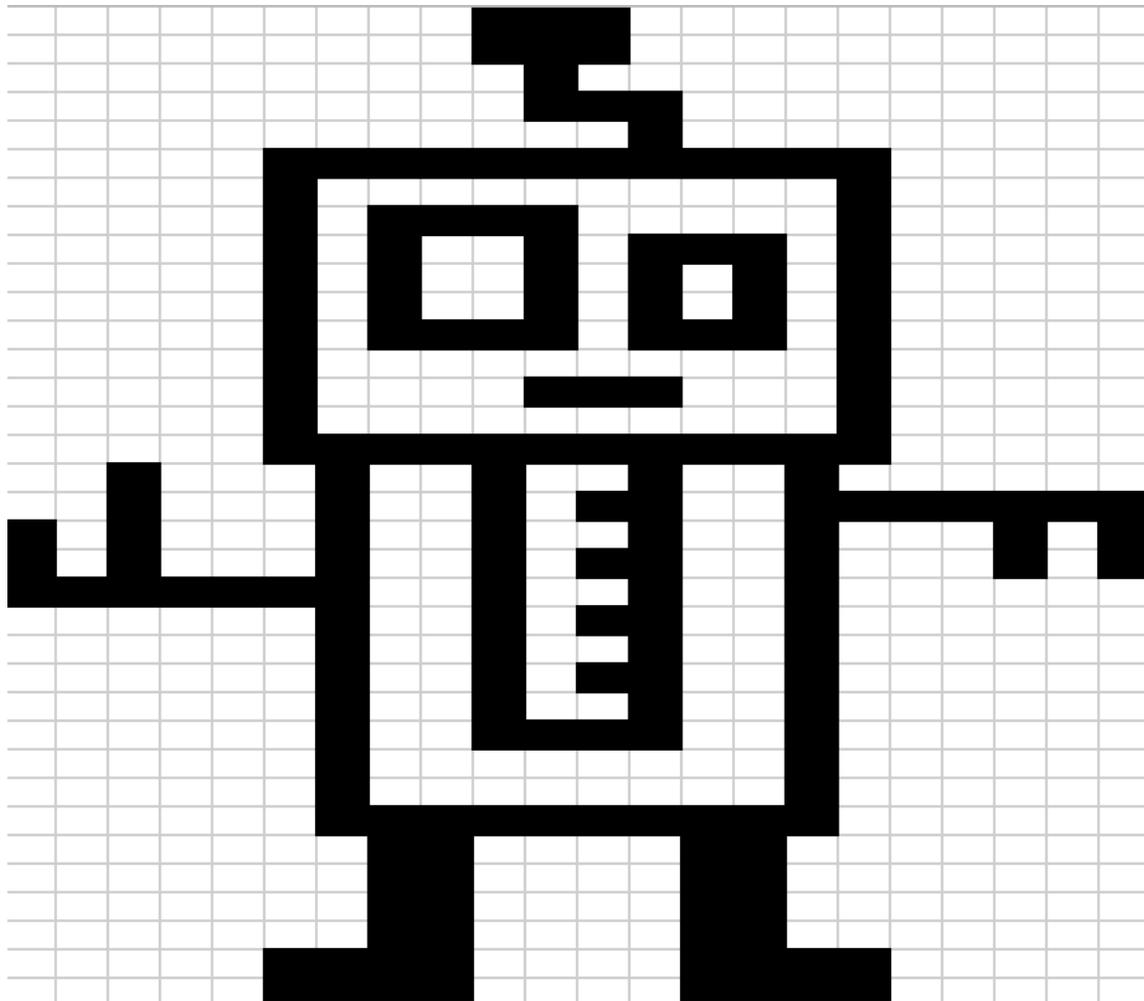
用数字表示图片

第1行有1个白像素，跟着3个黑像素，再跟着一个白像素，所以表示为 1, 3, 1
如果第一个像素是黑像素，比如第4行，就先加一个0，再开始标记像素数量

破解密电码 时间：5分钟

																	0,3,1,3,1,3,1,5
																	0,3,1,3,1,3,1,5
																	0,1,4,1,2,1,3,1,1,1,1,1
																	0,1,4,1,2,1,3,1,1,1,1,1
																	0,1,4,1,2,1,3,1,1,1,1,1
																	0,1,4,1,2,1,3,1,1,1,1,1
																	0,1,4,1,2,1,3,1,1,1,1,1
																	0,1,4,1,2,1,3,1,1,1,1,1
																	0,3,2,1,2,3,1,1,1,1,1,1
																	0,3,2,1,2,3,1,1,1,1,1,1
																	2,1,2,1,2,1,3,1,1,1,1,1
																	2,1,2,1,2,1,3,1,1,1,1,1
																	2,1,2,1,2,1,3,1,1,1,1,1
																	2,1,2,1,2,1,3,1,1,1,1,1
																	2,1,2,1,2,1,3,1,1,1,1,1
																	2,1,2,1,2,1,3,1,1,1,1,1
																	0,3,2,1,2,3,1,1,1,1,1,1
																	0,3,2,1,2,3,1,1,1,1,1,1

破解密电码 时间：5分钟



STEM

版权声明

- 本课程由【IBM】授权提供，选自IBM AI启蒙季系列课程。在此，特别致谢IBM对“云支教”助学计划的支持，以及对乡村儿童教育发展所做出的贡献。
- 本课程仅用于有爱有未来大学生志愿者及企业志愿者，针对乡村学校开展教育帮扶项目；同时，支持乡村学校用于开展课程教学。
- 未经授权许可，对课程内容进行摘取、复制、传播、修改、出租、售卖，或以其他方式进行处理及衍生其他作品的行为，均构成侵权。
- 任何企业、机构和个人，不得将本课程应用于商业用途。

有爱有未来企业志愿行动
2021年1月19日



云支教助学计划 Online Education Volunteer Project

为响应“乡村振兴”战略，有爱有未来企业志愿行动联合多家专业机构于2018年发起“云支教助学计划”，以“互联网+教育扶贫”的创新模式，为企业志愿服务和大学生助学支教提供可持续的教学直播平台，帮助乡村中小学校接触更多优质教育资源，助力乡村教育发展。

截至2020年，共有6家企业志愿者及139所高校大学生响应“云支教”，在安徽、青海、云南、海南、甘肃、四川、宁夏等地共计100所乡村学校已开展云支教各类型课程，直接受益学生1,193,372人次。



有爱有未来企业志愿行动 Share the Care Volunteer Organization

有爱有未来企业志愿行动创立于2006年，是国内领先的企业青年志愿服务平台和枢纽机构，服务于FESCO系统4万家中外企业和400万白领员工。有爱有未来长期关注和帮扶弱势儿童群体，是乡村儿童和城市特殊儿童教育援助领域的主要行动者，工作方向涉及乡村教育、环境保护、健康卫生、灾难救助四大领域，志愿服务项目和公益资源网络覆盖全国。

截至目前，有爱有未来已发动超过300家中外企业，在7个省市200所乡村学校开展超过600项/次公益项目和志愿服务活动，参与企业员工及高校大学生志愿者超过155万人次，近50万名乡村师生受益。

联系我们

联系人: Ms. Alina Wang王公佑
Office: 8610-82193926
Mobile: 135 2057 6052
Email: alina_wang@pmacasia.com





让每个孩子都能拥有更好的教育机会

有爱有未来企业志愿行动

中国北京市海淀区中关村南大街6号中电信息大厦614室

Web: www.sharethecare.cn