

苹果公司发布首份人工智能报告，AI 视觉是研究新方向

摘要

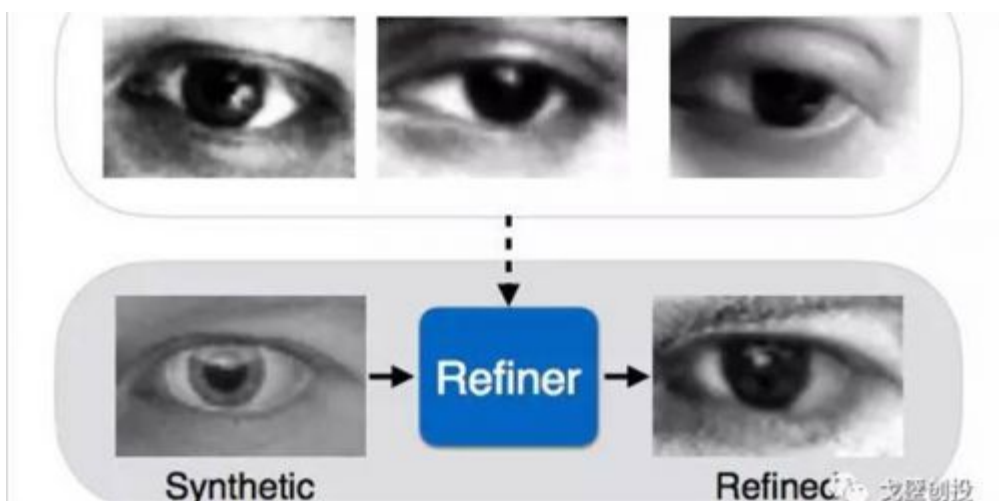
随着图形技术不断进步，利用合成图像训练机器学习模型变得越来越容易，这可以帮助避免注释图像的昂贵支出。然而，通过合成图像训练机器学习模型可能无法达到令人满意的效果，因为合成图像与真实图像之间毕竟存在区别。为了减少这种差异，苹果公司提出“模拟+无监督”学习方法，即通过计算机生成图像或合成图像来训练算法的图像识别能力。

事实上，这种“模拟+无监督”学习需要将无标注的真实图像数据与已经注释的合成图像相结合。在很大程度上，它需要依赖生成式对抗网络 (GAN) 的新机器学习技术，它可通过两个神经网络相互对抗以生成更加真实的图像。苹果公司对标准 GAN 算法进行了多处关键性修改，以保留注释、避免伪影以及稳定性训练：自正则化(self-regularization)-局部对抗性损失-使用精炼图像升级鉴别器。

苹果公司发现，这个过程可以产生高度逼真的图像，在质量上和用户研究方面都已经获得证实。苹果公司已经通过训练模型评估视线水平和手势姿态，对计算机生成图像进行定量评估。通过使用合成图像，苹果公司的图像识别算法已经取得了巨大进步。在没有使用任何标准真实数据的情况下，苹果公司在 MPIIGaze 数据集中获得了最高水平的结果。

引言

随着最近高容量深度神经学习网络的崛起，大规模标注训练数据集正变得日益重要。可是，标准数量庞大的数据集成本非常高，而且相当耗费时间。为此，使用合成图像而非真实图像训练算法的想法开始出现，因为注释已经可实现自动化。利用 XBOX360 外设 Kinect 评估人体姿势以及其他任务，都是使用合成数据完成的。



(图 1：“模拟+无监督”学习：通过计算机生成图像或合成图像来训练算法的图像识别能力)

然而，由于合成图像与真实图像之间存在差距，使用合成图像训练算法可能产生很多问题。因为合成图像通常不够真实，导致神经网络学习仅仅能够了解到合成图像中的细节，并不能完整地识别出真实图像，进而也无法为算法提供精确的学习。一种解决方案就是改进模拟器，可是增加真实性的计算往往非常昂贵，渲染器的设计也更加困难。此外，即使最顶级的渲染器可能也无法模仿真实图像中的所有特征。因此，缺少真实性可能导致算法过度拟合合成图像中不真实的细节。

在这篇论文中，苹果公司提出“模拟+无监督”学习的方法，其目的就是使用未标注真实数据的模拟器提高合成图像的真实性。提高真实性可更好地帮助训练机器学习模型，而且无需收集任何数据，也无需人类继续标注图像。除了增加真实性，“模拟+无监督”学习还应该保留注释信息以用于训练机器学习模型，比如图一中的注视方向应被保留下来。此外，由于机器学习模型对合成数据中的伪影非常敏感，“模拟+无监督”学习也应该产生没有伪影的图像。

苹果公司为“模拟+无监督”学习开发出新的方法，称之为 SimGAN，它可以利用称之为“精炼器网络 (refiner network)”的神经网络从模拟器中提炼合成图像。图二中展示了这种方法的概述：第一，黑盒子模拟器中生成合成图像，然后利用“精炼器网络”对其进行提炼。为了增加真实性，也就是“模拟+无监督”

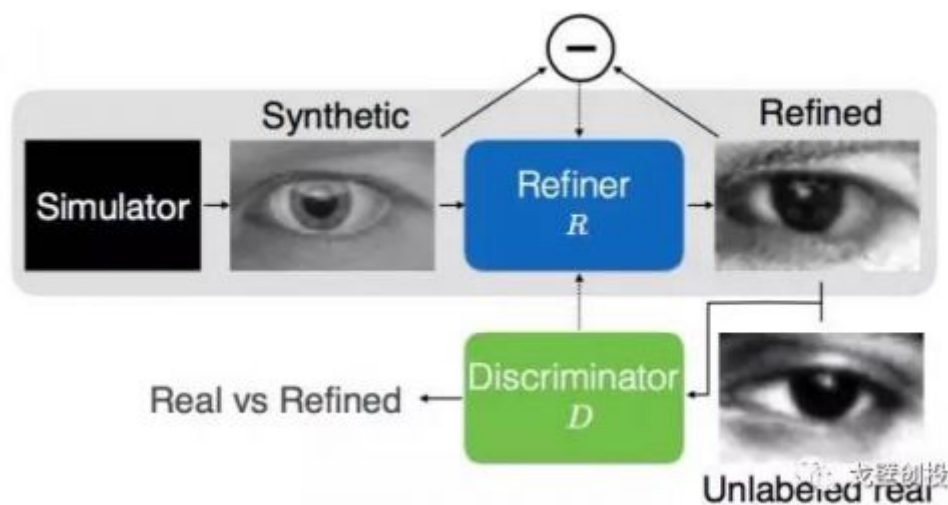
学习算法的首要需求，我们需要利用类似生成式对抗网络(GAN)来训练“精炼器网络”，进而产生判别网络无法区分真假的精炼图像。

第二，为了保留合成图像上的注释信息，苹果公司需要利用“自正则化损失”弥补对抗性损失，在合成图像和精炼图像之间进行修改。此外，我们还利用完全卷积神经网络，在像素水平方面进行操作，并保留全局结构，而非整体修改图像的内容。

第三，GAN 框架要求训练 2 个神经网络进行对抗，它们的目标往往不够稳定，倾向于产生伪影。为了避免漂移和产生更强的伪影，导致甄别更困难，我们需要限定鉴别器的接收区域为局部接收，而非整张图片接收，这导致每张图像都会产生多个局部对抗性损失。此外，苹果公司还引入提高训练稳定性的方法，即通过使用精炼图像而非当前“精炼器网络”中的现有图像升级鉴别器。

1.1 相关工作

GAN 框架需要 2 个神经网络竞争损失，即生成器与鉴别器。其中，生成器网络的目标是在真实图像上绘制随机向量，而鉴别器网络的目标则是区分生成图像与真实图像。GAN 网络是由古德弗罗 (I. Goodfellow) 等人首先引入的，它可以帮助生成逼真的视觉图像。自从那以来，GAN 已经有了许多改进，并被投入到有趣的应用中。



(图 2 : SimGAN 概观 : 我们利用 “精炼器网络” 提炼模拟器产生的输出图像，并最大限度地减少局部对抗性损失，并进行自正则化。

对抗性损失可以欺骗鉴别器网络，从而令其将合成图像误认为真实图像。而自正则化则会最大限度地减少合成图像与真实图像的差异，包括保留注释信息，并让精炼图像被用于训练机器学习模型。“精炼器网络”与鉴别器网络也会交替升级。)

王 (X. Wang) 与古普塔 (A. Gupta) 利用结构化 GAN 学习表面法线，然后将其与 Style GAN 相结合，以生成天然的室内场景。我们提议使用对抗性训练进行递归生成模型(recurrent generative model)训练。此外，最近推出的 iGAN 能够帮助用户以交互模式改变图像。刘 (M.-Y. Liu) 等人开发的 CoGAN 结合 GAN 可多模态联合分布图像，无要求应对图像的元组，这有利于开发出联合发布解决方案。而陈 (X. Chen) 等人开发的 InfoGAN 是 GAN 信息理论的扩展，允许有意义的陈述学习。

恩杰尔·图泽尔 (Oncel Tuzel) 等人利用 GAN 解决人脸图像超分辨率问题。李 (C. Li) 和王 (M. Wand) 提议 Markovian GAN 进行有效的纹理合成。洛特 (W. Lotter) 等人在 LSTM 网络中利用对抗性损失进行视觉序列预测。于 (L. Yu) 等人提议 SeqGAN 框架，利用 GAN 强化学习。许多近来的问题都显

示出与生成模型领域相关的问题，比如 PixelRNN 可利用 RNN 的 softmax 损失预测像素顺序。生成网络侧重于使用随机噪声矢量生成图像，与我们的模型相比，其生成的图像没有任何标注信息，因此也无法用于训练机器学习模型。

许多努力都在探索使用合成数据进行各种预测任务，包括视线评估、RGB 图像文本检测与分类、字体识别、对象检测、深度图像中手部姿态评估、RGB-D 场景识别、城市场景语义分割以及人体姿态评估等。盖伊登 (A. Gaidon) 等人的研究表明，利用合成数据训练深度神经网络，可以提高其表现。我们的工作是对这些方法的补充，我们使用未标记真实数据提高了模拟器的真实性。

嘉宁 (Y. Ganin) 与莱姆皮茨基 (V. Lempitsky) 在数据域适应设置中利用合成数据，了解合成图像与真实图像域的变化过程中保持不变的特征。王 (Z. Wang) 等人利用合成和真实数据训练层叠卷积码自动编码器，以便了解其字体检测器 ConvNet 的低级表示形式。张 (X. Zhang) 等人学习多通道编码，以便减少真实数据与合成数据的域的转变。与经典域适应方法相反，它采用与特定的特征以便适应具体的预测任务，而我们可通过对抗性训练，弥合图像分布之间的差距。这种方法使我们能够生成非常逼真的图像，可以用于训练任何机器学习模型，并执行潜在的更多任务。

2 “模拟+无监督”学习

模拟+无监督学习的目标是使用一组未标记的真实图像 $y_i \in \mathcal{Y}$ ，学习可提炼合成图像 X 的 refiner $R_\theta(x)$ ，其中 θ 属于函数参数。让我们用 $X^?$ 表示精炼图像，然后会得出 $X^? : \theta = R(X)$ 。在“模拟+无监督”学习中，最关键的要求就是精炼图像 $X^?$ ，以便于其看起来更像真实图像，同时保留来自模拟器的注释信息。为此，我们建议通过最大化减少两个损失的组合来学习：

$$\mathcal{L}_R(\theta) = \sum_i \ell_{\text{real}}(\theta; \tilde{x}_i, \mathcal{Y}) + \lambda \ell_{\text{reg}}(\theta; \tilde{x}_i, x_i)$$

其中, x_i 是 i 个合成训练图像, x 是相应的精炼图像。第一部分是真实性成本, 即向合成图像中增加真实性的成本。第二部分则代表着通过最小化合成图像精炼图像的差异保存注释信息的成本。在下面的章节中, 我们会展开这个公式, 并提供优化 θ 的算法。

2.1 对抗性损失

为了向合成图像中添加真实性, 苹果公司需要在合成图形和真实图像的分部之间建立起联系。在理想情况下, 精炼机可能无法将给定的图像分类为真实图像还是高度精炼图像。这就需要使用对抗性的鉴别器, 网络 D_ϕ , 它可训练分辨图像到底是真实图像还是精炼图像, 而 ϕ 是鉴别器网络参数。对抗性损失训练 refiner network R , 它负责欺骗 D 网络, 令其将精炼图像误认为是真实图像。利用 GAN 方法, 苹果公司建造了 2 个神经网络参与的极限博弈模型, 并升级 “精炼器网络” R_θ 和鉴别器网络 D_ϕ 。接下来, 苹果公司更精确地描述这种模型。鉴别器网络通过最大限度地减少以下损失来更新参数:

$$\mathcal{L}_D(\phi) = - \sum_i \log(D_\phi(\tilde{x}_i)) - \sum_j \log(1 - D_\phi(y_j)).$$

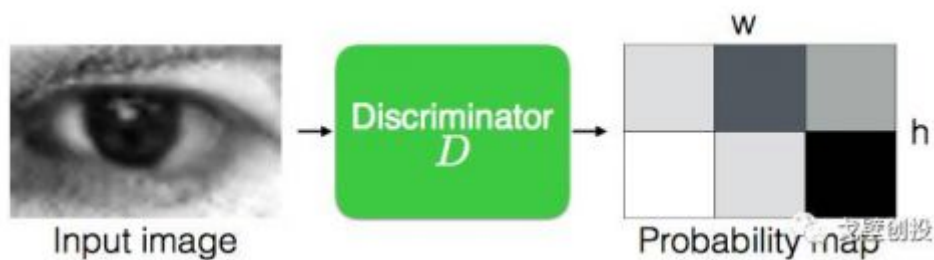
这相当于两级分类问题产生的交叉熵误差, 其中 $D_\phi(\cdot)$ 输入的是合成图像, 而 $1 - D_\phi(\cdot)$ 则是真实图像。至此, 我们实现了 D_ϕ 作为 ConvNet 的最后输出层, 样本很可能是精炼图像。为了训练这个网络, 每个小批量随机抽取的样本都由精炼合成图像和真实图像组成。对于每个 y_j 来说, 交叉熵的目标标签损耗层为 0, 而每个 x_i 都对应 1。然后通过随机梯度下降 (SGD) 方式, ϕ 会随着小批量梯度损失而升级。在苹果公司的实践中, 真实性损失函数使用训练有素的鉴别器网路 D 如下:

$$\ell_{\text{real}}(\theta; \tilde{x}_i, \mathcal{Y}) = - \sum_i \log(1 - D_\phi(R_\theta(\tilde{x}_i))).$$

通过最小化减小损失函数，“精炼器网络”的力量促使鉴别器无法分辨出精炼图像就是合成图像。除了产生逼真的图像外，“精炼器网络”应该保存模拟器的注释信息。举例来说，用于评估视线的学习转变不该改变凝视的方向，手部姿势评估不应该改变肘部的位置。这是训练机器学习模型使用配有模拟器注释信息的精炼图像的必要组成部分。为了实现这个目标，苹果公司建议使用自正则化，它可以最大限度地减少合成图像与精炼图像之间的差异。

<p>Algorithm 1: Adversarial training of refiner network R_θ</p> <p>Input: Sets of synthetic images $\mathbf{x}_i \in \mathcal{X}$, and real images $\mathbf{y}_j \in \mathcal{Y}$, max number of steps ($T$), number of discriminator network updates per step (K_d), number of generative network updates per step (K_g).</p> <p>Output: ConvNet model R_θ.</p> <pre> for $t = 1, \dots, T$ do for $k = 1, \dots, K_g$ do 1. Sample a mini-batch of synthetic images \mathbf{x}_i. 2. Update θ by taking a SGD step on mini-batch loss $\mathcal{L}_R(\theta)$ in (4). end for $k = 1, \dots, K_d$ do 1. Sample a mini-batch of synthetic images \mathbf{x}_i, and real images \mathbf{y}_j. 2. Compute $\tilde{\mathbf{x}}_i = R_\theta(\mathbf{x}_i)$ with current θ. 3. Update ϕ by taking a SGD step on mini-batch loss $\mathcal{L}_D(\phi)$ in (2). end end </pre>

(算法 1)

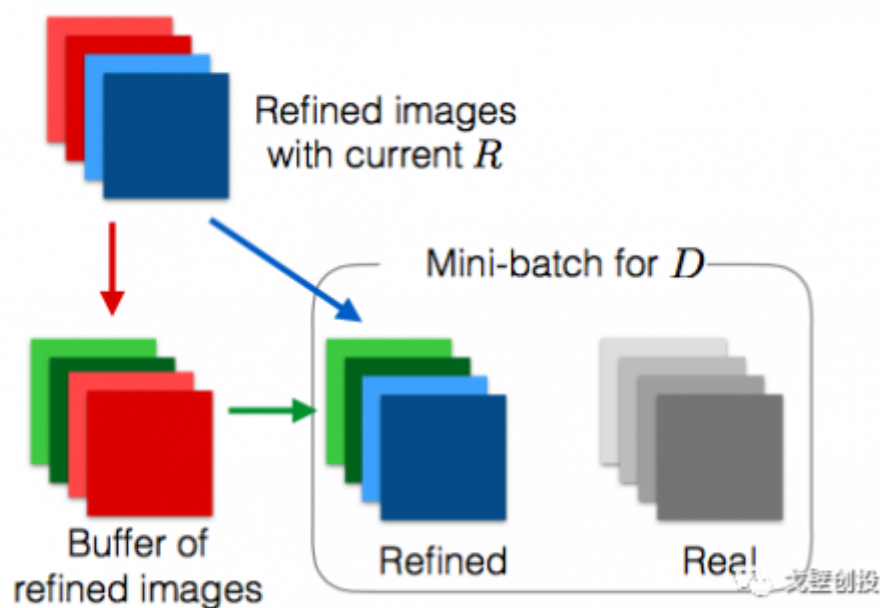


(图 3：局部对抗性损失的图示。鉴别器网络输出 $w \times h$ 概率图。对抗损失函数是局部块上的交叉熵损失的总和。)

因此在苹果公司的执行中，整体精炼损失函数 (1) 为：

$$\mathcal{L}_R(\theta) = - \sum_i \log(1 - D_\phi(R_\theta(\mathbf{x}_i))) + \lambda \|R_\theta(\mathbf{x}_i) - \mathbf{x}_i\|_1$$

(4) 在 $\|\cdot\|_1$ 是 L1 常模时，将 R_θ 作为一个完全卷积的神经网络，而无需跃进或池化。在像素级别上修改合成图像，而不是整体地修改图像内容。例如在完全连接地编码器网络中便会如此，保留全局结构合注释。苹果公司通过交替地最小化 $\mathcal{L}_R(\theta)$ 和 $\mathcal{L}_D(\phi)$ 来学习精化器和鉴别器参数。在更新 R_θ 的参数时，保持 ϕ 固定不变，而在更新 D_ϕ 时，则要保持 θ 不变。我们在算法 1 中描述了整个训练过程。



(图 4：使用精细图像历史示意图。相关信息请参阅文本描述。)

2.2 本地对抗损失

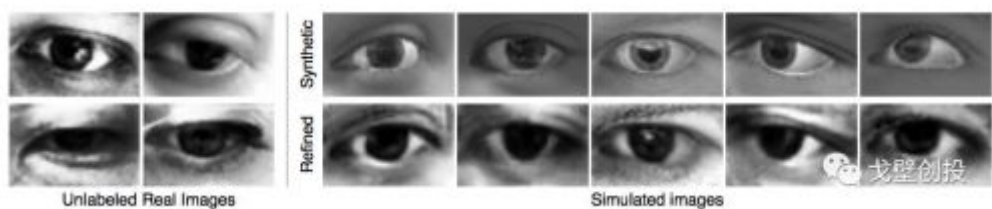
精炼网络另一个关键要求是 ,它应该学习模拟实际图像特性 ,而不引入任何伪影。当训练强鉴别器网络时 ,精炼网络倾向于过度强调某些图像特征以欺骗当前鉴别器网络 ,导致偏差和产生伪影。关键是任何我们从精化图像中采样的本地补丁都应该具有与真实图像相类似的统计。由此可以定制本地鉴别器网络对本地图像补丁进行分类 ,而不是定义全局鉴别器网络。

这不仅限制了接受域 ,还因此限制了鉴别器网络的容量 ,并且为每个图像提供更多样本以供学习鉴别器网络。同时由于每个图像由多个实际损失值 ,它还改善了精炼网络的训练。

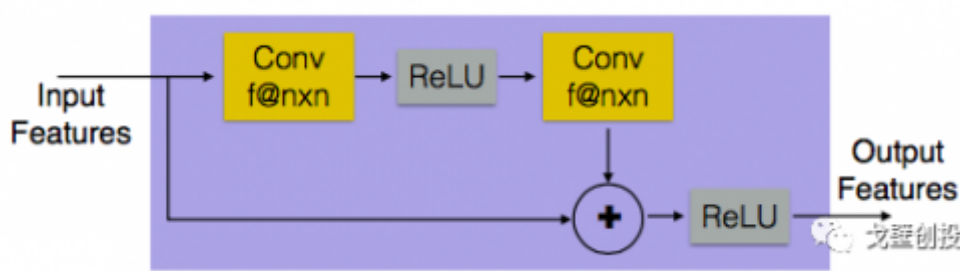
在执行中 ,将鉴别器器 D 设计成一个完全卷积网络 ,输出伪类 $w \times h$ 概率图。在后者中 $w \times h$ 是图像中本地补丁的数量。在训练精炼网络时 ,将 $w \times h$ 本地补丁的交叉熵损失值求和 ,如图 3 所示。

2.3 使用精化图像的历史记录更新鉴别器器

对抗训练的对抗训练的另一个问题是鉴别器网络只关注最新的精细图像。这可能导致 (i) 与对抗训练分歧, 以及 (ii) 精炼网络重新引入鉴别器已经忘记的工件。在整个训练过程中的任何时间由精炼网络生成的任何精细图像对于鉴别器来说都是伪造的图像。因此, 鉴别器应该有能力将这些图像识别为假。基于这一观察, 我们引入了一种方法, 通过使用精细图像的历史来提高对抗训练的稳定性, 而不仅仅是在当前小批次中小修小改。我们对算法 1 稍作改进, 增加对先前网络产生的精细图像的缓冲。设 B 为此缓冲的大小, 设 b 为算法 1 中使用的迷你批次大小。



(图 5 : SimGAN 输出的图像示例。左为 MPIIGaze 采集的实拍图像, 右为优化后的 UnityEye 合成图像。从图中可以看出精细合成图像中的皮肤纹理和虹膜区都更近似真实而非合成图像。)



(图 6 : 带有两个 nxn 卷积层的 ResNet 块, 每个都具有 f 特征图。)

在鉴别器训练的每次迭代中, 通过对来自当前精炼网络的 $b/2$ 图像进行采样, 以及从缓冲区采集额外 $b/2$ 图像, 从而更新参数 ϕ 。保持缓冲区 B 大小固定, 在每次迭代之后, 苹果公司使用新产生的精细图像随机替换缓冲区中的 $b/2$ 样本。该过程在图 4 中标示出。

3.实验

我们使用 MPIIGaze [40 ,43]上的外貌估计数据集和纽约大学的手势数据集 [35]来评估我们的方法。我们在所有实验中使用完全卷积的精炼网络与 ResNet 块 (图 6)。

3.1 基于外貌的注视估计

注视估计是许多人机交互 (HCI) 任务的关键因素。然而，直接由眼睛图像进行估计是有挑战性的，特别是在图像质量不佳时。例如智能手机或笔记本电脑前置摄像头采集到的眼睛图像。因此，为了生成大量注释数据，最近几次方法[40,43]用大量合成数据训练它们的模型。在这里，我们展示使用 SimGAN 生成的精细合成图像训练显著提高了任务的表现。

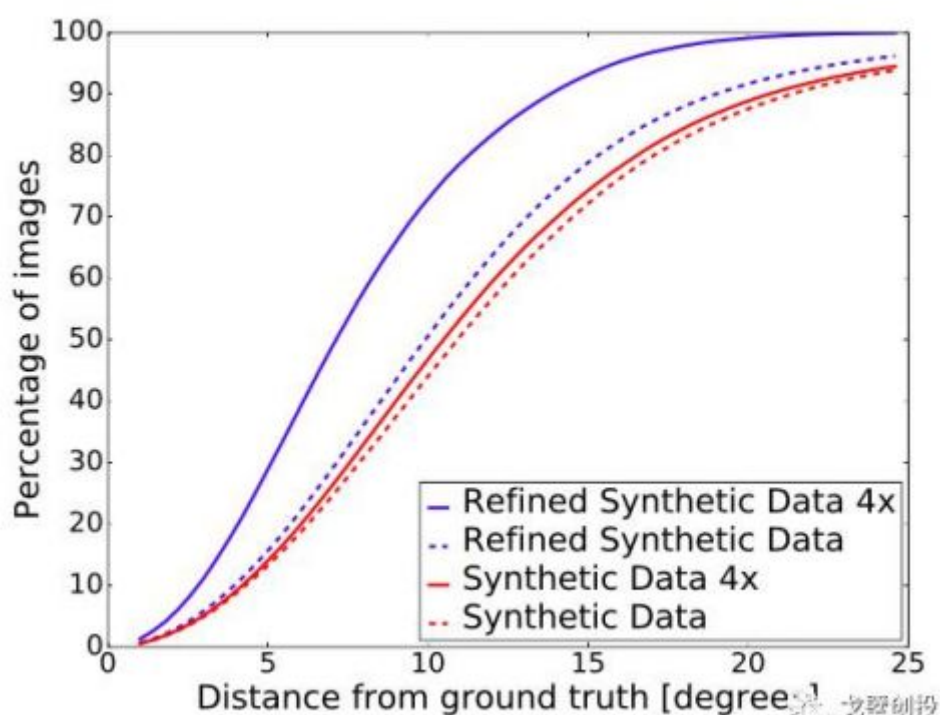
注视估计数据集包括使用眼睛注视合成器 UnityEyes 生成的 1200 万样本，以及来自 MPIIGaze 数据集的 21,000 实拍样本。MPIIGaze 的图像样本都是在各种照明条件不理想情况下捕捉到的图像。而 UnityEyes 图像都是在同一渲染环境下生成。

定性结果：图 5 展示了合成生成的眼睛注视图像以及经过处理的实拍图像。如图所示，我们观察到合成图像的显著质量改进：SimGAN 成功捕获皮肤纹理，传感器噪点以及虹膜区的外观。请注意，我们的方法在改善真实性的同时保留了注释信息（注视方向）。

‘视觉图灵测试’：为了定量评估精细图像的视觉质量，我们设计了一个简单的用户研究，要求受试者对图像是属于实拍或是合成进行区分。每个受试者被展示 50 张实拍图像和 50 张合成图像。在试验中，受试者不断观看 20 个真假混杂的图像，最终受试者很难分辨真实图像和精细图像之间的差异。在我们的总体分析中，10 个受试者在 1000 次试验中正确率只有 517 次 ($p=0.148$)，跟随机选差不多。表 1 展示了混淆矩阵。相比之下，当使用原始图像和实拍图像进行测试时，我们给每个受试者展示 10 个实拍和 10 个合成图像，这种情况下受试者在 200 此实验中选对了 162 次 ($p\leq 10^{-8}$)，结果明显优于随机选择。

	Selected as real	Selected as synt
Ground truth real	224	276
Ground truth synt	207	293

（表 1：采用真实图像和合成图像进行的“视觉图灵测试”。平均人类分类精度为 51.7%，表明自动生成的精细图像在视觉上已经达到以假乱真的程度。）



（图 7：使用 MPIIGaze 实拍样本进行的眼球注视估计的定量结果。曲线描述了不同测试数量下系统预估的误差。图示中使用精细图像而不是合成图像能显著提高系统表现。）

定量结果：我们训练了一个与[43]类似的简单的卷积神经网络 (CNN) 来对眼睛的注视方向进行预测。我们在 UnityEyes 上训练，并在 MPIIGaze 上进行测试。图 7 和表 2 比较了 CNN 分别使用合成数据以及 SimGAN 生成的精细数据的不同表现。我们观察到 SimGAN 输出训练的表现有很大的提高，绝对百分比提高了 22.3%。我们还发现训练结果和训练数据呈正相关——此处的 4x 指的是培训

数据集的 100%。定量评估证实了图 5 中观察到的定性改进的价值，并且表明使用 SimGAN 能使机器学习模型有更好的表现。表 3 展示了同现有技术的比较，在精细图像上训练 CNN 的表现优于 MPIGaze 上的现有技术 相对改善了 21%。这个巨大的改进显示了我们的方法在许多 HCI 任务中的实际价值。

实施细节：精炼网络 R0是一个残差网络 (ResNet) 。每个 ResNet 块由两个卷积层组成，包含 63 个特征图，如图 6 所示。大小为 55x35 的输入图像和 3x3 的过滤器进行卷积，输出 64 个特征图。输出通过 4 个 ResNet 块传递。最后 ResNet 块的输出被传递到 1x1 卷积层，产生对应于精细合成图像的 1 个特征图。

Training data	% of images within d
Synthetic Data	62.3
Synthetic Data 4x	64.9
Refined Synthetic Data	69.4
Refined Synthetic Data 4x	87.2

(表 2: 使用合成数据和 SimGAN 输出进行比较。在无需监督真实数据的情况下，使用 SimGAN 输出的图像进行训练表现出 22.3%的优势。)

Method	R/S	Error
Support Vector Regression (SVR) [30]	R	16.5
Adaptive Linear Regression ALR) [21]	R	16.4
Random Forest (RF) [33]	R	15.4
kNN with UT Multiview [43]	R	16.2
CNN with UT Multiview [43]	R	13.9
k-NN with UnityEyes [40]	S	9.9
CNN with UnityEyes Synthetic Images	S	11.2
CNN with UnityEyes Refined Images	S	7.8

(表 3: SimGAN 与 MPIIGaze 现有技术进行比较。R=真实图像，S=合成图像。误差是以度为单位的平均眼睛注视估计误差。对精细图像的训练带来了 2.1 度的改进，相对现有技术提高了 21%。)

鉴别器网络 $D\phi$ 包含 5 个扩展层和 2 个最大合并层, 分别为: (1) Conv3x3, stride = 2, 特征映射 = 96, (2) Conv3x3, stride = 2, 特征映射 = 64, (3) MaxPool3x3, stride = 1, (4) Conv3x3, stride = 1, 特征映射 = 32, (5) Conv1x1, stride = 1, 特征映射 = 32, (6) Conv1x1, stride = 2, (7) Softmax。

对抗网络是完全卷积的, 并且已经被设计为使 $R\theta$ 和 $D\phi$ 中的最后层神经元的接受域是相似的。我们首先对 $R\theta$ 网络进行 1000 步的自正则化损失训练, $D\phi$ 为 200 步。然后对于 $D\phi$ 的每次更新, 对应在算法中更新 $R\theta$ 两次。即 K_d 被设置为 1, K_g 被设置为 50。

眼睛注视估计网络同[43]类似, 不过略作修改以使其更好地利用我们的大型合成数据集。输入是 35x55 的灰度图, 通过 5 个卷积层, 然后是 3 个完全连接的层, 最后一个编码三维注视向量: (1) Conv3x3, 特征图 = 32, (2) Conv3x3, 特征映射 = 32, (3) Conv3x3, 特征映射 = 64, (4) Max-Pool3x3, stride = 2, (5) Conv3x3, 特征映射 = 80, (6) Conv3x3, MaxPool2x2, stride = 2, (8) FC9600, (9) FC1000, (10) FC3, (11) Euclidean loss。所有网络都使用恒定的 0.001 学习速率和 512 批量大小进行训练, 直到验证错误收敛。

3.2 关于深度图像的手势图像模拟

下一步, 苹果公司将用这宗方法对各种手势的深度图像进行模拟。在研究中, 主要使用了纽约大学所提供的 NYU 手势数据库, 其中包含 72757 个训练样本以及使用 3 台 Kinect 相机所采集的 8251 个测试样本, 其中每个测试样本包括一个正面手势图像以及两个侧面手势图像。而每一个深度图像样本都对手势信息进行了标记, 从而生成了合成图像。图 10 展示了手势数据库中的一个样本。对数据库样本进行了预处理, 利用合成图像从真实图像中提取了相应的像素点位。在使用深度学习网络 ConvNet 进行处理之前, 每个图像样本的分辨率大小被统一调整为 224*224, 背景值设置为零, 前景值设置为原始深度值减 2000。(此时假设背景分辨率为 2000)。



图 10：NYU 手势数据库。左图为深度图像样本；右图为处理后的合成图像。

定性描述：图 11 显示了“生成对抗网络”（ SimGAN ）对手势数据库的计算结果。由图可知，真实深度图像的噪声已经边缘化，且分布不连续。SimGAN 能够有效对原有图像噪声进行学习并模拟，从而产生出更加真实精细的合成图像，且不需要在真实图像上做出任何标记或注释。

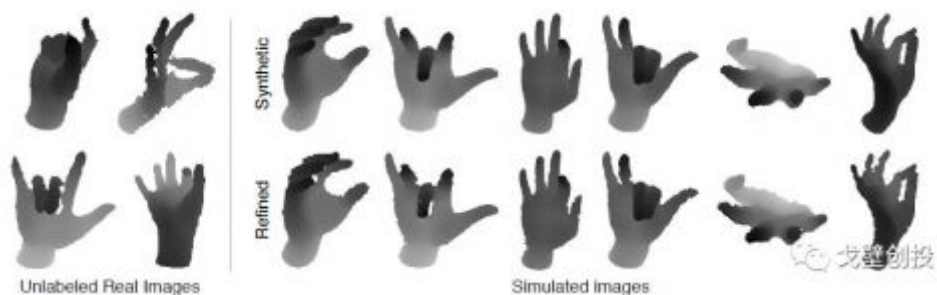


图 11： NYU 手势数据库的精细测试图像示例。左图为真实图像实像，右图上为合成图像，右图下为来自苹果生成对抗网络的相应精细化输出图像。

实际图像中的主要噪声源是非平滑的边缘噪声。学习网络能够学习模拟真实图像中存在的噪声，重要的是不需要任何标记和注释。

定量分析：

苹果公司采用一种类似于 Stacked Hourglass 人体姿态算法的 CNN 仿真算法应用于真实图像、合成图像以及精细化合成图像处理，与 NYU 手势数据库中的测试样本进行对比。通过对其中 14 个手关节变换进行算法训练。为了避免偏差，

我们用单层神经网络来分析算法对合成图像的改进效果。图 12 和表 4 显示了关于算法对手势数据库进行处理的定量结果。其中由 SimGAN 输出的精细化合成图像显著优于基于真实图像进行训练而产生的图像，其更为逼真，比标准合成图像高出了 8.8%，其中仿真模拟输出的注释成本为零。同时要注意的是，3X 代表图像训练选取了所有角度。

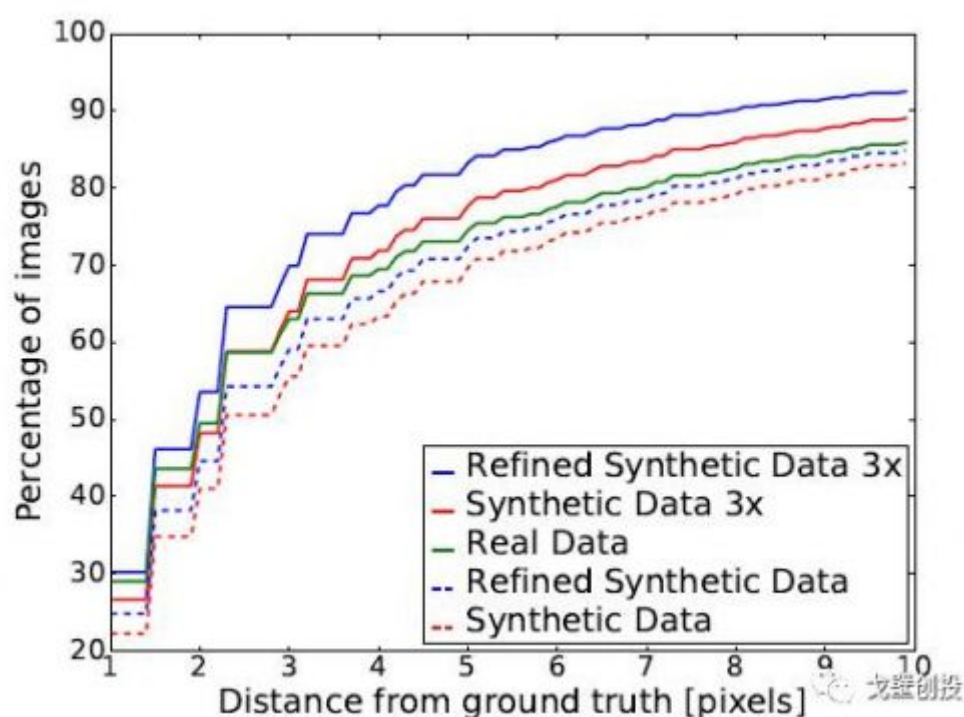


图 12：手势估计的定量结果，关于 NYU 手势测试集的真实深度图像。图表显示图像与背景之间函数的累积曲线。可见，SimGAN 输出的精细化合成图像显著优于基于真实图像进行训练而产生的图像，其更为逼真，比标准合成图像高出了 8.8%。重要的是，我们的学习网络不需要对真实图像进行标记。

Training data	% of images within d
Synthetic Data	69.7
Refined Synthetic Data	72.4
Real Data	74.5
Synthetic Data 3x	77.7
Refined Synthetic Data 3x	83.3

表 4：通过训练生成各种手势图像的相似度。

Synthetic Data 为一般网络训练产生的合成图像，Real Data 为真实图像，Refined Synthetic Data 为生成对抗网络 SimGAN 输出的精细化合成图像。3X 表示对真实图像进行多角度模拟。

实现细节：关于手势图像判别的架构与眼睛图像相同，但输入图像分辨率为 224×224 ，滤波器大小为 7×7 ，残差网络值为 10。判别网络 D 如下 (1) Conv 7×7 , stride=4, feature maps=96, (2) Conv 5×5 , stride=2, feature maps=64, (3) MaxPool 3×3 , stride=2, (4) Conv 3×3 , stride=2, feature maps=32, (5) Conv 1×1 , stride=1, feature maps=32, (6) Conv 1×1 , stride=1, feature maps=2, (7) Softmax。

首先，会对 R 网络进行自我规则化训练 500 次，随后引入 D 网络训练 200 次；随后，每更新一次 D 网络，就相应将 R 网络进行两次更新。在手势估计中，我们采用 Stacked Hourglass Net 人体姿态算法输出大小为 64×64 的热点图。我们在网络学习中引入 $[-20, 20]$ 的随机数据集来对不同角度的图像进行训练。直至验证误差有效收敛时，网络训练结束。

3.3 对抗训练的修正分析

首先苹果公司比较了本地化对抗训练以及全局对抗训练的图像偏差。在全局对抗中，判别网络使用了完全连接层，从而使整个图像相对于更加精细。而本地化对抗训练则使得生成的图像更为真实，如图 8 所示。

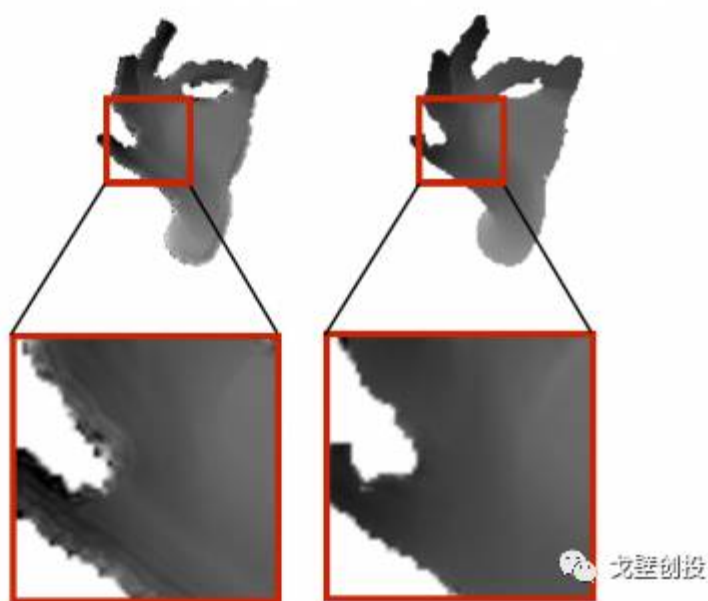


图 8：左为全局对抗训练结果，右为本地化对抗训练结果。
显示了全局对抗训练与本地化对抗训练的结果偏差。左图生成的图像更精细但不真实，右图生成的图像相对真实度更高。

接下来，在图 9 中，显示了使用反复训练的历史精细化图像对判别网络进行更新，并将其与标准对抗生成的合成图像进行比较的结果。如图所示，使用反复训练的历史精细化图像刻产生更为真实的阴影，譬如在标准对抗训练中，眼睛角落里没有阴影。

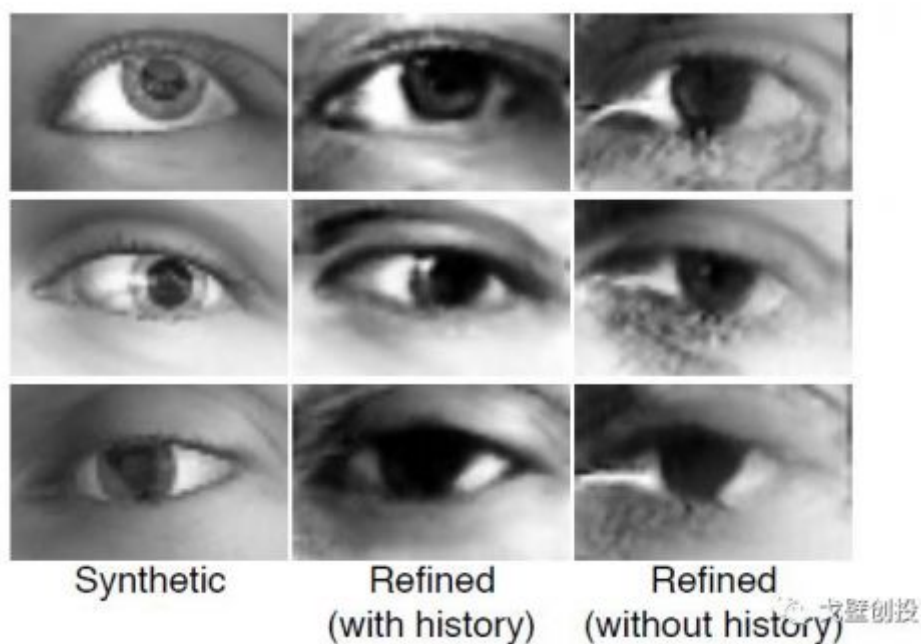


图 9：使用历史精细图像对判别网络进行更新的结果。

左图：标准合成图像；中图：使用历史数据对判别网络进行更新后的图像结果；右图：使用近期历史数据对判别网络进行更新的图像结果。如图所示，使用反复训练的历史精细化图像刻产生更为真实的阴影。

4.结论以及下一步工作

在文中，苹果公司提出了一种“模拟+无监督”的机器学习方法，能够有效提高模拟图像的真实感。并描述了一种新的生成对抗网络 SimGAN，并将其应用于未经标注的真实图像，获得了最好的结果。下一步，苹果公司将继续探索为合成图像建立更为逼真的精细图像，同时研究如何对视频进行处理。

[本文作者戈壁创投（微信 ID：gobivc）授权 i 黑马发布。文中所述为作者独立观点，不代表 i 黑马立场。推荐关注订阅号（ID:iheima）图片来自 www.123rf.com.cn]

苹果 机器 人工智能

0

分享到：

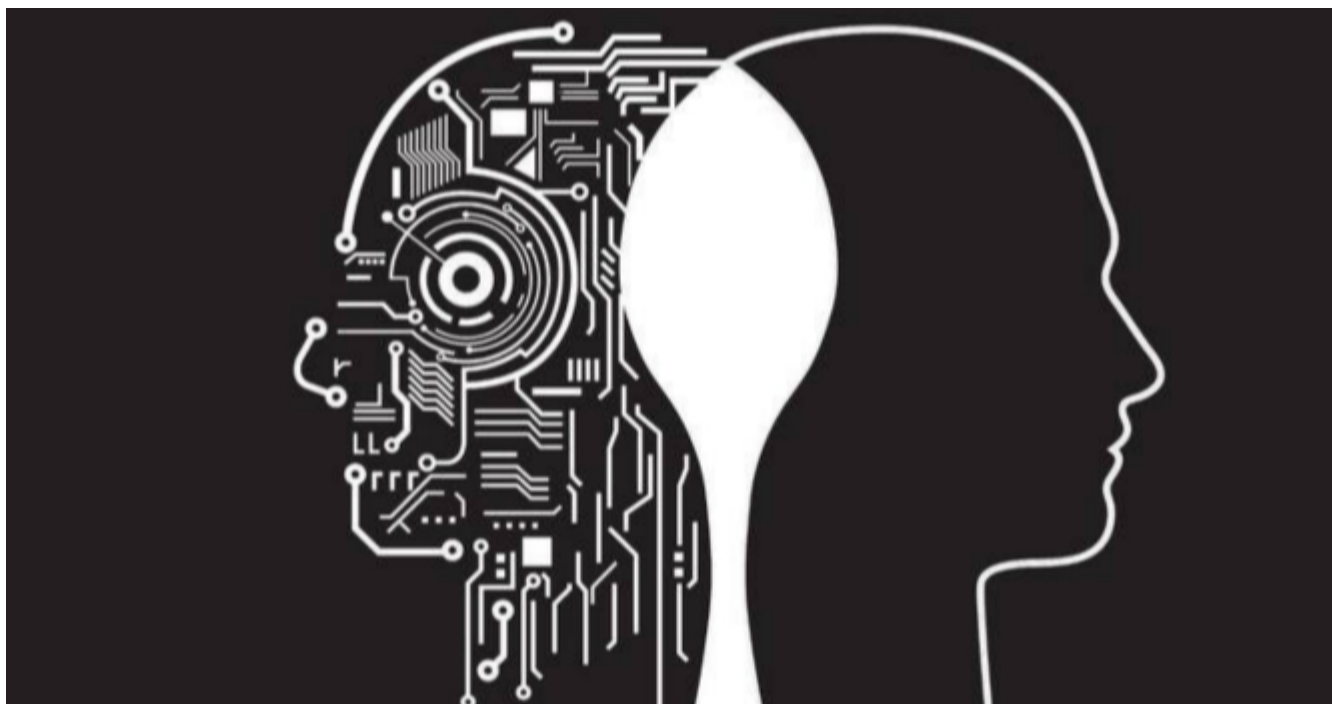


VR 历史上曾失败过，在未来三五年...

•

用打麻将的精神去工作 这世上恐怕就没有什么 干不好的工作了。麻将 精神是什么？

- 一 谁叫随到 从不拖拖拉拉
- 二 不在乎工作环境 专心致志
- 三 不抱怨 经常反省自己 唉又错了！
- 四 永不言败 推倒再来
- 五 牌好牌坏都努力往更好的方向整
- 六 不管跟谁搭档 照样努力
- 七 对于工作中使用的工具从不
挑剔 一样顺手
- 八 最主要的是从不嫌弃工作时间长



MIT 研发出脑控机器人：可使用脑...

文章评论



网友



发布

百度联合创始人徐勇：从建团队、找融资、
应对危机来看百度早期创业



投资人说 2016-12-29 10:29



1999 年，美国全国上下都笼罩在互联网的热潮之中，所有人都在谈论「网」，我们预感到中国的互联网今后将会有一个非常好的发展。

本文由投资人说（微信 ID:touzirenshuo）授权 i 黑马发布，作者徐勇。

1、相识李彦宏

1982 年，我就读于北京大学生物系，1989 年完成生物学硕士学位后，我便赴美留学，随后任加州大学伯克利分校博士后。后来，我先后任职于两家跨国高新技术公司，并在其中担任高级销售经理。

1990 年，在一场同行聚会上，我认识了马东敏（李彦宏的妻子），当时她刚毕业，在一家生物公司做销售。由于我们两人既是中国人，又是同行，所以很自然地成为了好朋友。

1997 年，我通过马东敏认识了他的丈夫李彦宏。Robin（李彦宏）和我都毕业于北京大学，后来他又在美国布法罗纽约州立大学完成了他的计算机科学硕士学位。毕业后，Robin 还担任了道·琼斯公司高级顾问，道·琼斯公司的实时金融系

统就是他设计的，而且这套系统至今仍被广泛地应用于华尔街各大公司的网站，其中包括《华尔街日报》的网络版，同时他还是国际互联网企业——INFOSEEK 的资源工程师，是新一代的互联网技术领域专家。

Robin 对互联网非常感兴趣，并且十分认可硅谷的创业文化和创业机制。1998 年，他出版了《硅谷商战》。这本书记录了 1994 年以来，硅谷几家大企业之间的商业竞争，并进行了一系列深入的分析。**比如，技术本身不是惟一的决定性因素，商战策略才是决胜千里的关键；要允许失败；让好主意有条件孵化；要容忍有创造性的混乱；要有福同享等。**

和 Robin 一样，我同样为硅谷散发出的创业气息而着迷。恰巧在 1998 年这一年，我成为了大型专题纪录片《走进硅谷》的制片人之一，而我们拍摄《走进硅谷》就是希望能够客观地反映硅谷的发展历程，并探求硅谷成功的基因。

于是我便经常找 Robin 一起探讨硅谷和互联网，两人逐渐加深了了解并有了志同道合的感觉。在我拍摄《走进硅谷》时，非常有意思的一次经历是，Robin 扮成了记者，旁听摄制组对雅虎创始人杨致远的采访。

2、创建百度

1999 年，美国全国上下都笼罩在互联网的热潮之中，所有人都在谈论「网」，我们预感到中国的互联网今后将会有有一个非常好的发展。

11 月的一天，Robin 约我到他家谈「大事」，第二天我应约而来。当时，Robin 拿出一份保密协议，并对我说我们回国做一家互联网公司吧。我有些惊讶，但很快又很坦然，因为这在美国是一种常见的商业习惯。于是，我便在保密协议上签了字。

创业的冲动，使我和 Robin 走到了一起。后来，我们两人便一起写好商业计划书，当时我们确定的盈利模式是：做出最好的中文搜索引擎，然后卖给门户网站。商业计划书写好后，我们便开始一同寻找风险投资。

经过一番努力,有三家投资机构表示愿意给我们投资,最终经过几轮艰苦的谈判,其中的两家投资机构决定给我们投资 120 万美元。敲定了投资,我们便从美国硅谷回到了中关村,并创立了百度在线网络技术(北京)有限公司。

最初,公司的办公地点设在北京大学的「资源宾馆」里。之所以把公司选在这里,主要有两个原因:

1) **可以节省办公费用**,当时融到的 120 万美元原本计划半年用完,但我们两人认为应当精打细算,争取能够支撑一年;

2) **北大资源宾馆离北大和其他高校都很近,便于找学校的老师和学生做兼职。**对于百度而言,当时我们的首要任务就是搭建一个技术研发团队,于是我们便把目标锁定在了北京大学计算机系副教授刘建国身上。刘建国曾主持开发过国内第一个大规模中英文搜索引擎系统——天网,所以当时他是中文搜索的先行者和领军人物。

那时让他选择加入我们两个人的初创公司的确需要非常大的勇气,毕竟在美国 95%以上的小公司都会在 3 年内自动消失,而且北大副教授的职位不是谁都能轻易放弃的,但后来我们问刘建国为什么加入时,**刘建国说他意识到只有创业才能推动搜索引擎技术的开发,并让技术实现价值,于是便加入了我们。**

随后,我们又招了一些北大、清华、中科院、交大的硕士毕业生,这些人个个都是写程序的高手,于是我们便开始从事软件体系的总体设计,并分工写程序。

我和 Robin 是比较典型的海归派创业者,和大多数创业者一样,带回来的是硅谷的创业文化。在公司创业初期,我们一直专注于技术,给自己的定位就是搜索服务后台供应商,我们还做了产品计划。我们把主要精力一直放在核心技术(高端)的开发。在我们的团队中,技术人员比例一直保持在 50%左右,而在做技术提供商时还高于这个数字。

3、公司转型

最初，我们将百度定位为是像美国 Inktomi 公司一样做门户网站的后台搜索技术提供商，也就是网民在门户网站上使用搜索时，搜索请求可以被转到百度的服务器，每个搜索收取一定的费用，Inktomi 正是靠这种模式成为了雅虎等门户网站的幕后英雄并获利颇丰。

经过 4 个多月的开发，「百度 1.0」终于问世。**开发出产品后，李彦宏和我便四处出击，做起了兼职销售**，但事情并没有太多进展。2000 年 5 月，在投资人的帮助下我们终于迎来了第一个客户——硅谷动力，每年收取 8 万美元的服务费。在拥有第一个客户后，百度的道路变得平坦了很多。2000 年 6 月，我们正式推出全球最大、最快、最新的中文搜索引擎，并全面进入中国互联网技术领域，凭借着先进的中文搜索引擎技术，我们陆续拿下了一系列的重量级客户：2000 年 8 月，开始服务搜狐；2000 年 10 月，开始服务新浪；2001 年 1 月，开始服务 263；2001 年 2 月，开始服务 TOM.....

因为研发和销售进行得都很顺利，在 2000 年 9 月百度成立 9 个月的时候，我们又进行了第二轮融资，获得美国风险投资机构 DFJ（德丰杰）和 IDG（国际数据集团）的 1000 万美元投资。

然而在 2000 年年底时，美国纳斯达克股市突然崩盘，一直被某些人诟病为「泡沫」的网络经济也在一夜间破裂，这对当时的中国互联网企业产生了巨大冲击。

虽然我们刚刚获得第二轮融资，在短期内不必太担心生存问题，但是我们并没有占据最大的市场，特别是专业搜索引擎门户的出现让我们不得不重新思考百度的商业模式。

而且，此时我们面临着一个非常苦恼的问题：**尽管百度在细分市场已做到中国第一，但却并未实现盈利**。这其中的主要原因是，重视内容的门户网站并不看重搜索，所以百度能收取的费用并不多；又由于受网络经济衰退的影响，很多客户都提出了降低费用的要求，某知名网站甚至还一再拖欠费用。

这时，我们面临着转型这一艰巨任务。尽管我们前后借鉴并尝试了两家美国公司的业务模式，但都失败了。后来经过一系列的考虑，Robin 决定借鉴 Overture 的商业模式，让自己的搜索引擎直接面对网民，靠竞价排名广告赢利。

2001 年 8 月，我们正式推出了面向终端的专业搜索引擎门户 Baidu.com；2001 年 9 月 20 日，Baidu.com 开始竞价排名业务：**第一天的收入是 1.9 元，第二天是 3 元多，第三天超过了 10 元，第五天超过了 200 元**，随着时间的推移，百度的收入始终保持着令人吃惊的增长曲线，在 2003 年的第二季度，百度现金流方面的各项指数全部为正，并开始全面盈利。

4、融资与花钱

现在想来，与百度同时期一起创立的公司一开始都发展得很好，但是网络经济泡沫破裂后，许多公司都倒闭了，最后只有少数几家存活下来并且发展壮大进而盈利了。我想，这应该是主要得益于在网络经济泡沫破裂出现后，**百度对商业模式的调整及对资金的良好控制**。

其实，对于任何一个创业公司来说，控制好融资及花钱都至关重要，我很多创业的朋友也都在资金控制方面踩过许多坑。在这里，我也想分享一下自己对于如何正确融资和花钱的看法。

1) 如何融资？

有些创业者常常认为，一个好的想法和市场前景的方案一定会有好的风险投资来投资，但是现实中他们经常碰壁。**这是因为投资人总是希望最大限度地降低风险，然后才是看盈利**。投资人在选择项目时往往采用风险排除法，他们每天面对大量的商业计划书，所以要迅速从商业计划书中扫描出缺点并且迅速判断筛选。

许多创业者在和投资人谈过以后，常常感觉很奇怪，为什么投资人只字不提项目优势呢？**这是因为创业者和投资人的角度是相反的，创业者往往看到的是项目的优点，而投资人往往是先看项目的缺点或者是评估风险有多大**，任何创业者在融资时都必须认识到这个问题。

2) 如何花钱？

我身边很多的创业朋友融资很成功，但后来因为烧钱太厉害，公司都垮掉了。关于烧钱，我一直坚持的理念就是：**这不仅仅是管理上的问题，更涉及到领导者责任感的问题**。因为投资人对公司投入资金，这是基于他们信任公司的领导者，也是信任公司整个团队，所以在资金控制上公司的领导者一定要负起自己的责任。

[本文由投资人说（微信 ID:touzirenshuo）授权 i 黑马发布，作者徐勇。文中所述为作者独立观点，不代表 i 黑马立场。推荐关注 i 黑马订阅号（ID:iheima）。]

百度创业

赞(1)

分享到：

文章评论



网友



发布

阿里前 CEO 卫哲：B2B 的三差、四率、两大坑！



盛景商业评论 2016-12-29 10:39



我们来看一下大企业为什么做生意容易，小企业为什么做生意难。大企业有大企业的难处，大企业有三个特性比小企业做生意要容易。

本文由盛景商业评论（微信 ID：盛景商业评论）授权 i 黑马发布。

卫哲提出中国 B2B 行业历史远超 B2C 电商，但始终没有迎来大爆发，很大因素是该行业的发展轨迹与实体经济发展规律相反，当经济萧条时反而 B2B 行业出现增长。

他还指出，B2B 是分蛋糕的行业而不是做蛋糕，平台切入垂直产业必须满足“拼”、“聚”的能力，B2B2.0 是化学反应，金融属性是天生，**需要满足“三差”——区域差、规模差、时间差；扩张时需要重视“四率”——覆盖率、转换率、重购率、渗透率。最后卫哲提醒 B2B 创业者不要掉入“两大坑”——补贴陷阱，KPI 只重视交易额。**



以下为卫哲演讲实录：

各位领导、同行，大家上午好！

我在阿里巴巴工作五年，基本上也是负责 B2B 的业务，离开阿里以后创建嘉御基金，我们有三分之一的资本投向了 B2B 领域，也算是从事 B2B 行业的老兵老司机，从业十年了，一旦有 B2B 的同行聚会，一定要和大家见见面，分享一下十年来，尤其是近期的一些感受。

■ 为什么 B2B 春天来得如此之慢？

互联网变化非常快，但是刚才说到了 BAT 当中的 A 阿里巴巴，起家的业务到现在还存在，并且还产生重大贡献的业务就是 B2B 的业务，1999 年开始到现在已经存在了十七年了，中国有多少互联网的商业模式能够真正持续十七年，所以 B2B 实际上是中国最有历史的互联网应用之一，但是为什么一直没有非常非常火，在过去的十二到二十四个月，开始逐渐受到了资本的关注。

首先大的环境时机没有到，B2B 必须诞生在经济的冬天，在经济的春天和夏天的时候 B2B 没有机会，真正经济到了冬天，B2B 才有机会，包括阿里巴巴每个业务上台阶，都和大环境的冬天有关，第一天让阿里巴巴 B2B 活下来的是非典，这个时候人不能见面了，不能坐飞机不能出差，这种活动有利于非典的传播，不

利于交易，所以非典是阿里巴巴一次难得的机遇，而恰恰是当时，因为非典带来的不方便，当时的广交会都没有办法办下去了。

阿里历史上迎来 B2B 第二个重要的发展时期，就是 2008 年的金融危机，对于外贸行业的冲击，使大量的外贸企业失去了以前旱涝保收的好日子，不得不拥抱互联网，刚才几位专家领导说到整个宏观经济的新常态，也就是宏观经济将进入比较漫长的冬天的时候，B2B 的春天反而来了，背后有一个很重要的原因，B2B 或者是 B2C，有了淘宝才有了双十一剁手党，因为你买了不想买的东西不该买的东西，买了不想买的不该买的实际上做了增量，可能有了滴滴和易到，大家打车的频率更高了量更大了，我们经常看到很多互联网 To C 的业务会带来行业增量，To B 是没有增量的。

■ B2B 是分蛋糕，不是做蛋糕

没有一个企业说因为以前传统的采购和销售，我把采购搬到了网上原来买十吨原料，现在买二十吨，不会的，你还是买十吨，而且有可能因为平台效率更高了，你可以买的更少了，你的周转速度更快了，所以 B2B 不是一个做蛋糕的互联网，B2B 是一个分蛋糕的互联网，B2B 只能使这个行业的效率做的更高，而不会把蛋糕盘子做的更大。经济不好的时候做蛋糕的难度大了，只能向内往前引，只能每个行业每个企业提高自己分蛋糕的水平和效率。

这就是为什么 B2B 十七年这么有生命力，生命力这么顽强，必须等到经济不那么好的时候才能爆发机会，或者行业不好的时候才有机会。四万亿、十万亿刺激经济，整个钢材行业“抢钢”的时代，哪有 B2B 电子商务进入钢材交易。像钢联的朱总一样赚点茶水钱服务费，没有机会进入交易的，因为经济不好。宏观经济不好，B2B 的宏观机会来了，任何一个垂直行业的冬天来了。

阿里在十七年前打造 B2B 的时候，诞生了阿里的集团使命“让天下没有难做的生意”，这句话跟实现共产主义差不多，都是一个美好的愿望，天下怎么可能什

么生意都好做，什么生意都好做就没有生意了，阿里的这句话解读一下，天下谁做生意呢？

天下中小企业做生意，所以做 B2B 的天然的我们就要服务中小企业，江南造船厂问宝钢买钢，和我们在座的大家没有关系，大 B 对大 B 是没有我们中间商什么机会的，他们门当户对，自己谈就行了，要我们第三者插足干什么，人家挺愉快的，我们插足没有任何机会。所以 B2B 最理想是买卖两头都是中小企业，那是最理想的状态，至少买卖有一头必须是中小企业。

■ B2B 是小三，插进产业链必须做到“拼”、“聚”

我们来看一下大企业为什么做生意容易，小企业为什么做生意难。大企业有大企业的难处，大企业有三个特性比小企业做生意要容易。

第一个特性，有规模，小企业的采购没有规模，大企业的采购有规模，我们都知道批发和零售有价差，所以对于一个卖家来说都喜欢接大单，所以有规模；第二个特性，大企业有计划，企业都是有计划，消费者有计划；第三个特性，企业越大，计划性越强，企业越小，计划性越小，一个夫妻老婆店开一个餐饮，可能决定每天买什么菜，再大一点可能决定这个月买什么菜，再大的企业可能决定这一年的计划，有计划对于一个卖方很重要，不能临时的做，因为不可能准备大量的库存随时准备你要什么。

第一个有规模、第二个有计划、第三个有稳定，每年都稳定，上游组织供货，有计划、有规模、有稳定，这是大企业的特性，小企业这三个特性都没有，这就是平台的价值。对于单个中小企业来说，三个有都不太可能，但是平台你要想办法把中小企业聚集起来，聚集起来以后，我们要合并同类项，你就可能有规模。要有“聚”和“拼”。

如何“聚”，现在 B2B 的交易额起来了，比 To C 更快，一下子几百亿，一下子幸福的不得了，没花多少钱，交易额起来了，但是最近资本又想明白了，这么热

闹不赚钱，在座的 B2B 能赚钱的不多，能赚大钱的更少。其实赚钱不赚钱取决于你有没有真实的创造多少价值。B 比 C 理性，B 比 C 会算帐，你不创造价值，为什么还要给第三者插足人的钱。我们做 B2B 的都是小三，没有 B2B 之前上下游做了好几百年了。

凭什么你一下子插进去，你要插进去一定要有价值？而价值的来源就是“拼”、“聚”，如何有价值，我们像水力发电一样，得有落差，落差做出来了就有价值，这个落差就是三个差，地方差、规模差、时间差。B2C，淘宝、京东、亚马逊，首先实现的是有组织带有技术含量的窗口，我以前做零售，全国不同城市分区域不同定价，北京卖的价格和上海卖的价格肯定不一样，甚至在全上海都不是同样的价值，利用信息不对称，消费者不知道上海什么价，知道了也没办法过来买。

一旦搬到网上 B2C 首先成功的彻底让全国全世界串货，去年全年流行的海淘，就是国际串货，国际上的货品卖到网上来，淘宝和全球一个价，京东和全球一个价，以前想分区域定价争取不同的价差，在消费者领域这是没有的。但是实际上我们 ToB 的各种分销，原来也是分地区定价，大家不要觉得做黄牛不好，我们做一个有价值观的黄牛，做一个有高科技的黄牛挺好的。

■ B2B 要有“三差”

第一个“差”，区域差。串货串起来了就可以。中国 To B 的产品在不同的地区，或者服务在不同地区价格是不一样的，很多人看不起黄牛，黄牛有存在的价值，更何况我们做一个有使命和有价值的黄牛，我们帮助中小企业做生意更容易，我们做这样的黄牛，所以要会“搬”，这是第一个。地区差，做 To B 的应该比 To C 还要敏感一点，做地方差。

第二个“差”，规模差，这个比较容易理想，大变小小变大，一个餐饮店，一个小的食品城，每个店用的量很少，要想办法聚起来，做 To B 的做到一定规模的，一定是标准化的容易做，容易做的特点是把产品聚集起来，以前我们叫采购联盟，现在不叫联盟，叫平台，利用平台的公信力，利用信息的快速化把需求迅速的聚起来，这过程当中，让中小企业逐渐把原来没有计划变得相对有计划。一个餐厅为什么每天供油供糖，理论上一个月供一次油供两次糖不行吗？

因为以前的随机性造成了很难聚集起来，双规，规定时间规定货物，再小的餐厅也有小仓库，多放几桶油没有问题，以前把他们惯坏了，尽可能贴近原有的采购行为，既对也不对，要转型要改变原来的态度“变态”一下，任何的企业都有可能有计划性，或者说必须有一定的计划性。所以第二个你要能够合并好，要相对把企业需求变得有计划性，每个月十五号订个糖，每个月二十号订个油完全可以做到，如果所有的企业都是这个时间点订油和糖，你就更容易做，这是如何解决规模差的问题。

第三个，时间差，中小企业的零星采购，使得今天的交易变得很辛苦，物流成本变得很高，今天大量的 B2B 交易，本质上是把原来的批发市场，一个专业批发市场搬到了网上，采购行为没有发生改变。做 B2B 的时候没有长短协议、没有现货中远期是不可能的，三个差的核心都是聚和拼，只有解决三个差，你才有价值，你解决的落差越大价值越大。

■ B2B2.0 是化学反应，金融属性是天生的

B2B 有两个自带，天生做 B2B 不需要考虑这两个角度，一个是 To B 既然有计划，有计划就有中远期，有计划就有长短协议，没有计划要想办理创造出来，把大的拆小，把小的拼大。第二个 B2B 自带金融属性，做 To B，不带金融和 To C 就一样了，中国大量 B2B 的低效，到今天几万亿几十万亿的交易，而 To B 交易没有互联网没有电子商务之前，自带金融属性，有预付款，有应收款，有账期，以前叫做贸易条线，这就是金融条线，原有的贸易是自带金融属性的，和自带中远期是一样的，我订一万双鞋子，你分三个也交货，这就是中远期协议，但是这是大企业，耐克向广东的企业一年订十万双，这叫长协中远期，耐克有账期，耐克提供信用证。

所以这是大企业天生的 B2B 自带金融自带长协自带远期合同，而没有金融，没有长协议的全部是中小企业，今天愿意去钢材批发市场，到那现款现货，这是落后的 B2B 贸易，今天做 B2B 贸易的电子化、互联网化，不能停留在刚才说的 1.0 版本。物理发生了变化，发生了物理反映，本来人需要去批发市场，现在人不用去搬在了网上，但是交易动作和行为和你去市场没有改变。消费者不愿意去超市，为了便利，To B 便利永远是不重要的。

阿里有一句话，小企业的钱不是省出来的，你给小企业提省钱省时间是没用的，小企业有的是时间就是没钱，大企业你要说提高效率省时间。小企业有的是时间去批发市场淘点货，我们知道物理反应没有化学反应那么剧烈那么有爆发力，你做的改变是物理变化还是化学变化，你想想交易形态和交易方法是否和以前一样，如果发生化学变化，我们提供的服务模式是传统的方式无法实现的，炸药就是化学反应，我们以前古代人打仗冷兵器拿刀砍你，拿枪捅你，很落后，冷兵器就是物理反应，现代战争化学反应，B2B2.0 要从物理反应变成化学反应。这是和大家分享 B2B 的基本属性，是没有互联网已经存在的，或者说没有互联网很多交易形态只存在于大企业和大企业之间，平台我们要聚集起两头，至少是一头是小企业，让交易条件和交易环境无限的接近大企业的水平。

这样中小企业做生意就没有那么难了。要无限接近大企业，我们说单个小企业没计划，平台应该有计划，单个小企业不稳定，平台应该稳定。单个小企业没有规模，平台应该有规模，平台采购行为应该是把中小企业三无变成三有，很难改变单个企业在平台，如果有几万几十万家中小企业在一起，这些三无是改变成三有，是可以做到的，三无变成三有就是化学反应。B2B 还有密度，B2B 的密度比 To C 还要重要，不是广度是密度，淘宝上卖东西京东上卖东西都是标准化产品，只要物流可以送到交易天津是一样的，可能就是物流配送的成本，To B 不一样，所以 To B 我们看到的是过早的全国化，甚至过早的喊出全球化，不需要这么做。

我们做 To B 买卖，通常有产地、销地，我说的密度是你要不在产地形成上游的密度，要么在销地形成销地的密度。B2B 能够成功，除了前面一些基本原理以外，很重要的是要知道，有没有把密度做起来，阿里巴巴很会做密度，我去阿里的时候阿里全国只有三个分公司，在外贸最发达的三个省份，我去这三个省份走走，先走浙江，浙江阿里巴巴打到了五线城市，阿里在浙江的密度做到了，之后做广州的密度，密到任何一个村任何一个镇。

因为 To B 业务有一个特点，离不开比较重的服务，甚至我说 To B 地推有一定的必须性。今天时间有限不展开讲了，网上有我的课题大家看看就行了，为什么 To B 不能离开地面部队，如果没有地面部队，也没有服务密度，成本就提高

了。物流也是一样，产地和销地不能达到一定的客户密度，仓储物流效率就非常低，所以一定要把密度做起来，把客户密度做起来。

如果做密度，今天和我们的同行见面，没有什么空的东西，都是实实在在把我们怎么做怎么想的怎么成功的和大家分享！

■ 四个率：覆盖率、转换率、重购率、渗透率

第一个月抓覆盖率，这个区域所有需要接触的客户全部扫一遍，这样才有密度，密度的基础就是覆盖率。什么是覆盖？按照电商说的就是开个案，做不做生意没关系，第一个阶段考核覆盖率。

第二个阶段转换率，上个月开户的，有多少人下订单，订单大小不重要，有买卖开张就好。

第三个复购率，上个月买过单的，这个月能不能买第二次，重复购买，再买一次。我们把客户黏性从第二笔生意，一回生二回熟，生意生意，陌生人做的有意义的事做出来了生意就来了。前三个月就抓这三个率。

第四个月开始，一直到未来只做渗透率。To B 的采购量通常很大，有些产品在你这儿买，有些产品为什么不在你这儿买？既然进去开户了，想办法在采购品类里面尽可能渗透，让客户不只买一个产品，一个月可以采购十万个原料，为什么只在你这里采购了一万个原料，另外九万个原料去哪了？这是四个率。

■ B2B 企业两大误区

第一个是太注重“交易额”，交易额是结果，B2B 简单做点交易额很容易，但是没有这四个率做支撑的交易额水中花、镜中月，空的。交易额等于全面的四个率加起来的，你刚开始就抓交易额，你也不知道怎么抓。第一个误区只注重交易额，不注重四个率，只注重广度不注重密度。

B2B 第二个误区，B2B 给你现金补贴。你刚才说消费端补贴很好，你看很多共享经济，都是靠补贴来的，确实是，消费者有的时候通过补贴会养成新的消费习惯，消费习惯变化了客户就变化了。B2B 不可能通过补贴养成新的采购行为，永远会算这种账，其实 B2B 用你的平台给你补贴点钱付的代价很大，如果不能真正带来价值，用现金补贴带来的任何交易没有意义，没有说不能不补贴，不能用现金补贴，或者不能对单笔交易进行补贴。

刚才说到 B2B 三个流，信息流、资金流、物流，通过交易改变物流习惯是可以的，给红包、给积分、给奖励，通过交易行为产生的，但是可能落在物流上，必须使用物流，或者必须使用金融服务，也有可能产生的物流在交易商，产生金融产生在物流上，这是一个铁三角，不能一个角补一块，用一个角产生的收益补另外两块，提升另外两块的使用习惯。当没有补贴的时候，你才会想真正没有带来改变采购销售习惯的价值，这是 To B 第二个非常大的误区。

讲 B2B 可以分享两三个小时，今天时间有限，把 B2B 的本质和大家分享一下，也把我们看到的比较多的 To B 企业犯的错误和大家分享一下。虽然我的题目是“B2B 是冬天的冬天”，但是我觉得春天还有一段时间，B2B 十七年来几乎都处于冬天，没有什么夏天和秋天，但是春天离我们并不远，春天在我们手里。B2B 还有一个好消息，B2B 和产业结合会比 To C 来的快，大家也看到了，我们的八爪鱼都和行业上市公司做了重组整合，这一点在 To C 里面也有，很多零售超市投资互联网。

To B 做 B2B 平台的，我们和行业大佬的关系先远后近，我们并不看好很多所谓的行业大佬，行业大佬自己孵化的 B2B 不太容易成功，你就是革他命的，他自己孵化的 B2B 就是革他原来贸易形态的命的，所以大企业现在都在搞自己的 B2B，昨天还在报道什么什么央企一个亿、十百个亿搞 B2B，你放心做不成的。

还是要脱离原有的大企业或者说行业龙头来合作，这叫先远，但是非常有意思，你做到一定规模，你就特别需要资源，一开始资源过多，反而让你做不成功，但是有一定规模以后，要积极拥抱行业资源，甚至拥抱行业资本，这和 To C 也不一样。