

CBF2018：比 GAN 快一万倍的 LAN 工程案例

作者：何海群 QQ: 357811718(字王)

LAN，是 Log-AI-NET 的缩写，中文名称是：基于逻辑的神经网络算法模型，比目前 GAN 生成式对抗网络、CNN 卷积神经网络等各种 AI 模型，在速度上快一万倍。

有关的工程案例：《中华大字典 2018》（简称 CBF2018），已经发布在 Github 项目网站：

<https://github.com/ziwang-com/CBF2018>

或参见字王网站：<http://ziwang.com/>

对标项目，是 Github 项目网站基于 GAN 的神经网络算法案例：

<https://github.com/kaonashi-tyc/Rewrite>

CBF2018，在纯 CPU(i7-4980HQ)环境下，无需任何 GPU 加速，每小时可生成全套国标二级中文字库，约 7000 个字符，512x512 像素，其中，50%时间是用于 truetype 格式编译，字符图像计算时间，大约 20-30 分钟。

对标案例 Rewrite，采用 GTX 1080 加速：小模型大约需要 20 分钟，而中型模型需要 80 分钟，大型模型需要 2 小时。每组数据只有 3000 个字符，输入数据为 128x128 像素，输出数据才 64*64 像素。

仅从单个汉字的像素尺寸，CBF2018 的运行速度，比对标案例 Rewrite 快 100 倍。

- CBF2018 单字是：512x512=262144
- Rewrite 单字是：64*64=4096

从单字角度对比，CBF2018 的计算量，是 Rewrite 的 64 倍；CBF2018 约 7000 个字符，Rewrite 只有 3000 个字符。

单组数据，运行时间都是 20 分钟左右，由此腿短，CBF2018 的运行速度，比基于 Rewrite 的快 150 倍。

考虑到 Rewrite 有 GTX 1080 加速，而 CBF2018，是纯 CPU(i7-4980HQ)运算，两者硬件的计算速度，大约也有 50-100 倍左右的差距。

因此，单从算法模型而言，基于 LAN 的 CBF2018，比基于 GAN 的 Rewrite 案例，速度快 1 万倍是可以接受的。

之所以出现如此大的差距，这是因为神经网络算法模型的先天缺陷：黑箱模型。

以 TensorFlow, pyTorch 为代表的新一代深度学习、神经网络算法模型，其底层核心是：BP 反向传播函数，以及更上层的 CNN 卷积神经网络。

目前，整个神经网络模型，仍然没有一个完整的理论基础，所有的案例，全部是基于：黑箱模型，或者说是基于：暴力运算。例如，AlphaGo 的简化版本，输入参数就超过一亿个。

因此，各种深度学习、神经网络算法模型，都需要借助大量的 GPU，进行饱和式“暴力运算”，才能获得一定的结果。

Geoffrey Hinton（杰弗里·辛顿），被誉为“神经网络之父”，“人工智能教父”，是 BP 反向传播算法的主要开发者，近年也表示：

他现在对 BP 反向传播算法“深感怀疑”，“我的观点是把它全部抛掉，重头再来。”

对标案例 Rewrite，其实本质上，还是源自 pyTorch 的经典案例：pix2pix 梵高画风，这方面 Github 案例很多，以下是两个相关的模型结构图：

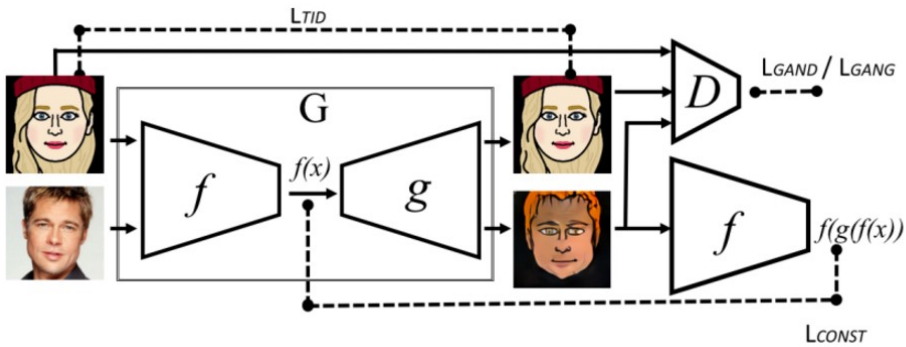


图 1，GAN 模型结构图 1

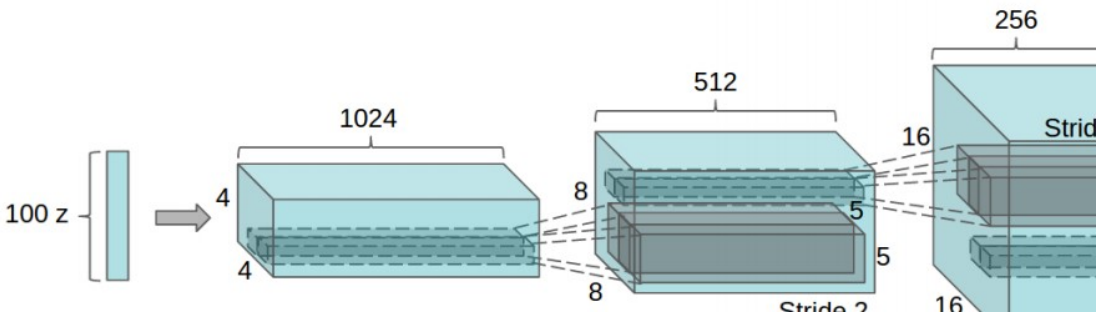


图 2，DC-GAN 模型结构图 2

由图 1、图 2 的 GAN 模型结构图当中可以看出:模型本身并没有相关的逻辑，而是基于两个子模型：G 模型、D 模型的反复迭代，不断修正误差值，来获取最终数据。



图 3，地毯式饱和轰炸

这个类似二战的地毯式饱和轰炸，所以需要大量的 GPU 加速卡，来进行超饱和的“暴

力运算”，破解最终结果。

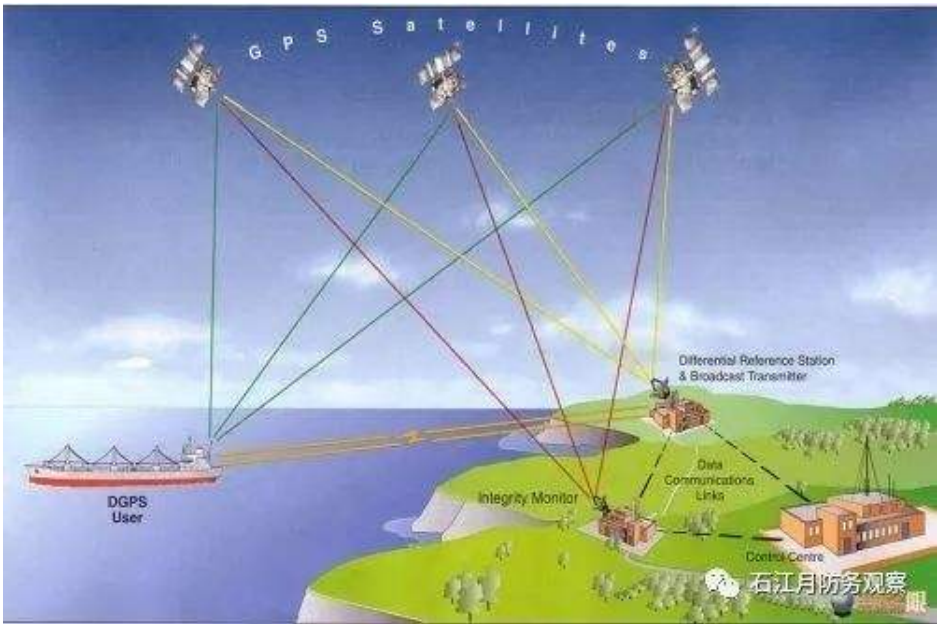


图 4，精确制导

而基于 LAN 的 CBF2018，则类似现代的高科技智能制导武器，精确打击，有的放矢，甚至可以千里之外，执行针对敌方领导人的斩首行动。

图 5 是简化版本的基于 CBF2018 项目的 LAN 算法模型图：

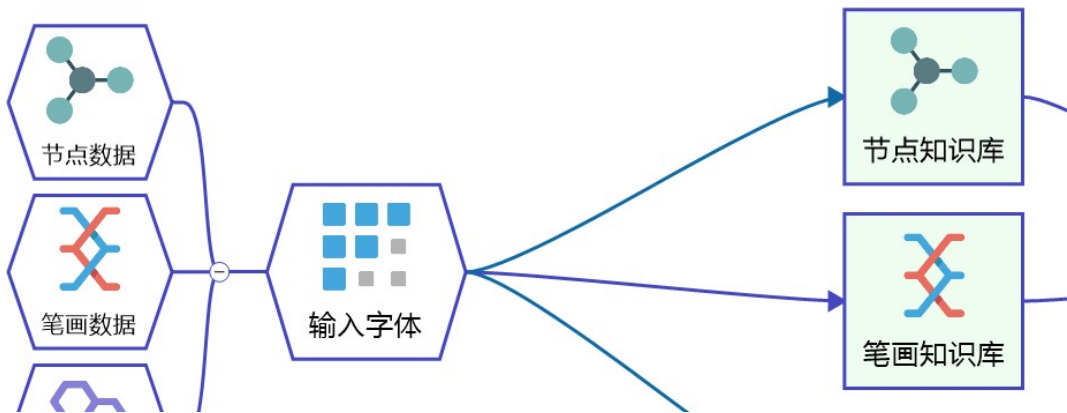


图 5 简化版基于 CBF2018 项目的 LAN 算法模型图

由图 5 可以看出，LAN 算法模型，更多接近古典 AI 人工智能，需要预先建立相关的各种知识库。

通过两个案例的对比，我们可以发现，由于专业知识库的建立，计算时可以根据项目逻辑，淘汰了 99.9% 的无效计算，从而大幅度提高算法模型的计算效率。

不过，LAN 算法模型，还是初级阶段，存在各种各样的缺点：

- 需要先建立相关的各种知识库，这方面可以借助新一代的深度学习算法，比传统手工标注模式，领先很多。
- 通用性差，各种项目，需要根据其内在逻辑，建立不同的算法模型。

不过，相比传统算法模型一万倍效率的提升，LAN 模型的这些缺点，都是可以逐渐克服

的。

附录：

《中华大字库》CBF2018，可以说是目前电脑 AI 中文字库的巅峰之作。

《中华大字库》是个性化中文字库的开创者，也是全球收录字体最多的中文电脑字库。
94 版《中华大字库》CBF1994，收录个性化中文整体 180 款，是全球首套中文字库突破百套大关的中文电脑字库。

2018 版《中华大字库》CBF2018，收录个性化中文整体 4435 款，是全球首套中文字库突破千套，4k 两大关口的中文电脑字库。

预计，2020 版《中华大字库》CBF2020，会推出一万套中文字体，成为全球首套中文字库突破“万套”大关的中文电脑字库。

更多资料，请参见：

字王网站：<http://ziwang.com/>

<https://github.com/ziwang-com/CBF2018>

科研成果，AI 人工智能是不是有效，一个最大的特点就是可以大规模投入应用，以及第三方验证：

CBF2018，采用的 Log-AI-NET（基于逻辑的神经网络算法模型），是目前国内唯一工业级的人工智能项目，无需人工干预，能够批量化、全自动生成“专家级”最终作品。

从创意、设计、到 ttf 字库设计，所有环节全部数字化，可完美仿真人工手写书法，以及泼墨、飞白等传统书法的特殊效果。

2018 版 CBF《中华大字库》升级工程，是国庆节后启动的，短短一个月，就完成了开发平台由 Delphi 到 Python 的移植，以及 4500 余款 AI 中文字库的全部设计工作。