# Title:

Model Complexity of Deep Learning: A Survey

# Abstract：

这篇文章是深度学习模型复杂度的综述；模型复杂度可以分成expressive capacity（表达能力） 和 effective model complexity（有效模型复杂度）；本文回顾了现有在这两个维度上关于四个重要特征的研究：模型框架、模型大小、优化过程、数据复杂度；之后作者讨论了模型复杂度的应用和未来方向。

# Intro：

先介绍背景，深度学习模型在很多领域表现比机器学习好，但是为什么？优化过程是什么？这些问题都和模型复杂度有关系。

之前机器学习对模型复杂度的研究并不能扩展到深度学习中，作者回顾了例如决策树、逻辑回归、贝叶斯层次模型的模型和复杂度，并介绍了他们和深度神经网络的关系。

最后说明分析ML模型的方法不能用在深度神经网络中，作者将在后面的研究中说明他的方法。

# Deep Learning Model Complexity

1. What is Deep Learning Model Complexity?

深度学习模型复杂度其实包含两重意思：

-Expressive capacity（表达能力）：

表达能力描述了模型族复杂度的上界

-Effective model complexity（有效的模型复杂度）：

也称为实际复杂性、实际表现力和可用容量，反映了具有特定参数化的深度模型所表示的函数的复杂性

它们俩的关系如下：

例如，考虑一个函数

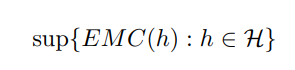


它的表达能力是x的二次方，也就是说，它无法表达比一元二次更复杂的情况，当其参数a，b，c取不同的值时，比如a=0，b=1，c=1，那么它就变成了



其有效复杂度变成了线性，明显低于模型表达能力

他们的关系可以表示为



有效模型复杂度是特定假设的复杂性，表示为EMC(h)，H为假设空间

深度学习模型可以被认为是“容器”，这个“容器”里装着从数据中学到的知识。作为“容器”的相同模型结构可能通过从不同的数据中学习而包含不同数量的知识，并因此配备不同的参数。表达力可以被认为是模型架构能够容纳的知识量的上限。有效模型复杂性关心的是对于一个特定的模型和特定的训练数据集，它包含了多少知识。

2. Important Factors of Deep Learning Model Complexity

深度学习模型由静态结构部分和动态参数部分组成。静态结构部分在学习过程之前由模型选择原则确定，一旦确定则保持不变。参数部分是优化的目标，由学习过程确定。静态和动态部分都对模型复杂性有贡献。

有四个重要的影响因子：

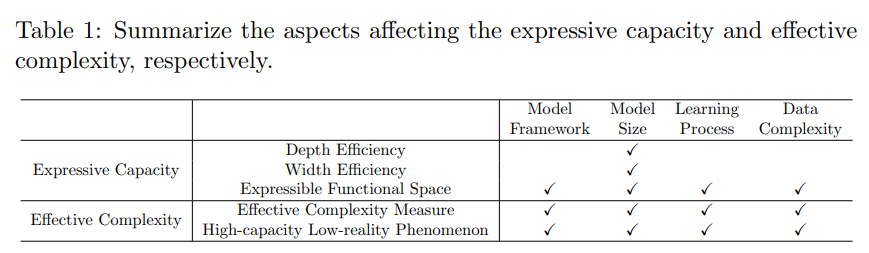
Model framework（模型框架）：模型种类，激活函数等

Model size（模型大小）：参数的数量，隐藏层的数量，filters的数量和大小

Optimization process（优化过程）

Data complexity（数据复杂性）

以下是它们各自对两个维度复杂性的影响

模型框架和模型大小主要影响静态结构部分，优化过程和数据复杂性主要影响动态结构部分

# Expressive Capacity：

表达能力主要从四个方面进行了探索

Depth efficiency（深度效率）：达到同一能力深度比浅层所需参数更少

Width efficiency（宽度效率）

Expressible functional space（可表达的函数空间）

VC Dimension and Rademacher Complexity（VC维数和Rademacher复杂性）：机器学习的经典方法

## 1. Depth efficiency（深度效率）

深度结构的表现要好于浅层结构

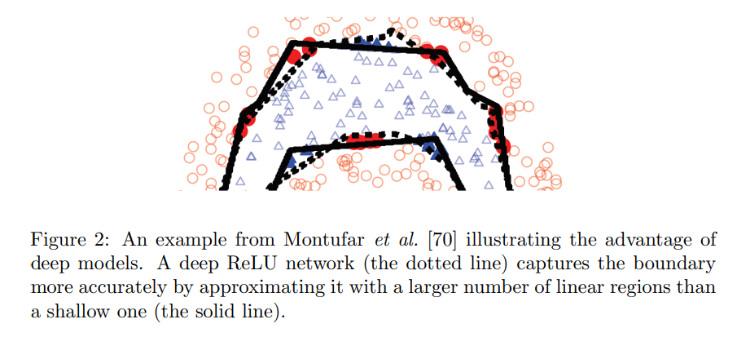
深度效率能分为两个子类别：模型约简方法和表达能力度量

1）Model Reduction

将深度模型简化成可理解的函数，并研究模型深度对表达能力的影响

Bengio和Delalleau对SPN的研究和Mhaskar et al的研究表明，表达同一函数，浅层网络的神经元呈指数增长，而深层网络则是线性增长

ReLU网络（虚线）更准确捕捉到了边界

比起浅层网络

2）Expressive Capacity Measures

深度神经网络具有更高的表达能力，能够学习比浅函数更复杂的函数。

Bianchini和Scarselli通过研究认为，深度模型的分层组成机制允许模型在输入空间的不同区域复制相同的行为，从而使深度比宽度更有效。

## 2. Width Efficiency（宽度效率）

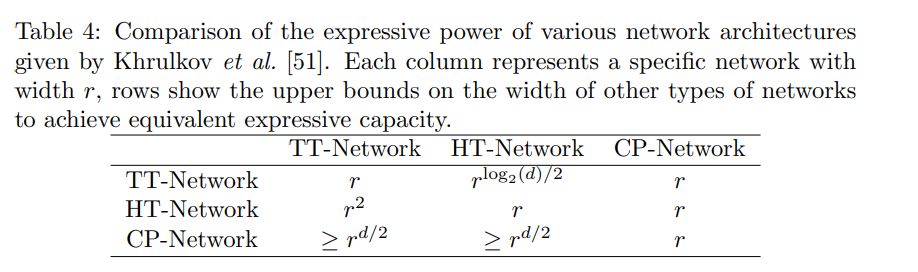
宽度效率分析宽度影响深度学习模型的表达能力。

存在一个浅层ReLU神经网络家族不能用深度受以下约束的窄网络来近似多项式边界。

近似深度线性增加的深层模型，浅层模型至少需要宽度呈指数级增长。一个浅而宽的模型宽度线性增加，一个深而窄的模型至少需要增加一个多项式深度。故宽度效率的多项式下界小于深度效率的指数下界。

Lu等人指出，深度不能严格地证明比宽度更有效，因为宽度仍然不足。深、窄的模型最多需要增加一个多项式深度确保逼近浅、宽的模型的多项式上界。

## 3. Expressible Functional Space（可表达的函数空间）

每列是宽度为r的各种网络，每行是要达到表达能力需要的宽度上限。

得出：浅层网络需要更大的宽度

## 4. VC Dimension and Rademacher Complexity（VC维数和Rademacher复杂性）

VC维数反映了假设空间可以粉碎的最大样本数量，更高VC维度意味着模型可以粉碎更多的样本，因此这个模型有更高的表达能力

Bartlett et al.研究了深度神经网络的VC维数，一个有L层和W个参数的神经网络VC下限是Ω(W Llog(W/L))，上限是O(W Llog W).

Rademacher复杂性是一种捕获假设空间适应随机标注的能力的度量，可以作为表达能力的衡量标准。Rademacher复杂性越大，意味着模型可以适应更多的随机标签，因此模型的表达能力越强。

Bartlett等人研究了具有ReLU激活函数的深度神经网络的Rademacher复杂性。对于具有L层的深度ReLU神经网络，设A;是第i层的参数矩阵，X∈Rn×d是数据矩阵，其中n是样本数，d是输入维度。Bartlett等人证明了此类网络的Rademacher复杂性的下界为Ω(||X||FQi||Aii​||σ)，其中||·||σ是谱范数，||·||F是弗罗贝尼乌斯范数。Neyshabur等人为两层ReLU神经网络证明了更紧的下界。设||A1||σ≤s1，||A2||σ≤s2，s1s2是网络表示函数的Lipschitz界。他们证明了Rademacher复杂性的下界为Ω(s1s2√m||X||F/n)，其中m是隐藏层的宽度。这个下界提高了√m倍。

# Effective Complexity of Deep Learning Model

他表示特定参数化的深度模型所表示的函数复杂性

研究有以下两个方面

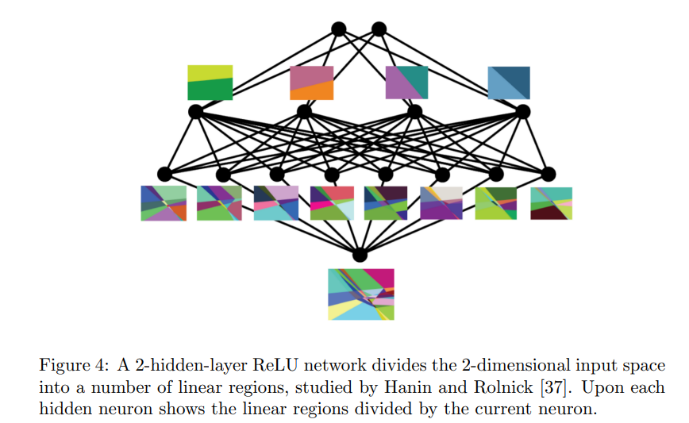
- General measures of effective complexity（有效复杂度的一般衡量标准）

- high-capacity low-reality phenomenon（高容量低真实性现象）：有效复杂度远远低于真实表达能力

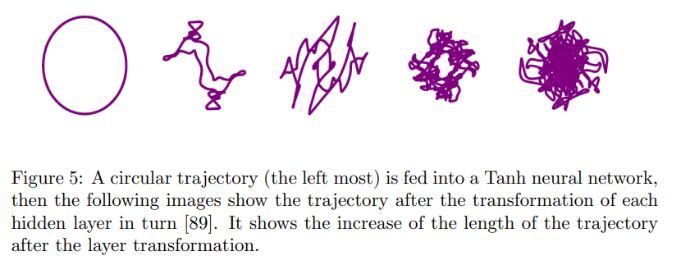
## 1. General Measures of Effective Complexity

与表达能力相比，有效模型复杂性的研究对敏感和精确的度量有更高的要求。这是因为有效复杂性不能仅从模型结构本身直接得出。相同模型结构的不同参数值可能导致不同的有效复杂性。

Piecewise Linear Property



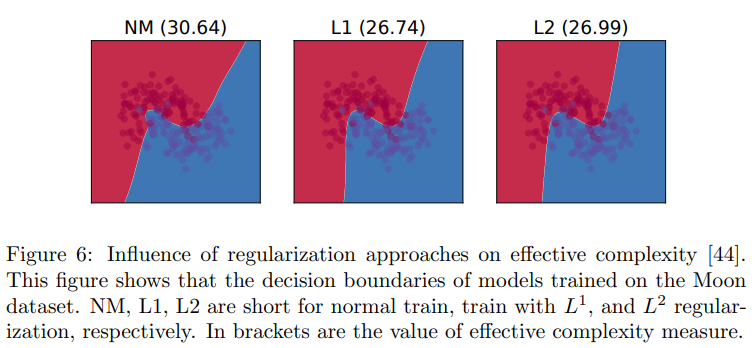
具有分段线性激活函数的神经网络在输入空间中生成有限个线性区域。该性质被称为分段线性性质，如图4所示。线性区域的数量以及这些区域的密度通常可以反映有效复杂性。因此，关于有效复杂性的研究系列始于分段线性激活函数(例如ReLU、Maxout)或基于分段线性性质



一个圆形轨迹（最左边）被送入Tanh神经网络中，下面的图则展示了依次经过每一层隐层转换后的轨迹情况。图中显示了轨迹在各层转换后长度的增长情况

首先，有效复杂性随着模型深度的增加而呈指数增长，随着宽度的增加而呈多项式增长，其次，参数的初始化方式影响有效复杂性。第三，给一个层注入扰动导致其余层以指数级方式出现更大的扰动。最后，正则化方法，例如批量标准化，有助于减少轨迹长度。这就解释了为什么批量标准化有助于模型的稳定性和泛化性

Novak et al证明了神经网络在训练数据流形附近具有较强的鲁棒性，此时深度模型具有良好的泛化能力。他们还表明，与较差的泛化性能相关的因素（例如全批量训练、随机标签）对应于较弱的鲁棒性，而与良好的泛化性能相关的因素（例如数据增强、ReLU）对应于较强的鲁棒性。



使用复杂性度量，Hu等人研究了模型在训练过程中复杂性的趋势。他们表明有效复杂性随着训练迭代次数的增加而增加。他们还证明了过拟合的发生与有效复杂性的增加正相关，而正则化方法（例如L1、L2正则化）会抑制模型复杂性的增加（见图6）

## 2.High-Capacity Low-Reality Phenomenon

一些研究探索了深度学习模型的有效复杂性和表现能力之间的差距。Ba和Caruana 表明，浅层全连接神经网络可以学习以前由深度神经网络学习的复杂功能，有时甚至只要求与深度网络相同的参数数量。特别是，给定一个训练良好的深度模型，他们提议训练一个基于深度模型输出的浅层模型，以模仿深度模型。他们表明，浅层模仿模型可以达到与深度模型一样高的精度。然而，浅层模型不能直接在原始带标签的训练数据上进行训练以达到相同的精度。这也被公认为知识蒸馏。

基于这一现象，Ba和Caruana [5]猜想，深度学习的优势在一定程度上可能来自于深度结构和当前训练算法之间的良好匹配。与浅层结构相比，深度结构可能更容易通过当前的优化技术进行训练。此外，他们提出，当能够使用浅层模型模仿由深度复杂模型学习的功能时，深度模型学习的功能并不是真的太复杂而无法学习。这项研究表明，深度学习模型的实际有效复杂性和其理论表达能力之间可能存在巨大的差距。我们称这种现象为“高容量低现实”现象。

深度神经网络的有效复杂性可能远低于理论界限。也就是说，深度神经网络学习的函数可能并不比浅层神经网络学习的函数更复杂。

# Discussion

有效模型复杂性是深度学习中相对较新、有前途且有用的一个问题。在训练过程中检测有效模型复杂性有助于研究优化算法的实用性、正则化的作用以及泛化能力。此外，有效模型复杂性可用于描述模型压缩比，因为有效模型复杂性可视为模型中信息量的反映。有效复杂性还可用于模型选择和设计，以平衡资源使用和模型性能。

除了有效复杂性度量和高容量低现实现象外，还有一些关于深度学习模型有效复杂性的有趣问题。例如，跨模型的比较值得探索。也就是说，如何比较具有不同架构的多个模型的有效复杂性，以及不同的模型架构选择如何影响有效复杂性。此外，能否指定有效复杂性度量的粒度？不同情况可能对有效复杂性有不同的要求。相应地，有效复杂性的应用范围和粒度应该被指定和阐明。通常情况下，通过非零参数的数量来衡量有效复杂性显然不足以研究优化过程。

# Conclusions and Future Directions

本文对深度学习中的模型复杂性进行了综述。我们总结了影响深度学习模型复杂性的四个方面，以及现有深度学习模型复杂性研究的两个角度。

我们讨论了深度学习模型复杂性的两个主要问题，即模型表达能力和有效模型复杂性。

我们概述了表达能力的最前沿研究，包括深度效率、宽度效率、可表达功能空间、VC维和Rademacher复杂度四个方面。

我们同时概述了有效复杂性的最前沿研究，包括一般有效复杂性度量和高容量低现实现象两个方面。

未来方向：

深度学习模型的表达能力：能否获得对于给定任务来说足够表达能力的深度学习模型的下界

当模型大小成为限制表达能力的瓶颈。

-在深度多项式网络中，如果有一层非常窄，无论其他层有多宽，网络都不能对应于凸函数空间。

但有效复杂性度量

-有效的复杂性度量必须能够捕捉到两个模型之间的细微粒度差异，例如相同模型架构的两个不同优化算法。

跨模型复杂度