

NeuNetS: Automating neural network model synthesis for broader adoption of AI

December 19 2018, by Cristiano Malossi

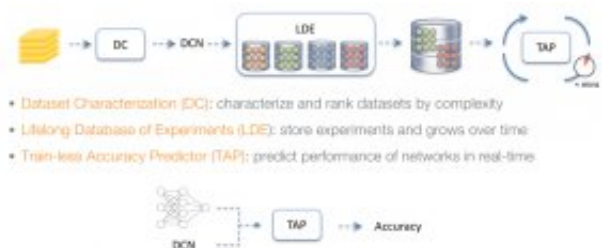


Figure 1: The TAPAS workflow. Credit: IBM

On December 14, 2018, IBM released NeuNetS, a fundamentally new capability that addresses the skills gap for development of latest AI models for a wide range of business domains. NeuNetS uses AI to automatically synthesize deep neural network models faster and easier than ever before, scaling up the adoption of AI by companies and SMEs. By fully automating AI model development and deployment, NeuNetS allows non-expert users to build neural networks for specific tasks and datasets in a fraction of the time it takes today—without sacrificing accuracy.

The need for automation

AI is changing the way businesses work and innovate. Artificial neural networks are arguably the most powerful tool currently available to data

scientists. However, while only a small proportion of data scientists have the skills and experience needed to create a high-performance neural [network](#) from scratch, at the same time the demand far exceeds the supply. As a result, most enterprises struggle to quickly and effectively get to a new neural network that is architecturally custom-designed to meet the needs of their particular applications, even at the proof-of-concept stage. Thus, technologies that bridge this skills gap by automatically designing the architecture of neural networks for a given data set are increasingly gaining importance. The NeuNetS engine brings AI into this pipeline to fast-track results. Using AI for the development of AI models brings a new and much-needed degree of scalability to the development of AI technologies.

Under the hood of NeuNetS

NeuNetS runs on a fully containerized environment deployed on the IBM Cloud with Kubernetes. The architecture is designed to minimize human interaction, automate user workload, and improve over usage. Users do not need to write code or have experience with existing deep learning frameworks: Everything is automated, from the dataset ingestion and pre-processing, to the architecture search training and model deployment. As the field of automating AI is moving at a fast pace, the system needs to be able to take in the latest approaches with minimal impact to the running service. As such, we have designed the NeuNetS framework to be flexible and modular, so that new powerful algorithms can be included at any time. NeuNetS leverages existing IBM assets, such as DLaaS, HPO, and WML. Neural Networks models are synthesized on the latest generation NVIDIA Tesla V100 GPUs.

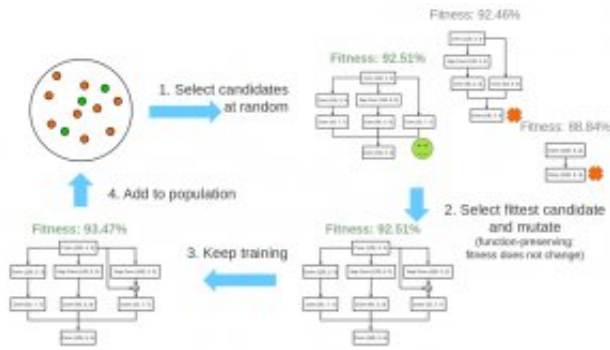


Figure 2: The NCEvolve workflow. Credit: IBM

Bleeding-edge research technology

NeuNetS algorithms are designed to create new neural network models without re-using pre-trained models. This allows us to explore a wide space of network architecture configurations and at the same time fine-tune the model for the specific dataset provided by the user.

The NeuNetS algorithm portfolio includes enhanced versions of recently published works, such as TAPAS [3], NCEvolve [4], and HDMS [5], as well as a fine-grain optimizer engine. These algorithms make a step forward with respect to the state-of-the-art in the literature and in practice, addressing fundamental problems such as dataset generality and performance scalability. TAPAS is an extremely fast neural-network synthesizer, performing close to transfer-learning approaches by relying on pre-generated ground-truth and smart prediction mechanisms. NCEvolve synthesizes top-performant networks, minimizing the amount of training time and resource needs. HDMS combines an improved version of hyperband with reinforcement learning to synthesize networks tailored for the less common datasets. Last but not least, our fine-grain synthesis engine uses an evolutionary algorithm for building custom convolution filters, leading to low-level fine-tuning of the neural

architecture.

Future of NeuNetS

Based on multiple optimization algorithms and a modular architecture, NeuNetS can accommodate a wide range of model synthesis scenarios. A next step is enabling users not only to update data, but to also decide how much time and how many resources to allocate for the model synthesis, as well as optionally the maximum size of the model, and the target deployment platform. In this respect IoT and time series analysis workloads will play a big role. To enable users to make effective use of the synthesized models, we are creating innovative visualization capabilities for comparing key model characteristics including performance, size and type. To continue assisting users once a model is deployed and furthering their trust in AI, we are working on techniques that improve visibility into the model's structure and behaviour across the AI lifecycle.

Try NeuNetS now

NeuNetS beta is available today as part of AI OpenScale product in Watson Studio, on the IBM Cloud. This first release offers [model](#) synthesis for image and text classification, with performance similar to that of hand-crafted [neural networks](#). Visual workloads have been the subject of intense research, development, and competitions over the past decade and thus represent a tough benchmark. In contrast, high accuracy models for text are not wide-spread today, and NeuNetS will help non-expert users to profit from the latest technology available in this domain.

You can get access at this link: dataplatfom.cloud.ibm.com/ml/neunets .

More information: NeuNetS demo: neunets.mybluemix.net

R. Istrate et al. TAPAS: Train-less Accuracy Predictor for Architecture Search. arXiv:1806.00250 [cs.LG]. arxiv.org/abs/1806.00250

Deep Learning Architecture Search by Neuro-Cell-based Evolution with Function-Preserving Mutations. [www.ecmlpkdd2018.org/wp-content ... loads/2018/09/108.pdf](http://www.ecmlpkdd2018.org/wp-content/uploads/2018/09/108.pdf)

This story is republished courtesy of IBM Research. Read the original story [here](#).

Provided by IBM

Citation: NeuNetS: Automating neural network model synthesis for broader adoption of AI (2018, December 19) retrieved 24 April 2024 from <https://phys.org/news/2018-12-neunets-automating-neural-network-synthesis.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.