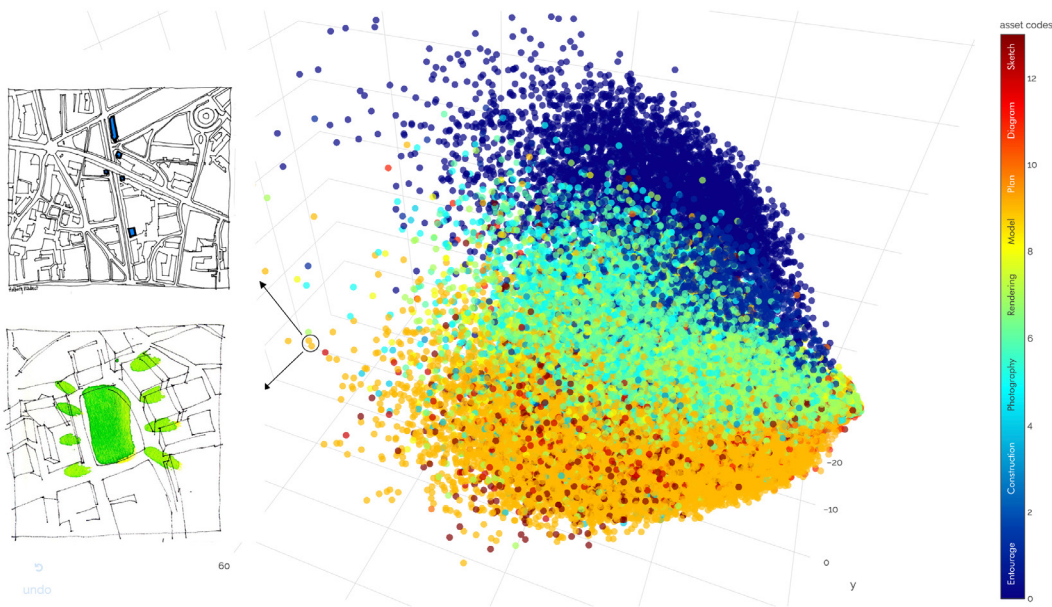


AI ACQUIRES AN EYE FOR ARCHITECTURE

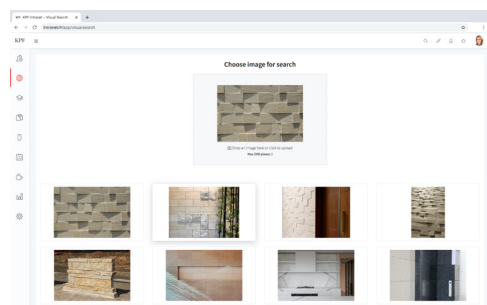
KPF Utilizes ResNet-50 and RAPIDS Nearest Neighbors for Architectural Classification



Visual asset clusters in the PCA-reduced 3D space

KPF realizes large performance gains with ResNet-50 and RAPIDS nearest-neighbors search algorithm for architectural classification on NVIDIA Quadro RTX 8000 and BOXX Data Science Workstation.

ABOUT KPF



Visual Search - User Interface

Kohn Pedersen Fox Associates (KPF) is a leading global architecture firm that designs buildings of all types and scales in all geographic regions. Founded in 1976 and headquartered in New York City, they have offices in London, Shanghai, Hong Kong, Seoul, Abu Dhabi, San Francisco, Singapore, and Berlin, providing architecture, interior, programming, and master plan services for both the public and private sectors.

THE CHALLENGE

KPF relies on a machine learning (ML) pipeline to perform visual searches on their vast image data store when a tag search isn't working well, for instance, when an asset isn't properly categorized or when it's incorrectly tagged. The KPF ML pipeline ran on a CPU-only Linux server based on TensorFlow ResNet-50 for image feature extraction and a scikit-learn k-nearest-neighbors (kNN) algorithm for proximity search. The process, although functional, didn't perform as quickly or efficiently as necessary. To accelerate it, enable further refinements, and improve overall pipeline performance, KPF needed a hardware solution purpose-built for their specific data science workflow.

CUSTOMER PROFILE

KPF

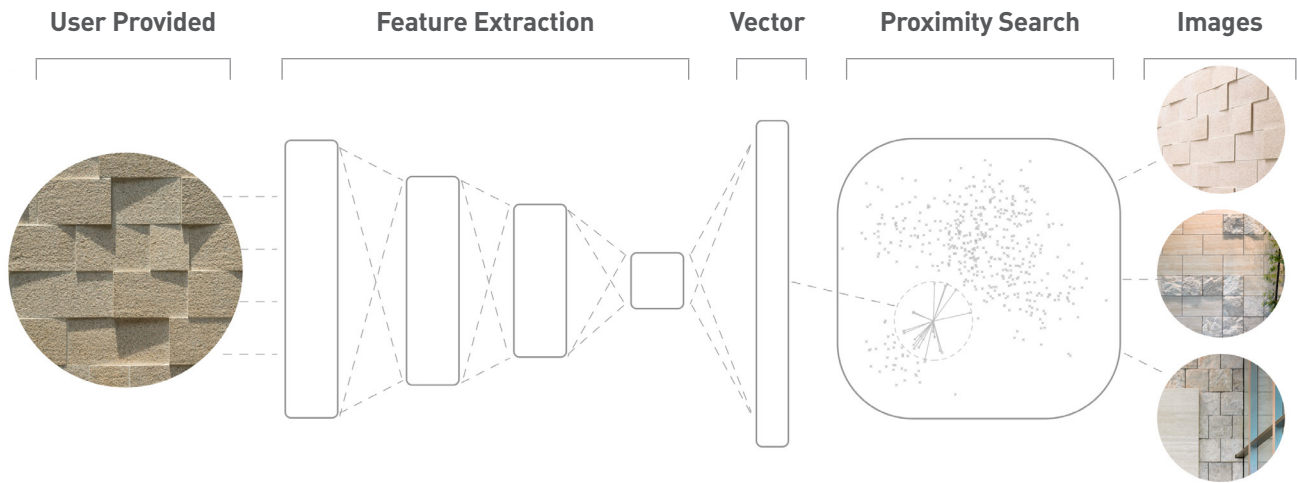
Organization
Kohn Pedersen
Fox Associates
(KPF)

Industry
Architecture

Location
New York City (Headquarters),
London, Shanghai, Hong
Kong, Seoul, Abu Dhabi, San
Francisco, Singapore, Berlin

Founded
1976

Website
kpf.com



Visual Search - Solution Architecture Diagram.

TESTING A POSSIBLE SOLUTION

KPF elected to test a BOXX Data Science Workstation (DSWS), a Linux-based system purpose-built for deep learning and equipped with an NVIDIA RAPIDS™ ML library and an NVIDIA® Quadro RTX™ 8000 GPU.

The complete product specifications are:

BOXX DATA SCIENCE WORKSTATION

GPU: NVIDIA Quadro RTX 8000

Processor(s): Dual Intel Xeon SP Silver 4210 2.2 GHz 13.75 MB cache, 9.6 UPI (Ten-Core)

System Memory: 128 GB DDR4-2933 MHz ECC REG (8-16 GB DIMMS)

Optical Drive: 20X Dual Layer DVD±RW Writer

M.2 Storage: 1.0 TB SSD M.2 PCIe Drive

Operating System: Ubuntu Desktop 18.04

Software Applications: NVIDIA CUDA-X AI™ Data Science Workstation Software Stack

One of the key advantages of the RTX 8000 GPU is its large memory capacity, which accommodates larger batches of data but also processes the data at a much faster rate. Because of this, KPF data scientists could run more experiments and then compare and evaluate the results of each one, selecting those that yielded the best outcomes. Most importantly, the BOXX test system arrived with a complete, preinstalled, GPU-accelerated data science software stack that offered significant advantages:

- > Eliminated set-up time with simple plug-and-play functionality
- > Eliminated the need for data scientists to perform IT tasks, such as manually building a data science software stack
- > Featured the popular data science software packages Docker, Python, pandas, scikit-learn, NumPy, and Numba
- > Included the GPU-accelerated libraries TensorFlow, PyTorch, RAPIDS, and XGBoost, which significantly accelerate data science workloads

“Exploring new ways to apply ML in architectural design practice at KPF, we needed to accelerate our capabilities to train and iterate models faster. With the NVIDIA RTX 8000-powered BOXX Data Science Workstation, we’re able to train larger models faster than ever before. Having the workstation arrive with the GPU-accelerated data science software stack pre-installed assisted us in getting running quickly with no additional configuration required.”

Cobus Bothma,
Director, Applied
Research, KPF

www.nvidia.com



© 2020 NVIDIA Corporation. All rights reserved.

BENCHMARKING

Loading the Dataset

To perform a benchmark study, KPF refactored the ML pipeline’s codebase to take advantage of RAPIDS ML cuDF, a GPU-enabled dataframe. That change made a solid improvement in the data read time: Using cuDF, it took 3 seconds to load the 2-gigabyte (GB) dataset—which was comprised of over 100,000 visual assets that translated to around 250 million data points— versus the 64 seconds it took using scikit-learn pandas.

Training

To perform proximity searches within the ML pipeline, the dataset needed to be fit into a kNN tree using the scikit-learn library. This process usually takes 35-60 seconds. To take advantage of faster GPU computational speeds, KPF replaced the scikit-learn k-NN model with a RAPIDS cuML nearest-neighbors model. Thanks to that change, this step in the pipeline was reduced to approximately 10-11 seconds.

Searching (Inference)

The last part of the ML pipeline, performing proximity searches, also improved with the cuML model. The process now takes approximately 300 milliseconds, down from 35–60 seconds. KPF noticed that the new cuML model performed best when the input data was converted to a NumPy array instead of cuDF.

NEXT STEPS

KPF builds their visual asset library—the dataset used in this pipeline—through feature extraction using the ResNet-50-based model, which takes 2.6 seconds. KPF’s data scientists believe they can accelerate this process by taking advantage of the GPU-enabled TensorFlow built into the BOXX DSWS.

KPF would also like to test principal component analysis (PCA) or temporal social network analysis (TSNA) for dimensionality reduction, to provide interactive visualization of all the assets in their catalog. Currently, KPF uses Plotly’s Dasher for 3D visualization and scikit-learn PCA and are incorporating techniques that will render results faster and enable smoother interaction.

