

Efficient Weather Forecasting using Quantized Latent-Diffusion Models

Type

Initial proposal

Principal investigator

PD Dr. Haojin Yang

Person of contact

Jona Otholt

Size

NHR Normal

Whitelist status

No

Short abstract

Accurate weather forecasts are instrumental to preventing or mitigating the impact of extreme weather events. Currently, numerical weather prediction methods are used to calculate such forecasts, but these methods require considerable amounts of compute time for each forecast, resulting in high energy consumption. Recently, data-driven methods relying on deep learning have achieved promising results, at significantly lower power consumption than numerical methods. As part of the BMUV-sponsored project "Neue energieeffiziente KI-Algorithmen für innovative Prognosemethoden von Extremwetterereignissen" (EKAPEX), we aim to take this trend to the next level by further increasing the efficiency of deep learning-based forecasting methods. The compute project consists of two major steps: First, we will develop a novel deep learning forecasting method based on latent diffusion models, then we will further improve the model's efficiency by employing low-bit quantization methods. Quantization reduces the model's memory footprint and increases its execution speed by replacing floating point parameters with quantized versions that use fewer bits to represent the model's parameters.