

GENERATIVE ADVERSARIAL NETWORKS QUANTIZATION

E. Mitrofanov *, *V. Grishkin* **

Saint Petersburg State University, Saint Petersburg, Russia

Generative models have become widespread over the past few years, taking valuable part in content creation. Generative adversarial networks (GAN) are one of the most popular generative model types. However, computational powers required for training stable, large-scale and high-resolution models can be enormous, making training or even running such models an expensive process. Study of neural network optimization proposes different techniques of lowering required GPU memory, fastening training time and creating more compact models without noticeable loss in generated sample quality. We apply quantization techniques to GAN and estimate results with a custom dataset.

Генеративные модели получили широкое распространение за последние несколько лет, занимая значительную часть в создании контента. Генеративно-сопоставительные сети (GAN) — один из самых популярных типов генеративных моделей. Однако вычислительные мощности, необходимые для обучения стабильных, крупномасштабных моделей с высоким разрешением изображений, могут быть огромными, что делает обучение или даже запуск таких моделей дорогостоящим процессом. Исследование оптимизации нейронных сетей открывает различные методы снижения требуемой памяти графического процессора, сокращения времени обучения и создания более компактных моделей без заметной потери качества генерируемых данных. В этом исследовании применяются методы квантования к GAN и оцениваются результаты с помощью собранного набора данных.

PACS: 44.25.+f; 44.90.+c

* E-mail: st054628@student.spbu.ru

** E-mail: valery-grishkin@yandex.ru