

Batching strategy untuk optimisasi Densely Connected Convolutional Networks pada GPU = Batching Strategy for Densely Connected Convolutional Networks Optimization on GPU.

Ari Nugroho, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20507953&lokasi=lokal>

Abstrak

ABSTRAK

Densely Connected Convolutional Networks (DenseNet) merupakan salah satu model arsitektur Deep Learning yang menghubungkan setiap layer beserta feature-maps ke seluruh layer berikutnya, sehingga layer berikutnya menerima input feature-maps dari seluruh layer sebelumnya. Karena padatnya arsitektur DenseNet menyebabkan komputasi model memerlukan waktu lama dan pemakaian memory GPU yang besar. Penelitian ini mengembangkan metode optimisasi DenseNet menggunakan batching strategy yang bertujuan untuk mengatasi permasalahan DenseNet dalam hal percepatan komputasi dan penghematan ruang memory GPU. Batching strategy adalah metode yang digunakan dalam Stochastic Gradient Descent (SGD) dimana metode tersebut menerapkan metode dinamik batching dengan inisialisasi awal menggunakan ukuran batch kecil dan ditingkatkan ukurannya secara adaptif selama training hingga sampai ukuran batch besar agar terjadi peningkatan paralelisasi komputasi untuk mempercepat waktu pelatihan. Metode batching strategy juga dilengkapi dengan manajemen memory GPU menggunakan metode gradient accumulation. Dari hasil percobaan dan pengujian terhadap metode tersebut dihasilkan peningkatan kecepatan waktu pelatihan hingga 1,7x pada dataset CIFAR-10 dan 1,5x pada dataset CIFAR-100 serta dapat meningkatkan akurasi DenseNet. Manajemen memory yang digunakan dapat menghemat memory GPU hingga 30% jika dibandingkan dengan native DenseNet. Dataset yang digunakan menggunakan CIFAR-10 dan CIFAR-100 datasets. Penerapan metode batching strategy tersebut terbukti dapat menghasilkan percepatan dan penghematan ruang memory GPU.

<hr>

ABSTRACT

Densely Connected Convolutional Networks (DenseNet) is one of the Deep Learning architecture models that connect each layer and feature maps to all subsequent layers so that the next layer receives input feature maps from all previous layers. Because of its DenseNet architecture, computational models require a long time and use large GPU memory. This research develops the DenseNet optimization method using a batching strategy that aims to overcome the DenseNet problem in terms of accelerating computing time and saving GPU memory. Batching strategy is a method used in Stochastic Gradient Descent (SGD) where the technique applies dynamic batching approach with initial initialization

using small batch sizes and adaptively increased size during training to large batch sizes so that there is an increase in computational parallelization to speed up training time. The batching strategy method is also equipped with GPU memory management using the gradient accumulation method. From the results of experiments and testing of these methods resulted in an increase in training time speed of up to 1.7x on the CIFAR-10 dataset and 1.5x on the CIFAR-100 dataset and can improve DenseNet accuracy. Memory management used can save GPU memory up to 30% when compared to native DenseNet. The dataset used uses CIFAR-10 and CIFAR-100 datasets. The application of the batching strategy method is proven to be able to produce acceleration and saving of GPU memory.