

Using GPU acceleration

GPU support in MatConvNet builds on top of MATLAB GPU support in the Parallel Programming Toolbox. This toolbox requires CUDA-compatible cards, and you will need a copy of the corresponding CUDA devkit (<https://developer.nvidia.com/cuda-toolkit-archive>) to compile GPU support in MatConvNet (see compiling (`../install#compiling`)).

All the core computational functions (e.g. `v1_nnconv`) in the toolbox can work with either MATLAB arrays or MATLAB GPU arrays. Therefore, switching to use the a GPU is as simple as converting the input CPU arrays in GPU arrays.

In order to make the very best of powerful GPUs, it is important to balance the load between CPU and GPU in order to avoid starving the latter. In training on a problem like ImageNet, the CPU(s) in your system will be busy loading data from disk and streaming it to the GPU to evaluate the CNN and its derivative. MatConvNet includes the utility `v1_imreadjpeg` to accelerate and parallelize loading images into memory (this function is currently a bottleneck will be made more powerful in future releases).