

Graphics processing units (GPUs) were originally used solely for the purpose of graphics rendering. This changed with the introduction of technologies like CUDA that enabled to use graphics processors as any other computing device. However, writing an efficient program for GPUs, also called GPU kernel, is one of the most difficult programming disciplines. The latest research in the field suggests that these difficulties could be potentially mitigated with machine learning methods. One especially successful approach is based on the utilization of recurrent neural networks (RNNs) over different representations of source code. In this work, we present two RNN-based solutions that are able to derive performance characteristics of a CUDA GPU kernel directly from its intermediate representation called PTX. We assess the applicability of our two methods in two GPU optimization tasks. In heterogeneous device mapping task, our methods are able to achieve accuracies of around 82%, results that are slightly worse than the current state of the art. In a more challenging achieved occupancy task, where the goal is to correctly predict one out of ten classes, our two methods achieve accuracies above 50%. These promising results indicate great potential in additional research focused in a similar direction.