

Grafické výpočetní jednotky (GPU) byly původně používány výhradně pro účely grafického vykreslování. To se změnilo zavedením technologií jako je CUDA, které umožnily použití grafických procesorů jako kterýchkoliv jiných výpočetních zařízení. Psaní efektivního programu pro GPU, nazývaného také GPU kernel, je však jednou z nejnáročnějších programovacích disciplín. Nejnovější výzkum v této oblasti naznačuje, že tyto problémy by mohly být potenciálně zmírněny metodami strojového učení. Jeden obzvláště úspěšný přístup je založen na využití rekurentních neuronových sítí (RNNs) nad různými reprezentacemi zdrojového kódu. V této práci představujeme dvě řešení založená na RNNs, která jsou schopna odvodit výkonové charakteristiky CUDA GPU kernelu přímo z jeho mezikódu, který se nazývá PTX. Použitelnost námi vyvinutých dvou metod vyhodnocujeme ve dvou optimalizačních úlohách pro GPU. V první úloze jsou naše metody schopny dosáhnout přesnosti okolo 82 %, což jsou výsledky, které jsou pouze o málo horší než kterých dosahují nedávné práce v této oblasti. Při druhé, náročnější úloze, kde je cílem správně klasifikovat jednu z deseti tříd, dosahují naše dvě metody přesnosti nad 50%. Tyto slibné výsledky naznačují velký potenciál dalšího výzkumu, který by se ubíral podobným směrem.