

Studio dello speedup della convoluzione usando CUDA

Progetto Parallel Programming for Machine Learning

A.A. 2021/2022

Eduard Marchidan

1 Convoluzione

La convoluzione è l'operazione matematica con la quale si aggiunge a ciascun elemento dell'immagine gli elementi nel suo intorno pesati da un kernel. L'equazione che rappresenta l'operazione è:

$$g(x, y) = k * f(x, y) = \sum_{dx=-a}^a \sum_{dy=-b}^b k(dx, dy) f(x - dx, y - dy)$$

Nello studio viene utilizzato il kernel gaussiano:

$$\frac{1}{16}K = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Usando l'algoritmo implementato il risultato è: L'effetto della sfocatura non



(a) Immagine iniziale



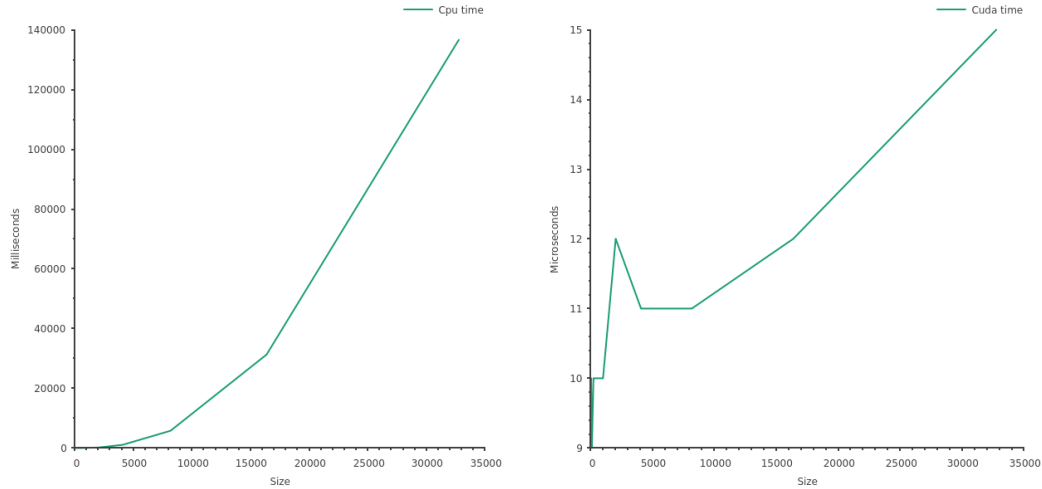
(b) Risultato della convoluzione

Figure 1

è molto forte ma possiamo notare che i numeri e soprattutto le scritte ai lati sono effettivamente meno visibili.

2 Implementazione e risultati

L'algoritmo è stato implementato sequenzialmente usando la cpu e poi utilizzando CUDA. I test sono effettuati su una matrice inizializzata casualmente di grandezza variabile tra (8×8) e (32768×32768) pixel. Importante cosa



(a) Tempo di esecuzione all'aumentare della grandezza dell'immagine

(b) Risultato della convoluzione

Figure 2

da notare, il tempo di esecuzione è in millisecondi mentre il tempo della gpu è in microsecondi! Inizialmente sono stati effettuati test fino a risoluzione (8192×8192) e la gpu eseguiva in tempo costante la convoluzione, probabilmente perchè erano disponibili ancora thread da utilizzare e quindi poteva parallelizzare ancora. Oltre quella soglia il tempo di esecuzione cresce in maniera lineare. La cpu invece scala in tempo pressochè esponenziale oltre al fatto che impiega secondi per eseguire la convoluzione.

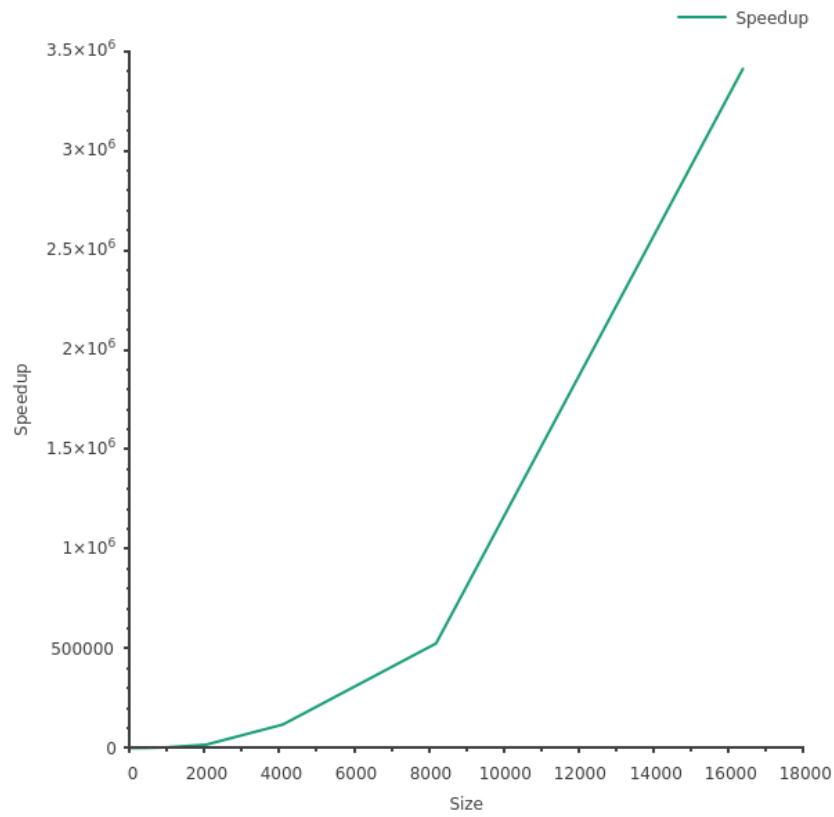


Figure 3: Speedup cpu/gpu

Lo speedup è lineare e la gpu è nell'ordine di 10^6 volte più veloce. E' evidente che le schede video sono un hardware specializzato nell'esecuzione di questo tipo di operazioni.