

# Vidění robotů - tréninková úloha - řešení

Jan Chleboun

29. října 2021

Mějme konvoluční síť s jednou vrstvou, která provádí konvoluci s třídímenzionálním jádrem  $4 \times 5 \times 3$ , padding = 1, stride = 1. Vstupem je obrázek  $100 \times 100$ , který má 4 složky - R, G, B a hloubkovou informaci. Spočítejte počet operací, které proběhnou při průchodu jednoho obrázku danou konvolucí (jedna operace je buď sečtení dvou prvků, nebo vynásobení dvou prvků). Má v takovémto případě 3D konvoluce smysl (pokud ne, navrhněte lepší řešení)?

- Nejprve si uvědomíme, že po aplikaci paddingu bude mít náš obrázek rozměry  $102 \times 102 \times 6$ . 3D konvoluci si rozdělíme do 3 směrů.
- V horizontálním směru se do obrázku vejde konvoluční jádro  $n_x$  krát, kde

$$n_x = (\text{šířka obrázku} - \text{šířka jádra}) + 1 = (102 - 4) + 1 = 99$$

- Ve vertikálním směru uděláme obdobnou úvahu, jádro se vejde  $n_y$  krát, kde

$$n_y = (\text{výška obrázku} - \text{výška jádra}) + 1 = (102 - 5) + 1 = 98$$

- Musíme vzít v potaz také průchod napříč jednotlivými kanály. Zde se jádro vejde  $n_z$  krát, kde

$$n_z = (\text{počet kanálů} - \text{hloubka jádra}) + 1 = (6 - 3) + 1 = 4$$

- Získáme počet "otisků jádra", jako

$$n = n_x n_y n_z = 99 \cdot 98 \cdot 4 = 38808$$

- Počet operací pro jeden otisk jádra je

Násobit 2 prvky budeme  $4 \cdot 5 \cdot 3 = 60$  krát

Tyto vynásobené prvky poté posčítáme, sčítat budeme tedy 59 krát.

- Na jeden "otisk jádra" tedy připadne  $60 + 59 = 119$  operací.
- Celkový počet operací je tedy

$$119 n = 119 \cdot 38808 = 4618152$$

- Nyní se zamyslíme nad smyslem zadané úlohy - provádíme zde 3D konvoluci obrázku, který obsahuje hloubkovou složku - tím vlastně "mícháme jablka a hrušky" - do informací o hranách v obrázku přidáváme informaci vzdálenosti, tyto spolu ale vůbec nemusí souviset. Dalším problémem je, že naším cílem je detekovat stejný objekt ať už je v jakékoli vzdálenosti, informace o ní nám tedy spíše škodí a vůbec ji nepotřebujeme. Lepším řešením by tedy byla například 2D konvoluce s jádrem  $3 \times 3$  sečtená přes barevné kanály, kde bychom hloubkovou informaci do sítě vůbec neposílali.