

Metódy hlbokého učenia pre segmentáciu obrazu

Autor: Maroš Plšík
Vedúci práce: Ing. Peter Bednár, PhD.

Technická Univerzita v Košiciach
Fakulta elektrotechniky a informatiky

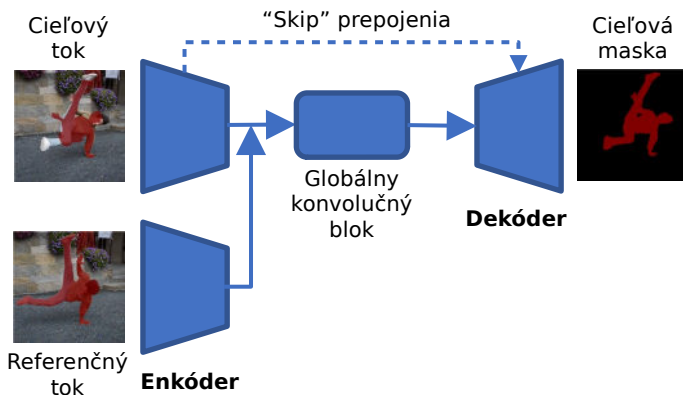


Motivácia

Cieľom segmentácie objektov vo videách (VOS) je oddeliť sledované objekty od pozadia na úrovni pixelov. Alternatíva riešenia s čiastočným dohľadom načíta na vstupe prvú anotovanú snímku videa, z ktorej sa následne naučí reprezentáciu sledovaných objektov a na výstupe segmentuje tieto objekty vo zvyšku videa. S novými technológiami ako virtuálna realita alebo autonómne autá, ktoré sa postupne stávajú súčasťou našich životov, úloha VOS takisto rýchlo naberať na dôležitosti. Cieľom tejto práce bolo zvýšenie kvality segmentácie u existujúceho riešenia použitím novej rozsiahlej trénovacej množiny a modifikáciou samotného trénovacieho procesu.

Riešenie

Riešenie navrhnuté v tejto diplomovej práci je založené na existujúcom a overenom prístupe, ktorý dosahuje vysokú kvalitu segmentácie pri zachovaní krátkeho času potrebného na spracovanie vstupných sekvencií. Tento prístup, nazývaný RGMP, je postavený na štruktúre známej ako enkóder-dekóder. V práci bola architektúra RGMP implementovaná najprv v jej pôvodnej forme, vďaka čomu mohla byť použitá ako referenčné riešenie pri následnom porovnávaní s upravenými verziami. Táto základná verzia bola upravená v dvoch krokoch.



Po prvé, trénovací proces modelu bol rozšírený o online trénovanie, ktoré využíva dáta vyprodukované metódou Lucid Data Dreaming - LDD. Táto metóda slúži na generovanie ľubovoľného počtu jedinečných anotovaných snímok iba na základe prvej anotovanej snímky videa. Tieto snímky sú následne použité na dodatočné online trénovanie modelu, ktorý sa tak lepšie prispôbi vstupu testovaciemu videu. LDD metóda bola v tejto práci implementovaná tak, aby bolo možné zvoliť si transformácie obrazu použité pri generovaní nových snímok. To následne umožňuje dynamicky meniť parametre generovania snímok a nájsť kombináciu s najlepšimi výsledkami.



vstupná snímka LDD generovaná snímka

Po druhé, model bol natrénovaný na novej dátovej množine - YouTube VOS, ktorá je niekoľkonásobne obsiahlejšia než akákoľvek doterajšia množina a ktorej obsah sa vyznačuje vysokou mierou diverzity. Teoreticky by väčší objem dát s vyššou komplexnosťou mal poskytnúť viac informácií pre model pri jeho učení, čo zabezpečí jeho lepšiu schopnosť generalizácie.

Výsledky a vyhodnotenie

Pre validáciu vylepšených modelov bolo všetko testovanie vykonávané na dátovej množine DAVIS, ktorá je dlhodobo používaná ako referenčná množina medzi ostatnými výskumníkmi na vzájomné porovnávanie ich riešení. Vylepšený model vo všeobecnosti podáva lepšie výsledky ako jeho pôvodná verzia, aj keď nie pre úplne všetky testovacie sekvencie. Pôvodný model RGMP podáva zmiešané výsledky - objekty v jednoduchých videách dokáže segmentovať na veľmi vysokej

úrovni, kvalita segmentácie však výrazne klesá pri zložitejších, tzv. hraničných prípadoch ako je rýchly pohyb sledovaného objektu, prekážky na scéne, neostrý obraz a ďalšie.

Online trénovanie v kombinácii s novou, obširnejšou dátovou množinou pozitívne ovplyvnilo schopnosť modelu spracovať videá aj so spomínanými hraničnými prípadmi. Tieto vylepšenia zároveň navýšili priemernú kvalitu segmentácie naprieč celou testovacou množinou, avšak za cenu mierne nižšej kvality segmentácie jednoduchých videí a takisto niekoľkonásobného predĺženia času potrebného na spracovanie jednej sekvencie.

Nakoniec bolo porovnaním výsledkov vylepšeného modelu s výsledkami iných modelov ukázané, že aj keď je možné daný model stále vylepšovať, celkový priestor na takéto zvyšovanie kvality je výrazne obmedzený a ďalší výskum by mal byť orientovaný na vývoj architektúry, ktorá dokáže poskytnúť modelu výrazne vyššiu schopnosť generalizácie pri zachovaní krátkeho času spracovania sekvencií.

Model	J Mean	J Recall	F Mean	F Recall	Rýchlosť
RGMPorigin	64.8	74.1	68.6	77.7	0.18s
RGMPnew	63.6	69.4	67.6	73.2	0.18s
RGMPnew+LDD	65.9	68.2	69.5	71.8	6s
PReMVOS	73.9	83.1	81.8	88.9	>20s
CINM	67.2	74.5	74.0	81.6	-
OSVOS-S	64.7	74.2	71.3	80.7	4.5s
OnAVOS	61.6	67.4	69.1	75.4	13s

Prínosy

- komplexný prehľad rozličných prístupov k riešeniu VOS problému
- zvýšenie kvality segmentácie u existujúceho modelu implementáciou online trénovania
- použitie novej dátovej množiny - YouTube VOS - na natrénovanie vylepšeného modelu