Video Frame Rate Upscaling Using Neural Networks

Ondřej Pleticha
ČVUT - FIT
pletiond@fit.cvut.cz

30. prosince 2019

1 Úvod

Každé video se skládá z obrázků, které běží v určité rychlosti za sebou a lidské oko to poté vidí jako video. Tato snímkovací frekvence je velmi důležitá, protože už při nižší frekvenci než 20 snímků za sekundu lidské oko vidí spíše obrázky než video a obraz je trhaný. V dnešní době je určitý standart okolo 60 snímků/s a v akčních hrách to je ještě více. Také pokud bychom se chtěli zaměřit na část video a přehrát si ho zpomaleně, potřebujeme, aby mělo vysokou frekvenci.

Ve své semestrální práci se zabývám interpolaci snímku ve videu, kde se snažím doplnit mezi jednotlivé dva snímky snímek nový a tím zdvojnásobit snímkovací frekvenci. Cílem je porovnat dva modely založené na odlišné architektuře a zhodnotit jejich výsledky.

2 Vstupní data

Pro experimenty jsem vybral celkem tři datasety, v tomto případě videa. Snaha byla vybrat odlišný styl videí a jejich délku. Vstupní data jsou ve formátu mp4 a v rozlišení 352x288 px. Video byla stažena z YouTube a upravena na stránce `https://ezgif.com/`. Použil jsem tyto videa:

Ball Jednoduché animované video, kde v kruhu obíhá kolo tvořené z koleček. Díky jednoduchému pohybu a přesným hranám, lze jednoduše porovnat výsledky experimentů. [1]

Tom Krátký úsek ze seriálu Tom & Jerry ukáže schopnosti modelů zpracovat animované video s rychlými pohyby, kde jsou poměrně velké rozdíly mezi jednotlivými snímky. [3]

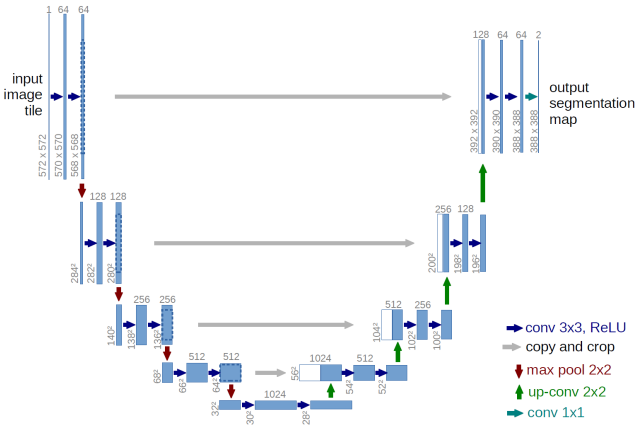
Gump Velmi krátké video běžícího člověka z profilu bude testovat výsledky modelů při malém množství trénovacích dat. [2]

3 Metody

Ve své práci jsem vybral dva modely, které se používají pro interpolaci snímku. První model je založený

Obrázek 1: Ukázka videí

na architektuře U-Net[7], konvoluční neuronová síť, která nejprve prokládá konvoluci s poolingem a následně provádí upsampling a spojuje výsledky z příslušných vrstev první části.2 Inspiroval jsem se prací Deep Motion[5]



Obrázek 2: Příklad U-Net

Druhý model je založený na architektuře GAN[4]. Pro experimenty jsem použil upravený kód k článku Frame Interpolation Using Generative Adversarial Networks[6].

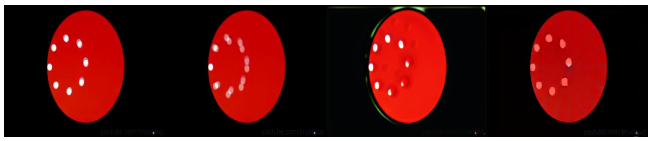
Při trénování modelů se vždy používá trojice po sobě jdoucích snímků x_1, x_2, x_3, kde x_1, x_3 byly vstupní hodnoty a x_2 cílová hodnota.

4 Výsledky

Na obou modelech jsem natrénoval všechny tři datasety. Výsledky budu porovnávat se skutečným snímkem a snímkem, který vznikl aritmetickým průměrem mezi snímky x_1, x_3, což se dá pokládat za nejjednodušší řešení.

Výsledky na datasetu s kruhem jsou vidět na

obrázku 3.



Obrázek 3: Výsledky 1 - zleva: skutečnost, průměr sousedů, U-Net, GAN

Nejblíže ke skutečnosti je zde nejspíše výsledek z GANu. Jako jediný nemá náznaky koleček nikde jinde než by měl mít, avšak nemají přímo bílou barvu. U modelu U-Net jsou podivné zelené okraje, které se vůbec nevyskytují v datasetu a je zde náznak koleček ze začátku videa.

Výsledky na datasetu z animovaného seriálu jsou vidět na obrázku 4.

Obrázek 4: Výsledky 2 - zleva: skutečnost, průměr sousedů, U-Net, GAN

V tomto případě byl velký rozdíl mezi sousedními snímky. Nejhorší výsledek byl u GANu, kde jsou postavy velmi rozmazané. Na první pohled má nejlepší výsledek U-Net, ačkoli i zde jsou nepřesné barvy a mírné rozmazání.

Výsledky na datasetu běžícího chlapce jsou vidět na obrázku 5. Tento dataset obsahoval pouze 73 snímků a chtěl jsem porovnat modely na základě toho, jak si s tím budou schopné poradit.

Obrázek 5: Výsledky 3 - zleva: skutečnost, průměr sousedů, U-Net, GAN

Oba modely zde mají problémy. GAN je možná přirozenější, ale v některých částech nohy splývají v pozadí. U-Net vypadá jako kdyby měl rozpité barvy. Pro lidské oko by byl určitě nejpřirozenější snímek s průměrem sousedů.

5 Závěr

Bylo zajímavé porovnat odlišné architektury na tomto úkolu. Nějaké výsledky byly pro mě překvapující, protože jsem od takto jednoduchých modelů neměl velké očekávání. Myslím si, že ani jeden z těchto modelů není vhodný pro komerční užití. Ať už z důvodu nekvality, tak velmi vysokým požadavkům na výpočetní výkon. State of the art modely dosahují velmi slibných výsledků, ale pořád nevidím jejich praktické využití.

Reference

[1] Crazy circle illusion! `https://www.youtube.com/watch?v=pNe6fsaCVtI`. Accessed: 2019-12-30.

[2] Run, forrest, run! - forrest gump (2/9) movie clip (1994) hd. `https://www.youtube.com/watch?v=x2-MCPa_3rU`. Accessed: 2019-12-30.

[3] Tom jerry | classic cartoon compilation | tom, jerry, spike. `https://www.youtube.com/watch?v=cqyziA30whE`. Accessed: 2019-12-30.

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[5] N. Joshi and D. Woodbury. Deep motion: A convolutional neural network for frame interpolation. 01 2017.

[6] Mark Koren, Kunal Menda, and Apoorva Sharma. Frame interpolation using generative adversarial networks. 2017.

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. 05 2015.