

# Generovanie Cinemagrafov pomocou hlbokých rekurentných Q-sietí

Erik Klembara

Fakulta Elektrotechniky a Informatiky, Technická Univerzita Košice, Slovensko

**Abstrakt**—Pri pohľade na fotografiu si človek dokáže predstaviť dynamický pohyb jednotlivých objektov fotografie oproti statickému pozadiu. Tento princíp bol realizovaný pomocou cinemagrafov, kde pohyb určitých objektov sa opakuje vytvárajúc tak dojem animácie. V tejto práci skúmame možnosti tvorby cinemagrafov kombináciou metódik sekvenčnej generácie a konvolučných neurónových sietí.

## I. ÚVOD

**L'**UDIA si na základe celoživotných skúseností vybudovali schopnosť predstavivosti vzhľadom statických a dynamických objektov v prirodzenom svete. Napríklad obraz padajúceho jablka ako je znázornené na Obr.1 si človek vie predstaviť ako jablko bude padať, odraziť sa, skotúľať ešte nejakú vzdialenosť a nakoniec zastaví.



Obr. 1. Padajúce jablká

Cinemagrafy sú fotografie v ktorých je opakovaný pohyb jedného alebo viacerých objektov vo forme animovaného GIFu. Tieto objekty sú väčšinou prírodné veci ako napríklad tečúca voda alebo stromy ohýbajúce sa vo vetre.

Cinemagrafy môžu byť chápané ako ilustrácia pohybu špecifických dynamických objektov v statických scénach.

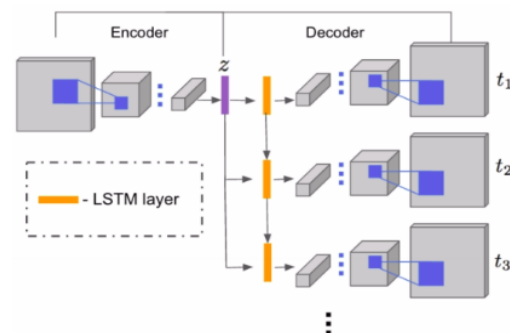
## II. METODIKA

Naším cieľom je vygenerovať cinemagraf s  $n$  snímkami  $Y = \{y_1, \dots, y_n\}$  so statickým obrázkom  $x$  na vstupe. To vyžaduje model, ktorý dokáže predpovedať budúci stav pohybujúceho objektu v čase. Model by sa mal naučiť, ktoré časti obrázku by mali byť animované a ako vytvoriť animáciu daného pohybu.

## III. REKURENTNÝ MODEL

V rekurentnej neurónovej sieti je každá skrytá vrstva charakterizovaná vlastnými váhami a biasmi. Tieto nezávislé vrstvy sa nasledovne zlúčia do rekurentnej vrstvy. Rekurentný neurón udržiava informáciu jej predošlého stavu a kombinuje ho s aktuálnou iteráciou. Teda dostávame vzťah medzi aktuálnymi a minulými stavmi, ktorý nám pôsobí na výsledok.

Model rekurentnej siete generuje Cinemagrafy zo statického obrázku. Dosiahneme to aplikovaním konvolučnej siete s LSTM vrstvou. Rekurzívne generujeme budúce snímky so skrytou premennou  $z$  ako môžeme vidieť na Obr. 2.



Obr. 2. Rekurentná architektúra

Chybové funkcie splňajú dve úlohy. Prvá kvadratická chybová funkcia je založená na pixeloch. Resp. rozdiel medzi generovaným pixelom a očakávaným pixelom.

$$loss_{mse} = |Y - G(x)|^2$$

Kde  $x$  je vstupný obrázok resp. prvý snímok v cinemagrafickej sekvencii,  $G()$  je výstup generovaného modelu a  $Y$  je očakávaný výstup. Druhá chyba je konfrontačná.

$$loss_{mse} = -\log(D[G(x)])$$

Kde  $D()$  je výstup diskriminátora. Táto chyba podporuje tvrdenie že generovaný obrázok je realistický. Naša dvojúlohová chybová funkcia je teda reprezentovaná vzťahom

$$loss = loss_{mse} + \lambda_{adv} * loss_{adv}$$

Pričom  $\lambda_{adv}$  je parameter, ktorý kontroluje dopad konfrontačnej chybovej funkcie.

Definujeme autoenkóder ako generátor a ako diskriminátor použijeme CNN klasifikátor. Diskriminátor klasifikuje vstupný obrázok ako skutočný alebo falošný.

#### IV. LSTM SIET'

Long Short-Term Memory (LSTM) sú špeciálnym druhom rekurentnej neurónovej siete schopnej učiť sa dlhodobým závislostiam, ktoré nám pomáhajú pracovať so sekvenciami, ktorých dĺžku nemôžeme poznať.

Sú vhodné pre modelovanie závislostí s veľkým rozstupom v čase. Je to spôsobené tým, že berú do úvahy viac predošlých stavov. Zavádza dodatočné regulujúce vstupy označované ako Gate. Tieto vstupy máme tri:

- Input Gate  $i_t$ : ovplyvňuje šírenie klasického vstupu do bloku,
- Forget Gate  $f_t$ : ovplyvňuje zapamätaný skrytý stav i vstup,
- Output Gate  $o_t$ : ovplyvňuje skrytý stav neurónu a teda výstup

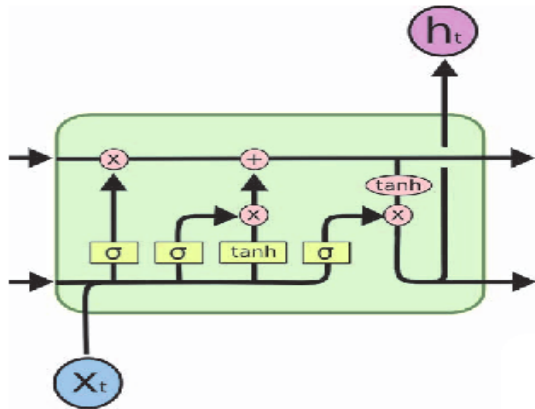
Skrytý stav  $h(t)$  sa spočíta pomocou nasledujúcich rovníc:

$$i_t = \delta(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + bi)$$

$$f_t = \delta(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + bf)$$

$$o_t = \delta(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + bo)$$

Medzi prepojeniami Gate existujú ďalšie váhy. Vektory teda tenzory týchto váh majú rovnakú veľkosť ako skrytý stav. Keď je výstup z Input Gate  $i_t$  blízky k hodnote nula, tak sa zabráňuje šíreniu vstupu  $x_t$  do siete.



Obr. 3. LSTM sieť

#### V. HLBOKÁ REKURENTNÁ Q-SIET'

Vstup pre rekurentnú Q-sieť je stav (napríklad obrázok) a výstupom siete je vektor Q. Rozmer Q-vektoru je rovný počtu možných akcií. Hodnota  $Q(s, a)$  pre každý rozmer je definovaná ako kumulatívne maximum očakávanej odmeny za vykonanie akcie  $a$  pre stav  $s$ .

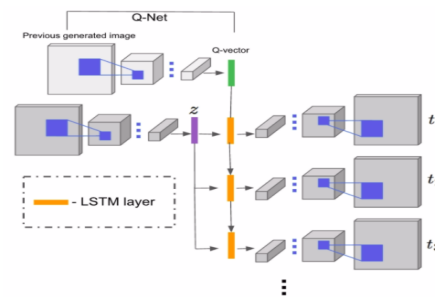
$$Q(s, a) = \max E[r_t + \delta r_{t+1} + \delta^2 r_{t+2} \dots | s_t, a_t = s, a, \pi]$$

Kde  $r_t$  je odmena v čase  $t$ ,  $\delta$  faktor zľavy a  $\pi$  je stratégia mapovania vstupných stavov pre akcie. Pri každej iterácii  $\pi$  vyberá akcie pre maximalizovanie očakávanej odmeny.

Ako prvé je nutné kvantovať lineárny akčný priestor do diskrétného priestoru. Generačná štruktúra je rovnaká ako pri rekurentnom modeli. Ako vstup použijeme vygenerovaný obrázok z predošlej iterácie alebo obrázok pre prvú iteráciu.

Pre výstup kvantujeme lineárny akčný priestor do diskrétného definovaním výstupného Q-vektoru ako  $N$  rozmerný vektor na určenie akcie pre daný stav. Ten enkódujeme do „one-hot“ vektora priradením 1 maximálnej hodnoty a 0 ostatným hodnotám Q-vektora.

Pre každú iteráciu vyrátame odlišný Q-vektor založený na predošlej iterácii a spojíme ho so skrytou premennou z LSTM vrstvy (Obr.3.).



Obr. 4. DQN s LSTM vrstvou

Spojený vektor pošleme do dekodovacej časti pre vytvorenie snímku nasledujúceho stavu.

#### VI. ZÁVER

V tomto článku sme sa venovali problematike vytvárania cinemagrafov z jediného obrázka. Ponúkli sme metódu, ktorá kombinuje rekurentné modely s Q-sieťou na učenie, ktorá oblasť obrázka by sa mala pohybovať a ako by tento pohyb mal vyzeráť v čase.

Budúca práca vyžaduje a zahrna vytvorenie komplexného datasetu s viacerými príkladmi a kategóriami. Taktiež je nutné podrobnejšie analyzovať čo sa model učí z obrázku z pohľadu pohybu objektov vo svete.

#### LITERATÚRA

- [1] H. Kopka and P. W. Daly, *A Guide to L<sup>A</sup>T<sub>E</sub>X*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
- [2] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computing*, vol. 9, pp. 1735–1780, Nov. 1997.
- [3] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *ICML*, 2017.
- [4] K. Gregor, I. Danihelka, A. Graves, D. Rezende, and D. Wierstra. Draw: A recurrent neural network for image generation. In *ICML*, 2015.