



PREDVIĐANJE EFEKATA TEMPORALNOG FILTRA NA KVALITET VIDEO SEKVENCI

PREDICTING TEMPORAL FILTER EFFECTS ON QUALITY OF VIDEO SEQUENCES

Miloš Kudus, *Fakultet tehničkih nauka, Novi Sad*

Oblast – ELEKTROTEHNIKA I RAČUNARSTVO

Kratak sadržaj – U ovom radu je opisan algoritam mašinskog učenja za predviđanje efekata temporalnog filtra na video sekvence.

Ključne reči: Kvalitet videa, Obrada slike, Mašinsko učenje

Abstract – This paper describes a machine learning algorithm for predicting temporal filter effects on video sequences.

Keywords: Video quality, Image processing, Machine learning

1. UVOD

Cilj rada jeste kreiranje algoritma za predviđanje efekata temporalnog filtra na video sekvence. Efekat se evaluira na osnovu promene mera šuma i oštine. Nakon primene temporalnog filtra mera šuma videa se popravlja, dok se mera oštine kvira. Amplituda promene mera zavisi od videa i od parametara filtriranja.

Generalno, sekvence koje imaju jak temporalni šum i koje su uglavnom statične imaju najviše benefita od obrade. Takvim videima se mera šuma popravlja značajno dok se mera oštine kvira za veoma male vrednosti.

Na dinamičnim sekvencama sa mnogo kretanja i sa malo šuma je kontraproduktivno primenjivati filtriranje jer je mera šuma već poprilično dobra i nema potrebe za poboljšanjem, a oština objekata koji se kreću se gubi u velikoj meri.

Za predviđanje efekata obrade je korišćena konvoluciona neuronska mreža za regresiju, čiji je izlaz promena kvaliteta video sekvence nakon obrade temporalnim filtrom. Kvalitet video sekvence se računa kao zbir mera kvaliteta šuma. Ulazi su četiri uzastopna frejma monohromatske sekvence. Predobrada je vršena u MATLAB-u, a neuronska mreža je implementirana u Python-u, pri čemu su korišćene biblioteke: *tensorflow* [4][5], *numpy*, *os*, *matplotlib*, *math*, *pandas*, *opencv*.

2. BAZE PODATAKA

2.1. Baza 1

Prvu bazu čini GMM baza fluoroskopskih videa. Sastoji se od 10 videa različite rezolucije i trajanja. Sekvence su snimljene digitalnim fluoroskopom.

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Vladimir Petrović, vanr. prof.

Iz smanjenog skupa izabrano je nasumično 25% sekvenci za test skup (ukupno 3 sekvence), a ostalih 75% je korišćeno kao trening skup (ukupno 7 video sekvenci).

2.2. Baza 2

Drugu bazu čini eVQA Video Dataset baza [3] koja je sakupljena godinama radom na eVQA projekatima. Sastoji se od dva podskupa (sirovi i izvedeni). Za ovaj zadatak je korišćen deo sirovog podskupa (njih 134 od 489).

Video sekvence su različite rezolucije i trajanja. Snimane su kamerama sa sensorima vidljive svetlosti (monohrom i u boji), IR spektra (LWIR, MWIR, SWIR), ili Visible intensified sensorima.

Sekvence su birane tako da u skupu ne bude više sekvenci koje prikazuju istu scenu, snimanu istim tipom senzora kako ne bi došlo do overfittinga na te video sekvence. Iz tog smanjenog skupa izabrano je nasumično 25% sekvenci za test skup (ukupno 34 sekvence), a ostalih 75% je korišćeno kao trening skup (ukupno 100 video sekvenci).

3. PREDOBRADA

Neuronska mreža zahteva ulazne podatke fiksne dimenzije [2]. To predstavlja problem, jer su video sekvence različite rezolucije i trajanja, sa različitim brojem kanala. Problem različitog broja kanala je rešen tako što su sve sekvence prebačene u monohromatske usrednjavanjem sekvenci po kanalima. Što se tiče različitog trajanja, problem je rešen tako što se mreži prosleđuju samo po četiri uzastopne sekvence. Time je smanjena računaska složenost modela jer se izbegla upotreba rekurentnih slojeva.

Problem različitih rezolucija mogao je biti rešen na različite načine.

Prvi način bi mogao da bude proširivanje sekvenci nulama kako bi svi videi imali iste prostorne dimenzije. Taj pristup je odbačen jer računski usloznjava problem.

Drugi način bi bio redimenzionisanje frejmova na neku fiksnu veličinu ali je to odbačeno zbog toga što bi to moglo uticati na mere šuma i oštine. Treći pristup, koji je i korišćen u radu, jeste da se iseku delovi frejma fiksne dimenzije (192x192).

Time se smanjila računaska složenost jer su uzorci koji ulaze u mrežu manjih dimenzija, a šum i oština nisu izmenjeni redimenzionisanjem. Dakle, uzorak za neuronsku mrežu se dobija isecanjem regiona 192x192 iz četiri uzastopna frejma i njihovim prebacivanjem u

monohromatsku sliku. Nakon toga evaluirane su mere šuma i oštine za taj uzorak. Potom je uzorak filtriran temporalnim filtrom i nakon toga su ponovo evaluirane mere šuma i oštine za taj uzorak čime su dobijene mere šuma i oštine nakon obrade temporalnim filtrom.

S obzirom da su RGB kanali uzoraka zapravo različiti frejmovi u vremenu, statični objekti su sivi jer je vrednost na pikselima sva tri kanala podjednaka. Kada imamo kretanje ili šum, vrednosti piksela se menjaju iz frejma u frejm što na ovim slikama vidimo kao boju.

4. UKLANJANJE TEMPORALNOG ŠUMA

Za uklanjanje temporalnog šuma korišćen je temporalni Gausov filtar. Obradeni frejm se dobija kao linearna kombinacija prethodnih frejmova (1).

Koeficijenti kojima frejmovi doprinose obrađenom frejmu su odbirci Gausove funkcije sa centrom u trenutnom frejmu i jediničnom standardnom devijacijom.

$$I_{obr}(i, j, n) = \sum_{k=0}^2 I(i, j, n - k) * TF(n - k) \quad (1)$$

I_{obr} predstavlja piksel obrađene sekvence, i i j su prostorne koordinate, a n je diskretno vreme. I je originalna video sekvenca, a TF su koeficijenti temporalnog filtra.

Direktna posledica uzimanja odbiraka Gausove funkcije za težinske koeficijente jeste ta da što je kadar vremenski udaljeniji od trenutnog kadra, manje doprinosi obrađenom kadru.

Radi smanjenja računске složenosti u obzir su uzeti samo trenutni i dva prethodna kadra koji doprinose sa oko 95% obrađenom kadru. Ovaj filtar smanjuje temporalni šum, ali objekti koji se kreću izgledaju zamućeno.

Za GMM skup srednja vrednost mere šuma posle obrade je 0.97, a standardna devijacija je 0.07. Srednja vrednost mere oštine posle obrade je 0.98, a standardna devijacija je 0.06. Za eVQA skup srednja vrednost mere šuma posle obrade je 0.95, a standardna devijacija je 0.11. Srednja vrednost mere oštine posle obrade je 0.63, a standardna devijacija je 0.26.

Na osnovu srednjih vrednosti mera vidimo da temporalni filtar uglavnom kviri meru oštine (vrednost mere oštine opada), ali istovremeno uglavnom popravlja meru šuma (vrednost mere šuma raste).

Kao konačna vrednost koju mreža treba da predvidi korišćena je promena kvaliteta sekvence nakon filtracije. Ona je data sa (2):

$$\Delta \text{kvalitet} = NMB + SMB - NMA - SMA \quad (2)$$

NMB i SMB predstavljaju meru šuma i oštine pre obrade, respektivno. NMA i SMA predstavljaju meru šuma i oštine nakon obrade, respektivno.

6. OBJEKTIVNE MERE

6.1. Mera šuma

Mera šuma je udeo frejma koji nema nestruktuisane nasumične varijacije između trenutnog i prethodnog

frejma [3]. Što je površina zahvaćena temporalnim šumom veća to je vrednost mere šuma manja. Opseg joj je od 0 do 1. Vrednost 1 znači da temporalni šum nije uočljiv u sekvenci, a vrednost 0 znači da je da je cela površina frejma prekrivena temporalnim šumom.

6.2. Mera oštine

Mera oštine je relativna snaga na visokoj frekvenciji u odnosu na niske frekvencije na lokacijama na kojima je najveća razlika između trenutnog i prethodnog frejma [3]. Idealno oštre slike imaju veću energiju na višim frekvencijama nego na nižim frekvencijama.

Više vrednosti ove mere znače da su u frejmu strukture oštih ivica, niske vrednosti ove mere znače da je slika zamućena. Opseg vrednosti joj je od 0 do 1. Vrednost mere 1 znači da su ivice struktura u frejmu idealno oštre, dok vrednost 0 znači da je frejm jako zamućen, tako da je energija na najvišim frekvencijama spektra 0.

7. NEURONSKA MREŽA

7.1. Objektivne funkcije

7.1.1. L2 norma

L2 ili Euklidska norma je jedna od najčešće korišćenih normi. Penalizuje velika odstupanja vrednosti (3).

$$\text{loss} = \frac{1}{\text{broj uzoraka}} \sum_{\text{uzorci}} \sqrt{\sum_{\text{izlazi}} (y - \hat{y})^2} \quad (3)$$

7.1.2. Tačnost klasifikacije

Tačnost klasifikacije se definiše kao procenat tačno klasifikovanih uzoraka. Da bi problem regresije pretvorili u problem klasifikacije, bilo je potrebno odrediti granicu promene kvaliteta za koju je potrebno vršiti obradu. U ovom radu je za taj prag odabrana vrednost 0,05. Dakle, sve ocene koje su veće od 0,05 su smatrane kao jedna klasa, a ostale kao druga klasa.

7.2. Optimizacija – Adam Optimizer

Adaptive moment estimation (Adam) je algoritam za efikasnu optimizaciju težina i ofseta u neuronskim mrežama [6]. Pokazao se dobro u praksi i široko je korišćen.

Taj algoritam poseduje dobre osobine metode opadajućih gradijenata sa inercijom i *RMS prop*-a. Algoritam prvo izračunava gradijente $\partial\omega$ i ∂b na beču na isti način kao što se to radi metodom opadanja gradijenata.

Beč je grupa uzoraka koji zajedno ulaze u mrežu. Potom se računaju vrednosti gradijenata sa inercijom $V\partial\omega$ i $V\partial b$. Zatim se računa *RMS* ažuriranje $S\partial\omega$ i $S\partial b$. Nakon toga se računaju korekcije tih vrednosti. I konačno, na osnovu toga ažuriramo vrednosti ω i b .

7.3. Trening

Za trening se koriste bečevi sa po jednim uzorkom iz svake trening sekvence. Pošto ima 7 (odnosno 100 za eVQA skup) trening sekvenci, svaki trening beč ima 7

(odnosno 100 za eVQA skup) uzoraka. Redosled kojim uzorci ulaze u mrežu se određuje posebno za svaki video. Redosled kojim uzorci jednog videa ulaze u mrežu je nasumičan sa najmanjim mogućim brojem ponavljanja nekog uzorka.

Permutuje se redosled kojim su izvučeni uzoci iz videa nasumično. Mreža se trenira sa maksimalno 5000 bečeva. Trening prekidamo i ako tačnost na test skupu počne značajno da opada. Za optimizaciju se koristi Adam algoritam i optimizuje se srednja L2 norma, ali računa se i L1 norma i tačnost klasifikacije. Brzina učenja opada linearno sa iteracijom.

7.4. Test

Za test se koriste bečevi sa po 90 (odnosno 10 za eVQA skup) uzoraka iz svake test sekvence. Pošto ima 3 (odnosno 34 za eVQA skup) test sekvenci svaki test, beč ima 270 (odnosno 340 za eVQA skup) uzoraka. Redosled kojim uzorci jednog videa ulaze u mrežu je nasumičan sa najmanjim mogućim brojem ponavljanja nekog uzorka. Permutuje se redosled kojim se izvlače uzorci iz videa nasumično.

Evaluacija jednog test beča se vrši na svakih 50 (odnosno 5 za eVQA skup) evaluacija treninga. Mere se tačnost klasifikacije, L1 i L2 norma. Mreža se snima ukoliko je vrednost tačnosti u trenutnom test beču veća od tačnosti na prethodnim test bečevima.

7.5. Opadanje brzine učenja

U prvim iteracijama treninga objektivna funkcija brzo opada. Kako iteracije odmiču, to opadanje usporava, da bi na kraju počelo da osciluje oko neke vrednosti. Objektivna funkcija ne može dalje da opada jer mreža pravi prevelike korake ažuriranja.

Bilo bi korisno da se brzina učenja smanjuje tokom vremena kako bi mreža, kada se približi lokalnom minimumu, pravila manje korake i time sprečila objektivnu funkciju da osciluje.

U ovom radu je implementirano linearno opadanje brzine učenja zbog svoje jednostavnosti [6]. Jednačina po kojoj brzina učenja opada je data sa (4). Na taj način brzina učenja se kontroliše sa 2 parametra: LRstart i LREndDecay.

$$LR = \frac{(LRstart - LREndDec)}{(maks. br. tr bečeva - 1)}it + LRstart \quad (4)$$

LR je brzina učenja, LRstart brzina učenja u nultoj iteraciji i LREndDec brzina učenja u iteraciji u kojoj želimo da zaustavimo opadanje brzine učenja, it je iteracija. Nakon te iteracije brzina učenja LREndDecay se zadržava do kraja obuke.

7.6. Uslovi za prekidanje treninga

Ukoliko tačnost klasifikacije videa na one koje je poželjno obraditi i na one koje je nepoželjno obraditi na test skupu počne da opada, to je znak overfitinga. Pošto je evaluacija tačnosti na test beču sklona šumu zbog konačnog broja test uzoraka koji se odabira nasumično iz

većeg test skupa, ne bi bilo dobro prekidati obuku na prvom test beču koji je imao manju tačnost od prethodnog. Umesto toga se čeka da trend opadanja tačnosti bude evidentan u više uzastopnih evaluacija na test bečevima. Obuka se prekida ako je gradijent tačnosti na prethodnih n test evaluacija negativan, ili ako je u više gradijenata u prethodnih n iteracija bilo negativno nego pozitivno.

8. REZULTATI

Vršene su obuke sa različitim parametrima sa različitim skupovima. Mreža je snimana u trenucima kada je zabeležena najveća tačnost na test beču do tog trenutka. U tabeli 8.1 su dati rezultati evaluacija na test skupu u nedostatku dovoljnog broja videa za podelu podataka u tri skupa. Podati su dati za GMM skup.

Tabela 8.1. Rezultati neuronske mreže

Mreža	1	2	3	4
Skup	GMM	GMM	GMM	eVQA
DKF	1	1	3	3
KS	32 16 8	32	32	32
MP	2 2 8	32	32	32
PPS	8 2 1	8 1	8 1	8 1
ACT	<i>LeakyReLU</i>	<i>LeakyReLU</i>	<i>LeakyReLU</i>	<i>LeakyReLU</i>
ACC	95.2%	95.2%	96.3%	71.2%
FPR	8.3%	1.3%	1.4%	23%
TPR	96%	90.6%	93.5%	63.5%

Sa DKF je označena dimenzija konvolucionih filtera; sa KS broj filtera po konvolucionom sloju; MP predstavlja dimenzije Max-Pooling prozora po slojevima; PPS označava broj potpuno povezanih neurona po sloju; ACT stoji za aktivacionu funkciju; ACC predstavlja tačnost (engl. Accuracy); FPR je stopa lažnih pozitivna (engl. False Positive Rate) i TPR stopa pravih pozitivna (engl. True Positive Rate).

9. ZAKLJUČAK

Iz rezultata vidimo da su mreže na GMM skupu ostvarile zadovoljavajući učinak. Najbolju tačnost od 96.3% je dala mreža s rednim brojem 3: sa jednim konvolucionim slojem sa 32 filtra dimenzija 3 x 3 i jednim skrivenim potpuno povezanim slojem sa 8 neurona.

Ta mreža je i računski jednostavna, što je bio jedan od zahteva zadatka. Mreža iste takve arhitekture je istrenirana na eVQA skupu, pri čemu je ostvarila mnogo lošiju tačnost od 71.2%.

U daljem radu je potrebno konstruisati mrežu koja daje bolje rezultate i na većem skupu i potencijalno podeliti eVQA skup na više podskupova na osnovu tipa senzora kojim su snimani. To može biti uzrok podbacivanja

10. LITERATURA

- [1] Vladimir Crnojević, “Prepoznavanje oblika za inženjere”, FTN Izdavaštvo, Novi Sad, 2014.
- [2] Josh Patterson, Adam Gibson, “Deep learning- a practitioner's approach“, Sebastopol, CA 2017
- [3] Dr. V. Petrovic, Mr. M. Willcox, Mr. C. B. Jackson, “Augmenting enhanced video quality assessment and enhanced Video quality control-milestone M2 report”, 2019.
- [4] Deeplearning.ai,
<https://www.coursera.org/learn/introduction-tensorflow>
- [5] Deeplearning.ai,
<https://www.coursera.org/learn/convolutional-neural-networks-tensorflow>
- [6] Deeplearning.ai,
<https://www.youtube.com/channel/UCcIXc5mJsHVYTZR1maL519w>

Kratka biografija:



Miloš Kudus rođen je u Somboru 1995. god. Bečelor rad na Fakultetu tehničkih nauka iz oblasti Biomedicinsko inženjerstvo odbranio je 2018.god. kontakt: miluskudas@gmail.com