

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

SEMINAR

Samonadzirano učenje optičkog toka SelfFlow metodom

Antonio Pavliš

Voditelj: prof. dr. sc. Siniša Šegvić

Zagreb, svibanj 2020.

Sadržaj

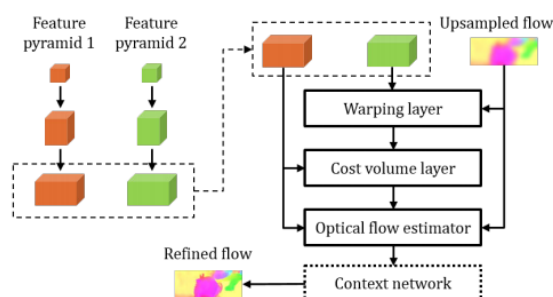
1	Uvod	1
2	Korišteni duboki konvolucijski model	2
3	Tehnike samonadziranog učenja	3
4	SelfFlow metoda samonadziranog učenja optičkog toka	5
4.1	Notacijska konvencija	5
4.2	Detekcija zaklonjenih slikovnih elemenata	6
4.3	Upravljanje učenjem i funkcije gubitka	9
5	Arhitektura modela	12
6	Eksperimenti	14
7	Zaključak	16
8	Literatura	17

1. Uvod

U području računalnog vida postignut je značajan iskorak razvojem automatskih sustava. Neki od tih sustava sustavi se danas mogu mjeriti s ljudskim sposobnostima prepoznavanja i detekcije objekata ili segmentiranja područja na složenim visualnim testovima. Ipak, zbog slikovnih elemenata pojedinih objekata koji su od strane drugih objekata u sceni zaklonjeni od pogleda kamere, javlja se problem precizne procjene optičkog toka pojedinih slikovih elemenata na sceni. Optički tok predstavlja obrazac kretanja vidljivih objekata, površina i rubova u sceni uzrokovano relativnim pomakom promatrača prema sceni. Postojeće metode za procjenu optičkog toka zasnivaju se na nadziranom učenju s obzirom na izmjereni tok. Takvo potpuno nadzirano učenje modela dubokih konvolucijskih neuronskih mreže zahtjeva ogromne količine označenih ulaznih podataka koje je vrlo teško prikupiti u svrhu učenja postupka procjene optičkog toka. Zbog nedostatka označenih primjera iz stvarnog svijeta pogodnih za učenje optičkog toka, metode samonadziranog učenja koriste umjetno označene primjere u svrhu učenja te je prednost ovakvih metoda upravo to što mogu učiti iz neoznačenih primjera. Pod neoznačene primjere smatramo slikovne isječke dovedene na ulaz modela koji nemaju prethodno definiran ispravan optički tok te se označavanje vrši tijekom procesa učenja. Zbog toga govorimo o metodama samonadziranog učenja. Osnovna ideja je minimizirati fotometrijski gubitak između referentnog i ciljnog slikovnog isječka dobiven tako da računamo razliku između tog referentnog isječka i isječka rekonstruiranog temeljem procjenjenog toka iz ciljnog isječka. U ovom radu biti će opisana metoda samonadziranog učenja optičkog toka nad neoznačenim podacima koji sadrže isječke sa zaklonjenim slikovnim elementima. Zaklanjanje pojedinih slikovnih elemenata će se provoditi uvođenjem nasumičnog šuma na određene regije unutar slike.

2. Korišteni duboki konvolucijski model

U ovom poglavlju biti će opisan duboki konvolucijski model koji se koristi u SelfFlow metodi opisanoj u ovom radu. PWC-Net [6] je duboki konvolucijski model razvijen 2018. godine u NVIDIA kompaniji. Model je prilagođen samonadziranom učenju i vrlo je učinkovit u radu s optičkim tokom, a temeljen je na jednostavnim i provjerenim principima, a to su piramidalna obrada, manipulacija nad slikama i traženje sličnosti među slikovnim elementima. PWC-Net model se sastoji od nekoliko dijelova. Piramidalno izlučivanje značajki generira piramide pripadnih značajki slikeu L razina. L je promjenjivi parametar. Najniža razina se sastoji od izvornih slika dovedenih na ulaz dok se svaka sljedeća razina generira primjenom konvolucijskih filtera na prethodnu razinu. Sloj manipulacije nad slikama na svakoj razini piramide povezuje reprezentativne značajke obje slike dovedene na ulaz. Nakon toga sloj za uspoređivanje pohranjuje vrijednosti funkcije gubitka pri pridruživanju slikovnog elementa njemu odgovarajućem u sljedećoj slici. Procjenitelj optičkog toka se sastoji od višeslojne neuronske mreže te na njegov ulaz dovodimo pohranjene vrijednosti gubitka pri pridruživanju i značajke slika te na izlazu dobivamo vrijednosti optičkog toka. Na kraju se koristi još jedan sloj za naknadno procesiranje dobivenog optičkog toka u svrhu preciznijeg predviđanja. Arhitektura PWC-Net modela je prikazana na slici 2.1.



Slika 2.1: Arhitektura PWC-Net modela [6]

3. Tehnike samonadziranog učenja

U svrhu samonadziraog učenja, potrebno je umjetno stvoriti podatke za učenje te ih pravilno označiti. Umjetno stvaranje podataka predstavlja zamjenski zadatak (eng. pretext taks) [1]. Zamjenski zadaci su predefinirani zadaci koje konvolucijska mreža treba obaviti te se vizualne značajke iz ulaznih podataka izlučuju obavljanjem zamjenskih zadataka. Proces učenja modela je vođen isključivo informacijama iz podataka koji nisu prethodno označeni. Tehnike samonadziranog učenja definiraju različite zamjenske zadatke neuronske mreže. Takve tehnike podrazumjevaju zadatke koji manipuliraju nad dostupnim podacima kao što su:

- **Rotacija** - Izrada 4 kopije rotiranjem izvorne slike za 0° , 90° , 180° i 270° te treniranje mreže nad svim orijentacijama izvornika. Želimo klasificirati svaku od ove 4 kopije u ispravnu klasu koja određuje primjenjenu rotaciju na izvorni podatak. Ovakvo razvrstavanje predstavlja jednostavan klasifikacijski problem između četiri klase. Ispravan model klasifikacije bi trebao prepoznati kanonska usmjerenja objekata u stvarnim slikama.
- **Primjerak** - Ova tehnika podrazumijeva da svaka pojedina slika pripada svojoj klasi kao uzorak te se ostali primjerci te klase generiraju podatkovnom manipulacijom nad uzorkom te klase. S obzirom da podaci za učenje nisu označeni, nastojimo izbjeći potrebu označavanja klasa pripadnosti te umjesto toga manipuliramo izvornim podacima te stvaramo nove primjerke svake klase. Manipulacije koje se mogu provoditi su translacija, skaliranje, rotacija, promjena kontraste te zamjena boja. Nove podatke dobivene manipulacijama promatramo u euklidskom prostoru te euklidska udaljenost dobivenog podatka od izvornog primjerka klase predstavlja gubitak koji želimo minimizirati ukoliko su podaci pripadnici iste klase, a istovremeno maksimizirati prema pripadnicima drugih klasa. Ovakva funkcija gubitka se naziva triplet loss [1].
- **Jigsaw** - Zadatak pri korištenju ove tehnike je otkrivanje relativne prostorne pozicije svake od 9 regija slike nakon njihove nasumične permutacije. Svaka od

tih regija se postavlja na ulaz modela te se izlazi ponovno spajaju i dovode na ulaz perceptrona koji treba otkriti permutaciju koja je korištena. U praksi se koristi unaprijed zadan skup od 100 permutacija. Svaka se regija neovisno jedna o drugoj pretvara u crno-bijelu sliku te se normalizira na srednju vrijednost 0 i jediničnu standardnu devijaciju. Po završetku postupka, reprezentaciju slike izlučujemo tako da usrednjavamo reprezentacije svake pojedine regije koje su uniformno uzorkovane i normalizirane. Funkcija gubitka koja se koristi tijekom učenja korištene permutacije temelji se na broju ispravno postavljenih regija. Što je broj ispravno postavljenih regija veći, vrijednost funkcije gubitka je manja zbog toga što je trenutno stanje bliže ispravnome.

- **Relativna lokacija regije** - Provođenje ove tehnike se sastoji od predikcije relativnog položaja promatrane dvije regije slike. Tehnike je slična tehnici Jigsaw, no ovdje je moguće samo 8 različitih relativnih položaja slikovnih regija koje treba predvidjeti s obzirom da su regije kvadratne te svaka može imati najviše 8 neposrednih susjeda. Izlučivanje reprezentacije slike je vrlo slično prethodnoj tehnici, usrednjavaju se slikovne reprezentacije svih 9 regija.

Primjer korištenja spomenutih tehnika je samonadzirano učenje semantičkog prognožiranja u videu. Prognožiranje budućih stanja na temelju informacija iz prošlosti je važno svojstvo inteligentnih sustava. Razine nad kojima je moguće provoditi predviđanje su predviđanje potpune slike, predviđanje značajki te semantičko predviđanje. Predviđanje potpune buduće slike je vrlo težak problem pa se bolji rezultati predviđanja dobivaju semantičkim prognožiranjem dijelova slika [5].

4. SelFlow metoda samonadziranog učenja optičkog toka

U ovom poglavlju biti će opisana SelFlow metoda samonadziranog učenja optičkog toka iz dostupnih slikovnih sekvenci koje nemaju pripadno označen ispravan optički tok. Takve uzastopne slike čine ulazni podatak. Za rješavanje ovog problema koriste se dvije konvolucijske neuronske mreže jednakih arhitektura koje će biti opisane u sljedećem poglavlju. Jedna mreža se koristi za učenje modela bez zaklonjenih slikovnih elemenata dok se druga mreža koristi za model sa zaklanjanjem pojedinih slikovnih elemenata kao što je slučaj u stvarnom svijetu. Informacije naučene jednostavnijim modelom bez zaklanjanja slikovnih elemenata koriste se za navođenje tijeka učenja drugog, složenijeg, modela. Iz tog razloga se ovakva tehnika učenja naziva samonadzirano učenje jer se tijekom učenja usmjerava upravo sirovim podacima nad kojima se uči. Tijekom faze testiranja predikcije optičkog toka koristi se složeniji model koji u obzir uzima slikovne elemente objekata koji su zaklonjeni drugim objektima na sceni. Za potrebe opisa postupka učenja optičkog toka, potrebno je konzistentno definirati korištene oznake.

4.1. Notacijska konvencija

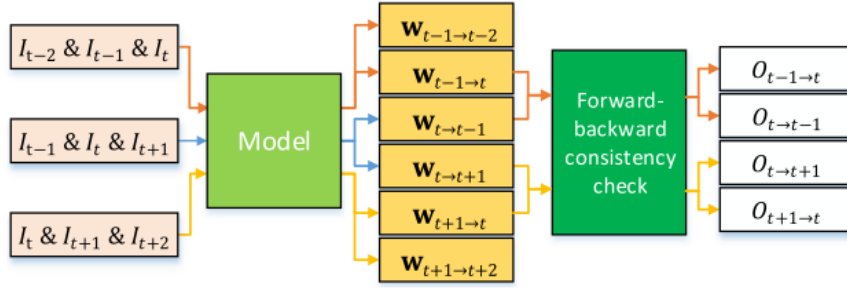
SelFlow metoda predikcije optičkog toka koristi tri uzastopna slikovna isječka u svrhu učenja. Slikovne isječke u obliku RGB informacije dovedene na ulaz modela označujemo sa I_{t-1} , I_t i I_{t+1} . Dakle, vidimo da metoda koristi jedan prethodni slikovni isječak za učenje optičkog toka između trenutnog i sljedećeg isječka. Prethodni isječak može sadržavati bitnu informaciju o položaju objekata na sceni koji su u narednim isječcima zaklonjeni od kamere te se pomoću takvih informacija može preciznije procijeniti optički tok u budućim isječcima. Optički tok između slika I_t i I_{t+1} označujemo s $\mathbf{w}_{t \rightarrow t+1}$. Općenito, optički tok između isječaka I_i i I_j označujemo $\mathbf{w}_{i \rightarrow j}$, tako $\mathbf{w}_{t \rightarrow t-1}$ predstavlja povratni tok između isječaka I_{t-1} i I_t . Koristeći poznati optički tok

$\mathbf{w}_{i \rightarrow j}$ između slika I_i i I_j , možemo iz slike I_j rekonstruirati sliku I_i . Tako dobivenu rekonstrukciju označavamo oznakom $I_{j \rightarrow i}^w$. Prilikom samonadziranog učenja, zaklanjanje slikovnih elemenata simuliramo dodavanjem nasumičnog podatkovnog šuma u slikovne isječke. Slučajan šum možemo dodati na bilo koji od tri slikovna isječka nad kojima procjenjujemo optički tok, ali praksa je šum dodati na slikovni isječak I_{t+1} te ga tada označujemo s $\tilde{I}_{t+1}$. U tom slučaju na ulaz modela dovodimo tri slikovna isječka: I_{t-1} , I_t i $\tilde{I}_{t+1}$. Oznakom $O_{i \rightarrow j}$ definiramo mapu zaklanjanja slikovnih elemenata između isječaka I_i i I_j . Vrijednost 1 u toj mapi sugerira da je pripadni slikovni element, koji je vidljiv u isječku I_i , zaklonjen u isječku I_j . Postavljanjem spomenutih slikovnih isječaka na ulaz modela, tijekom postupka koristimo optički tok oznake $\tilde{\mathbf{w}}$, mapu zaklanjanja $\tilde{O}$ i rekonstruirani isječak $\tilde{I}^w$.

4.2. Detekcija zaklonjenih slikovnih elemenata

Detekcija zaklonjenih slikovnih elemenata temelji se na usporedbi izvornog slikovnog isječaka i isječaka rekonstruiranog na temelju procjene povratnog optičkog toka. Za generiranje unaprijednog i povratnog optičkog toka, jednostavno zamjenimo slikovne isječke na ulazima modela te tako dobijemo optički tok u željenom smjeru. Model koji se koristi za procjenu optičkog toka među slikovnim isječcima je konvolucijska mreža čija će arhitektura biti objašnjena u sljedećem poglavlju. Zbog korištenja tri uzastopna slikovna isječka za procjenu unaprijednog toka ovom metodom kao što je objašnjeno ranije, potrebne su nam mape zaklanjanja među pojedinim parovima uzastopnih slika, to su mape: $O_{t-1 \rightarrow t}$, $O_{t \rightarrow t-1}$, $O_{t \rightarrow t+1}$ i $O_{t+1 \rightarrow t}$. Za generiranje pojedine mape zaklanjanja, potreban je optički tok među pripadnim slikama u oba smjera. Tako je primjerice za generiranje mape zaklanjanja $O_{t-1 \rightarrow t}$ potreban optički tok $\mathbf{w}_{t-1 \rightarrow t}$ kojeg možemo dobiti opisanim modelom tako da na ulaz dovedemo 3 uzastopne slike I_{t-2} , I_{t-1} i I_t . Dobivene tokove koristimo u svrhu određivanja mape zaklanjanja među susjednim slikovnim isječcima. Iz tog razloga je samo zbog generiranja mape zaklanjanja potrebno dovesti ukupno 5 uzastopnih slika od I_{t-2} do I_{t+2} . Postupak je prikazan na slici 4.1. Provodimo unaprijednu i povratnu provjeru konzistentnosti koja definira slikovni element kao zaklonjen u slučaju da je razlika unaprijednog i preokrenutog unaprijednog optičkog toka prevelika. Primjerice, prilikom generiranja mape zaklanjanja $O_{t \rightarrow t+1}$, najprije računamo preokrenuti optički tok kao:

$$\hat{\mathbf{w}}_{t \rightarrow t+1} = \mathbf{w}_{t \rightarrow t+1}(\mathbf{p} + \mathbf{w}_{t \rightarrow t+1}(\mathbf{p})) \quad (4.1)$$



Slika 4.1: Postupak generiranja mapa zaklanjanja korištenjem optičkih tokova dobivenih iz pet uzastopnih slikovnih isječaka. [4]

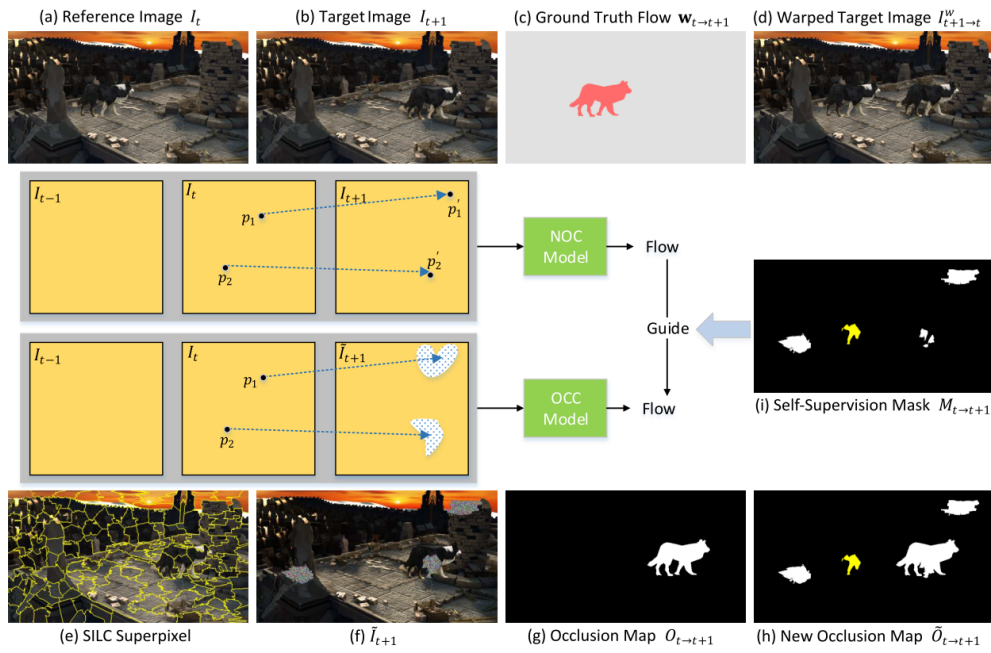
$\mathbf{p}$ je slikovni element u kojem računamo optički tok. Konačno, slikovni element smatramo zaklonjenim, te se njegova vrijednost postavlja na 1, ukoliko krši zadanu nejednakost:

$$|\mathbf{w}_{t \rightarrow t+1} + \hat{\mathbf{w}}_{t \rightarrow t+1}|^2 < \alpha_1(|\mathbf{w}_{t \rightarrow t+1}|^2 + |\hat{\mathbf{w}}_{t \rightarrow t+1}|^2) + \alpha_2 \quad (4.2)$$

Ostale mape zaklanjanja među slikovnim isječcima se generiraju na analogan način prikazan na slici 4.1. Vrijednosti parametara su postavljene tijekom faze validiranja te iznose $\alpha_1 = 0.01$ i $\alpha_2 = 0.05$.

U svrhu što vjernije simulacije stvarnog svijeta, tijekom faze učenja modela potrebno je umjetno zakloniti pojedine slikovne elemente na način kao što se to događa u stvarnoj sceni. SelfFlow metoda učenja optičkog toka dodaje nasumični šum na pojedine regije slikovnih isječaka čime pripadni slikovni elementi postaju zaklonjeni. Način dodavanja slučajnog šuma je proizvoljan no postoji nekoliko u praksi uobičajenih metoda. Najjednostavnija metoda je dodavanje šuma u obliku pravokutnih područja, ali takva pravilnost se vrlo rijetko javlja u stvarni scenama te je ispravnije model učiti na nepravilnim oblicima podatkovnog šuma. SelfFlow metoda koristi tzv. superpiksele koji predstavljaju nasumičnu povezanu skupinu slikovnih elemenata, a ideja je izabrati nekoliko takvih nakupina i na njih dodati podatkovni šum. Takvi nepravilni oblici vjernije oponašaju stvarni šum u podacima, a dijelovi tih regija su često unutar granica promatranih objekata što predstavlja djelomičnu zaklonjenost objekata na stvarnoj sceni. Osim toga, pošto se nakupine povezane i sastoje se isključivo od susjednih elemenata bez rupa unutar njih, regije koju pokrivaju obično imaju slična svojstva ili čak pripadaju istom objektu. Jedino ograničenje na slučajni šum koji dodajemo je raspon vrijednosti koji treba biti u rasponu postojećih slikovnih elemenata.

Slika 4.2 prikazuje postupak dodavanja nasumičkog šuma te generiranje mape zak-



Slika 4.2: Samonadzirano učenje u slikovnim elementima koji su u sljedećoj slici zaklonjeni. Zaklanjanje simuliramo dodavanjem nasumičnog podatkovnog šuma u obliku superpiksela. Na osnovu generiranih optičkih tokova iz jednostavnog modela, konstruiramo mapu zaklanjanja te ju koristimo za upravljanje učenja modela predikcije optičkog toka uz zaklonjene slikovne elemente [4]

lanjanja kao posljedice dodanog šuma. SelFlow metoda prvo trenira model nad dostupnim podacima bez dodavanja šuma. Neka je referentna slika prikazana pod (a), a ciljna slika pod (b). Tada učimo model minimiziranjem klasičnog fotometrijskog gubitka između referentne slike (a) i rekonstruirane slike (d). Međutim, slika (d) je rekonstruirana temeljem unaprijed poznatog ispravnog optičkog toka prikazanog na slici (c). Učenje je vođeno pripadnom mapom zaklanjanja prikazanoj na slici (g). Ideja je prvo naučiti model NOC koji procjenjuje optički tok objekata na sceni koji nisu zaklonjeni te kasnije dobiveni model koristiti prilikom učenja modela OCC za procjenu optičkog toka objekata koji se jednim djelom zaklanjaju na sceni. Za potrebe učenja takvog, složenijeg modela, nasumično generiramo područja među kojima ćemo izabrati nekoliko te na njih dodati podatkovni šum. Taj postupak je prikazan na slikama (e) i (f). Treba primjetiti da je dodavanje slučajnog šuma utjecalo na izvornu mapu zaklanjanja (g) zbog toga što su slikovni elementi prekriveni šumom narušili prethodno navedenu nejednakost te se smatraju zaklonjenima. Zbog toga pripadni element mape zaklanjanja ima vrijednost 1, odnosno obojan je bijelo kao što je prikazano na slici (h). Informacija koja upravlja učenjem složenijeg modela se predstavlja u obliku maske M prikazane na slici (i) čiji elementi imaju vrijednost 1 ako je pripadni element zaklonjen u (h), ali nije bio zaklonjen u (g). Žuto područje u masci (i) je dio pokretnog objekta, u ovom slučaju psa, čiji slikovni elementi nisu zadovoljili navedenu nejednakost pa su klasificirani kao zaklonjeni. Na ovakav način SelFlow metoda može učiti procjenu optičkog toka za pokretne objekte koji zaklanjaju pozadinu na sceni.

4.3. Upravljanje učenjem i funkcije gubitka

Metoda učenja optičkog toka se oslanja na minimizaciju fotometrijskog gubitka nezaklonjenih slikovnih elemenata jer je takva tehnika vrlo učinkovita na objektima čiji slikovni elementi nisu nimalo zaklonjeni. Ideja SelFlow metode je iskoristiti tako naučene informacije za učenje složenijeg modela. Slika 4.2 prikazuje slučaj kada slikovni element p_1 i njemu odgovarajući slikovni element u sljedećem isječku p'_1 nisu zaklonjeni, no nakon dodavanja umjetnog šuma u isječak I_{t+1} on postaje $\tilde{I}_{t+1}$ te slikovni element koji odgovara slikovnom elementu p_1 u izvornoj slici sada postaje zaklonjen. Korištenjem naučenih informacija možemo procijeniti optički tok promatranog piksela između slikovnih isječaka I_t i $\tilde{I}_{t+1}$ i u tom slučaju. Ovakva ideja je preuzeta iz stvarnog svijeta gdje je moguće procijeniti optički tok promatranih objekata, iako su neki njegovi dijelovi zaklonjeni, na temelju njegove okoline i njegovih ostalih vidljivih dijelova. Kao informacija pri učenju koristi se maska prikazana na slici 4.2 pod (i) koja

sadrži slikovne elemente zaklonjene na prijelazu između slikovnih isječaka I_t i $\tilde{I}_{t+1}$.

Minimizacija fotometrijskog gubitka temelji se na minimizaciji funkcije:

$$L_p = \sum_{i,j} \frac{\sum \psi(I_i - I_{j \rightarrow i}^w) \odot (1 - O_i)}{\sum (1 - O_i)} \quad (4.3)$$

Gdje je $\psi(x) = (|x| + \varepsilon)^q$ osnovna funkcija gubitka, a operator $\odot$ predstavlja množenje po elementima (Hadamardov produkt). Prilikom testiranja su korištene konstante sa vrijednostima $\varepsilon = 0.01$ i $q = 0.4$. U eksperimentima se za učenje jednostavnog modela bez zaklonjenih slikovnih elemenata koristi samo navedena funkcija L_p za minimizaciju fotometrijskog gubitka. Za složeniji model sa umjetno zaklonjenim slikovnim elementima koristimo funkciju samonadziranog gubitka L_o definiranu kao:

$$L_o = \sum_{i,j} \frac{\sum \psi(w_{i \rightarrow j} - \tilde{w}_{i \rightarrow j}) \odot M_{i \rightarrow j}}{\sum M_{i \rightarrow j}} \quad (4.4)$$

Gdje je M samonadzirana maska koja predstavlja umjetno zaklonjene slikovne elemente.

$$M_{i \rightarrow j} = \text{clip}(\tilde{O}_{i \rightarrow j} - O_{i \rightarrow j}, 0, 1) \quad (4.5)$$

Funkcija $\text{clip}(x, a, b)$ ograničava vrijednost argumenta x unutar intervala $[a, b]$. Ukoliko je vrijednost argumenta x izvan tog intervala, funkcija će vratiti vrijednost bliže od dvaju granica prema argumentu x .

Složeniji model učimo korištenjem funkcije gubitka oblika $L_p + L_o$ kako bismo precizno procijenili optički tok vidljivih kao i onih zaklonjenih slikovnih elemenata. Ovakve funkcije gubitka se mogu koristiti za procjenu optičkog toka na temelju dva slikovna isječaka kao i na temelju više uzastopnih slikovnih isječaka kao što je slučaj korištenjem ove metode gdje se koriste tri uzastopna slikovna isječaka.

Nakon učenja na neoznačenim podacima prethodno opisanom metodom, model je potrebno fino ugoditi (engl. fine tune) na dostupnim označenim stvarnim podacima. S obzirom da je na stvarnim podacima dostupan samo unaprijedni tok $\mathbf{w}_{t \rightarrow t+1}$, povratni tok $\mathbf{w}_{t \rightarrow t-1}$ ne uzimamo u obzir pri računanju fotometrijskog gubitka. Uz ispravno označen optički tok $\mathbf{w}_{t \rightarrow t+1}$, dostupna nam je i maska V koja govori koji slikovni elementi imaju ispravnu oznaku optičkog toka, koristimo ju kao filter pri odabiru slikovnih elemenata za računanje fotometrijskog gubitka L_s .

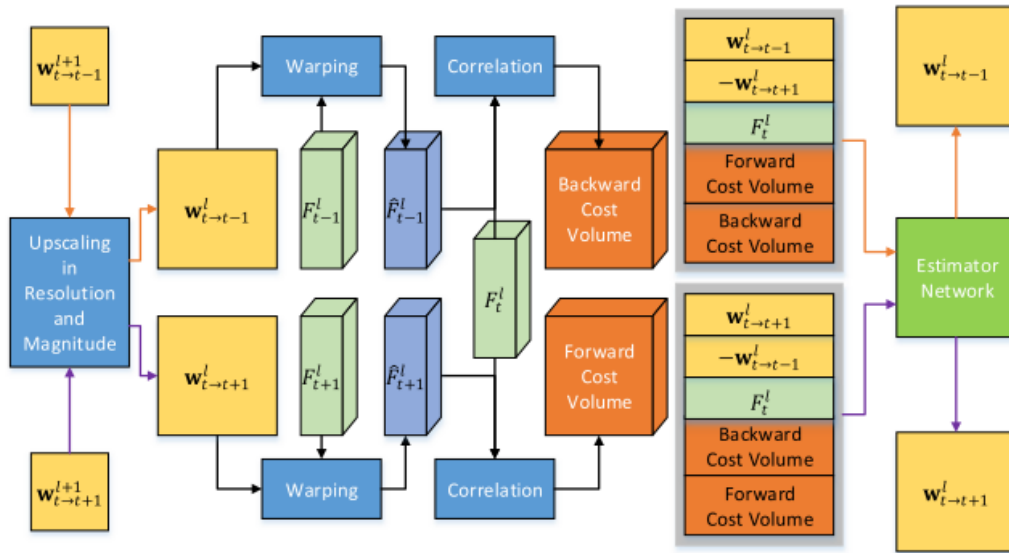
$$L_s = \frac{\sum \psi(\mathbf{w}_{t \rightarrow t+1}^g t - \mathbf{w}_{t \rightarrow t+1}) \odot V}{\sum V} \quad (4.6)$$

Gdje je $\mathbf{w}_{t \rightarrow t+1}^g t$ ispravan unaprijed označen tok.

Učenje se dakle svodi na minimiziranje ove funkcije gubitka. Inicijalno, model učimo tako da minimiziramo funkciju gubitka oblika $L_p + L_o$ kao što je objašnjeno te nakon toga optimiramo model minimizacijom funkcije L_s korištenjem stvarnih označenih podataka.

5. Arhitektura modela

Arhitektura konvolucijske neuronske mreže korištene u SelFlow metodi procjene optičkog toka je slična arhitekturi mreže PWC-Net. Spomenuta PWC-Net mreža koristi piramidalnu obradu te izlučivanje značajki iz slikovnih podataka izvodi od grublje prema finijoj segmentaciji podataka. Mreža korištena u metodi SelFlow izvedena je korištenjem PWC-Net mreže uz nekoliko ispravaka kako bi takvu mrežu prilagodili za rad s tri uzastopna slikovna isječka kako je prethodno opisano. Dovođenjem tri uzastopna slikovna isječka na ulaz modela, posljedično izlučujemo tri skupa značajki pripadnih slikovnih isječaka. Izlučene značajke označimo F_{t-1}^l , F_t^l i F_{t+1}^l . Arhitekturu modela možemo prikazati dijagramom na slici 5.1.



Slika 5.1: Arhitektura sloja mreže modela korištenog u SelFlow metodi procjene optičkog toka [4]

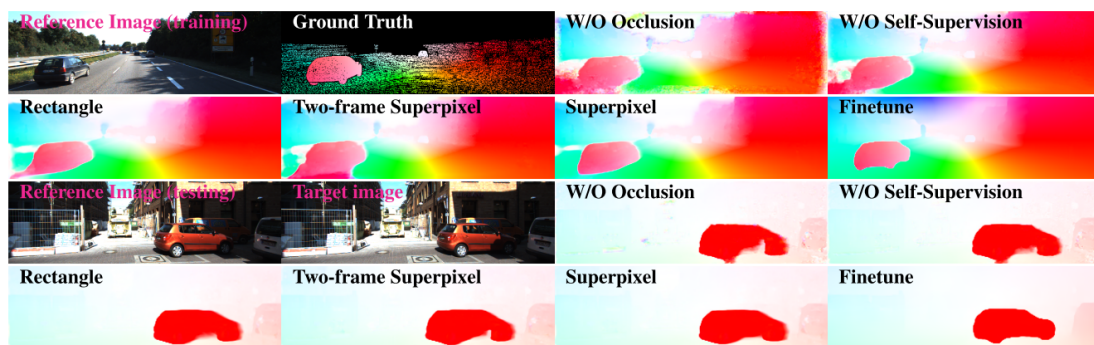
Na temelju izlučenih značajki, mreža procjenjuje unaprijedni tok $w_{t \rightarrow t+1}$, ali usporedno tome, procjenjuje i povratni tok $w_{t \rightarrow t-1}$. Kao što je spomenuto ranije, ideja računanja povratnog toka inspirirana je čovjekovim sustavom predviđanja gdje infor-

macije iz prošlosti značajno utječu na procjenu budućeg optičkog toka. Informacija o slikovnom elementu koji je prethodno bio vidljiv, ali je sada zaklonjen, unosi dodatna znanja o zaklonjenom dijelu scene te omogućuje precizniju procjenu budućeg stanja scene. Iz tog razloga koristimo informacije dobivene povratnim tokom $\mathbf{w}_{t \rightarrow t-1}$. Mreža usporedno koristi informacije o inicijalnom unaprijednom optičkom toku $\mathbf{w}_{t \rightarrow t+1}$ i inicijalnom povratnom toku $\mathbf{w}_{t+1 \rightarrow t}$ sa negativnim predznakom dobivene od prethodne razine mreže te značajkama izlučenim iz izvornog slikovnog isječka kako bi precizno procjenila optički tok na svakoj razini mreže. Svaka sljedeća razina mreže procjenjuje optički tok na višoj rezoluciji od prethodne razine. Mreža za procjenu unaprijednog i povratnog toka imaju identičnu strukturu i parametre.

6. Eksperimenti

Učenje i testiranje opisanog modela optičkog toka provedeno je na javnim testnim skupovima podataka Sintel i KITTI. Prilikom samonadziranog učenja modela, vrijednosti slikovnih elemenata koje su inicijalno bile u intervalu $[0, 255]$ skaliraju se u interval $[0, 1]$ te se svaki kanal boje normalizira kako bi se postigla normalna distribucija. Normalizirani slikovni podaci koji se dovode na ulaz modela robusniji su na promjene osvjetljenja u sceni te unose manji šum kod učenja optičkog toka. Za mjeru ispravnosti procjene optičkog toka u fazi testiranja koristi se EPE mjera za Sintel skup podataka odnosno Fl mjera na KITTI skupu podataka. Fl mjera izračunava postotak pogrešno klasificiranih slikovnih elemenata.

Slika 6.1 prikazuje procijenjeni optički tok tijekom faze učenja te u fazi testiranja dobiven SelFlow metodom. Vidljivo je da tehnike kao što su detekcija zaklonjenih slikovnih elemenata i umjetno zaklanjanje doprinose ispravnoj procjeni optičkog toka promatranog automobila na sceni uz postojanje sjene. Umjetno dodavanje zaklonjenih regija u obliku pravokutnika daje očito lošije rezultate pri učenju procjene optičkog toka od generiranja nasumičnih regija zaklonjenih slikovnih elemenata koji očekivano više sliče stvarnoj situaciji na promatranj sceni pa model bolje generalizira u ovom slučaju. Također, korištenje tri uzastopna slikovna isječka pri samonadziranom učenju modela doprinosi boljoj procjeni optičkog toka od modela sa dva uzastopna isječka zbog toga što prethodni slikovni isječki sadrže informacije koje se mogu upotrijebiti pri procjeni budućeg optičkog toka. Konačno, završno učenje modela u odnosu na izmjereni optički tok daje najbolje rezultate u procjeni optičkog toka što je prikazano na slici 6.1 pod oznakom Finetune. Postupak koji uči samo uz označene podatke zahtjeva ogromne količine označenih podataka, no prethodnim učenjem uz samonadzor, vrlo dobri rezultati mogu se postići čak i s vrlo malim brojem točno označenih podataka.



Slika 6.1: Procjena optičkog toka metodom SelfFlow. Slika prikazuje doprinos pojedinih elemenata metode. Ti elementi su: maska zaklanjanja, samonadzirano učenje toka u zaklonjenim slikovnim elementima, superpikseli, korištenje više od dvije uzastopne slike te fino ugađanje. [4]

7. Zaključak

U ovom radu opisana je SelfFlow metoda učenja procjene optičkog toka na temelju uzastopnih slikovnih isječaka dovedenih na ulaz modela. Posebnost metode je umjetno zašumljivanje podataka kako bi se simuliralo zaklanjanje slikovnih elemenata objekata u stvarnoj sceni. Metoda se sastoji od učenja procjene optičkog toka na jednostavnoj sceni bez zaklanjanja slikovnih elemenata uz poznatu mapu zaklanjanja. Naučene informacije o optičkom toku koriste se pri učenju procjene optičkog toka na scenama s umjetno zaklonjenim regijama. Metoda uspješno koristi informacije dobivene iz prethodnih slikovnih isječaka za vođenje procesa učenja složenijeg modela. S obzirom da se signal upravljanja procesom učenja izlučuje izravno iz podataka nad kojima se uči, govorimo o samonadziranom učenju. Samonadzirano učenje je područje nenadziranog učenja te je vrlo korisno u slučajevima nedostatka prethodno ispravno označenih podataka nad kojima bi se moglo učiti. Opisana metoda pruža izvrsne rezultate na javno dostupnim skupovima za testiranje te je značajno bolja u procjeni optičkog toka nad tim skupovima od drugih metoda nenadziranog učenja.

8. Literatura

- [1] Alexander Kolesnikov, Xiaohua Zhai, i Lucas Beyer. Revisiting self-supervised visual representation learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2019. doi: 10.1109/cvpr.2019.00202. URL <http://dx.doi.org/10.1109/CVPR.2019.00202>.
- [2] Yann LeCun. *Self-Supervised Learning*. NYU - Courant Institute Center for Data Science, Facebook AI Research. URL https://drive.google.com/file/d/1r-mDL4IX_hzZLDBKp8_e8VZqD7fOzBkF/view.
- [3] Pengpeng Liu, Irwin King, Michael R. Lyu, i Jia Xu. Ddflow: Learning optical flow with unlabeled data distillation. U *AAAI*, 2019.
- [4] Pengpeng Liu, Michael R. Lyu, Irwin King, i Jia Xu. Selfflow: Self-supervised learning of optical flow. U *CVPR*, 2019.
- [5] Josip Saric, Marin Orsic, Tonci Antunovic, Sacha Vrazic, i Sinisa Segvic. Single level feature-to-feature forecasting with deformable convolutions. *CoRR*, abs/1907.11475, 2019. URL <http://arxiv.org/abs/1907.11475>.
- [6] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, i Jan Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. *CoRR*, abs/1709.02371, 2017. URL <http://arxiv.org/abs/1709.02371>.

Samonadzirano učenje optičkog toka SelFlow metodom

Sažetak

Optički tok predstavlja obrazac kretanja vidljivih objekata, površina i rubova u sceni uzrokovano relativnim pomakom promatrača prema sceni. Problem precizne procjene optičkog toka pojedinih slikovih elemenata na sceni važan je u području računalnog vida, posebno u automatskim sustavima koji trebaju predvidjeti i izbjeći potencijalno neželjeno ponašanje. U ovom radu opisana je SelFlow metoda učenja procjene optičkog toka na temelju uzastopnih slika dovedenih na ulaz modela. SelFlow metoda se temelji na PWC-Net dubokom konvolucijskom modelu prilagođenom za rad s optičkim tokom. Podatke nad kojima učimo naš model generiramo iz dostupnih podataka, zbog toga govorimo o samonadziranom učenju. U tu svrhu, potrebno je definirati zamjenske zadatke u modelu koji će omogućiti ispravnu generalizaciju naučene procjene toka na stvarnim podacima. U stvarnim primjenama često je vizualno zaklanjanje pojedinih slikovnih regija zbog sjene ili drugih objekata. Da bi naučili procijeniti optički tok i u takvim situacijama, uvodimo umjetna zaklanjanja slikovnih elemenata dodavanjem nasumičnog šuma. Opisana metoda postiže najbolje rezultate procjene optičkog toka nad javnim testnim skupovima među do sada poznatim metodama.

Ključne riječi: Optički tok, Samonadzirano učenje, SelFlow, PWC-Net