Funkcije izgube za detekcijske števce z malo učnimi primeri

Jer Pelhan, Alan Lukežič, Vitjan Zavrtanik, Matej Kristan

Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, Večna pot 113, 1000 Ljubljana
E-pošta: {jer.pelhan, alan.lukezic, vitjan.zavrtanik, matej.kristan}@fri.uni-lj.si

Loss Functions for Few-Shot Detection-Based Counters

Most successful few-shot counting methods are adapted from density-based counters. They are trained to fit a unit Gaussian over the area of each target object using the L2 loss. This approach has significant drawbacks: it unnecessarily requires fitting a hyperparameter σ, and is prone to annotation noise. Additionally, it assumes that the geometric center of the object is the optimal point for predicting the bounding box. We analyze different loss functions designed to optimize detection for few-shot counting, addressing these limitations directly. This way method can autonomously learn the best locations for predicting bounding boxes and incorporate hard-negative mining to reduce false positives. We demonstrate significant performance variations in both total count estimation and object localization accuracy.

1 Uvod

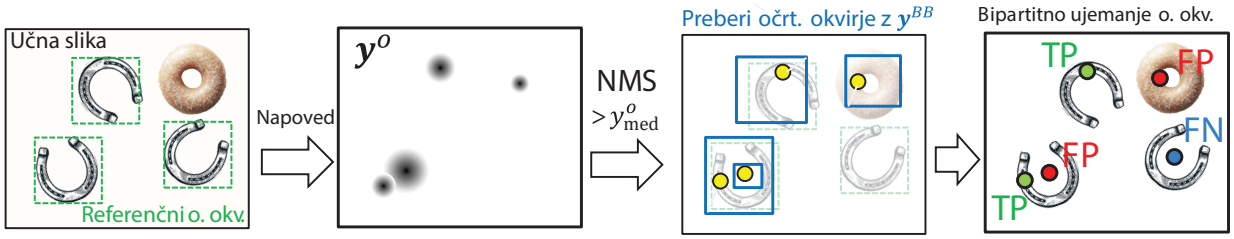
Štetje z malo učnimi primeri naslavlja problem ocenjevanja števila objektov novih kategorij na podlagi le nekaj označenih primerov v sliki [15]. Trenutno najsodobnejši števci [2, 7, 21, 17, 15, 20, 3, 20] napovedujejo gostotno mapo, seštevek katere predstavlja števila objektov ciljnega razreda. Metode na osnovi gostotne mape dosegajo najboljše rezultate, ampak so nepraktične za mnoge aplikacije [22, 19], ker ne nudijo dodatnih informacij, kot so velikost in lokacija objektov. Zato se je razvoj pred kratkim preusmeril v detekcijske števce [13, 10, 23, 12], ki napovedujejo očrtane okvirje (tj. velikost in lokacijo), končno število pa ocenijo kot število napovedanih očrtanih okvirjev. Sodobni detekcijsko snovani števci [13, 23, 10] sicer zaostajajo za na gostotnih mapah zasnovanimi metodami.

Pri metodah z malo učnimi primeri, se običajno najprej konstruirajo prototipi objektov na podlagi podanih učnih primerov, ki se potem korelirajo z značilkami slike [13, 23, 12, 10]. Za pokrivanje celotnega vizualnega obsega kategorije objektov je ključno posploševanje značilnic. To sicer omogoča visok priklic, povzroča pa tudi visoko stopnjo lažno pozitivnih detekcij. Metode za verifikacijo detekcij in zniževanje števila lažno pozitivnih detekcij že obstajajo [13, 23], vendar njihova večstopenjska formulacija onemogoča enostavno učenje. Trenutno najsodobnejši detekcijski števci [13, 23] napovedujejo lokacije objektov na lokalnih maksimumih v gostotni mapi ali mapi odzivov. Med učenjem je napoved gostotne mape nadzorovana z enotskimi Gausovimi funkcijami, ki se raztezajo čez površino vsakega izmed ciljnih objektov. Slednje vodi do težav z anotacijskim šumom in potrebo po ne-trivialni izbiri velikosti Gaussovega jedra, kar privede do prednostne detekcije kompaktnih, atomskih struktur. Nedavno predlagan števec [10], navdihnjen po arhitekturi transformerskega detektorja objektov DETR [1] se izogne tem težavam med učenjem, saj ne napoveduje gostotne mape, vendar iz istega razloga v gostih regijah objektov dosega visoko stopnjo lažno negativnih detekcij.

V članku naredimo pregled različnih funkcij izgube za detekcijske števce, ki delujejo na osnovi mape odzivov. V analizi uporabimo preprosto enostopenjsko metodo za štetje z malo učnimi primeri DSC [14]. Pokažemo, da je obstoječ način učenja z optimizacijo mape odzivov za napoved Gaussove porazdelitve čez vsak objekt neprimeren in zato raziskujemo bolj optimalne načine učenja detekcijskih števcev.

2 Sorodna dela

Sodobne detekcijske števce lahko glede na njihovo arhitekturo razdelimo na dve vrsti. Prva vrsta, trenutno najbolj uspešnih detekcijskih števcev [13, 23], napoveduje lokacijo objektov na podlagi lokalnih maksimumov gostotnih map. V prvi fazi ocenijo gostotno mapo, nato pa detektirajo potencialne centre objektov z iskanjem lokalnih maksimumov (NMS). Na teh lokacijah metoda v ločeni veji nevronske mreže napove odmike očrtanih okvirjev. Metoda DAVE [13] dobro ocenjuje število objektov, vendar je manj uspešna pri ocenjevanju natančnih očrtanih okvirjev. Pred kratkim predstavljen detekcijski števec [23] najprej napove centre objektov, na katere z dodatnimi hevristikami ocen velikosti objektov aplicira segmentacijsko metodo SAM [5] in izračuna očrtane okvirje objektov. Glavna slabost omenjenih metod je, da so med učenjem njihove gostotne mape optimizirane za napoved enotskih Gausovih porazdelitev za vsak objekt na sliki, kar ni nujno optimalno za napovedovanje očrtanih okvirjev. Druga vrsta detekcijskih števcev [10] sledi transformerski arhitekturi detektorjev DETR [1], ki za vsako izmed N poizvedb (*angl.* "query") napove očrtan okvir in pripadnost ciljnemu razredu. Takšni detektorji so uspešni



Slika 1: Prikaz cevovoda učenja. Najprej se izvede napoved očrtanih okvirjev, katerim se poišče bipartitna ujemanja z referenčnimi okvirji, katere klasificira kot resnične pozitivne (TP) lažno pozitivne (FP) in doda lažno negativne (FN) primere.

pri napovedi natančnih očrtanih okvirjev za večje objekte, manj uspešni pa v gostih regijah majhnih objektov. Metoda DETR ima vnaprej definirano število poizvedb, kar pomeni, da je to zgornja meja števila objektov, ki jih lahko detektira. C-DETR [10] doseže najboljši rezultat pri številu poizvedb 600, kar pomeni, da v primerih, kjer je prisotnih več kot toliko objektov, ne zmore prešteti vseh. DSC [14] je detekcijski števec, ki napoveduje mapo odzivov, torej ni omejen s številom poizvedb, poleg tega pa ni učen za napoved enotskih Gaussovih porazdelitev.

3 Povzetek metode DSC

Metoda DSC [14] (Detect, Segment and Count) je enostopenjski detekcijski števec, ki za štetje in detekcijo in segmentacijo objektov poljubne kategorije potrebuje zgolj nekaj anotiranih primerkov. DSC temelji na hrbtenici metode SAM [5], kar omogoča napoved segmentacijskih mask brez velike dodatne računske zahtevnosti, saj je dekodirnik SAM računsko nezahteven.

Metoda DSC za vsak označen primerek objekta izlušči dve vrsti prototipov (videz in oblika). Prototipi videza $\mathbf{p}^A \in \mathbb{R}^{k\times d}$ se izluščijo z operacijo RoI-pooling [4] nad značilkami slike $\mathbf{f}^I$ iz lokacij očrtanih okvirjev primerkov. Po zgledu [2], se prototipi oblike $\mathbf{p}^S \in \mathbb{R}^{k\times d}$ izračunajo kot $\mathbf{p}^S_i = \Phi([W_{\mathbf{b}_i}, H_{\mathbf{b}_i}])$, kjer sta $W_{\mathbf{b}_i}$ in $H_{\mathbf{b}_i}$ širina in višina očrtanega okvirja i-tega primerka, $\Phi(\cdot)$ pa je plitka MLP mreža. Konkatenacija $\mathbf{p}^A$ in $\mathbf{p}^S$ da končne prototipe $\mathbf{p} \in \mathbb{R}^{2k\times d}$.

Informacija prototipov $\mathbf{p}$ se nato prenese v značilke celotne slike $\mathbf{P}_0 = \mathbf{f}^I$ z operacijo medpozornosti (*angl.* cross-attention) $\mathbf{P}_i = \text{CA}(\mathbf{P}_{i-1}, \mathbf{p}, \mathbf{p})$ za kreiranje gostih prototipov. Ti se v naslednjem koraku transformirajo v goste poizvedbe (*angl.* dense queries) za detekcijo objektov. Goste poizvedbe $\mathbf{Q}$ se napovedo z operacijo medpozornosti med $\mathbf{P}_i$ in značilkami slike $\mathbf{f}^I$, ki se ponovi trikrat. Proces dekodiranja gostih poizvedb obsega tri konvolucijske faze povečevanja, vsaka izmed teh je sestavljena iz 3×3 konvolucije, Leaky ReLU in $2\times$ bilinearnega povečevanja. DSC iz povečanih poizvedb $\mathbf{Q}^{HR}$ s preprosto transformacijo za vsako pozicijo napove verjetje, da se tam nahaja objekt ciljnega razreda in ali je lokacija objekta tam zanesljivo napovedana, tj. mapo odzivov $\mathbf{y}^{\text{O}} = \text{LRelu}(\mathbf{W}_O \cdot \mathbf{Q}^{HR})$, kjer je $\mathbf{W}_O$ naučena projekcijska matrika, $\text{LReLU}(\cdot)$ pa aktivacijska funkcija Leaky ReLU. Vsaka poizvedba se dekodira tudi v lokacije objektov, tj. parametre očrtanih pravokotnikov v *tlrb* formatu [18] $\mathbf{y}^{\text{BB}} \in \mathbb{R}^{H\times W\times 4}$. Dekodiranje se izvede kot $\mathbf{y}^{\text{BB}} = \sigma(\text{MLP}(\mathbf{Q}^{HR}))$, kjer je $\text{MLP}(\cdot)$ tri-nivojski polno povezani perceptron, $\sigma(\cdot)$ pa sigmoidna funkcija.

Končne detekcije se pridobijo na sledeč način. Parametri očrtanih okvirjev se preberejo iz $\mathbf{y}^{\text{BB}}$ na lokacijah lokalnih maksimumov mape odzivov $\mathbf{y}^{\text{O}}$ z uporabo 3×3 NMS operacije. Očrtani okvirji so kot iztočnice (*angl.* prompts) skupaj s prej pridobljenimi značilkami slike $\mathbf{f}^I$ podani v dekoder SAM za napoved segmentacijskih mask. Izpopoljnjene očrtane okvirje DSC izračuna s segmentacijskih mask z min-max operacijo. Na koncu se uporabi NMS z $\text{IoU} = 0.5$ za odstranitev podvojenih detekcij kar nam da končno množico omejitvenih okvirjev $\mathbf{B}^P$ in pripadajoče segmentacijske maske $\mathbf{M}^P$.

4 Funkcija izgube

Učenje detekcijskih števcev, ki temeljijo na napovedi mape odzivov, zahteva nadzorovano učenje parametrov modela za napoved očrtanih okvirjev $\mathbf{y}^{BB}$ in mape odzivov $\mathbf{y}^{O}$. Zaželeno je, da se nevronska mreža nauči napovedovati mapo odzivov, na kateri se lahko z uporabo filtra NMS zanesljivo oceni lokalne maksimume. Poleg tega pa je zaželeno tudi, da se maksimumi nahajajo na optimalnih predelih objektov za napoved očrtanih okvirjev. V nadaljevanju bomo opisali nekaj funkcij izgube, ki zasledujejo navedene lastnosti.

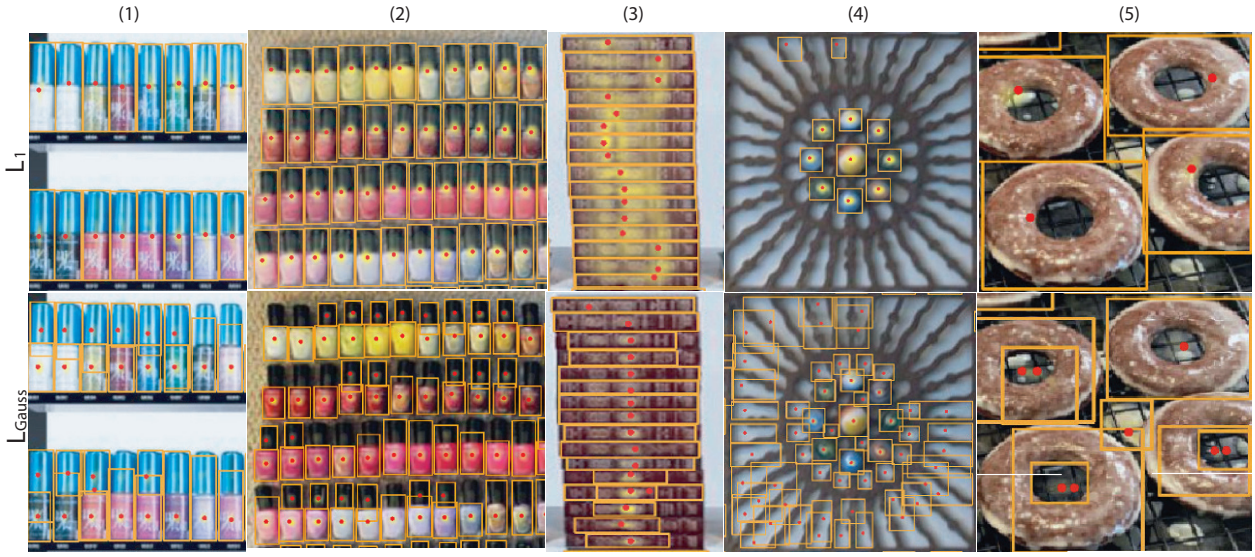
Po prehodu naprej (*angl.* forward pass) se s filtrom NMS na mapi odzivov $\mathbf{y}^{O}$ identificira niz lokalnih maksimumov $i{=}\,1:N_{\text{D}}$ in obdrži vse maksimume višje od mediane mape odziva, da se zagotovi čim višji priklic detekcij. Maksimumi se nato označijo kot *resnično pozitivni* (TP) in *lažno pozitivni* (FP) z bipartitnim ujemanjem [6] isto-ležečih očrtanih okvirjev ${\mathbf{y}_i^{BB}}_{i=1:N_{\text{D}}}$ z referenčnimi očrtanimi okvirji ${\mathbf{B}_j^{GT}}_{j=1:N_{\text{GT}}}$. Centri referenčnih omejitvenih okvirjev, ki ne prejmejo ujemajoče detekcije se označijo kot *lažni negativni* primeri (FN). Nova kriterijska funkcija je tako definirana kot:

$$\mathcal{L}^{(l)} = \mathcal{L}_{\mathbf{y}^O} + \mathcal{L}^{(l)}_{\mathbf{y}^{BB}}, \tag{1}$$

kjer je prvi člen definiran kot

$$\mathcal{L}_{\mathbf{y}^O} = \sum_{i\in\text{TP}\cup\text{FN}} (\mathbf{y}^{\text{O}}_i - 1)^2 + \sum_{i\in\text{FP}} (\mathbf{y}^{\text{O}}_i - 0)^2, \tag{2}$$

in nadzira učenje mape odzivov $\mathbf{y}^{\text{O}}$. Drugi člen lahko definiramo na več načinov – v nadaljevanju predstavimo



Slika 2: Metoda učena s funkcijo izgube $\mathcal{L}^{(1)}$ je prikazana v prvi vrsti, v drugi pa metoda učena na standarden način [13, 2, 23] tj. s funkcijo izgube $\mathcal{L}_{\text{Gauss}}$. Mape odzivov so prikazane z rumeno barvo, izbrane lokacije za napoved lokacije objektov (oranžnih očrtanih okvirjev) pa z rdečimi točkami.

štiri take načine, prvi je sledeč:

$$\mathcal{L}^{(1)}_{\mathbf{y}^{BB}} = -\sum_{i \in \text{TP}} \text{gIoU}(\mathbf{y}i^{BB}, \boldsymbol{B}^{GT}_{\text{HUN}(i)}), \tag{3}$$

kjer je $\text{gIoU}(\cdot,\cdot)$ generaliziran IoU [16], $\text{HUN}(i)$ pa je indeks ujemanja referenčnega očrtanega okvirja z i-tim napovedanim omejitvenim okvirjem. Drugi način je sledeč:

$$\mathcal{L}^{(2)}_{\mathbf{y}^{BB}} = \sum_{i=1:N_{\text{D}}} \sum_{j=1:N_{\text{GT}}} -\mathbb{1}(i, \boldsymbol{B}^{GT}_j)\text{gIoU}(\mathbf{y}i^{BB}, \boldsymbol{B}^{GT}_j), \tag{4}$$

kjer je $\mathbb{1}(i, \boldsymbol{B}^{GT}_j)$ indikacijska funkcija, ki je aktivirana, v primeru da je i-ti maksimum znotraj $\boldsymbol{B}^{GT}_j$ očrtanega okvirja, sicer zavzema vrednost nič. Torej, učenje napovedi očrtanih okvirjev je nadzorovano na vseh lokalnih maksimumih znotraj referenčnih očrtanih okvirjev, ne glede na to ali so uvrščeni med TP, ali FP.

V gostih regijah, kjer se objekti nahajajo blizu drug drugemu, so sosednji centri objektov le nekaj slikovnih elementov narazen. Zato funkciji izgube dodamo kriterij $\mathcal{L}^{(k)}_N$, ki omogoča, da so lokalni maksimumi dobro izraženi, kljub bližini objektov. Funkciji izgube $\mathcal{L}^{(3)}$ in $\mathcal{L}^{(4)}$ sta torej definirani kot

$$\mathcal{L}^{(k)} = \mathcal{L}_{\mathbf{y}^O} + \mathcal{L}^{(1)}_{\mathbf{y}^{BB}} + \mathcal{L}^{(k)}_N, \tag{5}$$

kjer $k \in \{3, 4\}$. V primeru $\mathcal{L}^{(3)}_N = \sum_{i \in \text{NEG}} (\mathbf{y}^{\text{O}}_i - 0)^2$ množica NEG predstavlja lokacije na sredini daljice med centri najbližjih sosedov – člen $\mathcal{L}^{(3)}_N$ torej zahteva nične vrednosti $\mathbf{y}^{\text{O}}_i$ na teh lokacijah. V primeru $\mathcal{L}^{(4)}_N$ množica NEG predstavlja vse lokacije izven referenčnih očrtanih okvirjev, kjer se prav tako zahteva čim nižja vrednost odziva $\mathbf{y}^{\text{O}}_i$.

Analizirane funkcije izgube neposredno optimizirajo nalogo detekcije brez učenja nadomestnih nalog, kot je napovedovanje Gaussovih gostotnih map. Take funkcije omogočajo avtonomno in neparametrično izbiro lokacije, ki je optimalna za napoved očrtanih okvirjev, kar vodi do bolj kvalitetnih napovedi (Slika 2). Funkcije izgube neposredno združujejo koristne lastnosti polno-konvolucijskih detektorjev s formulacijo funkcij izgube transformerskih detektorjev DETR [1]. Uporabljajo se lahko pri kateri koli detekcijski metodi, ki v enem koraku napove mapo odzivov in očrtane okvirje za vsako lokacijo v mapi odzivov.

5 Eksperimenti

5.1 Učenje

Z zamrznjenimi parametri hrbtenice metode SAM [5] je DSC najprej učen s klasično funkcijo izgube [13, 2, 7], nato pa treniran 200 epoh z vsako od štirih verzij funkcije izgube. Metoda je trenirana z velikostjo paketa 8, optimizatorjem AdamW [8], začetno stopnjo učenja 10^{-4} in razpadom uteži (*angl.* weight decay) 10^{-4}. Trening poteka na dveh A100 GPU-jih z uporabo standardnih metod za bogatenje podatkov [13, 2] in skaliranja slik na ločljivost 1024×1024.

5.2 Evalvacijski protokol

Kriterijske funkcije smo ovrednotili na standardni zbirki FSCD-147 [10], ki vsebuje 6135 slik s 147 kategorijami. Slike so razdeljene na učno, validacijsko in testno množico, ki obsegajo 3659, 1286 in 1190 slik tako, da so vsi primerki ene kategorije znotraj ene množice. Število objektov na posamezni sliki variira med 7 in 3731. Na vsaki sliki so z očrtanimi okvirji podani tudi trije objekti ($k = 3$), ki opredelijo kategorijo objektov za štetje.

Uporabimo standardni meri štetja [9], t.j., povprečna absolutna napaka (MAE) in korenjena povprečna kvadratna napaka (RMSE), ter detekcije [10], t.j., povprečno natančnost pri vrednosti IoU (Intersection Over Union) 0.5 torej AP50, in povprečno natančnost AP, ki zajema pov-

prečje natančnosti pri upragovanih vrednostih IoU med 0.5 in 1.0 z razmiki 0.05. Več podrobnosti o detekcijskih merah lahko bralec najde v [11].

5.3 Rezultati

Rezultati sposobnosti štetja in detekcije objektov metode DSC z malo učnimi primeri, ki je trenirana z različnimi funkcijami izgube so predstavljeni v Tabeli 1. Iz rezultatov je razvidno, da funkcija izgube igra pomembno vlogo pri uspešnosti štetja in detekcije.

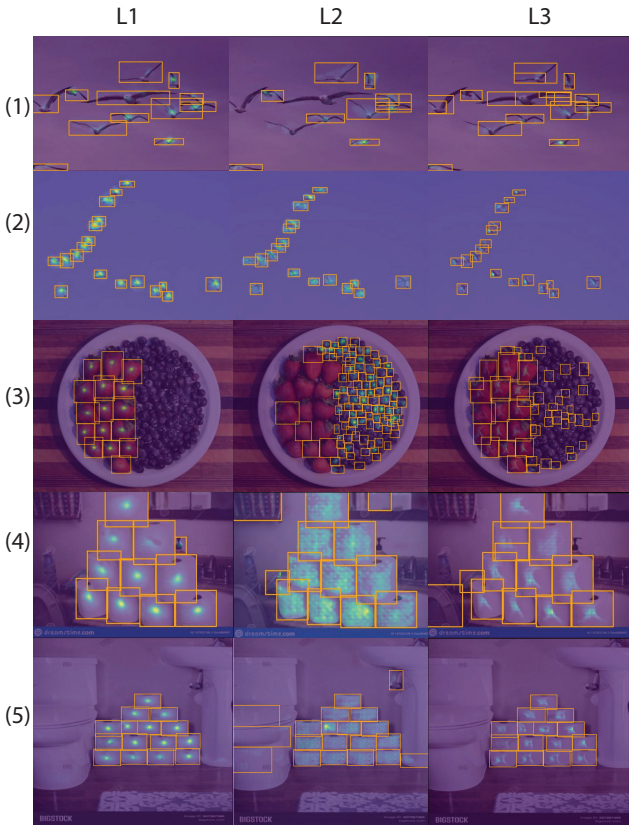
Osnovna verzija metode DSC je trenirana s standardno funkcijo izgube [13, 23] $\mathcal{L}_{\text{Gauss}}$ in dosega slabše rezultate, kar je razvidno tudi iz kvalitativnih rezultatov na Sliki 2. Kot je prikazano v stolpcih 3 in 5, $\mathcal{L}_{\text{Gauss}}$ optimizira mrežo za napovedovanje lokacije objektov s sredine očrtanih okvirjev, kar ni nujno optimalno. Vse različice predlagane funkcije izgube pa omogočajo, da mreža avtonomno izbere optimalno točko za napoved očrtanega okvirja. Iz stolpcev 1 in 2 je razvidno, da metoda učena s funkcijo izgube $\mathcal{L}^{(1)}$ optimalno izbere točko za napoved očrtanega okvirja tudi na objektih, ki so sestavljeni z več komponent, kar je pogost problem pri števcih z malo učnimi primeri [13, 23, 10]. Poleg tega pa funkcija izgube $\mathcal{L}^{(1)}$ omogoča samodejno iskanje težkih negativnih primerov preko identifikacije FP.

Tabela 1: Primerjava zmogljivosti štetja in detekcije metode DSC učene z različnimi funkcijami izgube.

	MAE(↓)	RMSE(↓)	AP(↑)	AP50(↑)
$\mathcal{L}_{\text{Gauss}}$	10.05	91.51	41.43	71.95
$\mathcal{L}^{(1)}$	**7.91**	**54.28**	**43.43**	**75.06**
$\mathcal{L}^{(2)}$	10.50	88.01	41.06	72.94
$\mathcal{L}^{(3)}$	12.09	81.01	43.37	74.32
$\mathcal{L}^{(4)}$	18.55	103.00	39.41	69.22

Model treniran z $\mathcal{L}^{(2)}$, kjer učenje očrtanih okvirjev poteka skozi vse maksimume znotraj očrtanih okvirjev, dosega slabše rezultate štetja in posledično tudi detekcije. Do poslabšanja uspešnosti pride, ker med učenjem lokacija optimalnega maksimuma za napoved očrtanega okvirja varira. Model treniran z $\mathcal{L}^{(3)}$, kjer na sredino med najbližjimi sosedi v mapi odzivov postavimo ničen odziv tudi dosega slabše rezultate, ker je sredinska točka med sosednjima objektoma večkrat del objekta, kar otežuje uspešnost učenja, saj v model vnaša dodaten šum. Pri funkciji izgube $\mathcal{L}^{(4)}$, kjer je mapa odzivov učena tako, da zunaj očrtanih okvirjev napoveduje ničen odziv, pa močno naraste število FN primerov, saj je model bolje ocenjen, če uspešno napove večino ničel izven objektov, kot pa če dobro napoveduje optimalno lokacijo za napoved očrtanega okvirja.

Na Sliki 3 kvalitativno predstavimo rezultate metode trenirane s tremi najbolj uspešnimi funkcijami izgube, tj. $\mathcal{L}^{(1)}$, $\mathcal{L}^{(2)}$ in $\mathcal{L}^{(3)}$. V primeru učenja s funkcijo $\mathcal{L}^{(1)}$ metoda dosega najmanj lažno pozitivnih detekcij (3-5). Pri učenju s funkcijo izgube $\mathcal{L}^{(2)}$, kjer se napoved očrtanih okvirjev uči na vseh maksimumih znotraj očrtanih okvirjev, med iteracijami učenja optimalni maksimum spreminja lokacije, zato se odzivi tako učene metode razprostrejo čez celoten objekt (5) in niso jasno izraženi (2). Metoda učena z $\mathcal{L}^{(3)}$ ima sicer bolj jasno izražene maksimume sploh v gostih regijah majhnih objektov (2), vendar zaradi tega napoveduje podvojene detekcije (1), ki povečajo napako štetja.



Slika 3: Mape odzivov in očrtani okvirji metode DSC učene s funkcijami izgube $\mathcal{L}^{(1)}$, $\mathcal{L}^{(2)}$ in $\mathcal{L}^{(3)}$.

6 Zaključek

Na problemu štetja in detekcije objektov z malo učnimi primeri smo raziskovali različne načine učenja in funkcije izgube za metodo DSC [14], ki podobno kot polnokonvolucijski detektorji [18] napoveduje mapo odzivov in očrtane okvirje za vsak slikovni element. Identificirali smo težave obstoječega standardnega načina za učenje števcev [13, 23, 2, 7, 17] za napoved gostotne mape, ki se je prenesel na detekcijske števce. Pokazali smo, da je napoved Gaussovih porazdelitev čez objekte neprimerna za detekcijo objektov in analizirali funkcije izgube, ki neposredno optimizirajo uteži modela za detekcijo.

7 Zahvala

Raziskave so bile delno podprte s programi in projekti ARIS P2-0214, J2-2506 in Z2-4459 ter superračunalniškim omrežjem SLING (ARNES, EuroHPC Vega - IZUM).

Literatura

[1] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European conference on computer vision. pp. 213–229. Springer (2020)

[2] Djukic, N., Lukezic, A., Zavrtanik, V., Kristan, M.: A low-shot object counting network with iterative prototype adaptation. In: ICCV (2023)

[3] Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. pp. 1126–1135. PMLR (2017)

[4] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)

[5] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)

[6] Kuhn, H.W.: The hungarian method for the assignment problem. Naval research logistics quarterly **2**(1-2), 83–97 (1955)

[7] Liu, C., Zhong, Y., Zisserman, A., Xie, W.: Countr: Transformer-based generalised visual counting. In: BMVC. BMVA Press (2022)

[8] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. ICLR (2019)

[9] Lu, E., Xie, W., Zisserman, A.: Class-agnostic counting. In: ACCV. pp. 669–684. Springer (2018)

[10] Nguyen, T., Pham, C., Nguyen, K., Hoai, M.: Few-shot object counting and detection. In: ECCV. pp. 348–365. Springer (2022)

[11] Padilla, R., Netto, S.L., da Silva, E.A.B.: A survey on performance metrics for object-detection algorithms. In: (IWSSIP). pp. 237–242 (2020)

[12] Pelhan, J., Lukežič, A., Zavrtanik, V., Kristan, M.: Detekcijska metoda za štetje objektov z malo učnimi primeri. p. 334–337. Slovenska sekcija IEEE; Fakulteta za elektrotehniko (2023)

[13] Pelhan, J., Lukežič, A., Zavrtanik, V., Kristan, M.: Dave – a detect-and-verify paradigm for low-shot counting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2024)

[14] Pelhan, J., Lukežič, A., Zavrtanik, V., Kristan, M.: Detekcijsko-segmentacijska metoda za stetje objektov z malo učnimi primeri (2024)

[15] Ranjan, V., Hoai, M.: Vicinal counting networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 4221–4230 (June 2022)

[16] Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S.: Generalized intersection over union. In: CVPR (2019)

[17] Shi, M., Lu, H., Feng, C., Liu, C., Cao, Z.: Represent, compare, and learn: A similarity-aware framework for class-agnostic counting. In: CVPR. pp. 9529–9538 (June 2022)

[18] Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9627–9636 (2019)

[19] Xie, W., Noble, J.A., Zisserman, A.: Microscopy cell counting and detection with fully convolutional regression networks. Computer methods in biomechanics and biomedical engineering: Imaging & Visualization **6**(3), 283–292 (2018)

[20] Yang, S.D., Su, H.T., Hsu, W.H., Chen, W.C.: Class-agnostic few-shot object counting. In: WACV. pp. 870–878 (2021)

[21] You, Z., Yang, K., Luo, W., Lu, X., Cui, L., Le, X.: Few-shot object counting with similarity-aware feature enhancement. In: WACV. pp. 6315–6324 (2023)

[22] Zavrtanik, V., Vodopivec, M., Kristan, M.: A segmentation-based approach for polyp counting in the wild. Engineering Applications of Artificial Intelligence **88**, 103399 (2020)

[23] Zhizhong, H., Mingliang, D., Yi, Z., Junping, Z., Hongming, S.: Point, segment and count: A generalized framework for object counting. In: CVPR (2024)