

# Računalni vid

Naučena detekcija i segmentacija

Siniša Šegvić, Josip Šarić

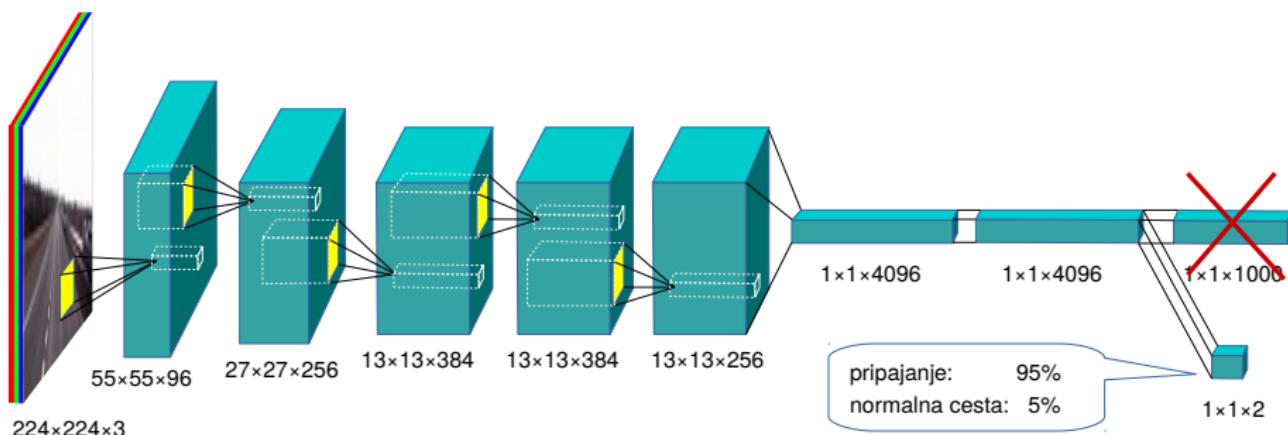
## SADRŽAJ

- Prijenos znanja s ImageNeta
- Detekcija objekata
- Gusta predikcija

# PRIJENOS ZNANJA S IMAGENETA: IDEJA

Klasifikacijski model naučen na ImageNetu prilagoditi za novi zadatak.

- odrezati posljednjih nekoliko slojeva
- spojiti preostale slojeve s prednjim krajem za novi zadatak
- trenirati (ugoditi) dobiveni model za novi zadatak
- naslijedeni slojevi već su naučeni pa sada možemo učiti s manje podataka (nekoliko tisuća slika)



# PRIJENOS ZNANJA S IMAGENETA: PREDNOSTI I IZAZOVI

Prednosti:

- brža konvergencija
- bolja točnost
- potrebno manje označenih slika za ciljni zadatak (posebno velika ušteda na skupim oznakama za guste zadatke)

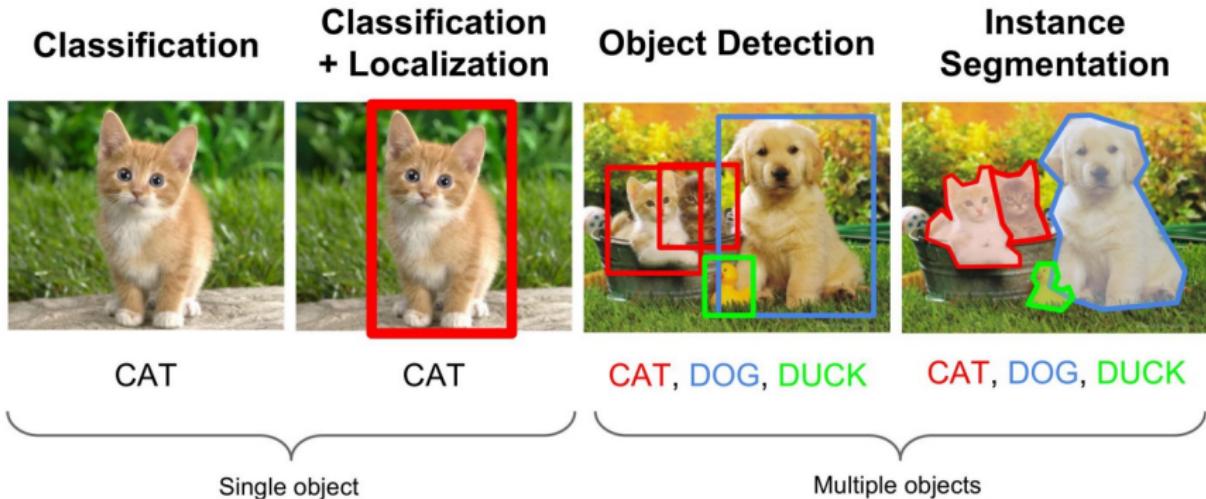
Izazovi prilikom dizajna modela za gustu predikciju:

- oporavak od poduzorkovanja
- povećanje receptivnog polja

# DETEKCIJA OBJEKATA: Uvod

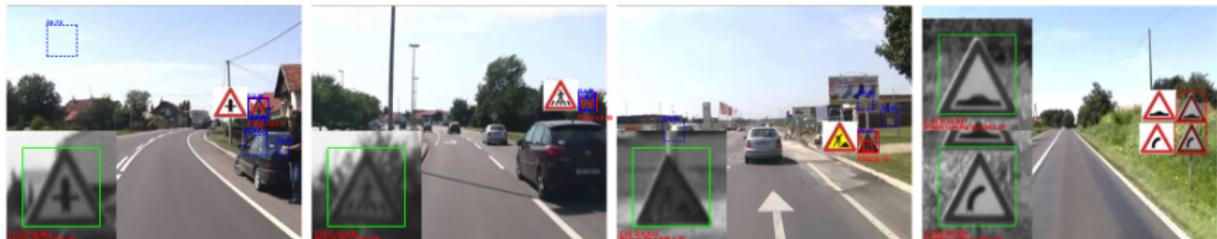
Detekcija objekata podrazumijeva lokalizaciju svakog objekta na slici u obliku opisujućeg pravokutnika i njegovu klasifikaciju u jedan od semantičkih razreda.

Naizgled jednostavan zadatak, ali varijabilan broj mogućih izlaza stvara izazove prilikom dizajna dubokih modela s kraja na kraj.



# DETEKCIJA OBJEKATA: POMIČNO OKNO

Klasična detekcija u pomičnom oknu zahtijeva relativno jednostavne značajke i klasifikatore



Takav pristup nije prikladan za istovremenu detekciju više razreda

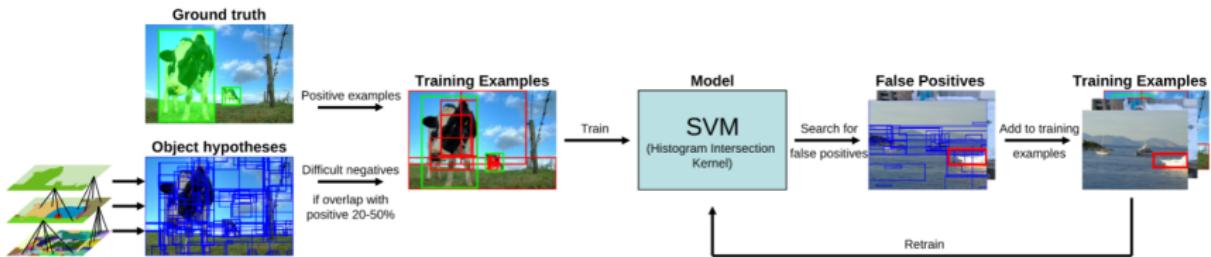
- npr: stol, tanjuri, pribor, mačke, kameleon, kotači, automobili



# DETKECIJA OBJEKATA: DVOPROLAZNI PRISTUPI

Zato su početkom 2010-ih godina popularizirani dvoprolazni pristupi

1. prvo pronaći kandidate ( $\sim 1000$ ) općenitim brzim postupkom
2. testirati i klasificirati kandidate teškom artiljerijom (BoW)



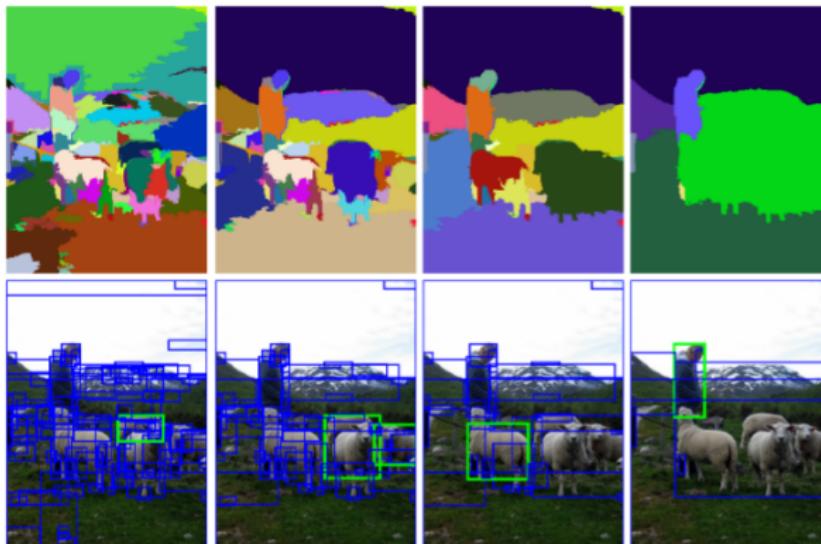
[uijlings13ijcv]

Ovo možemo promatrati i kao najnapredniji klasični pristup i kao prijelazni oblik prema dubokim modelima

## DETEKCIJA OBJEKATA: SELEKTIVNO PRETRAŽIVANJE

Ideja [uijlings13ijcv]: detektirati kandidate primjenom i) segmentacije u superpixele te ii) ručno dizajniranih strategija grupiranja

- kriteriji grupiranja: boja, tekstura, veličina, nadopunjavanje

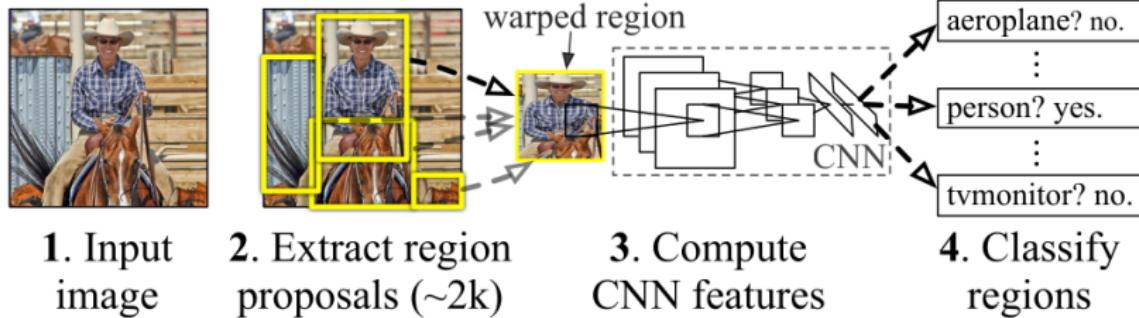


- točnost na VOC'12 test: 35.0% mAP@0.5

# DETEKCIJA OBJEKATA: R-CNN

Glavna ideja: features matter

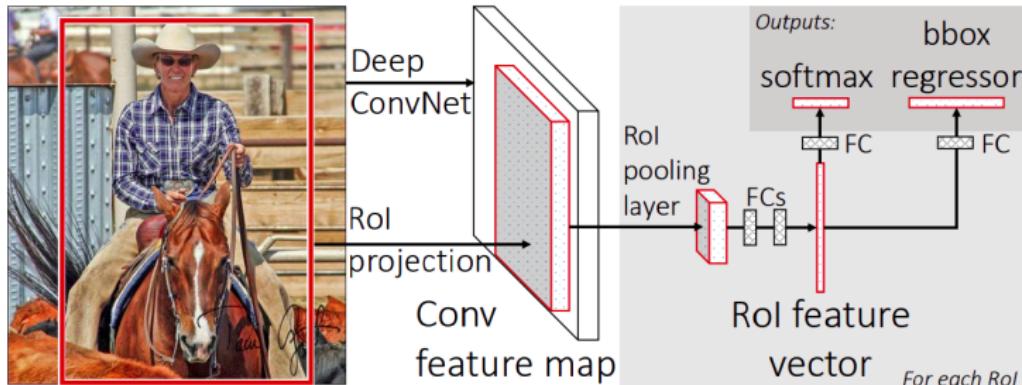
- zamijeniti colourSIFT-BoW predtreniranim dubokim modelom
- ugoditi duboki model za klasifikaciju u C+1 razred na grupama od 32 pozitivnih i 96 negativnih kandidata
- konačnu klasifikaciju provesti SVM-om (validacija: +4pp mAP@0.5)
- VOC'12 test: 53.3% mAP@0.5



# DETEKCIJA OBJEKATA: FAST R-CNN

Glavna ideja: učenje s kraja na kraj, kandidate još uvijek generira spori klasični postupak [uijlings13ijcv]

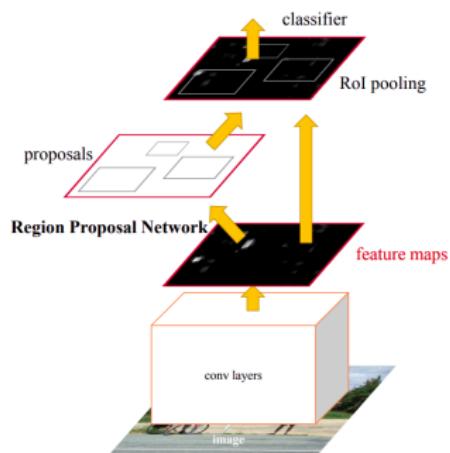
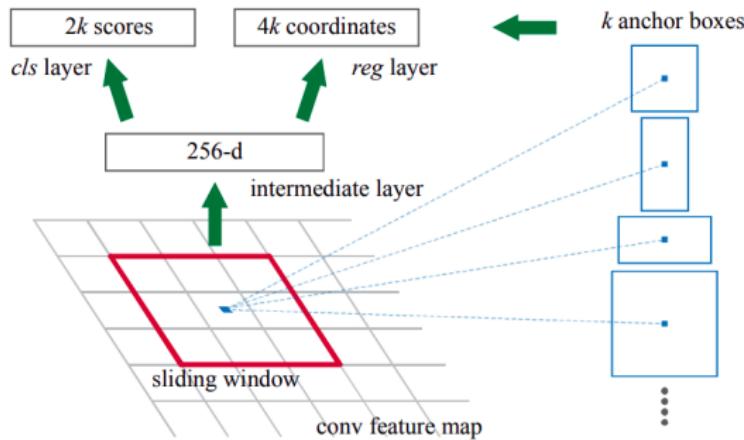
- provući sliku kroz duboki model, izlučiti značajke  $h/32 \times w/32 \times D$ 
  - izraziti regiju svakog kandidata deskriptorom  $7 \times 7 \times D$  (ROI pool)
  - svaki element ROI poola je kopija odgovarajuće značajke
  - nema interpolacije, backprop je jednostavan jer je okno fiksno
- klasifikaciju i popravljanje okvira provodi zasebno naučeni model
  - VOC'12 test: 66% mAP@0.5; COCO test-dev: 19.7% mAP COCO



# DETEKCIJA OBJEKATA: FASTER R-CNN

Glavna ideja: integrirano generiranje kandidata i detekcija objekata

- provući sliku kroz duboki model, izlučiti značajke  $h/32 \times w/32 \times D$
- iz tih značajki, gusto detektirati kandidate i pomake za k sidara
  - modul za predlaganje kandidata (RPN - region proposal network)
- izlučiti deskriptor svakog pozitivnog kandidata ( $7 \times 7 \times D$ , ROI pool) i proslijediti ga klasifikacijskom modulu (slično Fast R-CNN)



## DETEKCIJA OBJEKATA: FASTER R-CNN (2)

Backprop kroz ROI pool sada je komplikiran jer okno nije fiksno

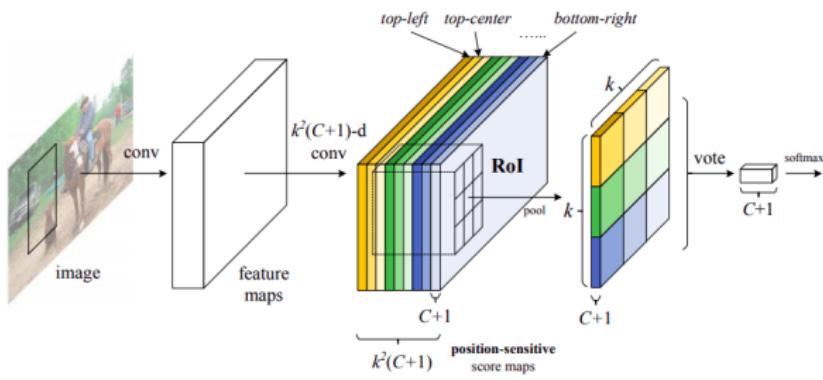
- zbog toga učenje alternira između optimiranja RPN i klasifikacijskog modula
- ovaj problem rješava bilinearno uzorkovanje deskriptora regije (ROI align)

VOC'12 test: 70.4% mAP@0.5; COCO test-dev: 21.9% mAP COCO

# DETKECIJA OBJEKATA: R-FCN

Ideja: cjelokupan posao provesti jednim modulom

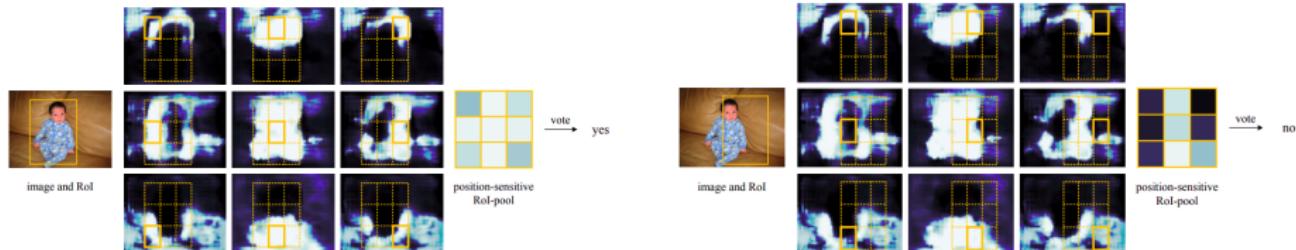
- glavni trik: pozicijski osjetljivo sažimanje
  - okvir kandidata se podijeli na  $k \times k$  zona;
  - svaka mapa agregira jednu od tih zona za svaki razred
  - puno veći domet od konvolucija
- jednako točno kao Faster R-CNN, ali značajno brže



[dai16nips]

## DETAKCIJA OBJEKATA: R-FCN (2)

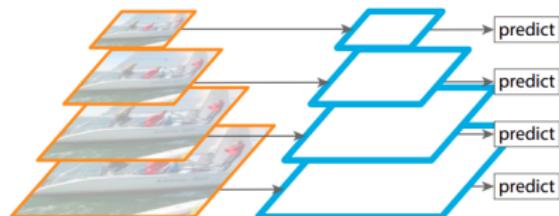
Pozicijski osjetljive mape aktiviraju se na odgovarajućim relativnim položajima u odnosu na objekt:



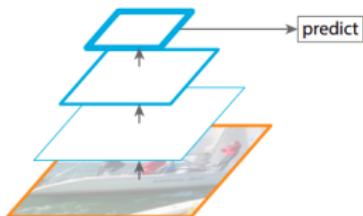
[dai16nips]

VOC'12 test: 77.6% mAP@0.5; COCO test-dev: 31.5% mAP COCO

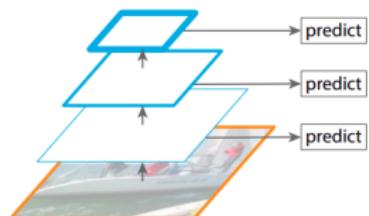
# DETJEKCIJA OBJEKATA: FPN



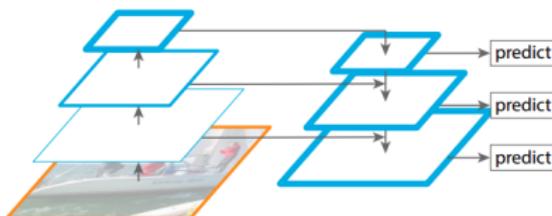
(a) Featurized image pyramid



(b) Single feature map



(c) Pyramidal feature hierarchy



(d) Feature Pyramid Network

[lin17cvpr]

Faster R-CNN + FPN, COCO test-dev: 36.2% mAP COCO

# DETKECIJA OBJEKATA: SOFTNMS

**Input** :  $\mathcal{B} = \{b_1, \dots, b_N\}$ ,  $\mathcal{S} = \{s_1, \dots, s_N\}$ ,  $N_t$

$\mathcal{B}$  is the list of initial detection boxes

$\mathcal{S}$  contains corresponding detection scores

$N_t$  is the NMS threshold

**begin**

$\mathcal{D} \leftarrow \{\}$

**while**  $\mathcal{B} \neq \text{empty}$  **do**

$m \leftarrow \text{argmax } \mathcal{S}$

$\mathcal{M} \leftarrow b_m$

$\mathcal{D} \leftarrow \mathcal{D} \cup \mathcal{M}; \mathcal{B} \leftarrow \mathcal{B} - \mathcal{M}$

**for**  $b_i$  in  $\mathcal{B}$  **do**

**if**  $iou(\mathcal{M}, b_i) \geq N_t$  **then**

$\mathcal{B} \leftarrow \mathcal{B} - b_i; \mathcal{S} \leftarrow \mathcal{S} - s_i$

**end**

NMS

$s_i \leftarrow s_i f(iou(\mathcal{M}, b_i))$

Soft-NMS

**end**

**end**

**return**  $\mathcal{D}, \mathcal{S}$

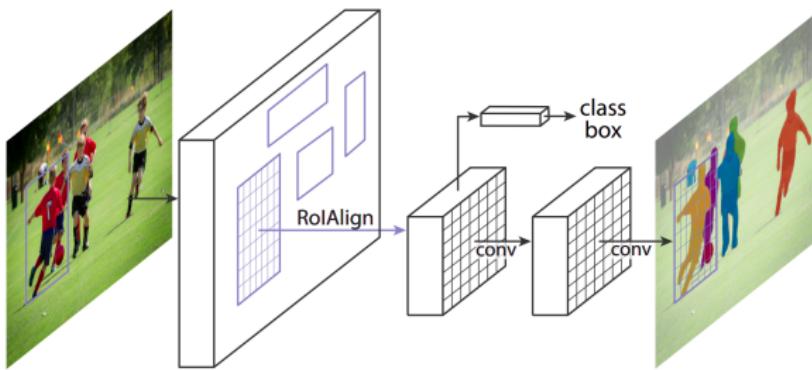
**end**

[bodla17iccv]

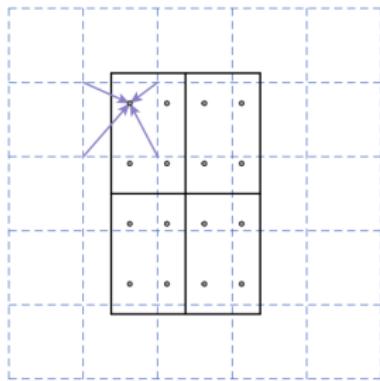
D-RFCN + MST + SoftNMS G, COCO test-dev: 40.9% mAP COCO

# DETKECIJA OBJEKATA: MASK R-CNN

Ideja: proširiti Faster glavom koja prediktira segmentaciju primjerka, dodati ROI align.



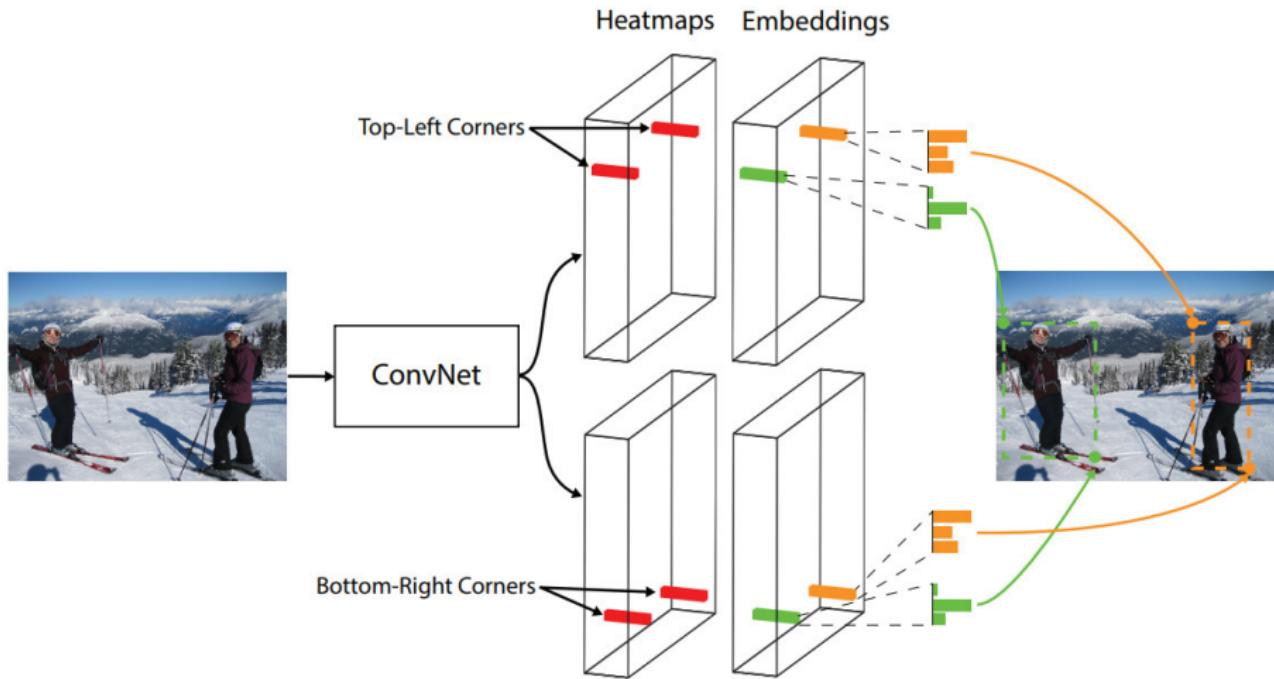
[he17iccv]



Mask R-CNN + FPN, COCO test-dev: 39.8% mAP COCO

# DETEKCIJA OBJEKATA: CORNERNET

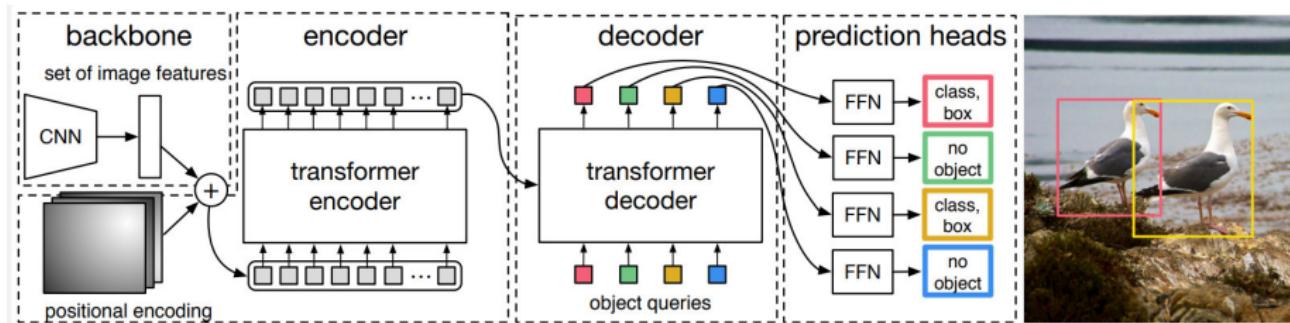
Ideja: Konvolucijski izračunati dvije toplinske mape za detekciju gornje-ljeve i donje-desne ključne točke. Ključne točke upariti prema predviđenom ugrađivanju.



# DETEKCIJA OBJEKATA: DETR

Ideja: Transformerom obraditi prepostavljeni skup upita koji predstavljaju objekte i izravno predviđati opisujući okvir i semantički razred za svaki upit.

- Tijekom učenja skup predviđenih okvira se uparuje s označenim okvirima na temelju bipartitnog gubitka uparivanja.
- Tijekom zaključivanja model može odbaciti neke upite/objekte tako da za njih predviđi poseban razred "no-object".



[carion20eccv]

# GUSTA PREDIKCIJA: SEGMENTACIJA

Razumijevanje slike na razini piksela (**semantička segmentacija**):

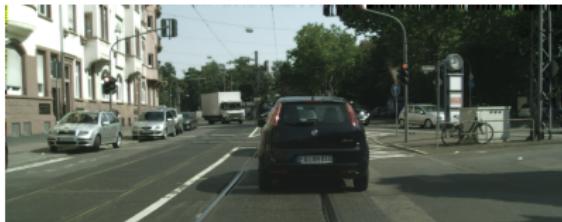
- svrstatи svaki slikovni element u odgovarajući razred

- razredi imaju značenje koje je važno za misiju agenta

**sudionici:** osoba, ciklist, auto, bicikl, kamion, autobus, vlak, motor

**signalizacija:** stup, znak, semafor

**okoliš:** cesta, nogostup, zgrada, bilje, teren, ograda, zid, nebo



## GUSTA PREDIKCIJA: ZADACI

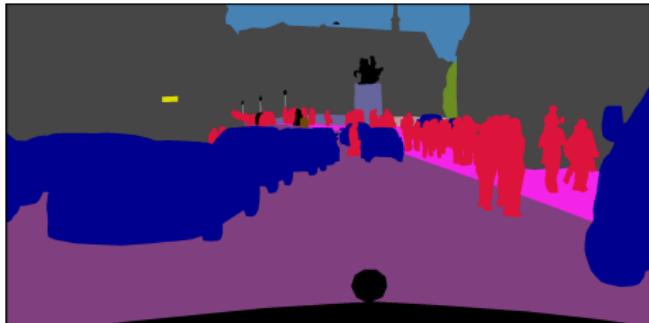
- semantička segmentacija: svakom pikselu dodjeljuje semantički razred, ne razlikuje primjerke
- segmentacija instanci (primjeraka): raspoznae samo prebrojive razrede (eng. things) i tim pikselima pridjeljuje semantički razred i identifikator primjerka, piksele neprebrojivih razreda zanemaruje
- panoptička segmentacija: svakom pikselu slike dodjeljuje semantički razred i identifikator primjerka, svi pikseli neprebrojivih razreda imaju isti identifikator

# GUSTA PREDIKCIJA: ZADACI

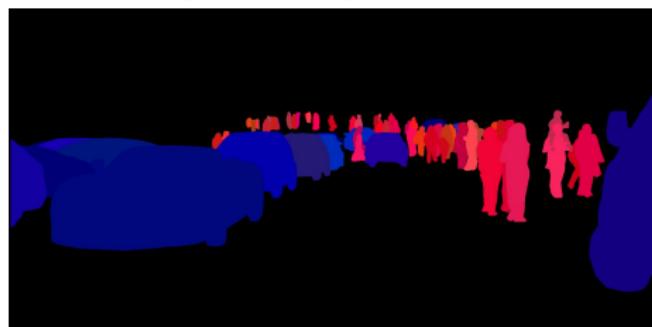
slika



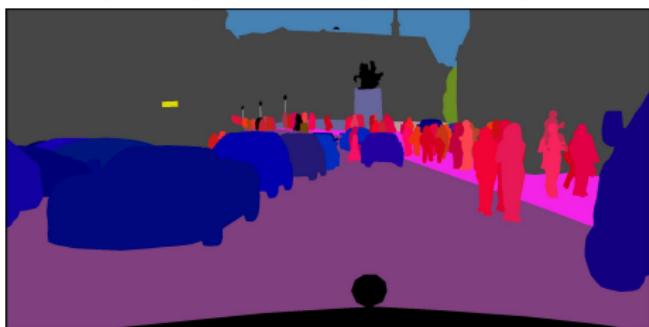
semantička segmentacija



segmentacija instanci

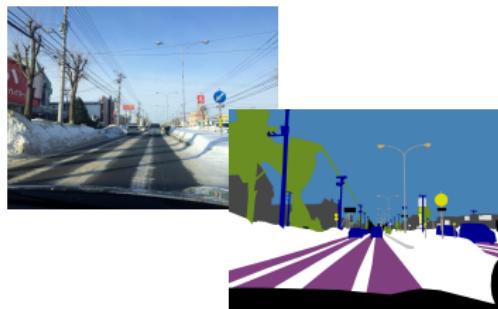
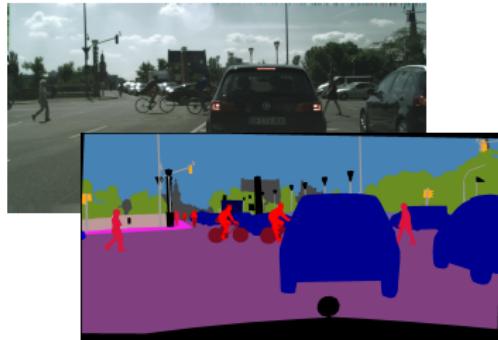


panoptička segmentacija



# GUSTA PREDIKCIJA: SKUPOVI PODATAKA

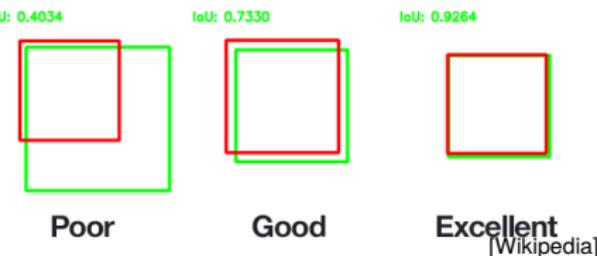
- Cityscapes [cordts16cvpr]:
  - perspektiva vozača, 19 razreda
  - 5000 stereo slika, 2MPixela
  - dobar odabir razreda i kategorija
  - 50 gradova, proljeće do jeseni
  
- Vistas [neuhold17iccv]:
  - perspektiva vozača, 100 razreda
  - 25000 slika, 2-8 MPixel
  - instance level annotations
  - širom svijeta, snijeg, magla, noć



# GUSTA PREDIKCIJA: TOČNOST

Široko korištena metrika: **omjer presjeka i unije (IoU)**

- skup A: označeni pikseli razreda c
- skup B: pikseli klasificirani u c
- $\text{IoU}_X = |A \cap B| / |A \cup B|$



Ukupnu uspješnost izražavamo kao srednji IoU preko svih razreda

- $mIoU = \frac{\sum_c \text{IoU}_c}{C}$
- ovo povećava utjecaj piksela rijetkih razreda
- primjeri: zid, ograda, stup, boca, sobna biljka

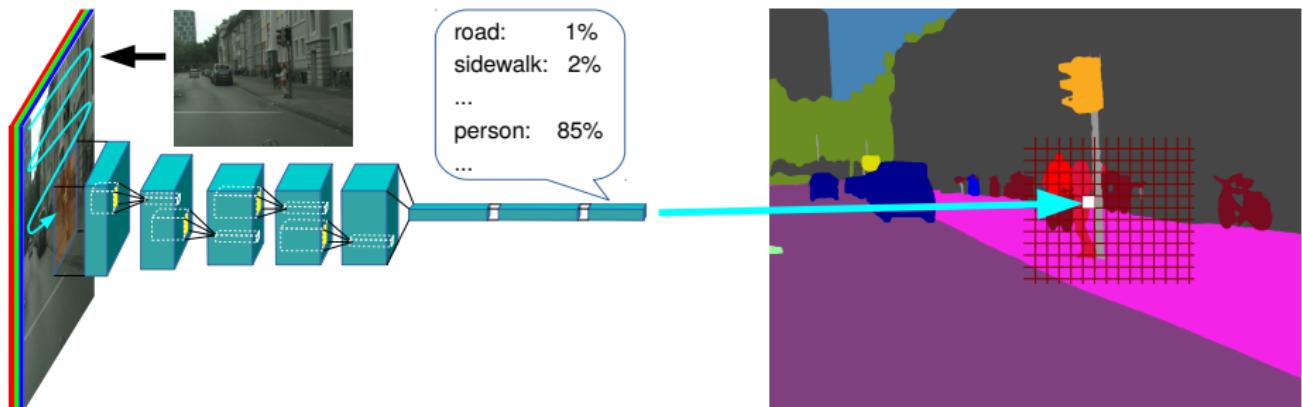
Oznake ispitnih slika nisu javno dostupne:

- ispitnu točnost određujemo podnošenjem na evaluacijski server
- naš mIoU po kategorijama: 89.7 (najbolji rezultat: 91.6)

# GUSTA PREDIKCIJA: POVRATAK POMIČNOG OKNA

Ideja: primijeniti klasifikacijski model u **pomičnom oknu**

- svako okno producira semantički razred piksela (ili okvir objekta)
- gusto **označene** slike omogućavaju učenje s kraja na kraj



U praksi potrebna optimizacija:  $10^6$  piksela  $\times 10^9$  množenja?

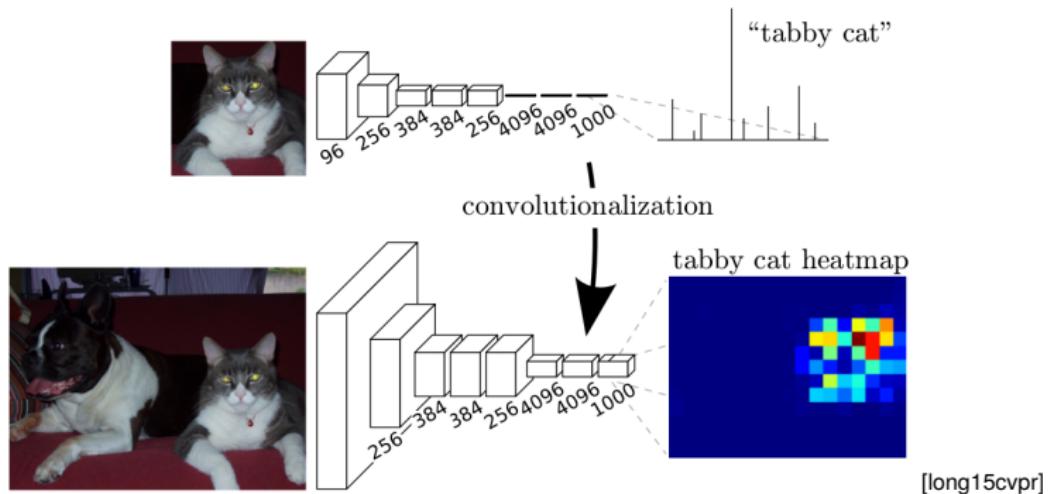
**Izazovi:** veliki objekti, mali objekti, računska složenost.

# GUSTA PREDIKCIJA: SEGMENTACIJA U PRAKSI

Obrada susjednih okana zahtijeva računanje istih latentnih aktivacija

**Optimizacija:** evaluirati pomicno okno **sloj po sloj** [long15cvpr]:

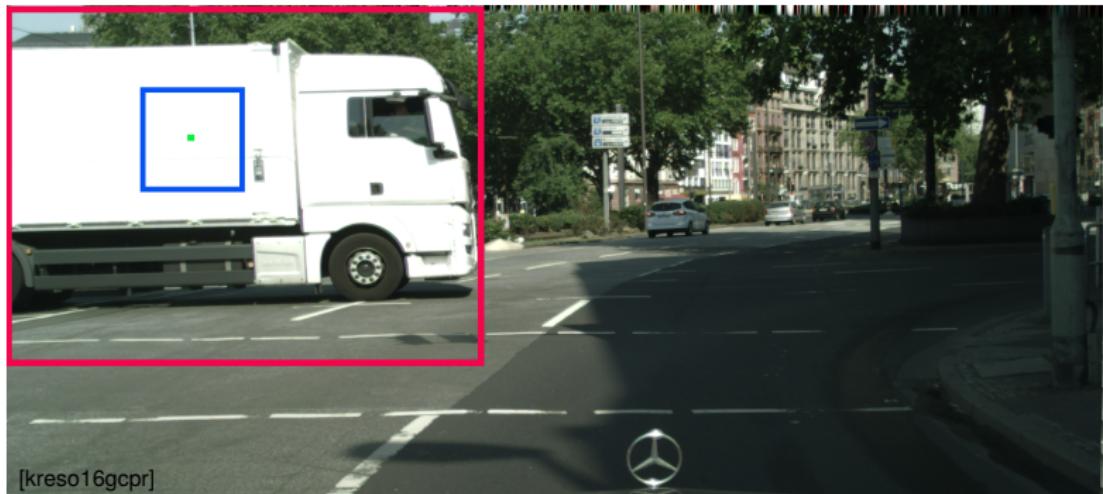
- izlazni tenzor je **poduzorkovan** zbog sažimanja



## GUSTA PREDIKCIJA: VELIKI OBJEKTI

Prepoznavanje velikih objekata zahtijeva gigantsko receptivno polje

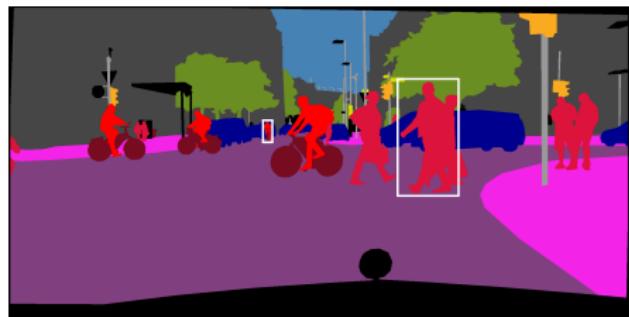
- velik broj lokalnih susjedstava nije dovoljno diskriminativan
- takva susjedstva mogu biti prepoznata samo u većem kontekstu
- problemi nastaju kada je kontekst veći od receptivnog polja



## GUSTA PREDIKCIJA: MALI OBJEKTI

Prepoznavanje **malih** objekata moćnim modelom s **velikim** receptivnim poljem rasipa resurse:

- mali objekti mogu biti prepoznati s malim brojem slojeva
- kasniji slojevi moraju prosljeđivati aktivacije bez doprinosa kvaliteti obrade
- to vodi do gubitka reprezentacijske moći i prenaučenosti modela

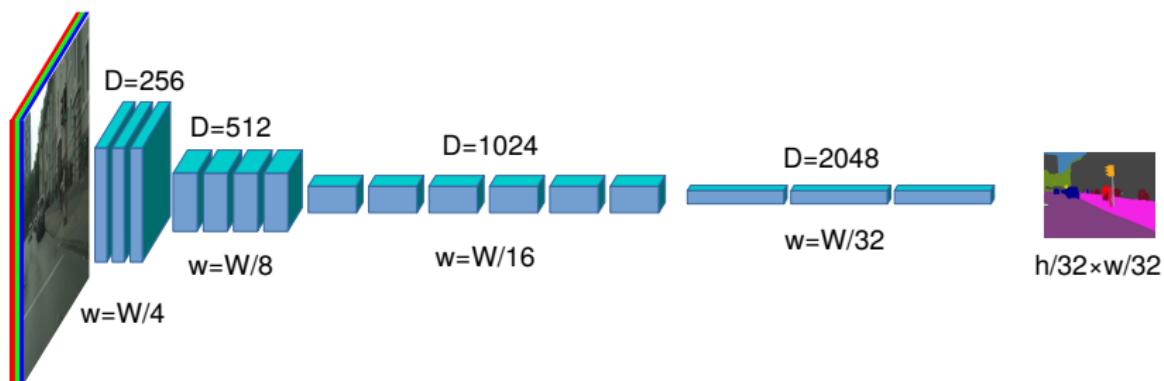


[kreso16gcpr]

## GUSTA PREDIKCIJA: SLOŽENOST

Uspješne segmentacijske arhitekture temelje se na prednaučenim klasifikacijskim modelima: mala ulazna i još manja izlazna rezolucija

- u segmentaciji trebamo veliku rezoluciju i na ulazu i izlazu
- to postavlja ogromne zahtjeve na GPU memoriju
  - prilikom učenja moramo pamtitи svih 100 latentnih tenzora
- to otežava evaluaciju modela na jednostavnim računalima
  - pola milijarde množenja po slici za najjednostavniji model



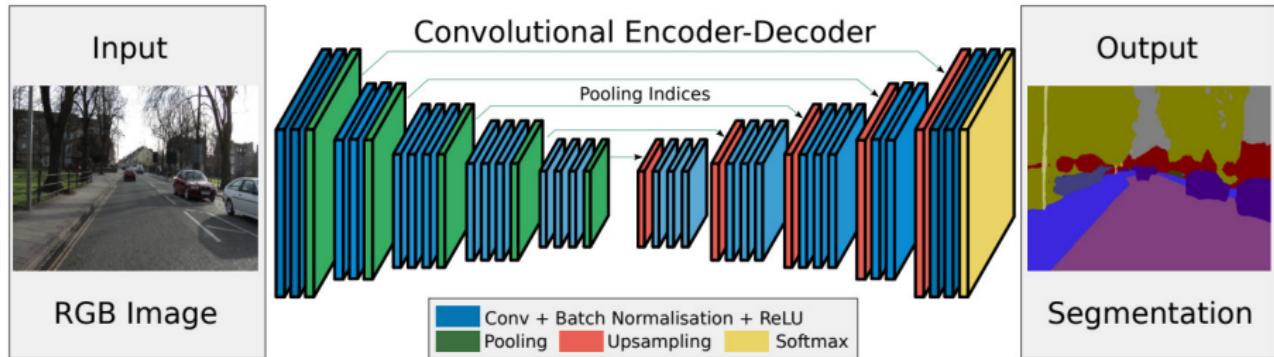
# GUSTA PREDIKCIJA: NADUZORKOVANJE

Za preciznu segmentaciju "oporavak" od poduzorkovanja iz klasifikacijskog modela je nužan.

Ideja: sačuvati indekse iz slojeva sažimanja i iskoristiti ih za naduzorkovanje.

Nedostaci:

- Slaba okosnica bez rezidualnih veza.
- Simetričan put naduzorkovanja (nepotrebno trošenje resursa).

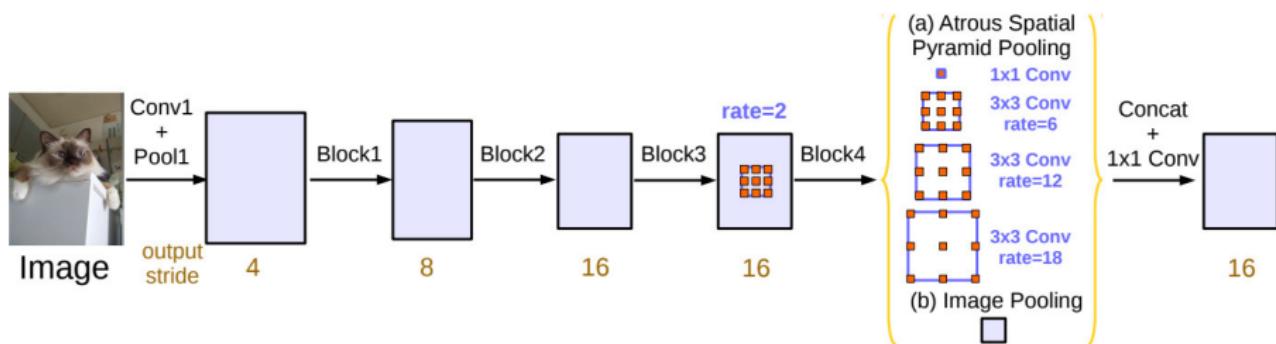


# GUSTA PREDIKCIJA: IZBJEGAVANJE PODUZORKOVANJA

Ideja: izbjegći poduzorkovanje na način da se korak konvolucije u nekim blokovima smanji na 1.

Nedostaci:

- Memorijski jako, jako skupo.
- Dilatirane konvolucije bitno sporije od regularnih.

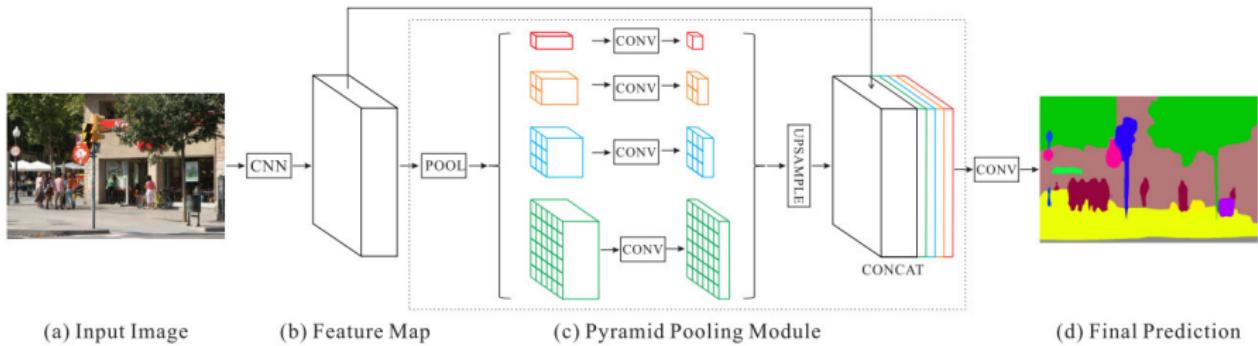


deeplab[chen17arxiv]

# GUSTA PREDIKCIJA: PIRAMIDALNO SAŽIMANJE

Ideja: ugrađivanje šireg konteksta u lokalne značajke.

- povećava receptivno polje i pospješuje raspoznavanje objekata različitih veličina
- vrlo popularan modul u modernim arhitekturama za gustu predikciju
- u literaturi se na slične module ponekad referira i s: *context* ili *SPP* (*spatial pyramid pooling*)

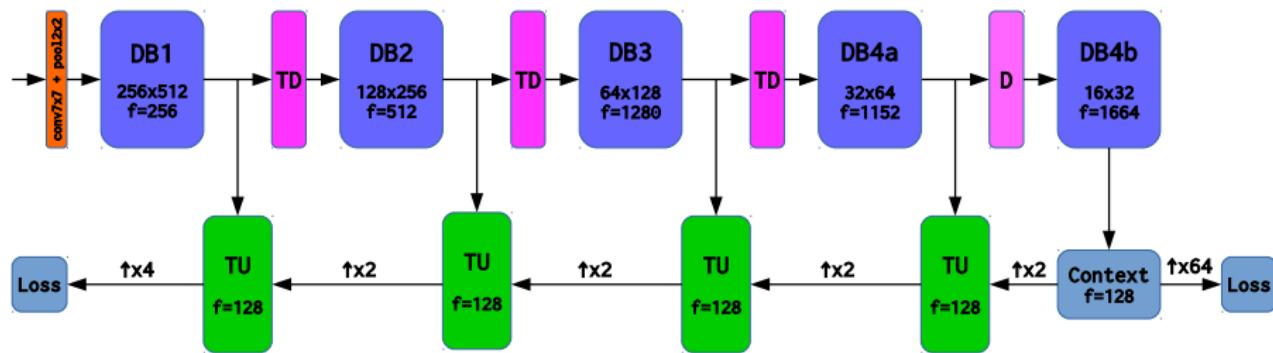


# GUSTA PREDIKCIJA: LJESTVIČASTO NADUZORKOVANJE

Ideja: nadoknaditi poduzorkovanje **miješanjem** slojeva na različitim dubinama [valpola14arxiv,ronneberger15arxiv,lin17cvpr]:

Prepoznavanje razreda je teže od finog podešavanja granica:

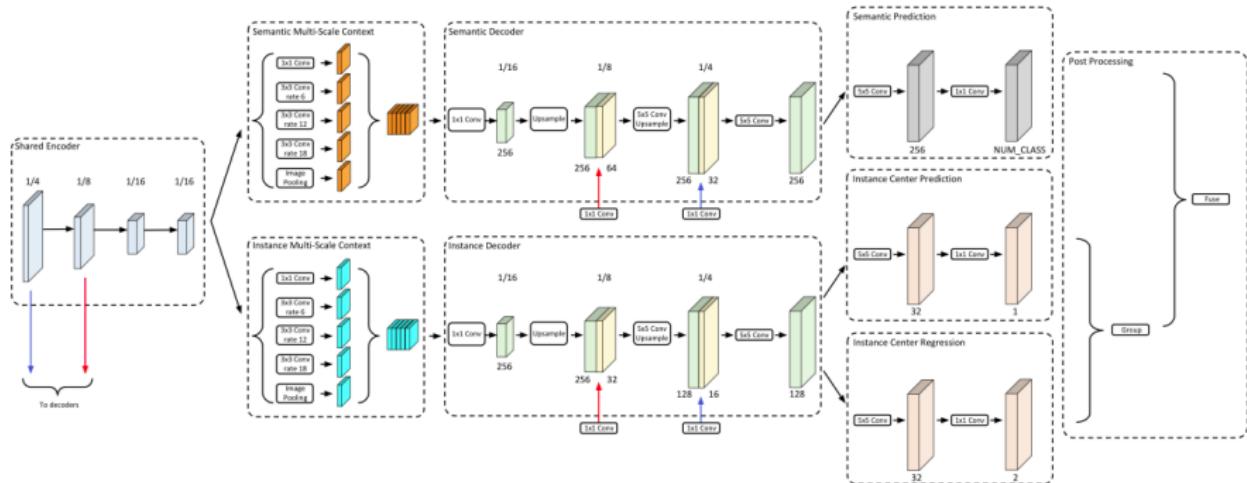
- na visokoj rezoluciji može se koristiti **malena** reprezentacija



[kreso17iccvw]

# GUSTA PREDIKCIJA: PANOPTIČKI DEEPLAB

Ideja: gusto predviđati i) semantički razred, ii) središta objekata te iii) vektore do središta. Instance se formiraju tijekom postprocesiranja



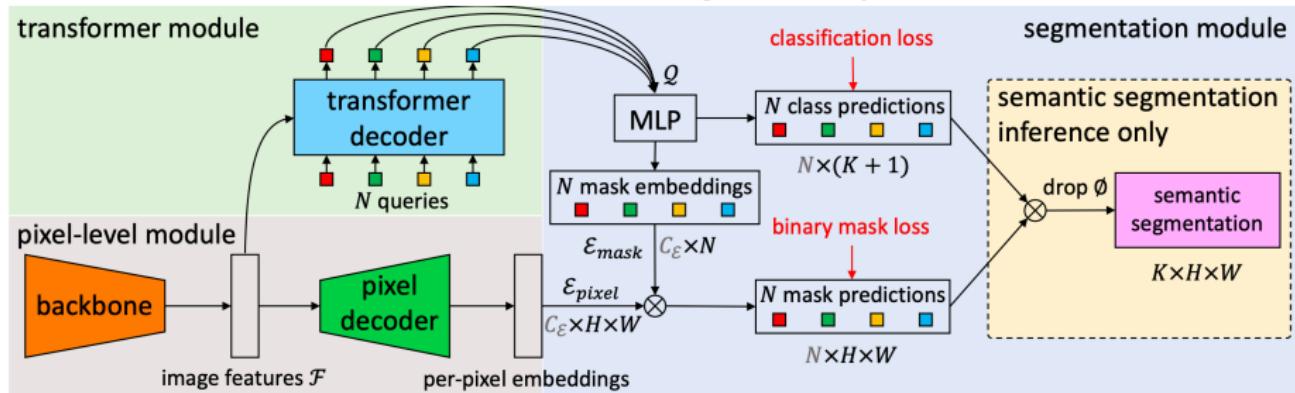
[cheng20cvpr]

COCO test-dev: 41.4% PQ COCO

# GUSTA PREDIKCIJA: MaskFormer

Ideja: raspoznavanje pomaknuti s razine piksela na razinu maski.

- Zasebni moduli računaju ugrađivanje za piksele i svaku masku.
- Pikseli se dodjeljuju maskama na temelju sličnosti ugrađivanja.
- Semantički razred se dodjeljuje maski, a posredno i svim njenim pikselima.
- Univerzalna arhitektura za sve segmentacijske zadatke.



## ZAHVALA

Ova predavanja proizišla su iz istraživanja koje je financirala Hrvatska zaklada za znanost projektom IP-2020-02-5851 Adept.