

Lokalizacija objekata postupkom SSD

Ivan Grubišić

Voditelj: Siniša Šegvić

Fakultet elektrotehnike i računarstva

8.7.2015.

Sadržaj

- ① Uvod
- ② Konvolucijske mreže
- ③ Pronalaženje objekata
- ④ SSD: Single Shot MultiBox Detector
- ⑤ Rezultati

Sadržaj

- ① Uvod**
- ② Konvolucijske mreže**
- ③ Pronalaženje objekata**
- ④ SSD: Single Shot MultiBox Detector**
- ⑤ Rezultati**

Uvod

- Razumijevanje sadržaja slike je važna klasa problema u računalnom vidu.
- Česti zadaci:
 - klasifikacija
 - pronalaženje (engl. *detection, localization*) objekata
 - semantička segmentacija
- U posljednje vrijeme postiže se velik napredak koji se temelji na primjeni dubokih konvolucijskih mreža.

Sadržaj

- ① Uvod
- ② Konvolucijske mreže
- ③ Pronalaženje objekata
- ④ SSD: Single Shot MultiBox Detector
- ⑤ Rezultati

Konvolucijske mreže

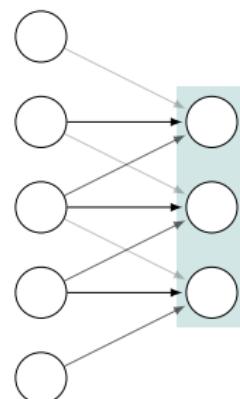
- Konvolucijske neuronske mreže se sastoje od više slojeva sa svojstvima koja su bolje prilagođena problemima u računalnom vidu.
- Slojevi su rijetko i lokalno povezani i dijele se parametri (slika 1).

Konvolucijski sloj

Konvolucijski sloj čini skup naučenih filtera od kojih svaki stvara jednu izlaznu matricu. Vrijedi

$$H_{lp} = \varphi \left(-\theta_{lp} + \sum_q w_{lpq} * H_{(l-1)q} \right),$$

gdje su H mape značajki, w konvolucijske jezgre, θ pragovi, l oznaka sloja, a q i p označke ulazne i izlazne mape značajki.



Slika 1: Jednostavni jednodimenzionalni konvolucijski sloj.

Učenje neuronske mreže

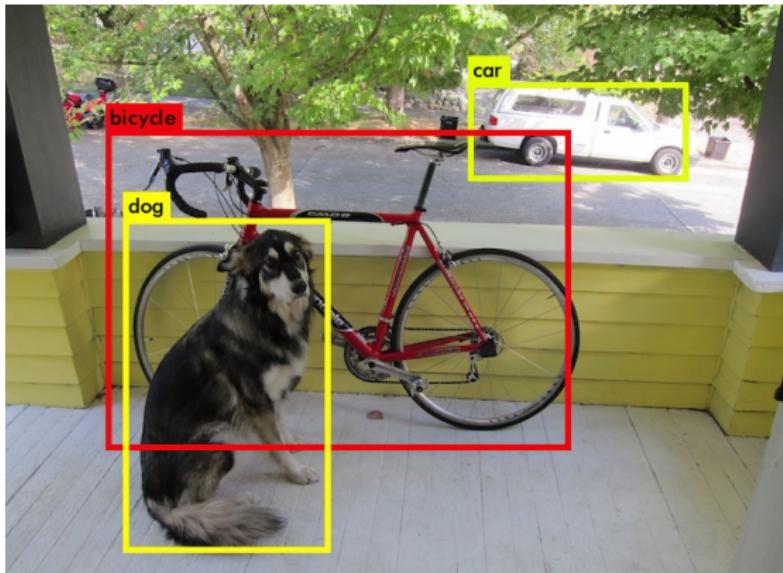
- Neuronska mreža se uči minimizacijom funkcije pogreške (gubitka).
- Parametri mreže se efikasno prilagođavaju gradijentnim spustom ili nekim algoritam koji se na njemu temelji, a gradijent se računa algoritmом propagacije pogreške unatrag.

Sadržaj

- ① Uvod
- ② Konvolucijske mreže
- ③ Pronalaženje objekata
- ④ SSD: Single Shot MultiBox Detector
- ⑤ Rezultati

Pronalaženje objekata

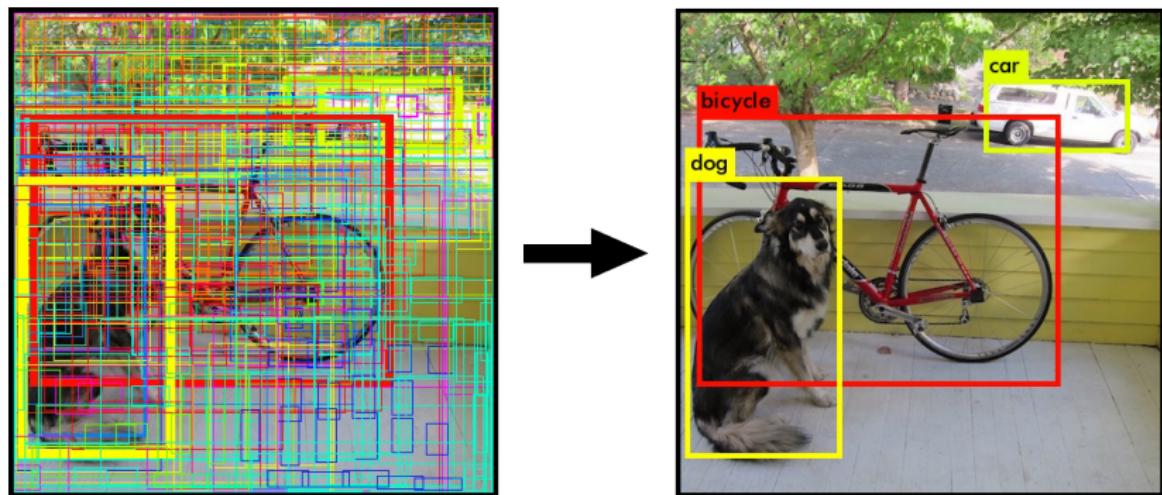
- Zadatak je prepoznati objekte i pridružiti im oznake razreda i najmanje opisane okvire.
- Sustav obično pronađe velik broj okvira od kojih treba izdvojiti najbolje.



Slika 2: Primjer slike s oznakama.

Pronalaženje objekata

- Zadatak je prepoznati objekte i pridružiti im oznake razreda i najmanje opisane okvire.
- Sustav obično pronađe velik broj okvira od kojih treba izdvojiti najbolje.



Slika 3: Odbacivanje okvira kandidata i dobivanje konačnih okvira.

Osnovni pristupi

- Najjednostavniji pristup pronalaženju objekata je korištenje pomičnih prozora različitih dimenzija i klasificiranje svake na taj način dobivene podslike.
- Do otprilike 2012. godine najuspješniji su bili modeli koji klasificiraju svaki dio slike dobiven pomičnim prozorom i objekte modeliraju dijelovima koji mogu biti različito raspoređeni (engl. deformable part model, DPM) [3] i histogramima gradijenata (engl. histogram of gradients, HoG) kao značajkama.
- Malo kasnije su se pokazali uspješnima postupci koji umjesto klasificiranja pomičnog prozora koriste neki način predlaganja okvira koji služi smanjivanju broja okvira koje treba potpuno klasificirati.

Mjere

- Mjera koja se koristi pri utvrđivanju koji okviri rezultati odgovaraju kojim cilnjim okvirima je omjer površine presjeka i unije (*IoU*, engl. *intersection over union*).

Jaccardov koeficijent sličnosti

Jaccardov koeficijent sličnosti definiran je kao omjer površine preklapanja i ukupne površine.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

Preciznost i odziv i mjera F_1

Preciznost P (engl. precision) je udio relevantnih rezultata u svim rezultatima, a odziv R (engl. recall) udio relevantnih rezultata u svim relevantnim objektima.

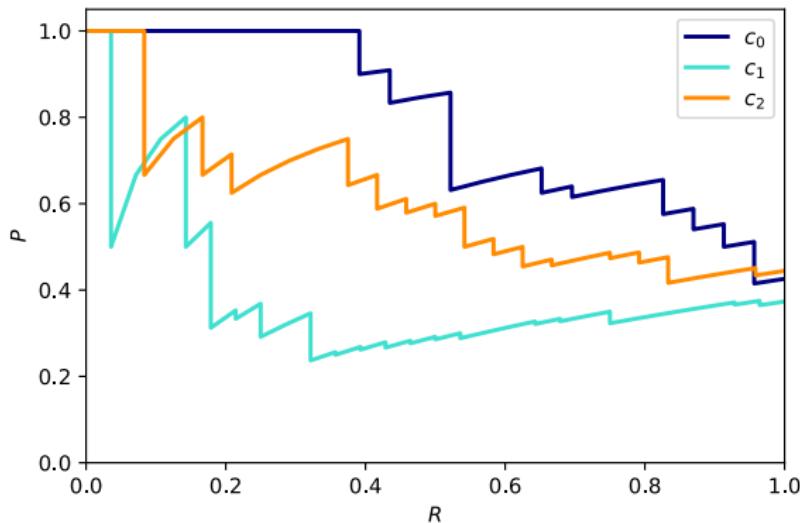
Mjera F_1 je harmonijska sredina preciznosti i odziva.

$$P = \frac{tp}{tp + fp}, \quad R = \frac{tp}{tp + fn}, \quad F_1 = (P^{-1} + R^{-1})^{-1}.$$

- Okvir rezultat se obično smatra ispravnim (tp) ako je njegov koeficijent sličnosti s obzirom na ciljni okvir odgovarajućeg razreda veći od 0.5 i najveći među koeficijentima sličnosti svih okvira rezultata za koje isto vrijedi.

Mjere

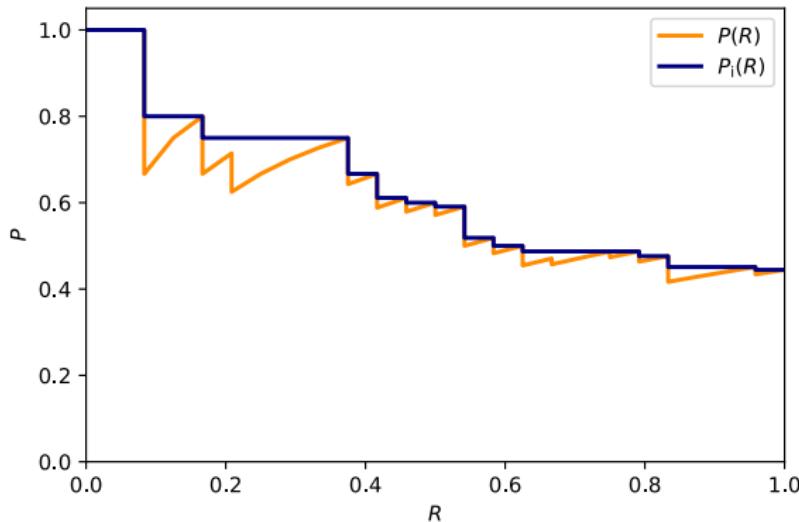
- Povećanjem skupa razmatranih rezultata redom od okvira za čiju je klasifikaciju sustav najsigurniji, preciznost se smanjuje, a povećava se odziv.



Slika 4: Krivulje preciznosti u ovisnosti o odzivu za 3 razreda.

Mjere

- Povećanjem skupa razmatranih rezultata redom od okvira za čiju je klasifikaciju sustav najsigurniji, preciznost se smanjuje, a povećava se odziv.



Slika 4: Krivulja interpolirane preciznosti u ovisnosti o odzivu.

Mjere

Srednja preciznost

Kao evaluacijska mjera za pronalaženja objekata s jednim razredom se često koristi prosječna (interpolirana) preciznost (AP) koja je jednaka površini ispod krivulje preciznosti i odziva:

$$AP = \int_0^1 P(r) dr.$$

Srednja prosječna preciznost

U slučaju više razreda koristi se srednja prosječna (interpolirana) preciznost (mAP):

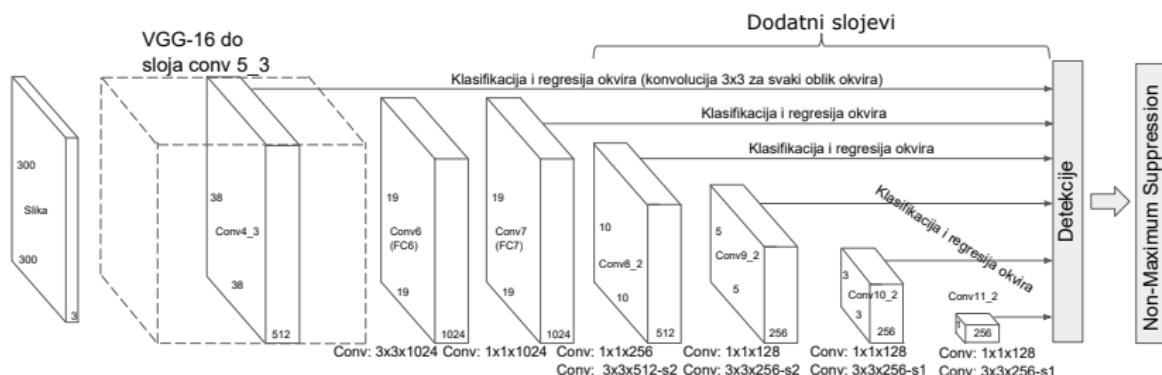
$$mAP = \frac{1}{|C|} \sum_{c \in C} AP_c.$$

Sadržaj

- ① Uvod
- ② Konvolucijske mreže
- ③ Pronalaženje objekata
- ④ SSD: Single Shot MultiBox Detector
- ⑤ Rezultati

SSD: Single Shot MultiBox Detector

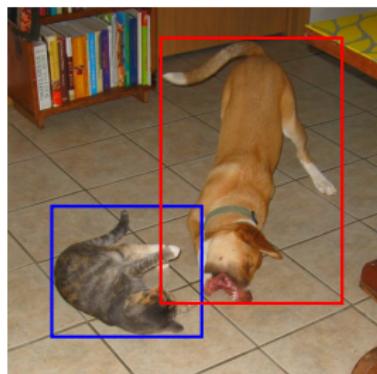
- SSD je model za pronalaženje objekata koji se temelji na dubokoj konvolucijskoj mreži koja u jednoj propagaciji unaprijed generira sve predikcije okvira.



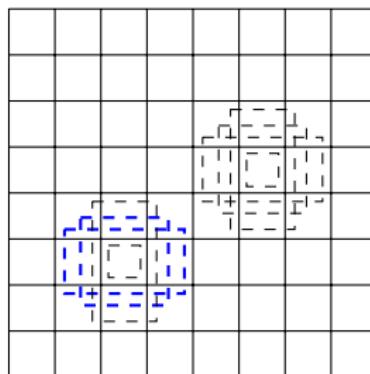
Slika 5: Ilustracija sustava s prikazanim slojevima značajki koje koriste slojevi za klasifikaciju i regresiju okvira.

SSD: Single Shot MultiBox Detector

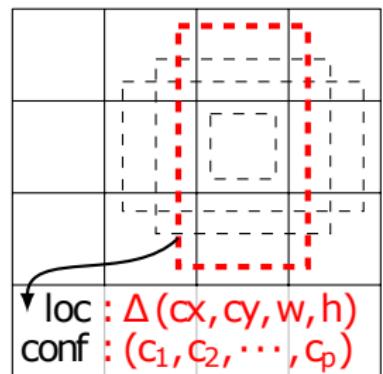
- Na nekim slojevima značajki konvolucijski se vrši klasifikacija i regresija (prilagodba) dimenzija i položaja unaprijed određenih okvira različitih oblika koji su dodijeljeni svakom položaju u sloju značajki.



(a) Slika sa željenim okvirima



(b) Mapa značajki dimenzija 8×8



(c) Mapa značajki dimenzija 4×4

Slika 6: Svakom pikselu nekih slojeva značajki pridruženo je nekoliko unaprijed određenih okvira različitih omjera stranica. Crtkani okviri najbolje odgovaraju cilnjim okvirima i zato su najpogodniji za učenje parametara.

- Za svaku sliku s ciljnim okvirima se iz skupa za učenje prvo među unaprijed određenim okvirima odaberu oni koji su najprikladniji za učenje pozitivnih i negativnih primjera rezultata, a onda se parametri korišteni za dobivanje odabralih okvira prilagođavaju algoritmom koji se temelji na gradijentnom spustu.
- S će označavati skup unaprijed određenih okvira, T skup ciljnih okvira, a C skup razreda koji uključuje i razred *ostalo*.

Odabir okvira za učenje pozitivnih primjera

Za učenje pozitivnih primjera svakom cilnjom okviru se pridružuje podskup unaprijed određenih okvira $S_+ \subseteq S$ u kojemu za svaki element vrijedi:

- koeficijent sličnosti J s nekim cilnjim okvirom je veći od 0.5 ili
- među svim unaprijed određenim okvirima najbolje se preklapa s nekim cilnjim okvirom.

Funkcija $a: T \rightarrow 2^S$ predstavlja to pridruživanje.

Odabir negativnih primjera

Umjesto korištenja svih negativnih primjera, za učenje se odabire samo njihov podskup koji daje najveći gubitak klasifikacije i nema više elemenata od $3|S_+|$. Na taj način se ostvaruje brža i stabilnija konvergencija.

- Neki okvir $u = (\mathbf{x}_u, \mathbf{d}_u, \mathbf{c}_u)$ određen je vektorima položaja \mathbf{x}_u , dimenzijama \mathbf{d}_u i pouzdanostima klasifikacije \mathbf{c}_u dimenzijske klasifikacije $|C|$ koje su u slučaju ciljnog okvira *one-hot*-vektor.
- Slojevi za klasifikaciju i regresiju okvira pomake i dimenzije okvira računaju u relativnim koordinatama u odnosu na unaprijed određene okvire.
- Funkcija $r_s(u)$ preslikava položaj i dimenzije okvira u iz koordinata slike u relativne koordinate unaprijed određenog okvira s :

$$r_s(u) = ((\mathbf{x}_u - \mathbf{x}_s) \odot \mathbf{d}_s, \ln(\mathbf{d}_u \odot \mathbf{d}_s), \mathbf{c}_u). \quad (1)$$

Gubitak

Gubitak je L definiran kao težinski zbroj gubitka položaja L_I i gubitka klasifikacije L_c :

$$L = \frac{1}{|S_+|} (L_I + \alpha L_c),$$

$$L_I = \sum_{t \in T} \sum_{s \in a(t)} \left(\| \hat{\mathbf{x}}_{\hat{h}_s} - \mathbf{x}_{r_s(t)} \|_{\tilde{1}} + \| \hat{\mathbf{d}}_{\hat{h}_s} - \mathbf{d}_{r_s(t)} \|_{\tilde{1}} \right),$$

$$L_c = - \sum_{t \in T} \sum_{s \in a(t)} \ln c_{s(\arg \max \mathbf{c}_t)} - \sum_{s \in S_-} \ln c_{s0}.$$

Sadržaj

- ① Uvod
- ② Konvolucijske mreže
- ③ Pronalaženje objekata
- ④ SSD: Single Shot MultiBox Detector
- ⑤ Rezultati

Rezultati

Model	FPS	Podaci za učenje	mAP/%	
			VOC2007	VOC2012
Fast-RCNN	0.5	VOC07	66.9	-
		VOC07, VOC12	70.0	68.4
Faster-RCNN	7	VOC07	69.9	-
		VOC07, VOC12	73.2	70.4
		VOC07, VOC12, COCO	78.8	75.9
YOLO	45	VOC07, VOC12	63.4	57.9
*YOLOv2-544	40	VOC07, VOC12	78.6	73.4
*PVANET-c	31	VOC07, VOC12, COCO	84.4	83.7
*PVANET	22	VOC07, VOC12, COCO	84.9	84.2
SSD300	46	VOC07	68.0	-
		VOC07, VOC12	74.3	72.1
		VOC07, VOC12, COCO	79.6	77.5
SSD512	19	VOC07	71.6	-
		VOC07, VOC12	76.8	74.9
		VOC07, VOC12, COCO	81.6	80.0

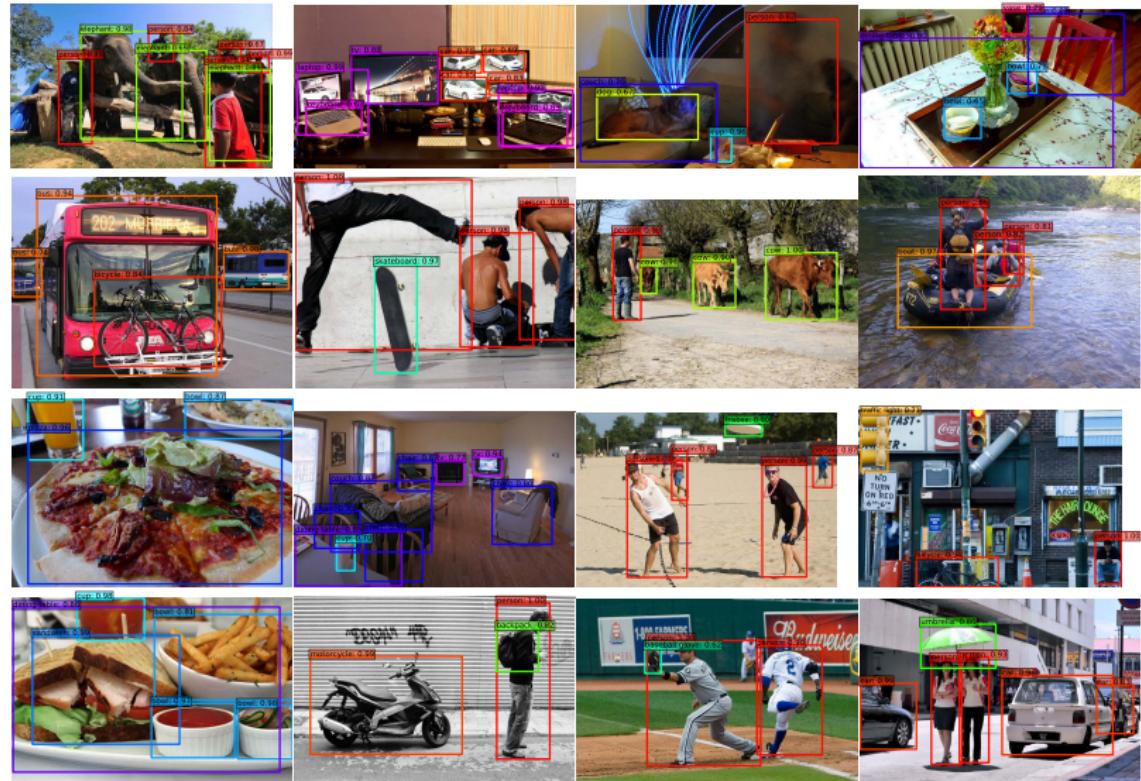
Tablica 1: Usporedba rezultata evaluacije na skupovima *VOC2007-test* i *VOC2012-test*

Rezultati

više slojeva značajki za traženje objekata	•	•	•	•	•	
više proširivanja podataka	•	•	•	•	•	
korištenje omjera $\left\{ \frac{1}{2}, 2 \right\}$	•	•	•	•	•	
korištenje omjera $\left\{ \frac{1}{3}, 3 \right\}$	•	•	•	•	•	
à trous	•	•	•	•	•	
<i>mAP/%</i>	62.4	65.5	71.6	73.7	74.2	74.3

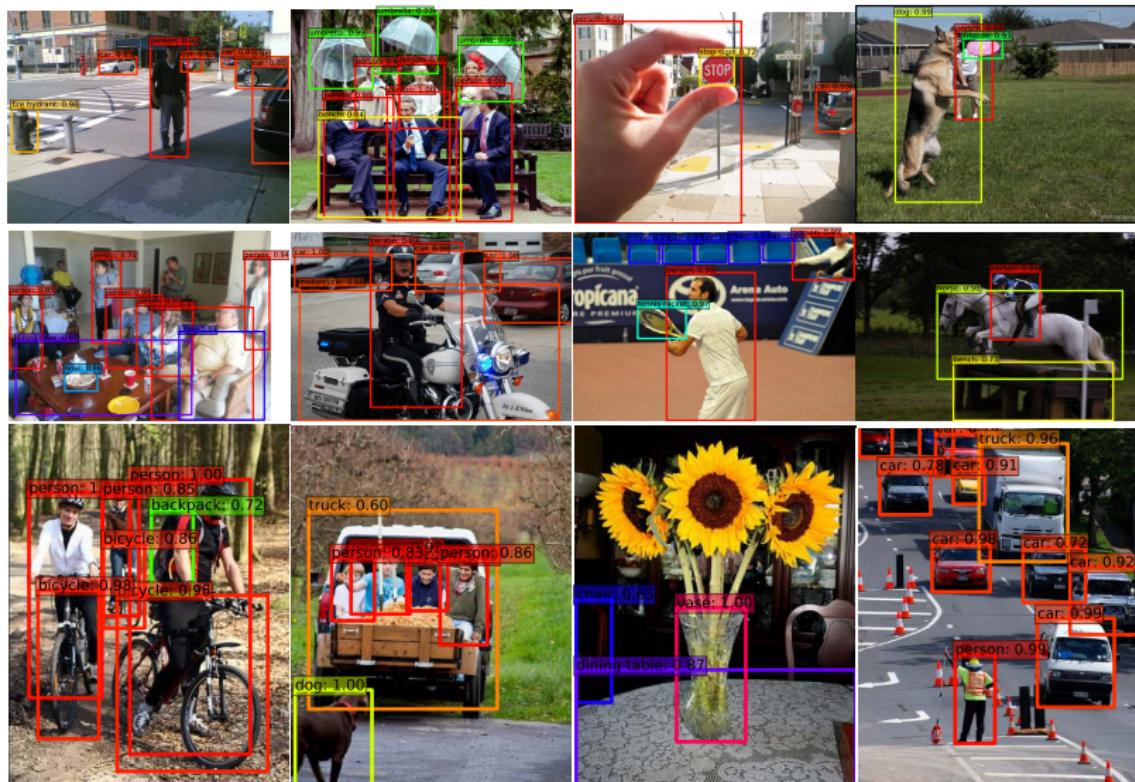
Tablica 2: Utjecaj različitih odabira komponenata na srednju prosječnu pogrešku kod modela SSD300 na skupu *VOC2007-test*. Redak "više slojeva značajki za traženje objekata" označava korištenje slojeva značajki conv4_3, conv7, conv8_2, conv9_2, conv10_2 i conv11_2 umjesto samo conv7 za pronalaženje objekata.

Rezultati



Slika 7: Primjeri rezultata pronalaženja objekata.

Rezultati



Slika 8: Primjeri rezultata pronalaženja objekata.