

رسالة محمد



مبانی بینایی کامپیوتر

مدرس: محمدرضا محمدی

۱۴۰۱

تشخيص اشیاء

Object Detection

تطبيق کلیشه



خواندن پلاک

۴



خواندن پلاک

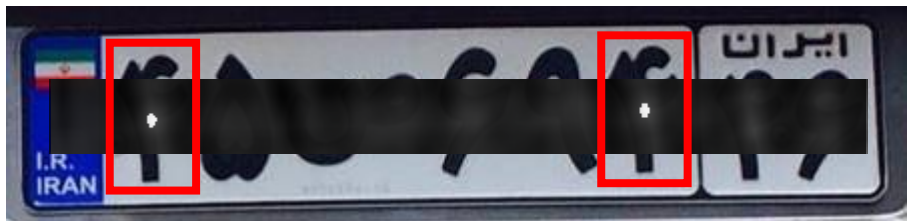
- برای خواندن کاراکترهای پلاک، می‌توان تطبیق کلیشه هر کاراکتر با تصویر را محاسبه کرد

۴



خواندن پلاک

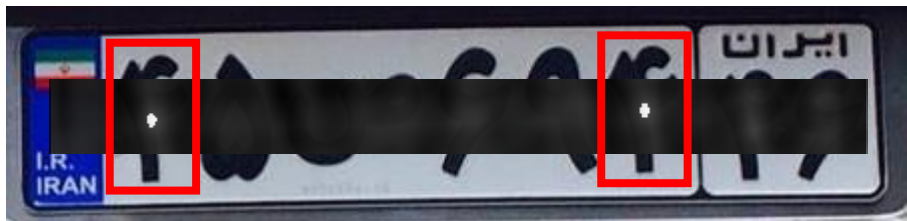
- برای خواندن کاراکترهای پلاک، می توان تطبیق کلیشه هر کاراکتر با تصویر را محاسبه کرد
- سپس، مکان هایی که نتیجه تطبیق کلیشه برای آنها بیش از حدی باشد را به عنوان مکان آن کاراکتر در نظر می گیریم



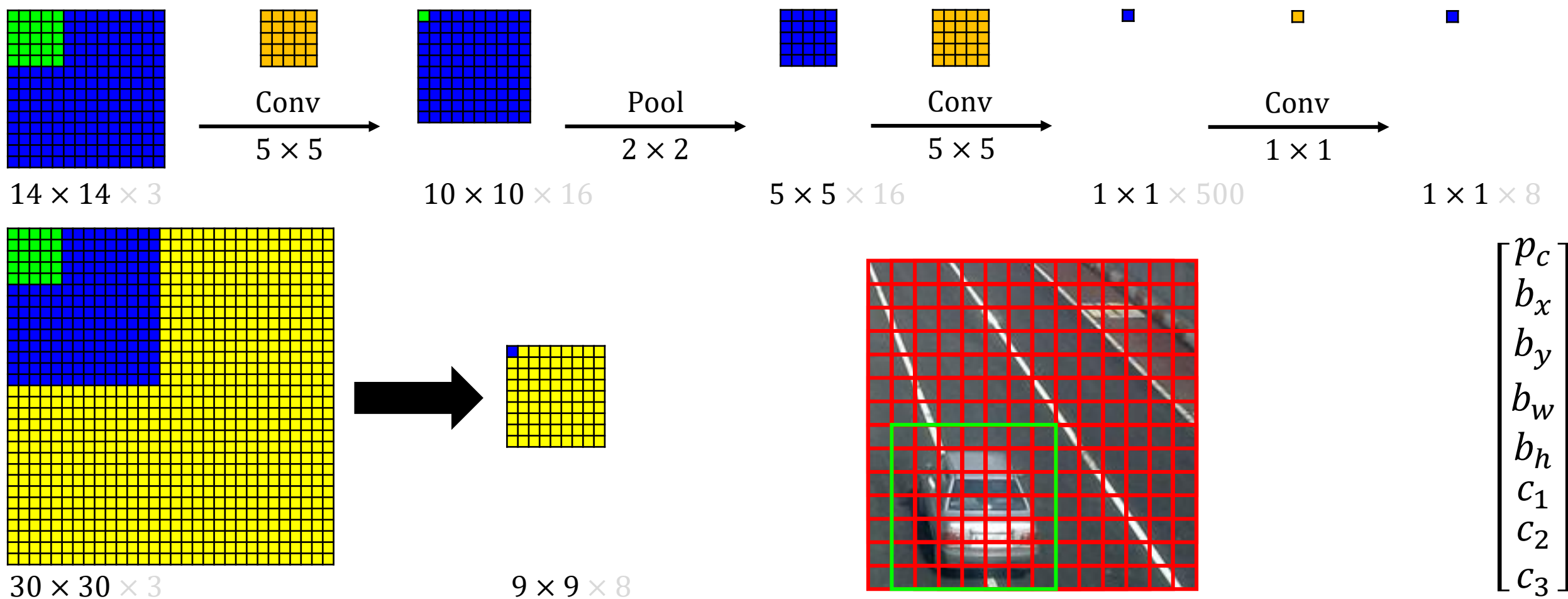
خواندن پلاک

- برای خواندن کاراکترهای پلاک، می توان تطبیق کلیشه هر کاراکتر با تصویر را محاسبه کرد
- سپس، مکان هایی که نتیجه تطبیق کلیشه برای آنها بیش از حدی باشد را به عنوان مکان آن کاراکتر در نظر می گیریم
- برای تشخیص کاراکترهای با ابعاد مختلف می توان از کلیشه های با ابعاد مختلف استفاده نمود

۴

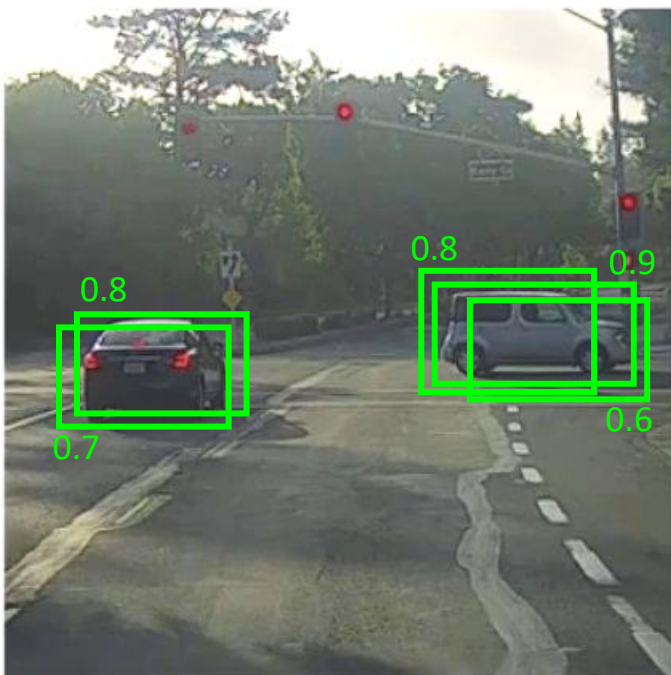


پیاده‌سازی کانولوشنی پنجره لغزان



حذف مقادیر غیربیشینه

- ممکن است یک شیء در چند نقطه تشخیص داده شود
- از میان ناحیه‌های دارای اشتراک زیاد، یکی انتخاب می‌شود

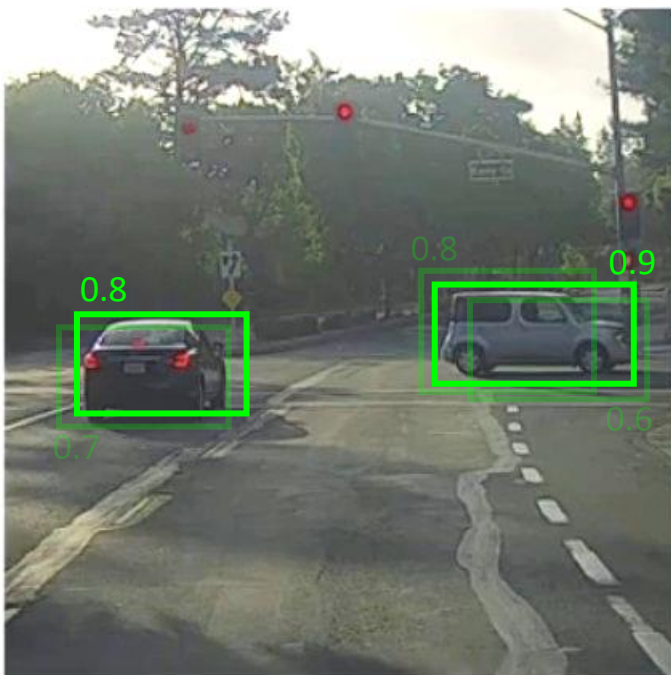


- پاسخ‌هایی که دارای $p_c < 0.5$ هستند حذف می‌شوند
- تکرار

- بیشترین p_c انتخاب می‌شود
- پاسخ‌هایی که دارای IOU بیش از ۰.۵ با این پاسخ هستند حذف می‌شوند

حذف مقادیر غیربیشینه

- ممکن است یک شیء در چند نقطه تشخیص داده شود
- از میان ناحیه‌های دارای اشتراک زیاد، یکی انتخاب می‌شود

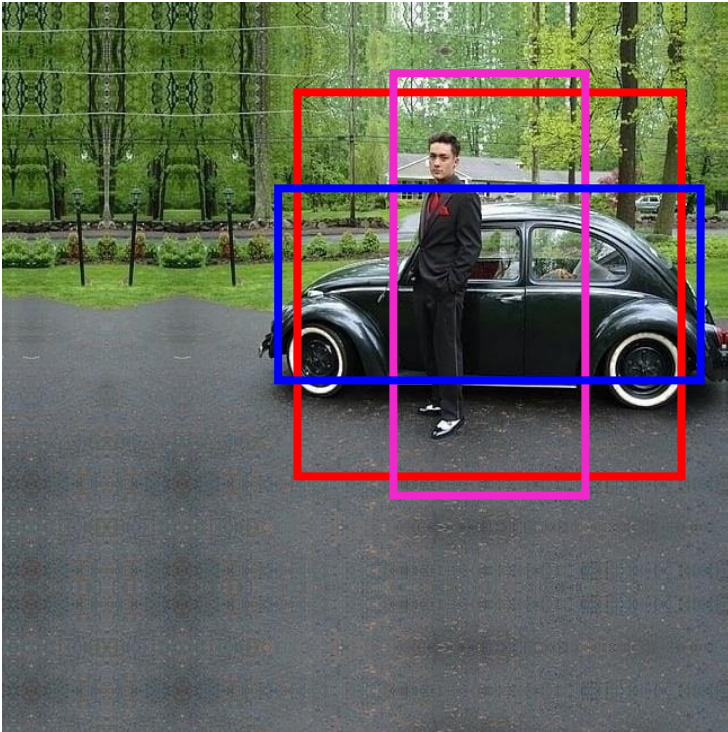


- پاسخ‌هایی که دارای $p_c < 0.5$ هستند حذف می‌شوند
- تکرار

- بیشترین p_c انتخاب می‌شود
- پاسخ‌هایی که دارای IOU بیش از ۰.۵ با این پاسخ هستند حذف می‌شوند

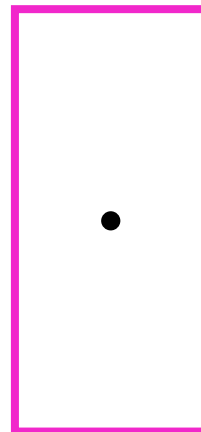
Anchor Boxes

- در هر ناحیه تنها یک شیء قابل تشخیص است
- برای اضافه کردن امکان تشخیص چند شیء، می توان در هر ناحیه چند خروجی قرار داد

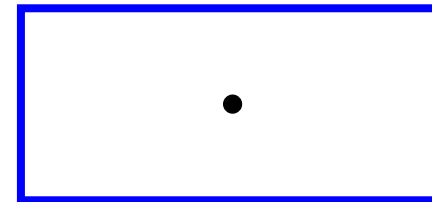


$$\begin{bmatrix} p_c \\ b_x \\ b_y \\ b_w \\ b_h \\ c_1 \\ c_2 \\ c_2 \\ p_c \\ b_x \\ b_y \\ b_w \\ b_h \\ c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

Anchor Box 1

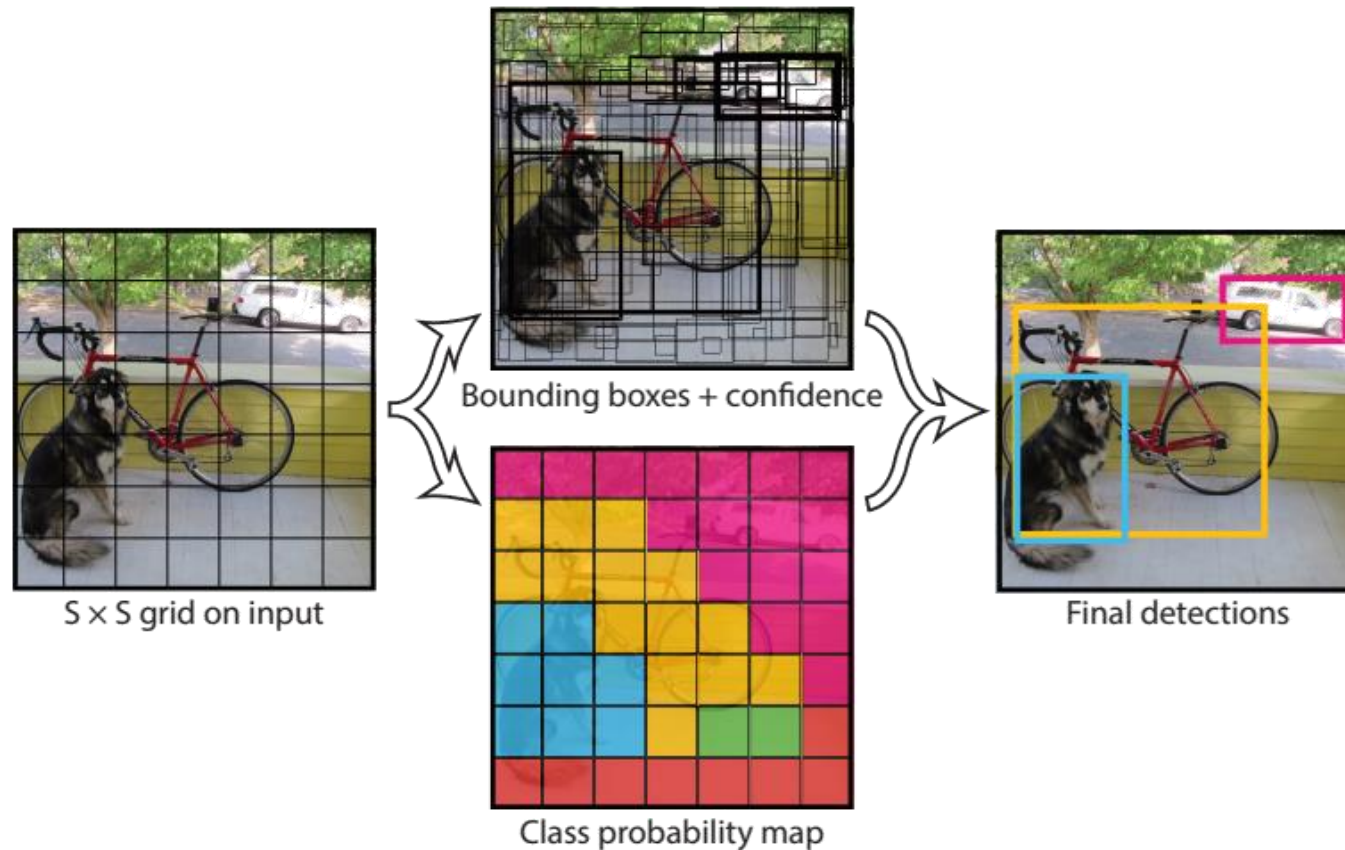


Anchor Box 2



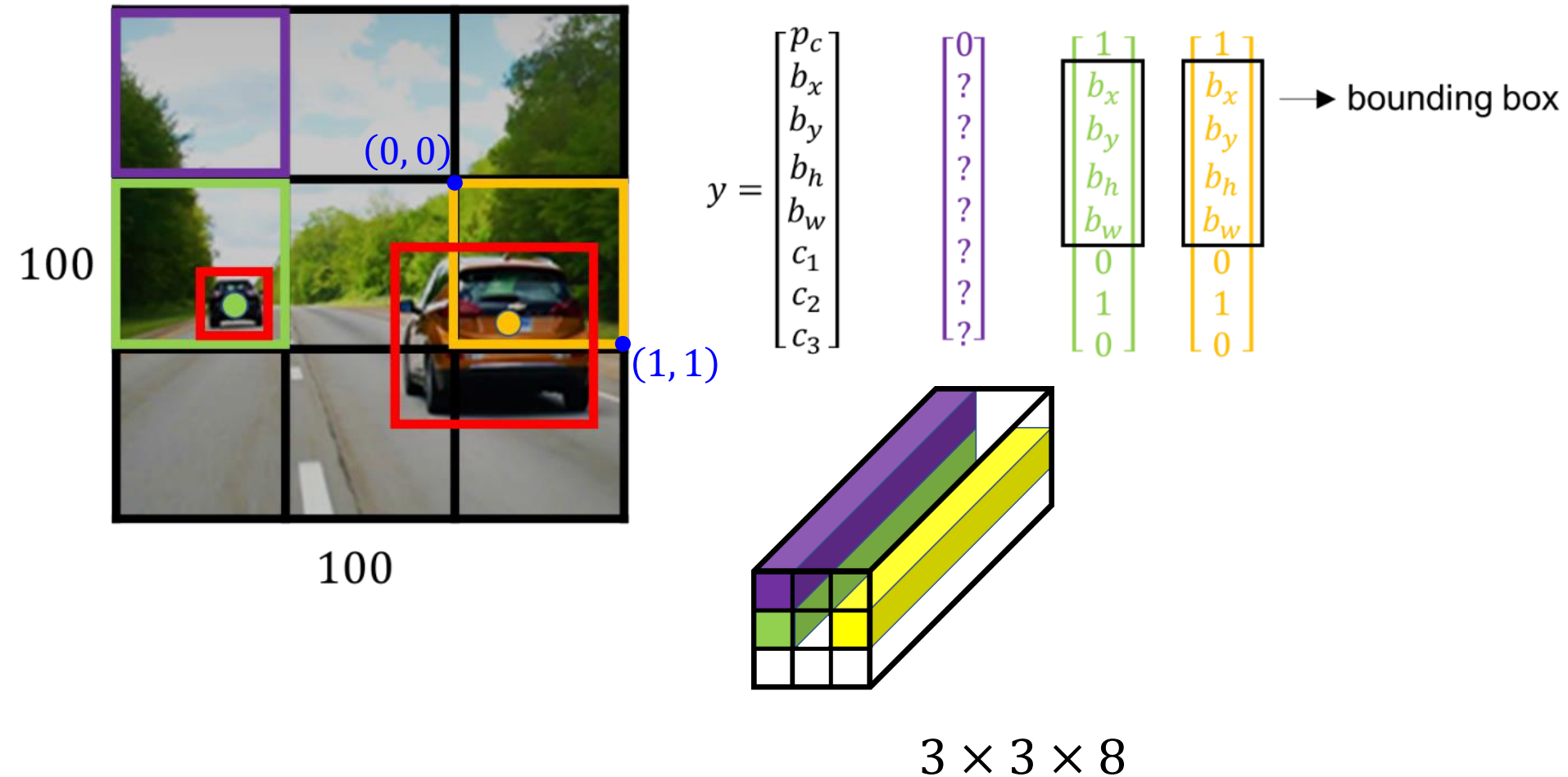
YOLO: You Only Look Once

- در روش YOLO، تصویر ورودی به تعدادی ناحیه کوچک تقسیم می‌شود و برای هر ناحیه یک دسته‌بند و یک تابع رگرسیون طراحی می‌شود

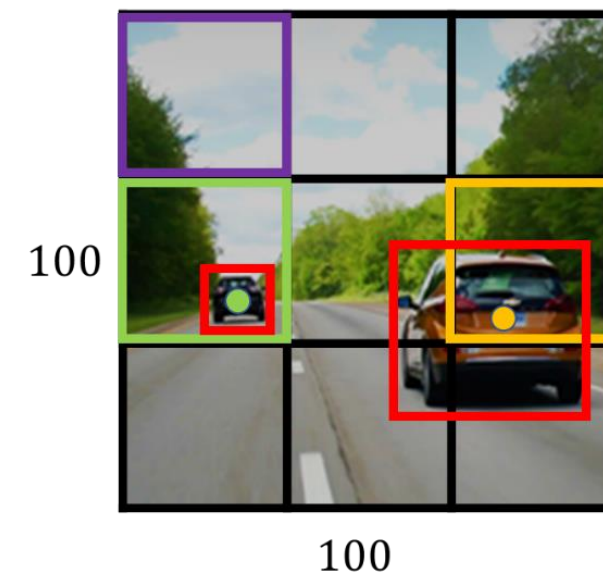
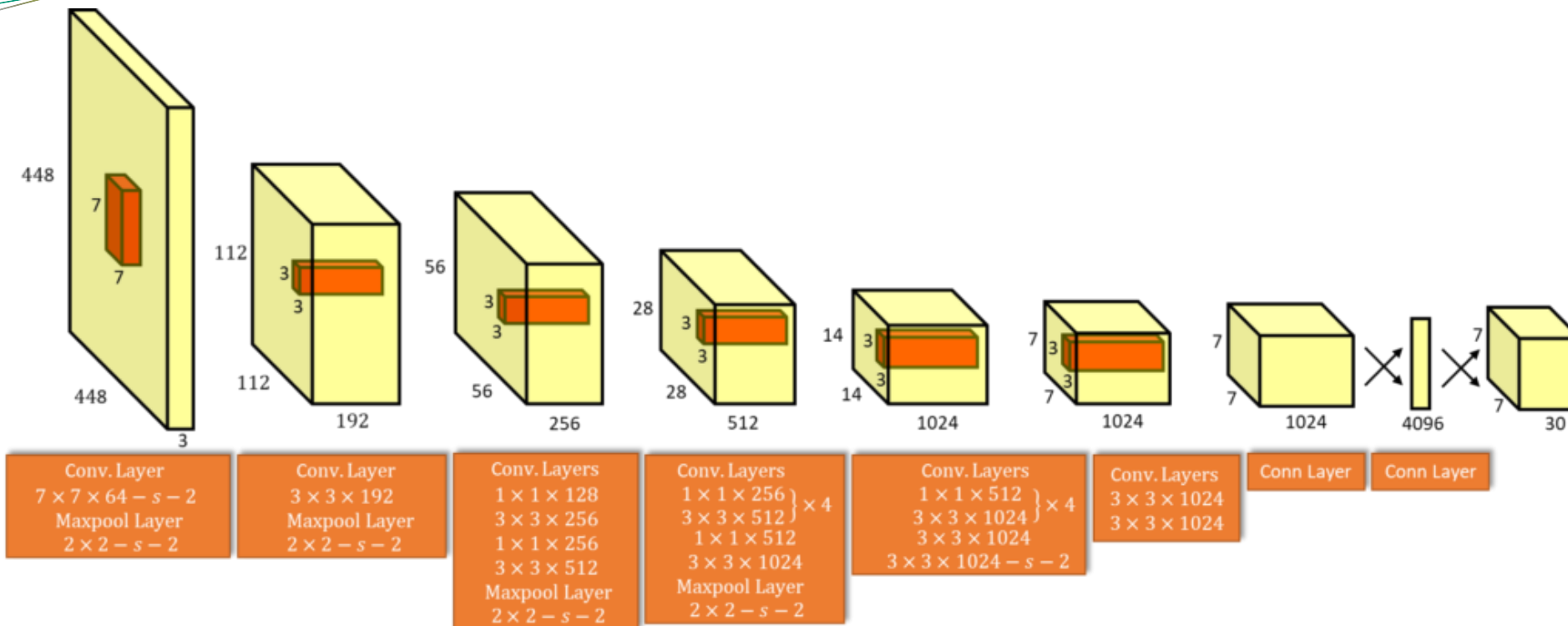


YOLO

Labels for training for
each grid cell:



YOLO



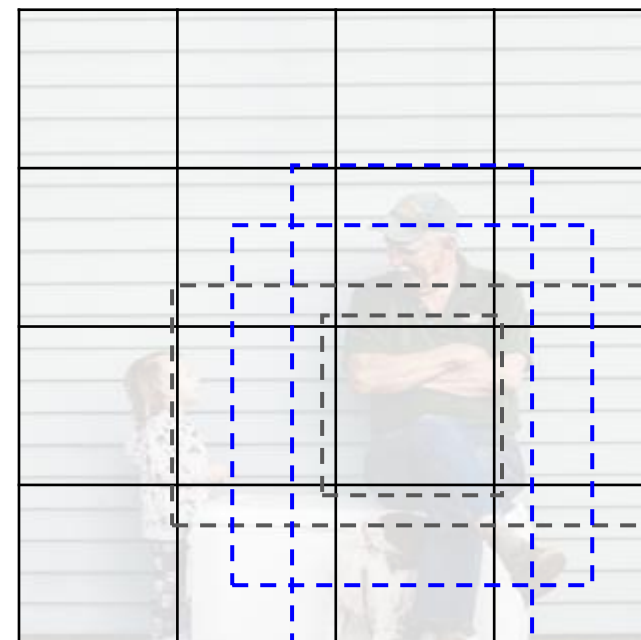
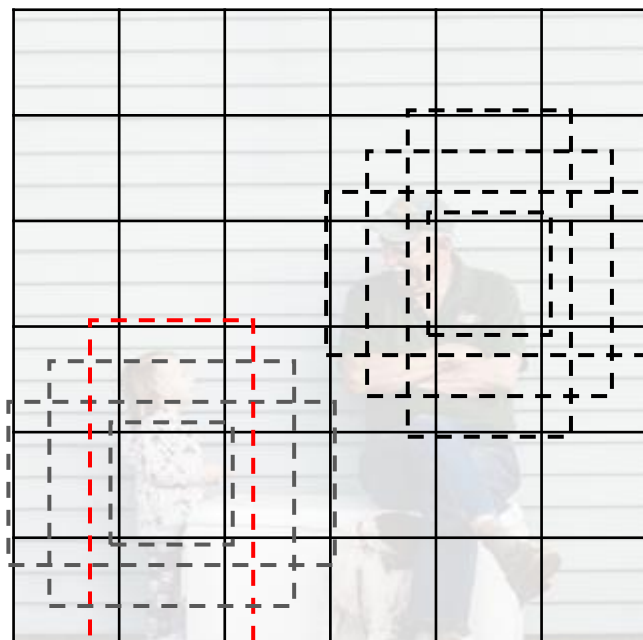
YOLO

Real-Time Detectors	Train	mAP	FPS
100Hz DPM [31]	2007	16.0	100
30Hz DPM [31]	2007	26.1	30
Fast YOLO	2007+2012	52.7	155
YOLO	2007+2012	63.4	45
Less Than Real-Time			
Fastest DPM [38]	2007	30.4	15
R-CNN Minus R [20]	2007	53.5	6
Fast R-CNN [14]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[28]	2007+2012	73.2	7
Faster R-CNN ZF [28]	2007+2012	62.1	18
YOLO VGG-16	2007+2012	66.4	21

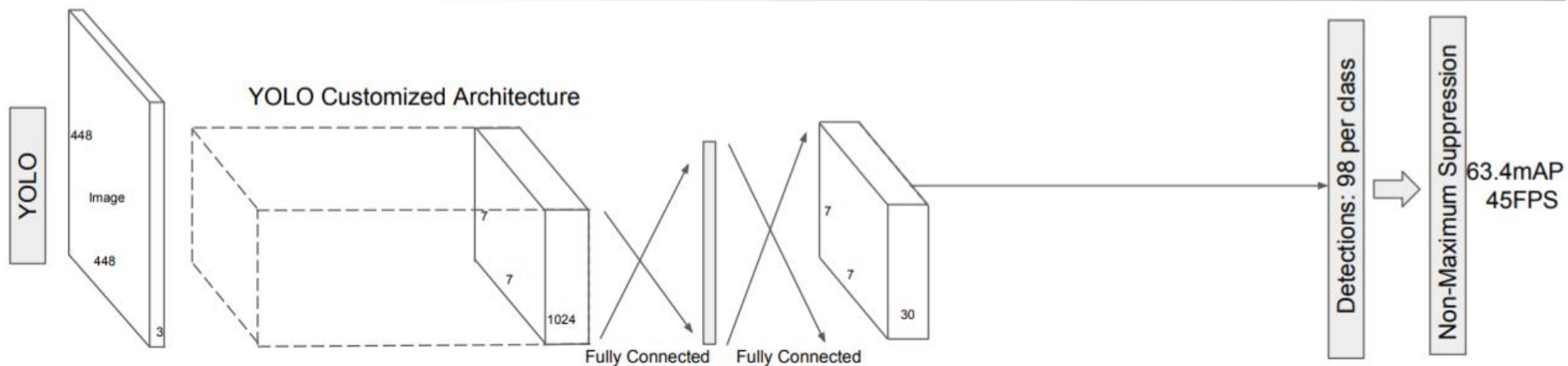
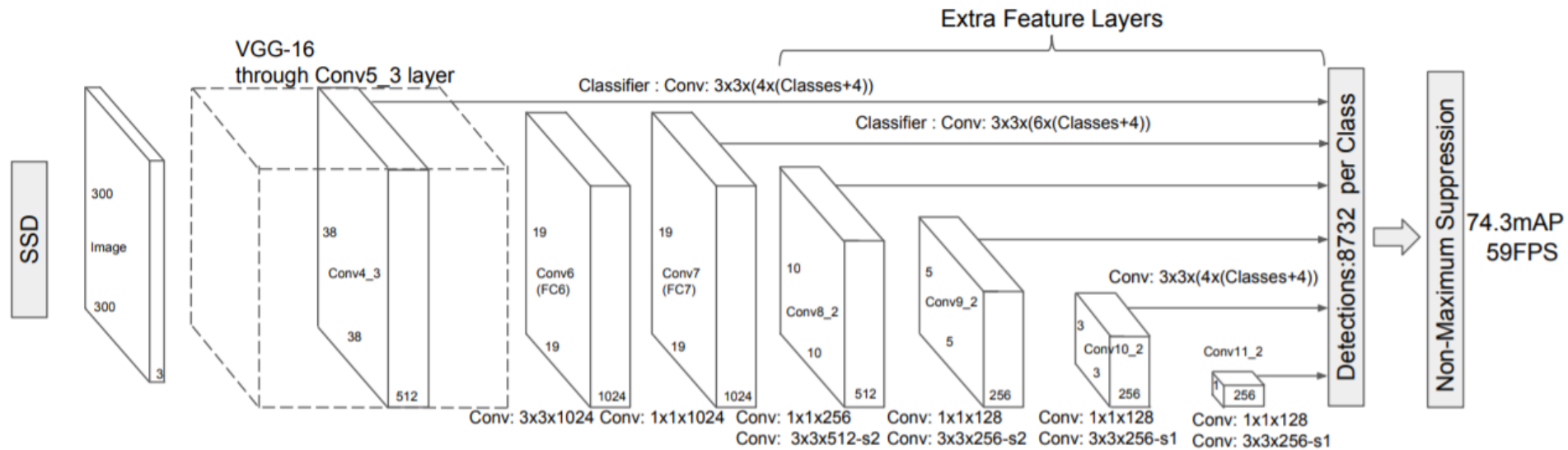
SSD: Single Shot MultiBox Detector

- روش YOLO یک روش آموزش end-to-end شبکه است که نیازی به تولید ناحیه پیشنهادی و تغییر ابعاد آنها ندارد
- سرعت روش YOLO از روش‌هایی که از بخش تولید ناحیه‌های پیشنهادی استفاده می‌کنند بهتر است اما دقت پایین‌تری دارد
- در مقاله SSD بهبودهایی داده شده است که در ضمن افزایش سرعت، دقت نیز افزایش یافته است (به خصوص برای اشیاء کوچک)
- مهمترین نوآوری SSD آن است که برای تشخیص اشیاء و محل آنها از چند لایه استفاده کرده است تا اشیاء با ابعاد مختلف قابل تشخیص باشند

SSD



SSD vs YOLO



RetinaNet

- یکی از مشکلات آموزش در شبکه‌های طراحی شده برای تشخیص اشیاء عدم توازن شدید میان نمونه‌های کلاس‌ها است
- به طور خاص، تعداد ناحیه‌هایی که هیچکدام از اشیاء مورد نظر در آن قرار ندارند به مراتب بیش از ناحیه‌های مربوط به کلاس‌های دیگر است
- ایده اصلی در مقاله RetinaNet پیشنهاد تابع هزینه‌ای است که بهینه‌سازی آن با استفاده از داده‌های نامتوازن منجر به عملکرد مناسب‌تری شود

Focal Loss

- تابع هزینه متداول در شبکه‌های عصبی cross entropy است

$$CE(p, y) = - \sum y_i \log(p_i)$$

$$CE(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1 - p) & \text{otherwise} \end{cases}$$

- حالت دو کلاسه:

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases}$$

$$CE(p, y) = CE(p_t) = -\log(p_t)$$

Focal Loss

- تابع هزینه cross entropy تلاش می کند تا تمام نمونه ها را با احتمال کامل درست بگوید
- به عبارت دیگر، دسته بندی با احتمال بالا کفایت نمی کند و این مقادیر هزینه کوچک زمانی که برای تعداد بسیار زیادی نمونه با یکدیگر جمع می شوند عدد قابل توجهی می شود
- در تابع هزینه focal، مقدار ضرر برای داده هایی که به خوبی شناسایی شده اند کاهش می یابد

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t)$$

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases}$$

$$CE(p, y) = CE(p_t) = -\log(p_t)$$

RetinaNet

