

شبکه‌های عصبی کانولوشن

سید ناصر رضوی www.snrazavi.ir

۱۳۹۶

- شبکه‌های عصبی کانولوشنی
- دسته‌بندی تصاویر
- دسته‌بندی و مکان‌یابی
- شناسایی نقاط عطف
- تشخیص اشیا
- الگوریتم YOLO
- تشخیص چهره و شبکه‌های سیامی
- تابع هزینه سه‌تایی
- انتقال سبک

گونه‌های مسائل در بینای ماشین

۳

□ دسته‌بندی تصاویر.



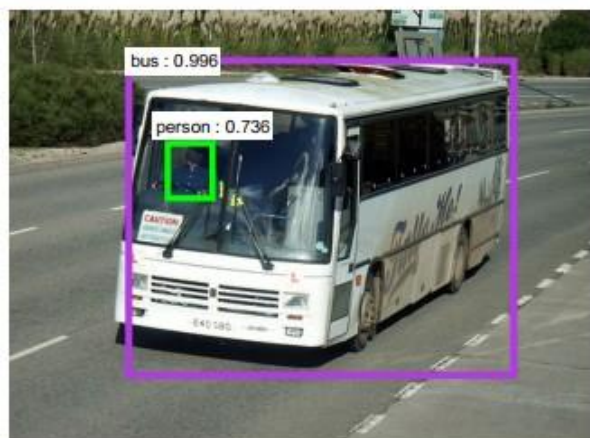
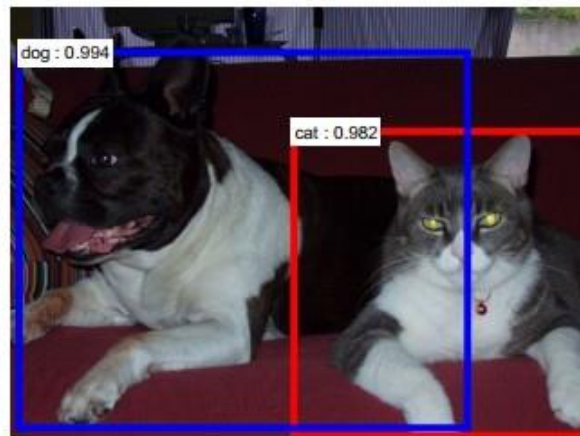
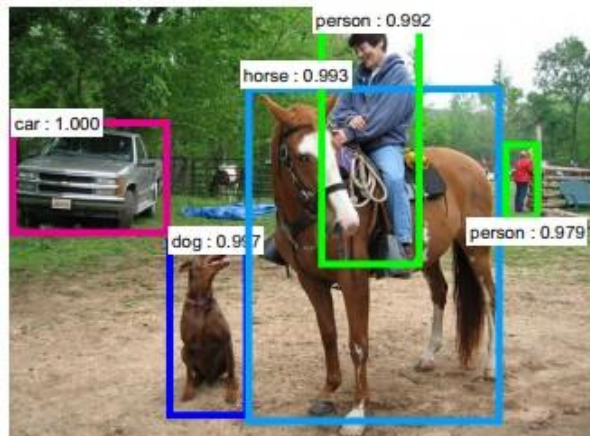
۳۰۰ × ۱۰۰ × ۳

هوایما
فودرو
پرنده
گربه
گوزن
سگ
قورباغه
اسب
کشتی
کامیون

گونه‌های مسائل در بینای ماشین

۴

□ شناسایی اشیاء.



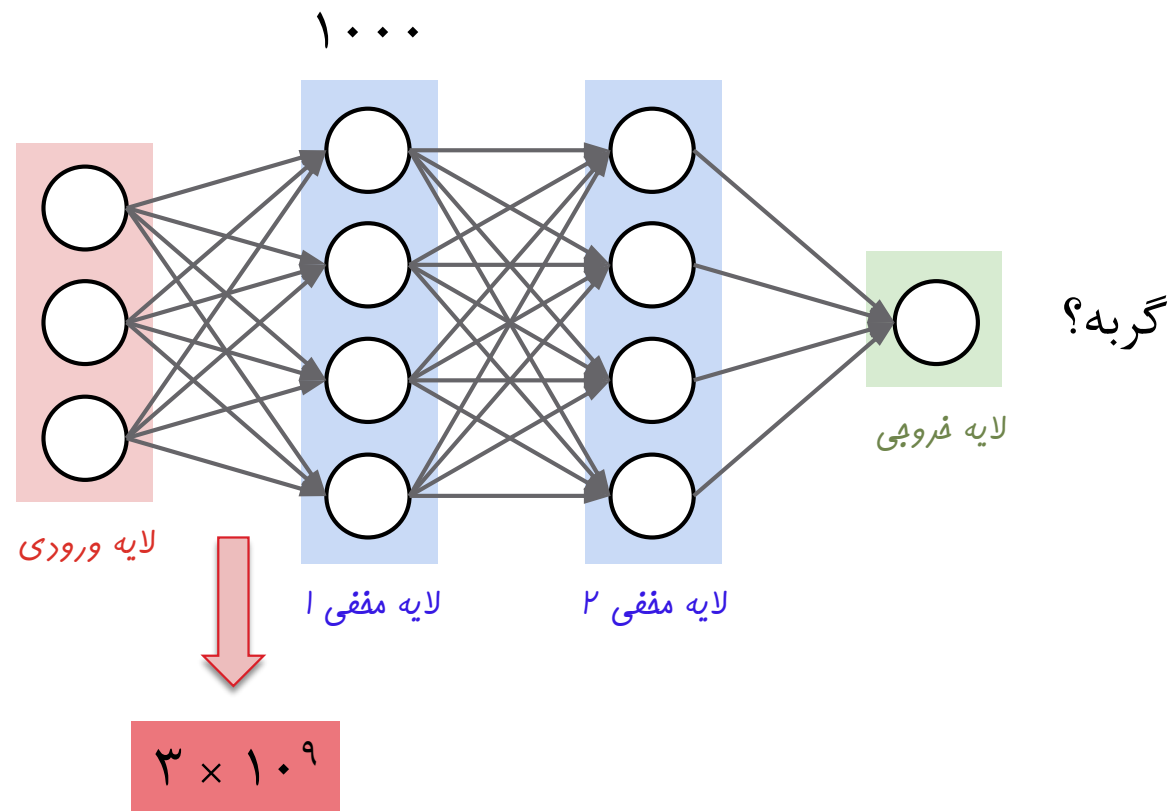
تصاویر بزرگ

۵

□ اندازه شبکه عصبی. تعداد زیاد پارامترها؛ مشکل بیش‌برازش و حجم محاسبات

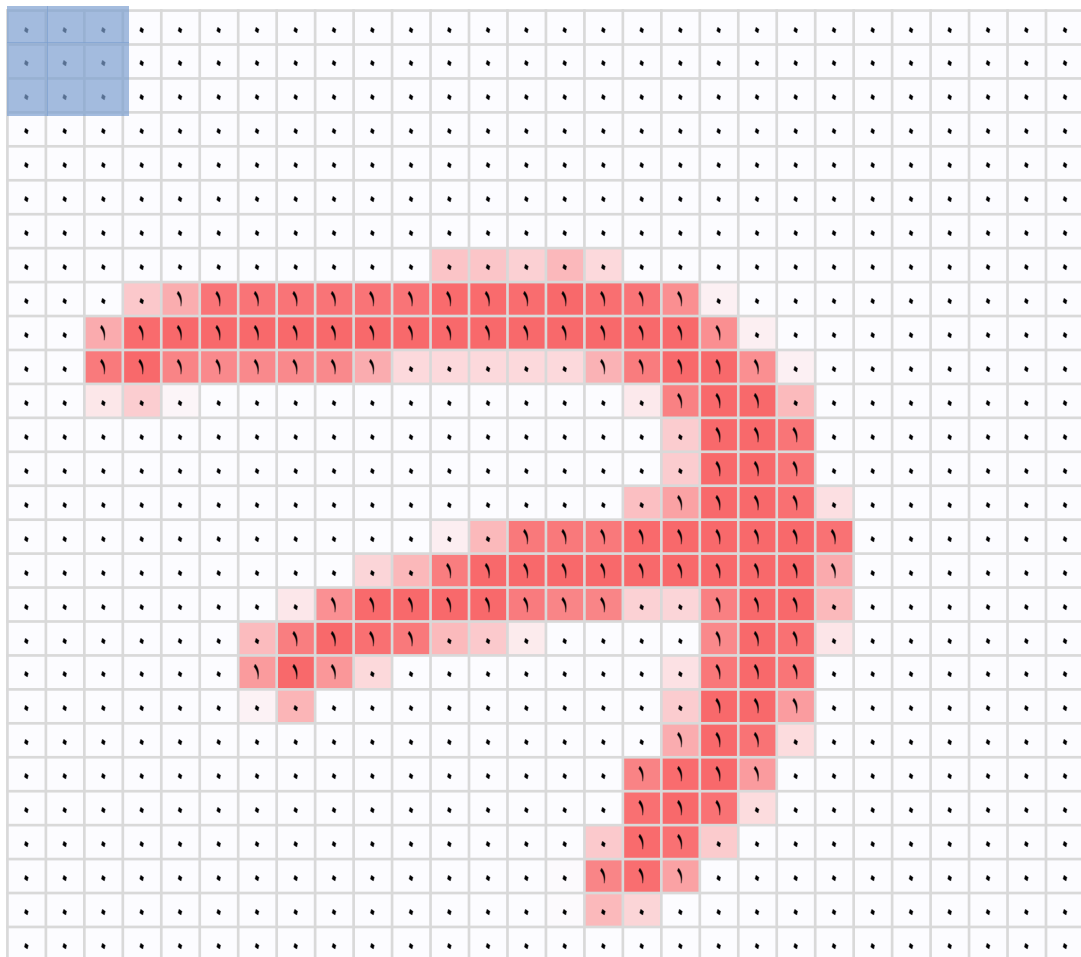


$$1000 \times 1000 \times 3$$

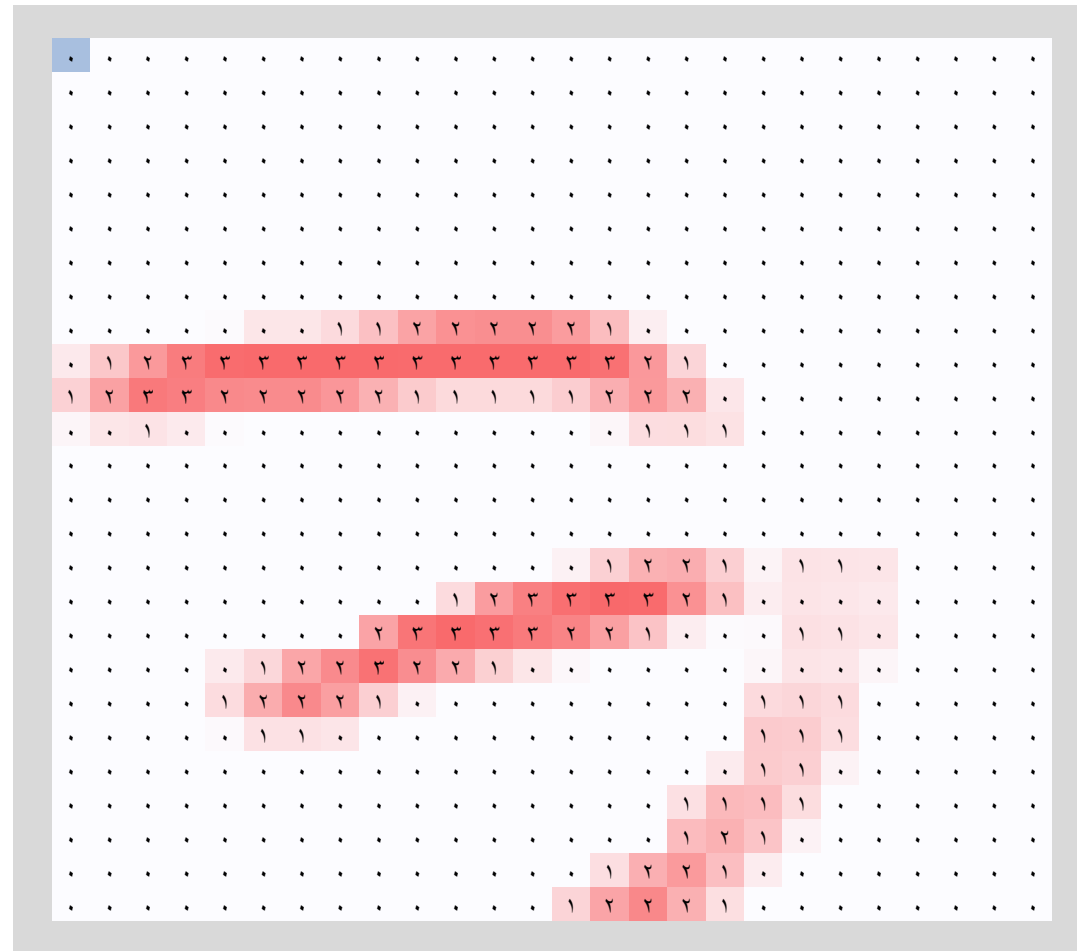


کانولوشن: فیلتر تشخیص لبه

۶

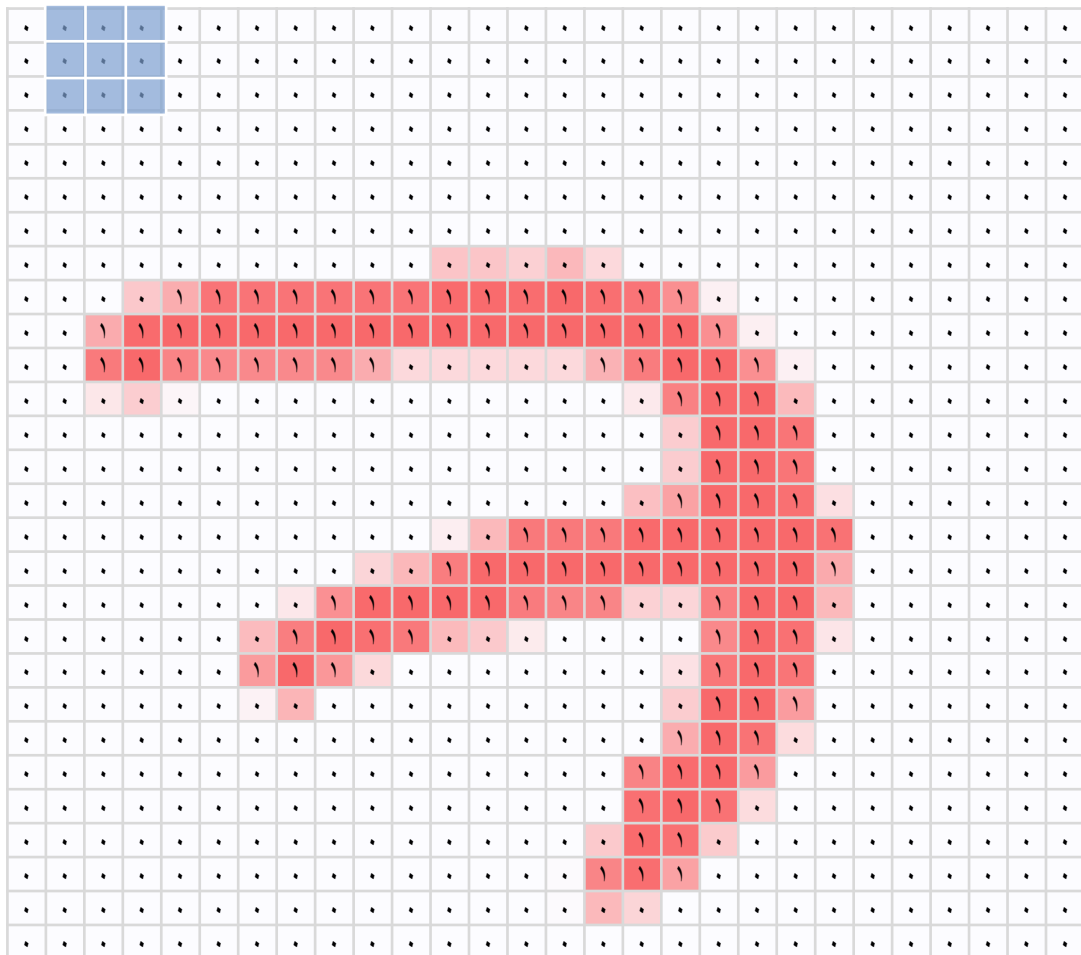


1	1	1
0	0	0
-1	-1	-1

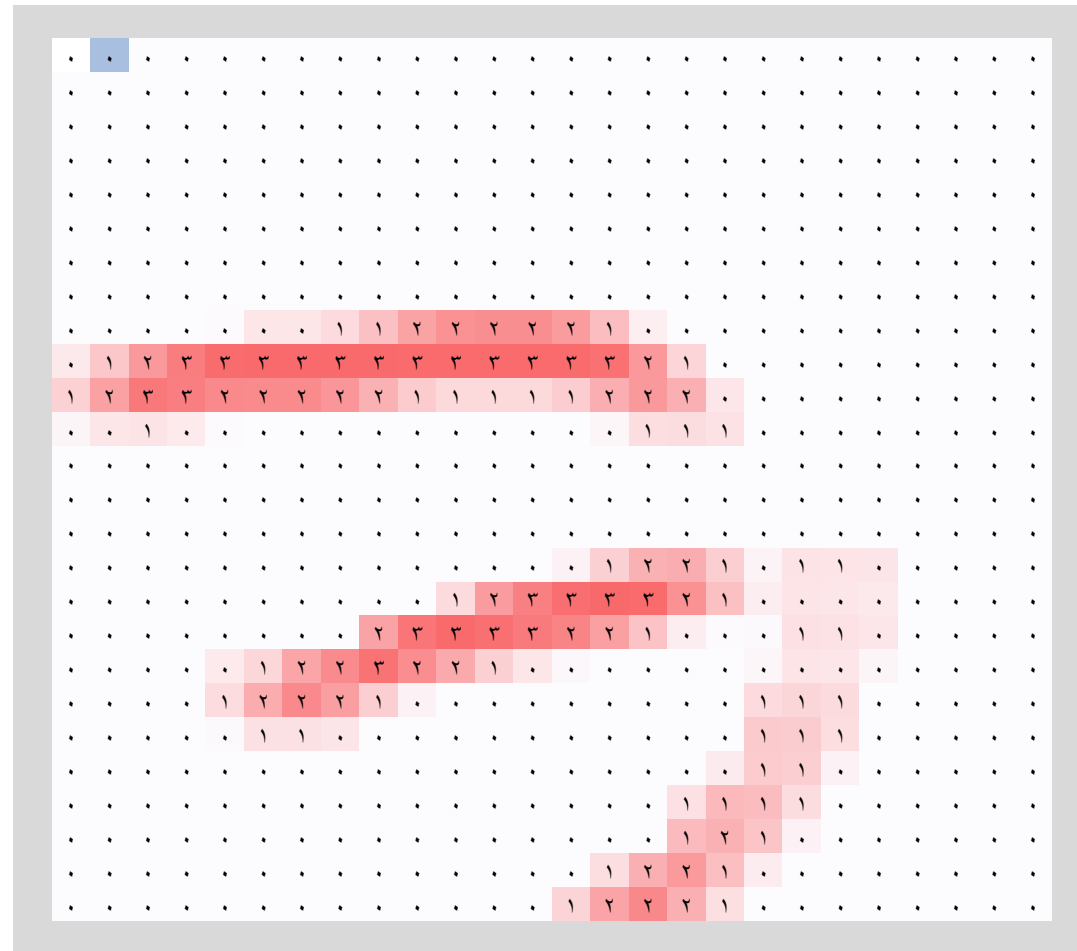


کانولوشن: فیلتر تشخیص لبه

۷

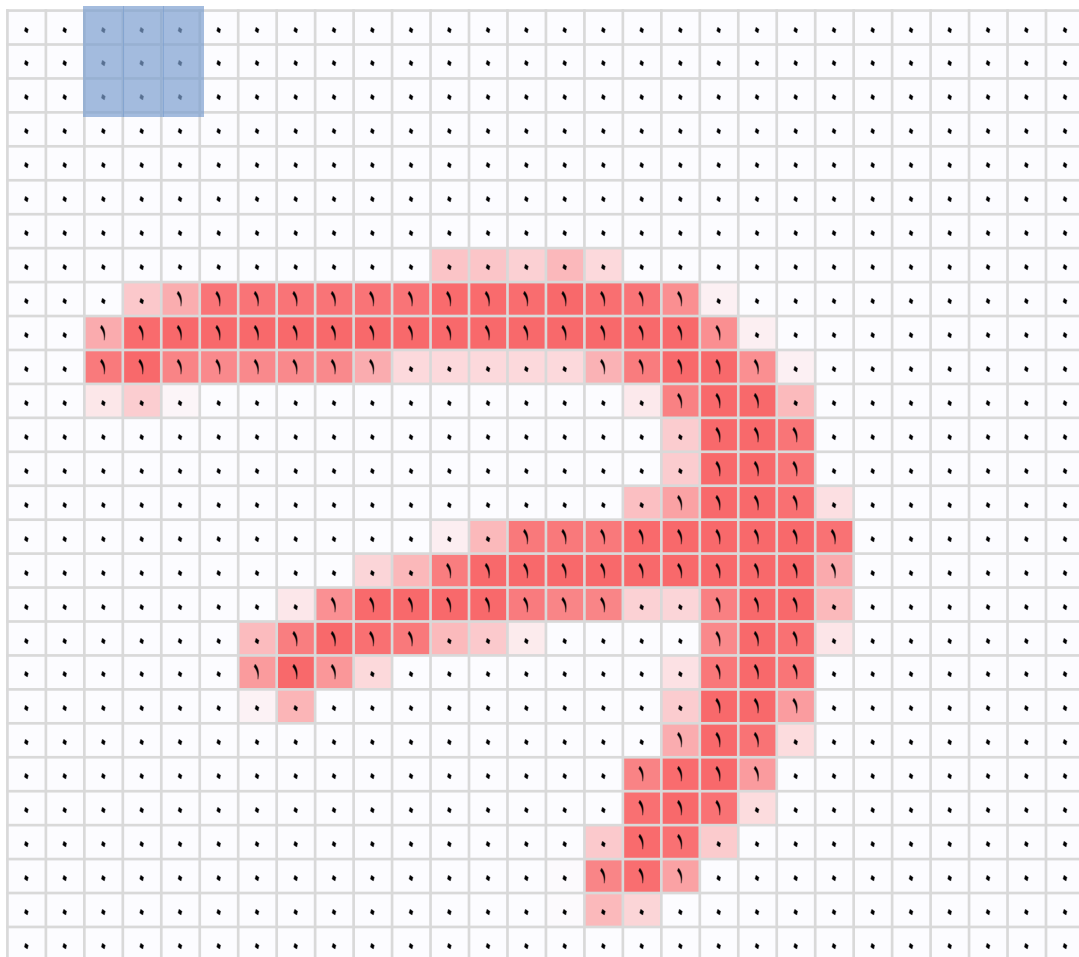


1	1	1
0	0	0
-1	-1	-1

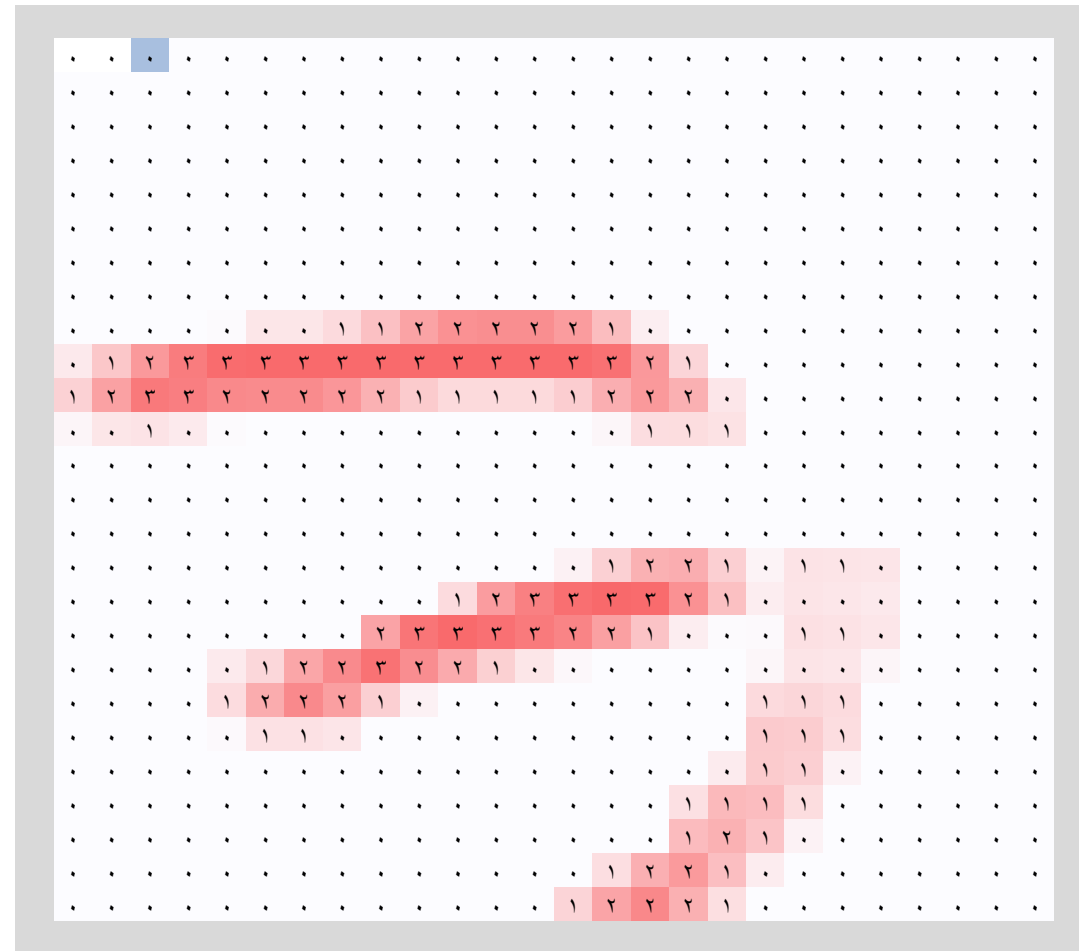


کانولوشن: فیلتر تشخیص لبه

۸

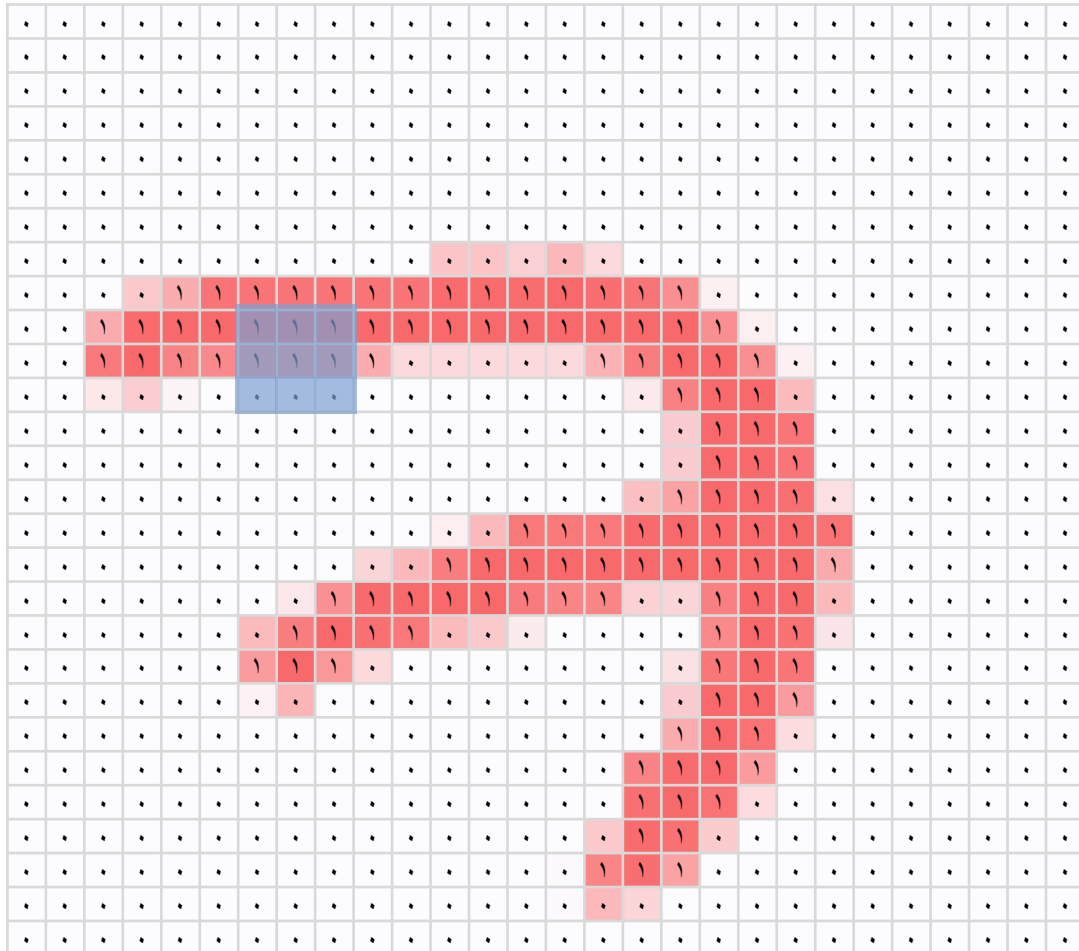


1	1	1
0	0	0
-1	-1	-1

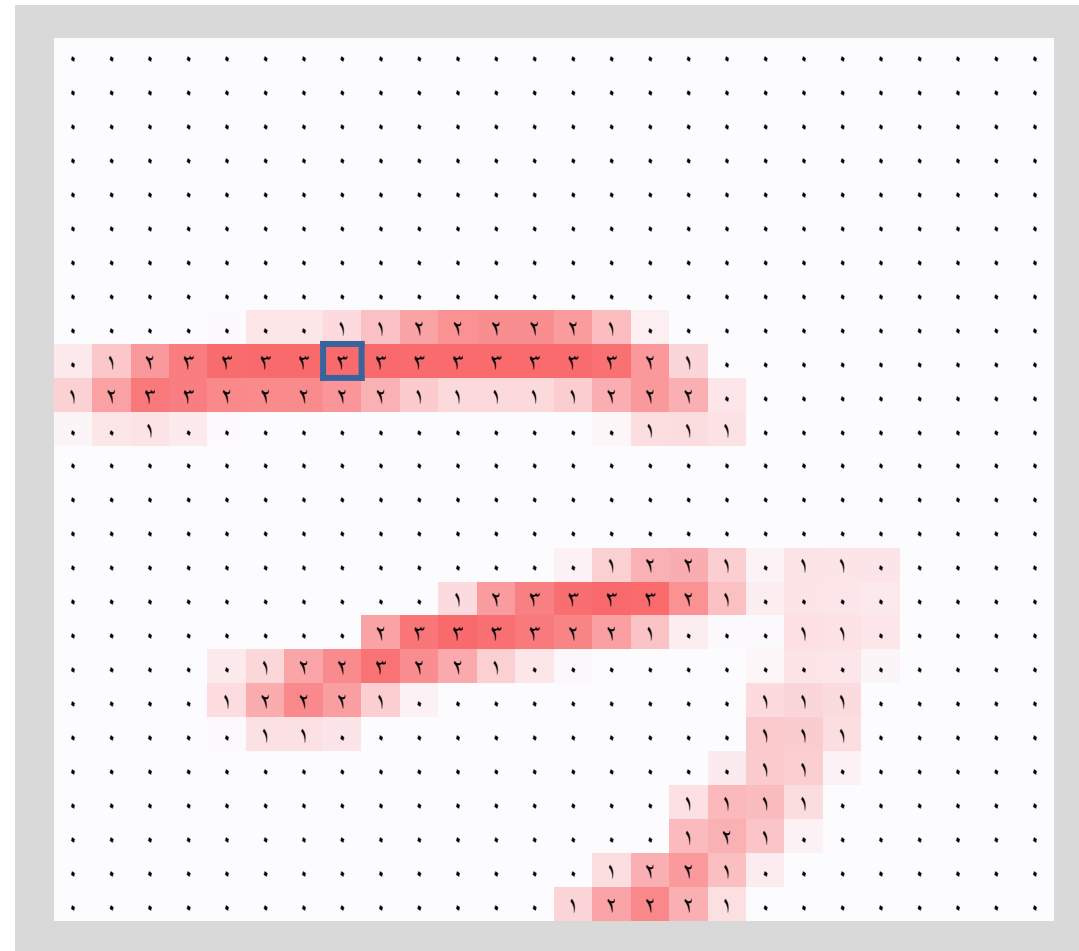


کانولوشن: فیلتر تشخیص لبه

۹

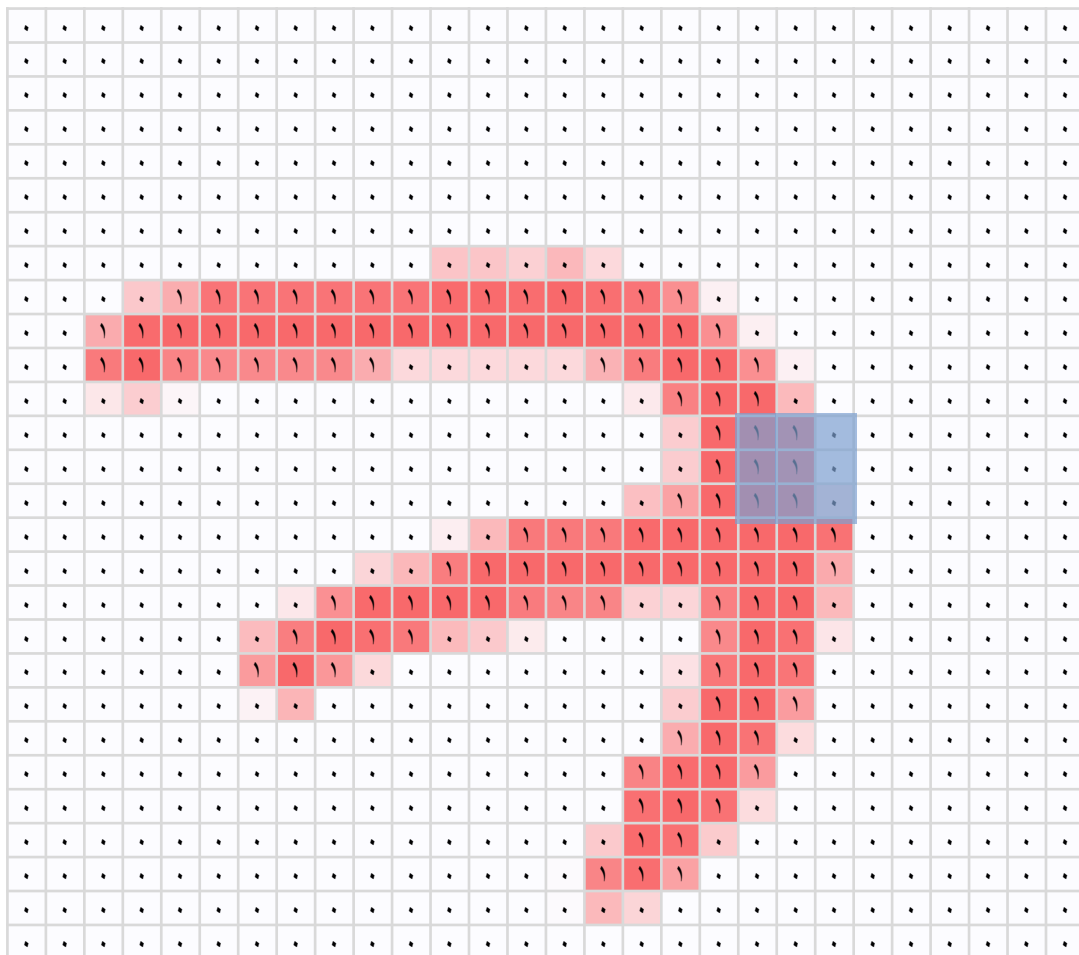


1	1	1
0	0	0
-1	-1	-1

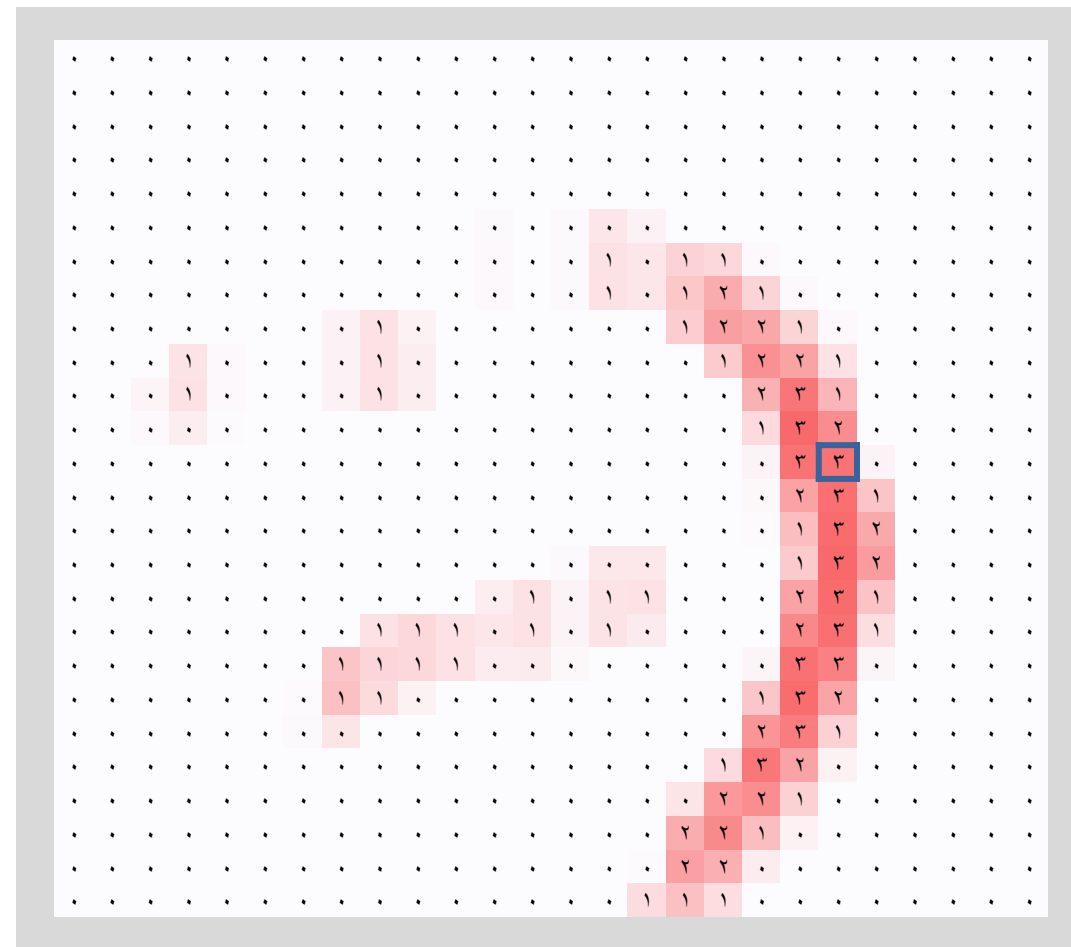


کانولوشن: تشخیص لبه‌های عمودی

۱۰

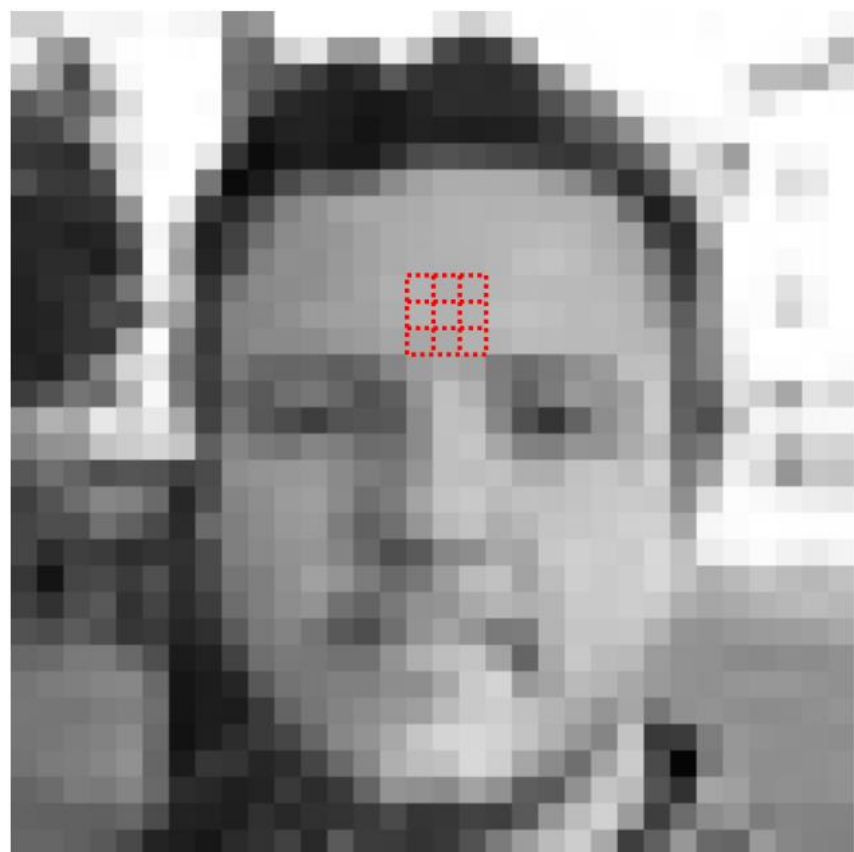


1	0	-1
1	0	-1
1	0	-1



لایه کانولوشن: یک نگاه دقیق‌تر

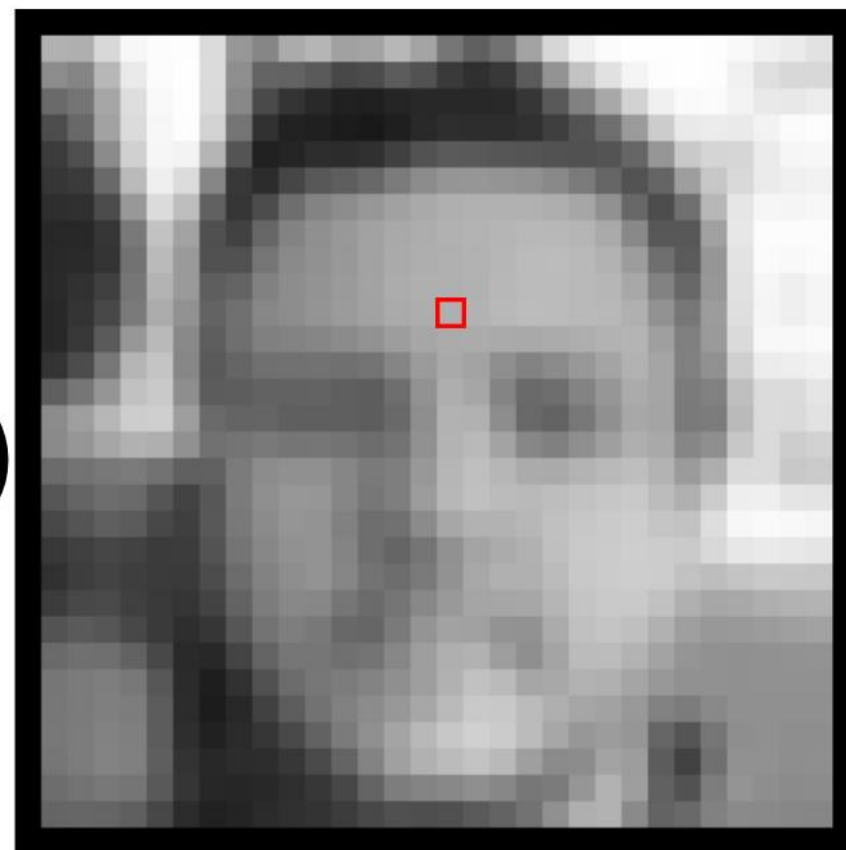
۱۱



input image

$$\left(\begin{array}{ccc} 178 & + & 174 & + & 177 \\ \times 0.0625 & \times 0.125 & \times 0.0625 \\ + & 179 & + & 178 & + & 179 \\ \times 0.125 & \times 0.25 & \times 0.125 \\ + & 168 & + & 171 & + & 178 \\ \times 0.0625 & \times 0.125 & \times 0.0625 \end{array} \right)$$
$$= 176$$

kernel:



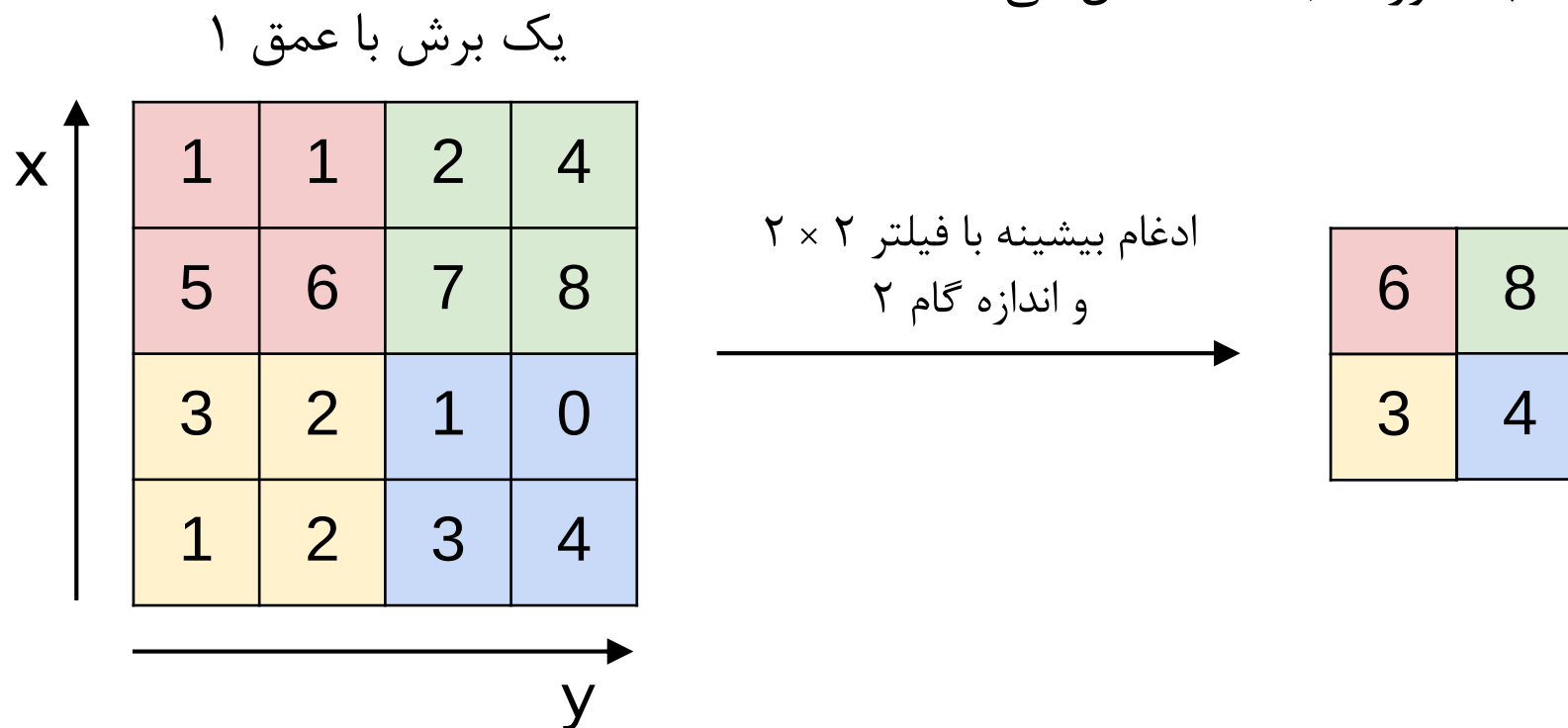
output image

ادغام بیشینه

۱۲

□ لایه ادغام. کوچک تر کردن بازنمایی!

□ بر روی هر نقشه فعالیت به صورت جداگانه عمل می کند.



تشخیص اشیا

۱۳

□ دسته‌بندی تصویر و شناسایی اشیا.



دسته‌بندی تصویر



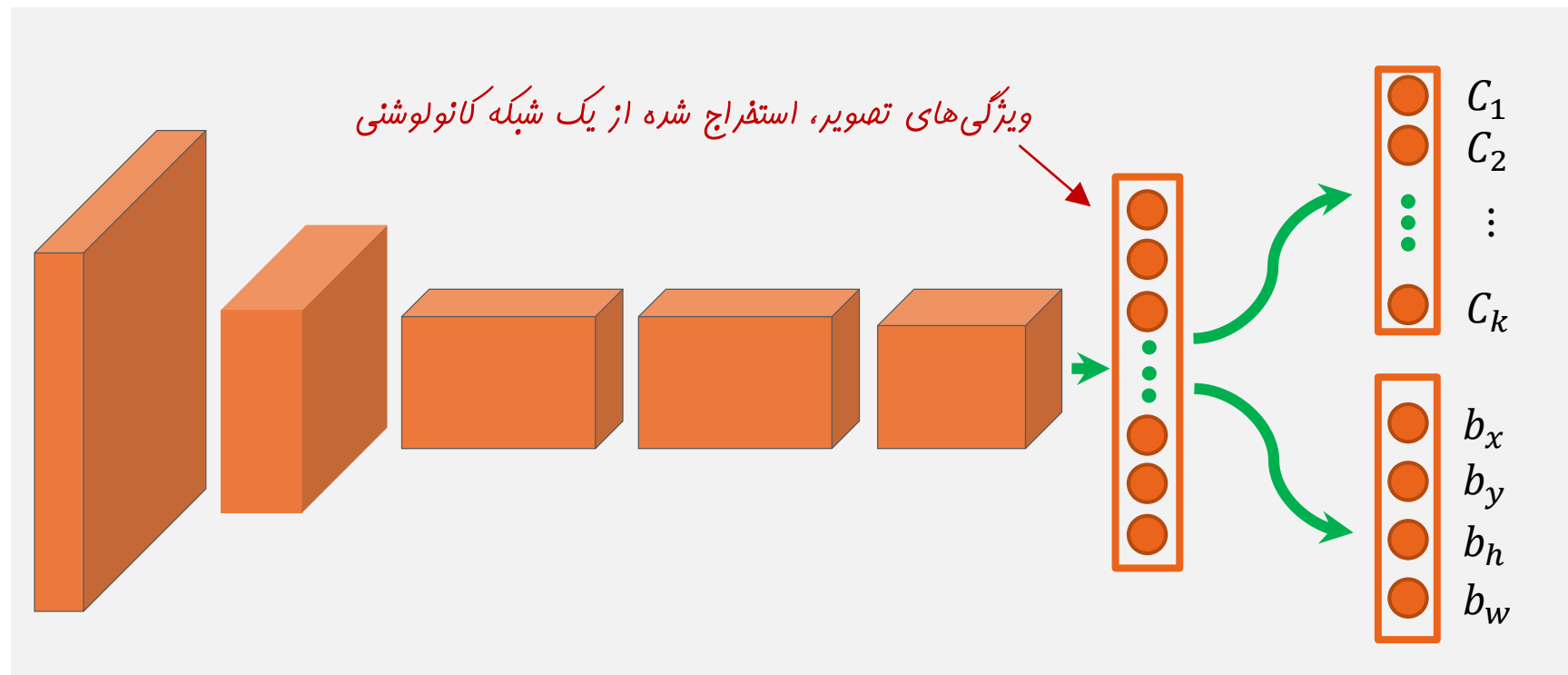
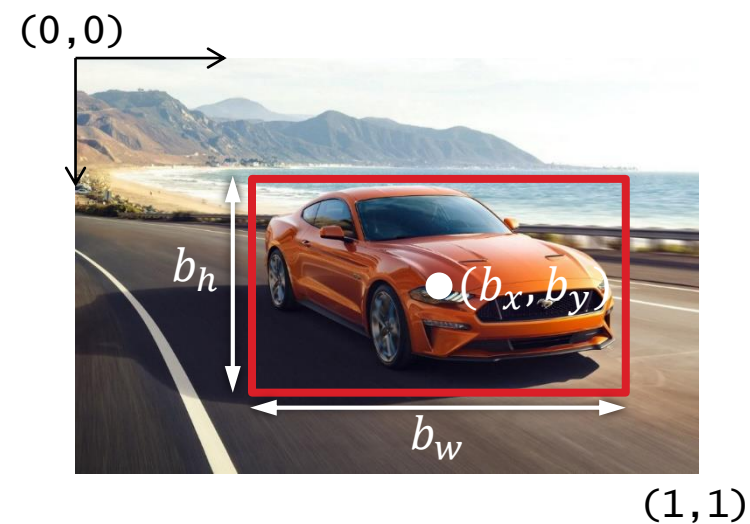
دسته‌بندی و مکان‌یابی



شناسایی اشیا

دسته‌بندی و مکان‌یابی

۱۴



تابع هزینه: مجموع مربعات خطا

۱۵



$$\begin{bmatrix} 1 \\ b_x \\ b_y \\ b_h \\ b_w \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ C_1 \\ C_2 \\ C_3 \end{bmatrix}$$

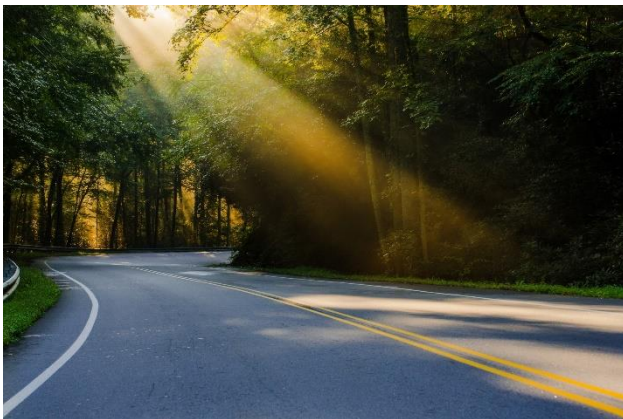
□ کلاس‌ها

۱ - رهگذر

۲ - خودرو

۳ - موتورسیکلت

۴ - پس‌زمینه



$$\begin{bmatrix} 0 \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \\ ? \end{bmatrix}$$

$$L(\hat{y}, y) = \begin{cases} \sum_{i=1}^f (\hat{y}_i - y_i)^2, & y_1 = 1 \\ (\hat{y}_1 - y_1)^2, & y_1 = 0 \end{cases}$$

تشخیص نقاط عطف: تشخیص احساسات

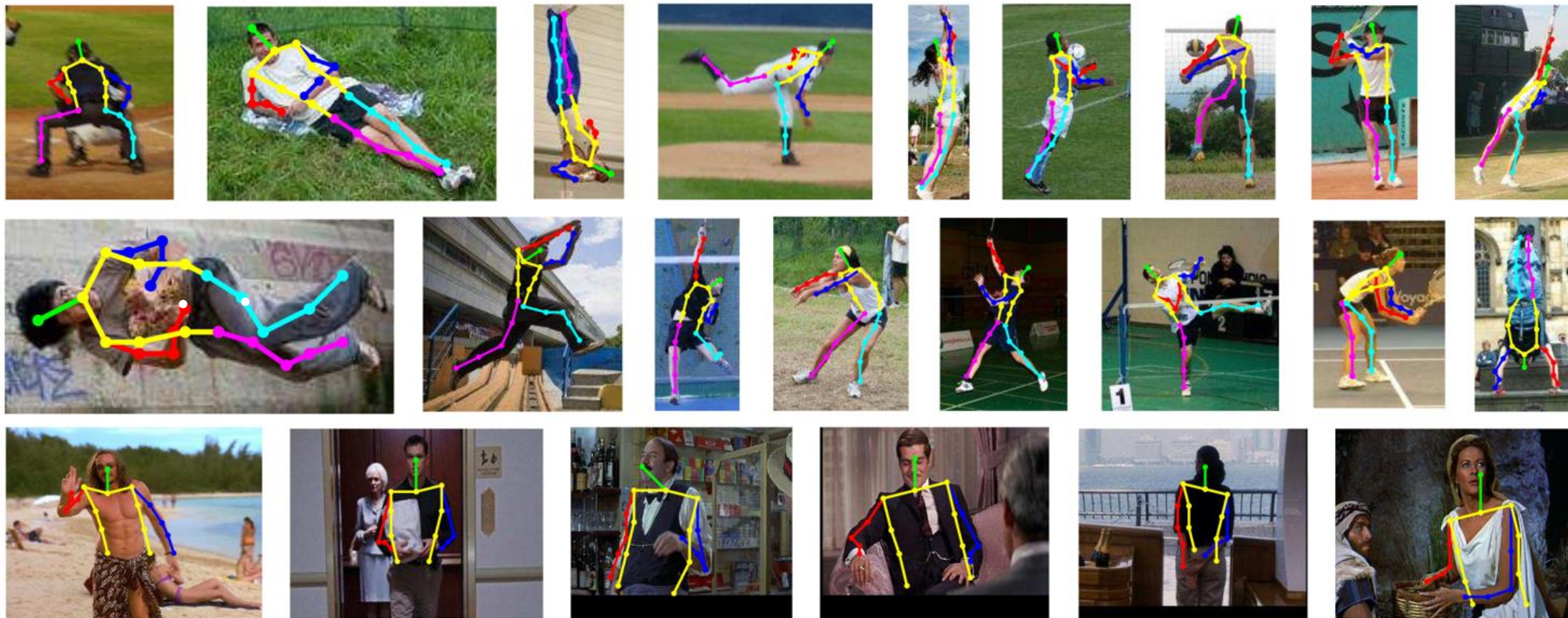
۱۶



Zhanpeng Zhang, et. al. Facial Landmark Detection by Deep Multi-task Learning, ECCV, 2014

تشخیص نقاط عطف: تشخیص وضعیت

۱۷

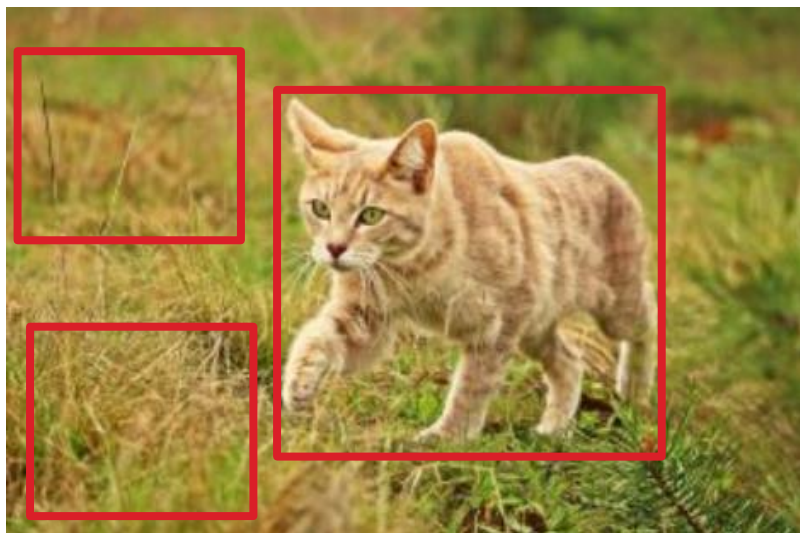


Xianjie Chen and Alan Yuille, Articulated Pose Estimation by a Graphical Model with Image Dependent Pairwise, NIPS, 2014.

شناسایی اشیاء: پنجره لغزان

۱۸

ایجاد مجموعه آموزشی



1

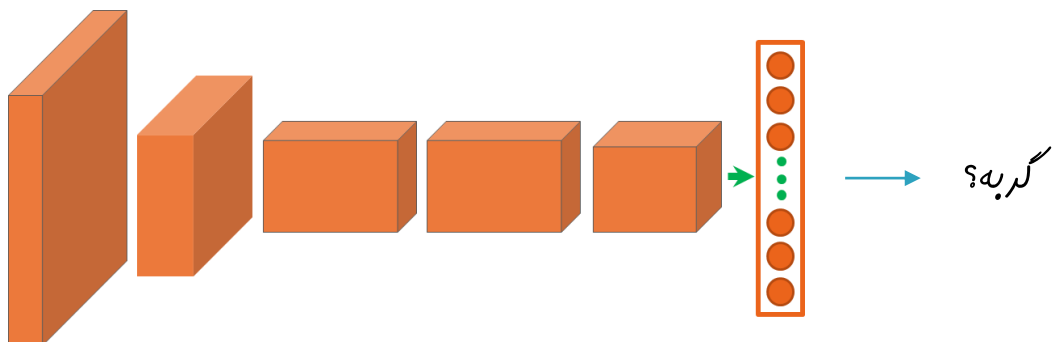


0



0

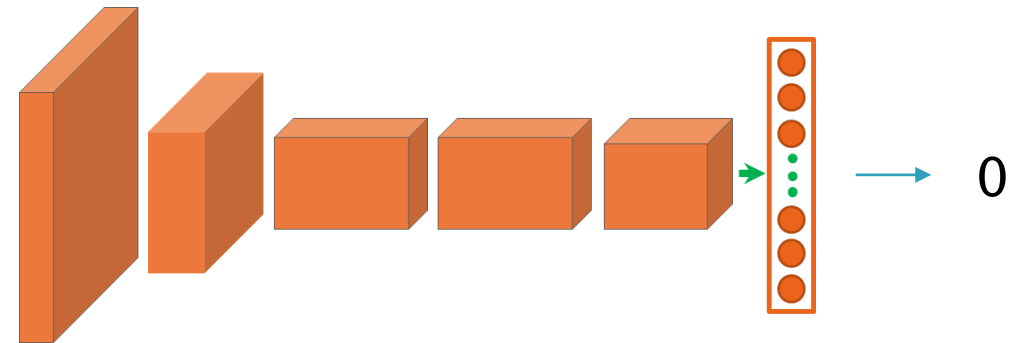
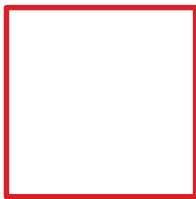
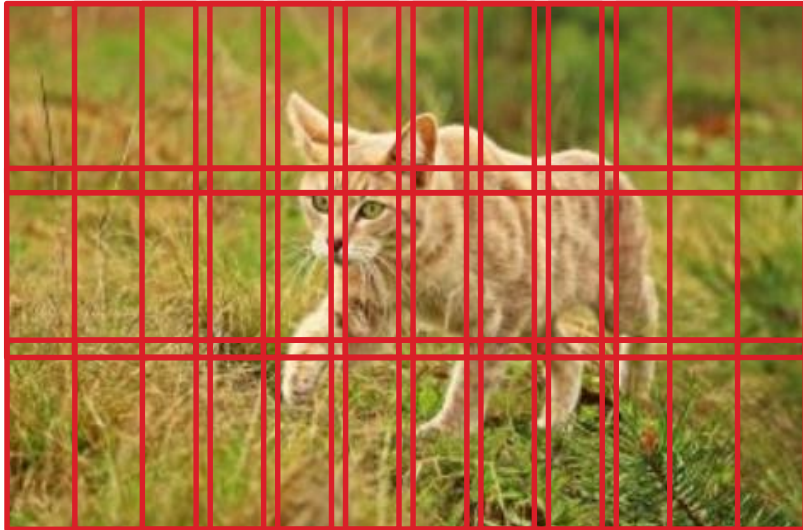
آموزش شبکه کانولوشنی



شناسایی اشیا: پنجره لغزان

۱۹

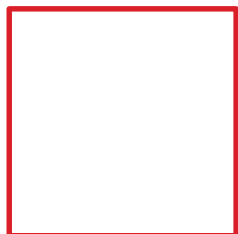
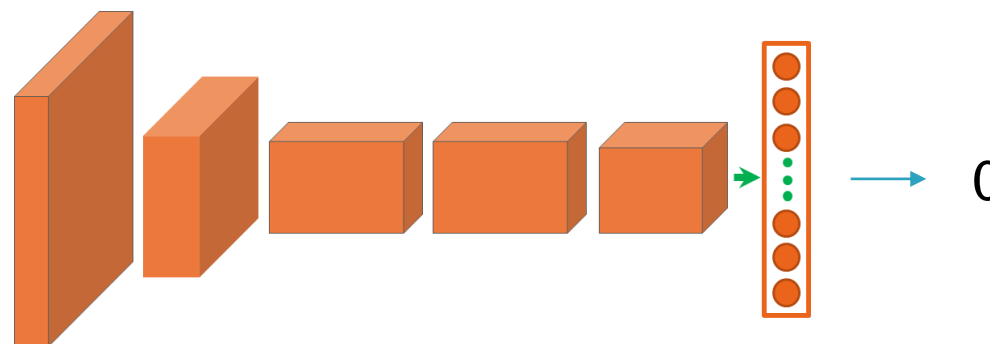
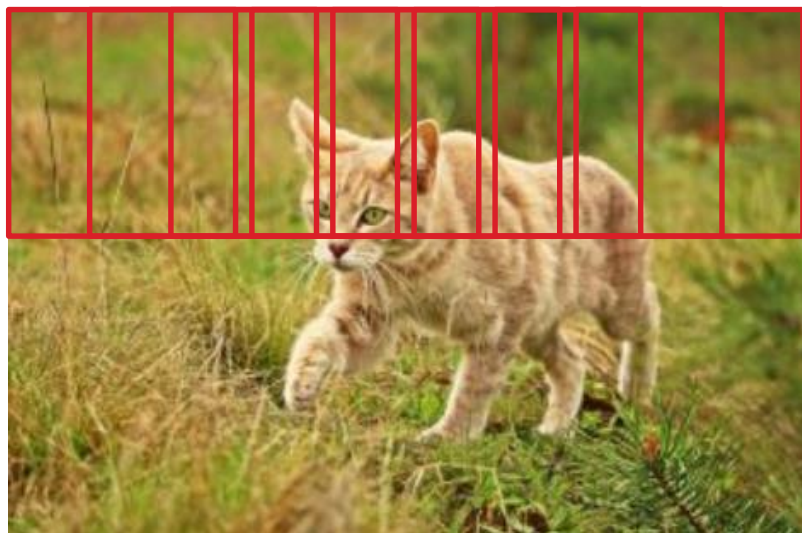
شناسایی اشیا با پنجره لغزان



شناسایی اشیاء: پنجره لغزان

۲۰

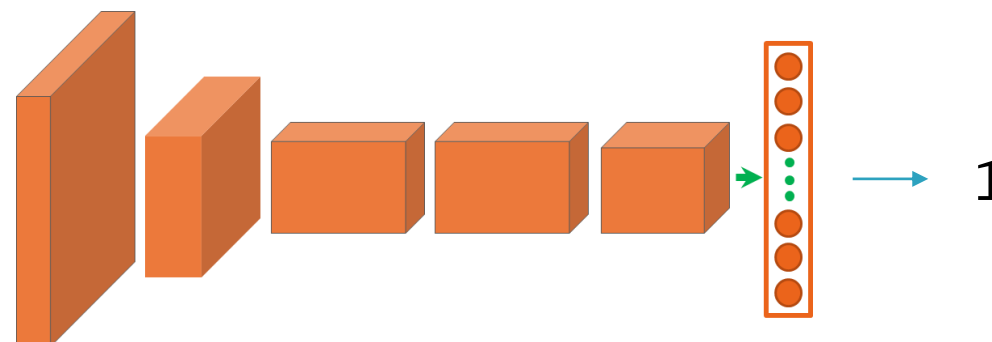
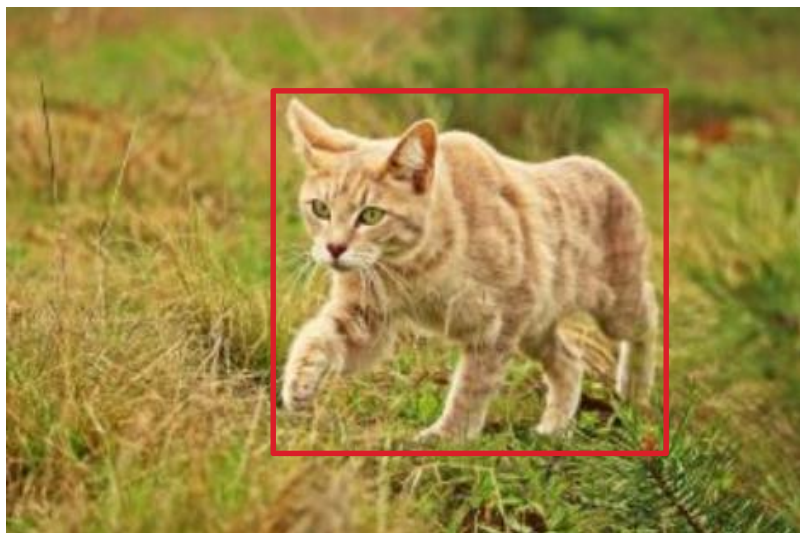
شناسایی اشیاء با پنجره لغزان



شناسایی اشیاء: پنجره لغزان

۲۱

شناسایی اشیاء با پنجره لغزان



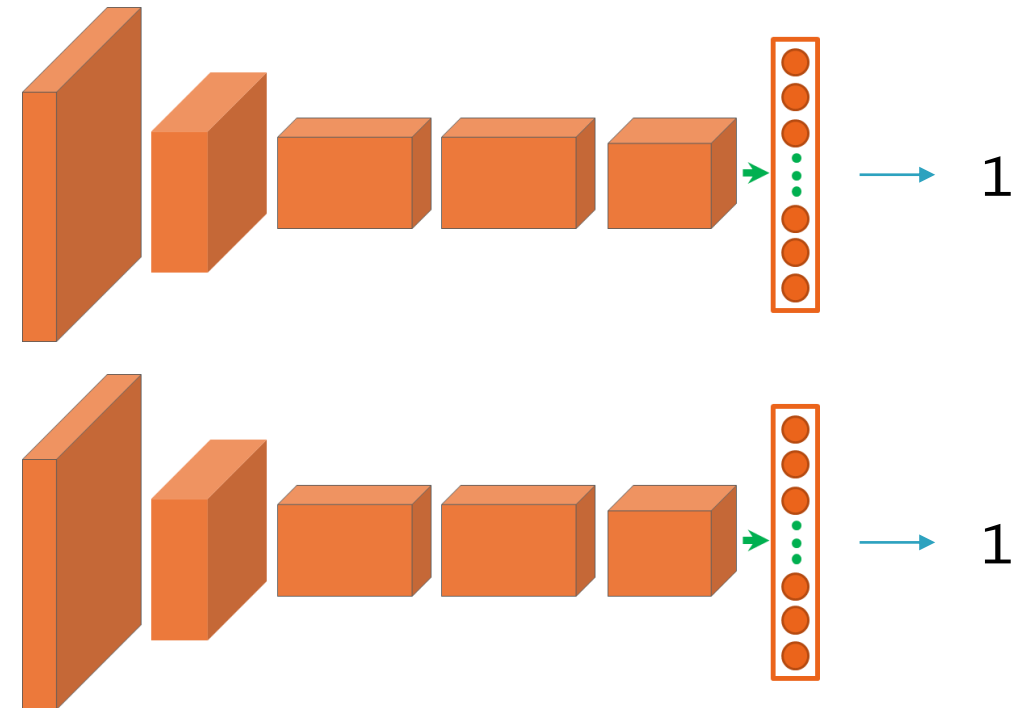
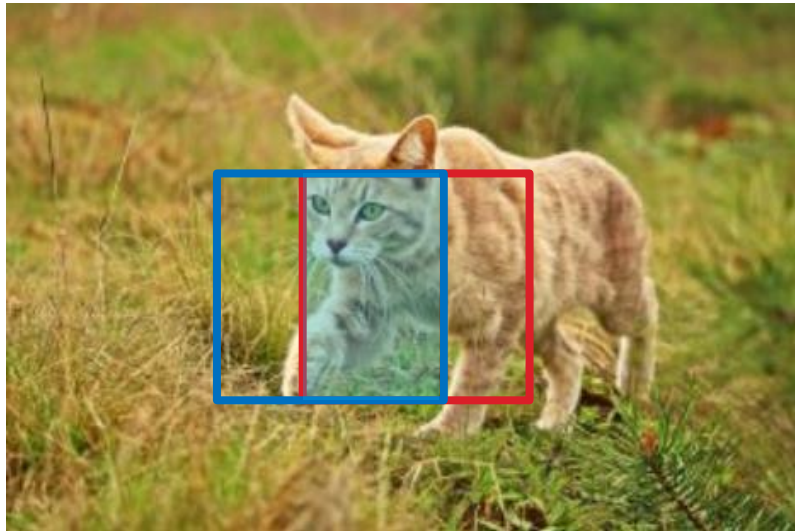
هزینه محاسباتی بسیار بالا!



پنجره لغزان: کاهش هزینه‌های محاسباتی

۲۲

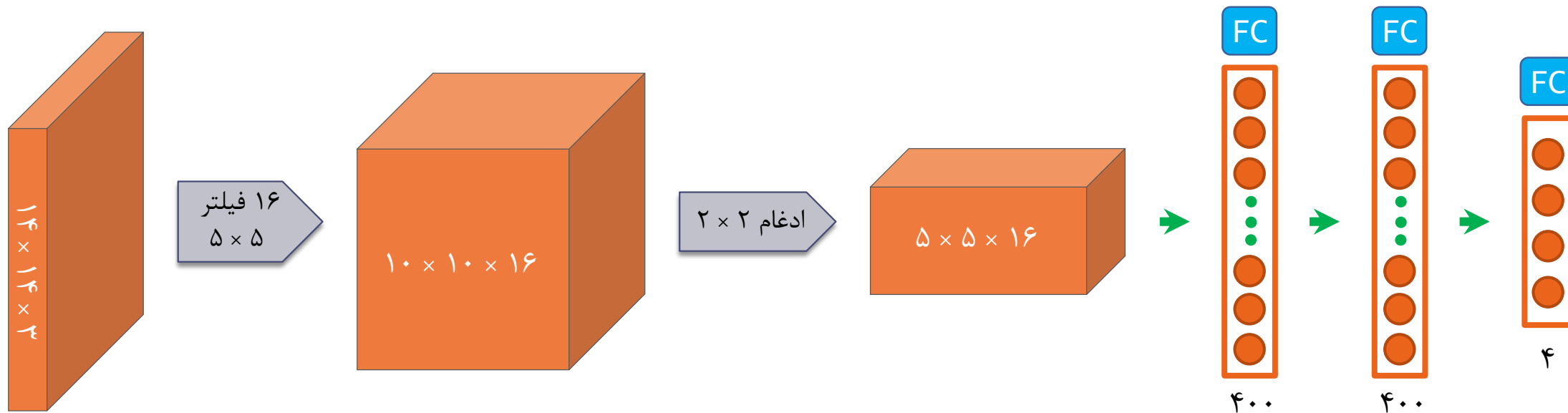
شناسایی اشیاء با پنجره لغزان



پنجره لغزان: پیاده‌سازی کانولوشنی

۲۳

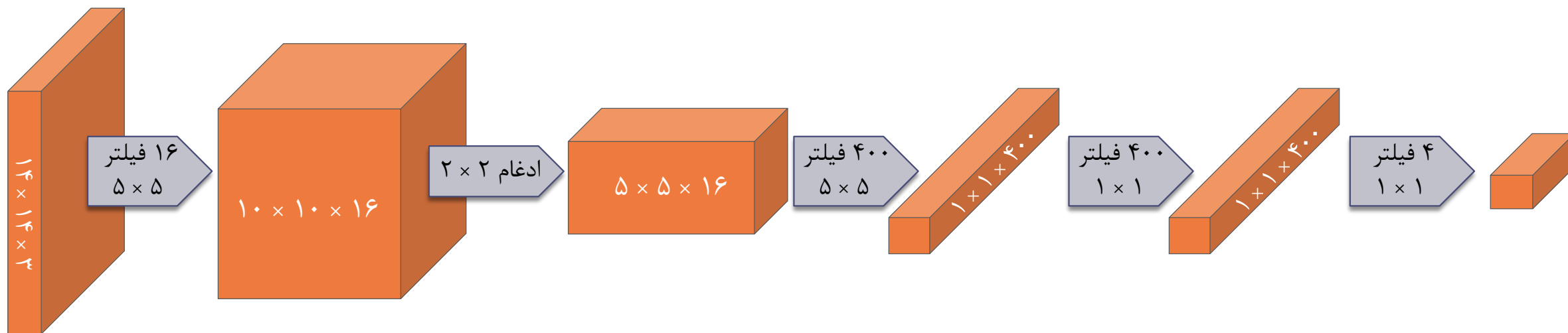
□ تبدیل لایه‌های کاملاً متصل به لایه‌های کانولوشنی.



پنجره لغزان: پیاده‌سازی کانولوشنی

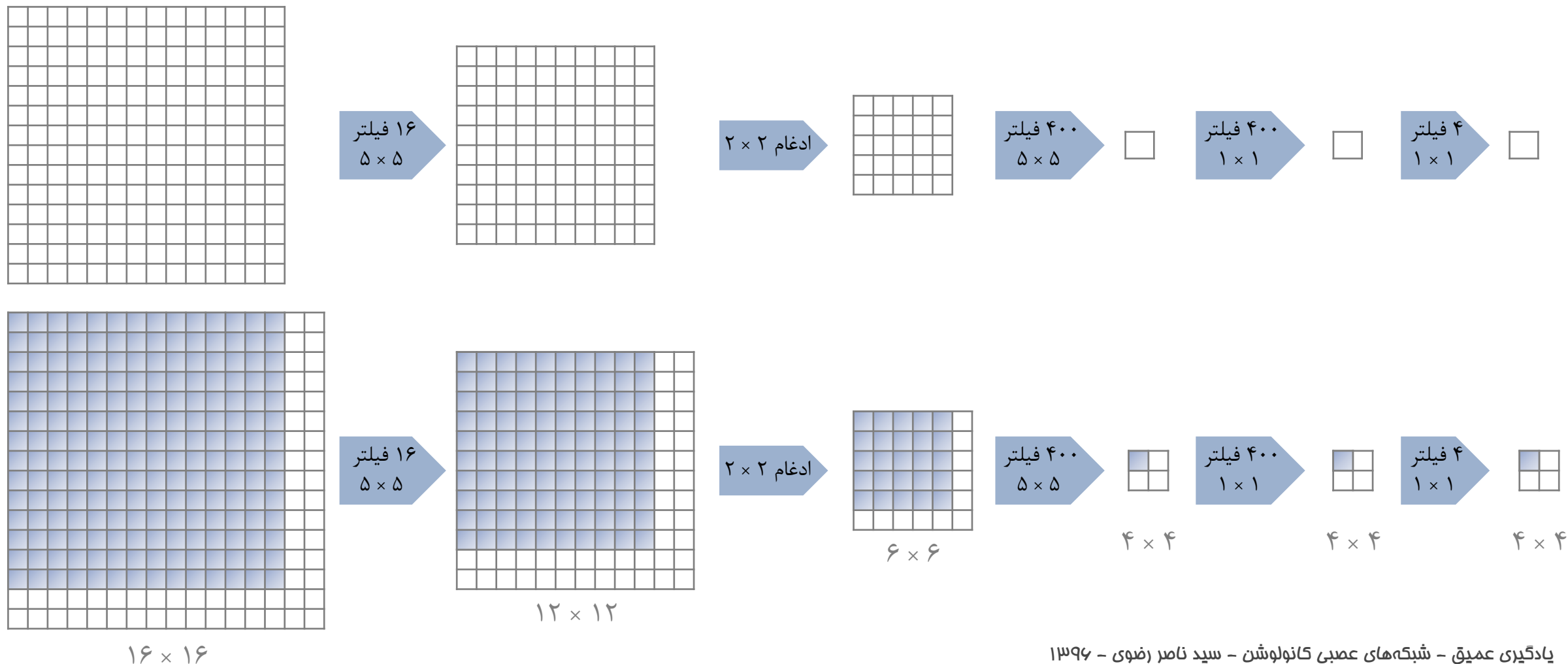
۲۴

□ تبدیل لایه‌های کاملاً متصل به لایه‌های کانولوشنی.



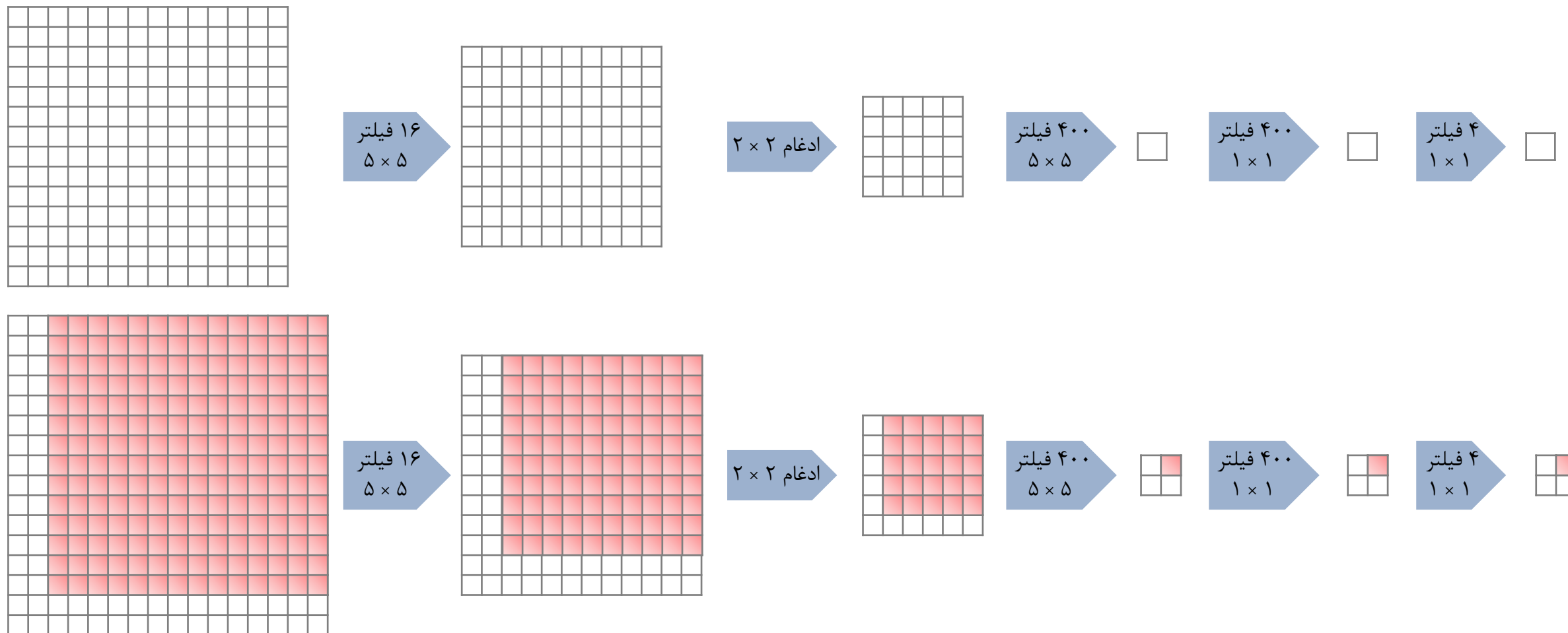
پنجره لغزان: پیاده‌سازی کانولوشنی

۲۵



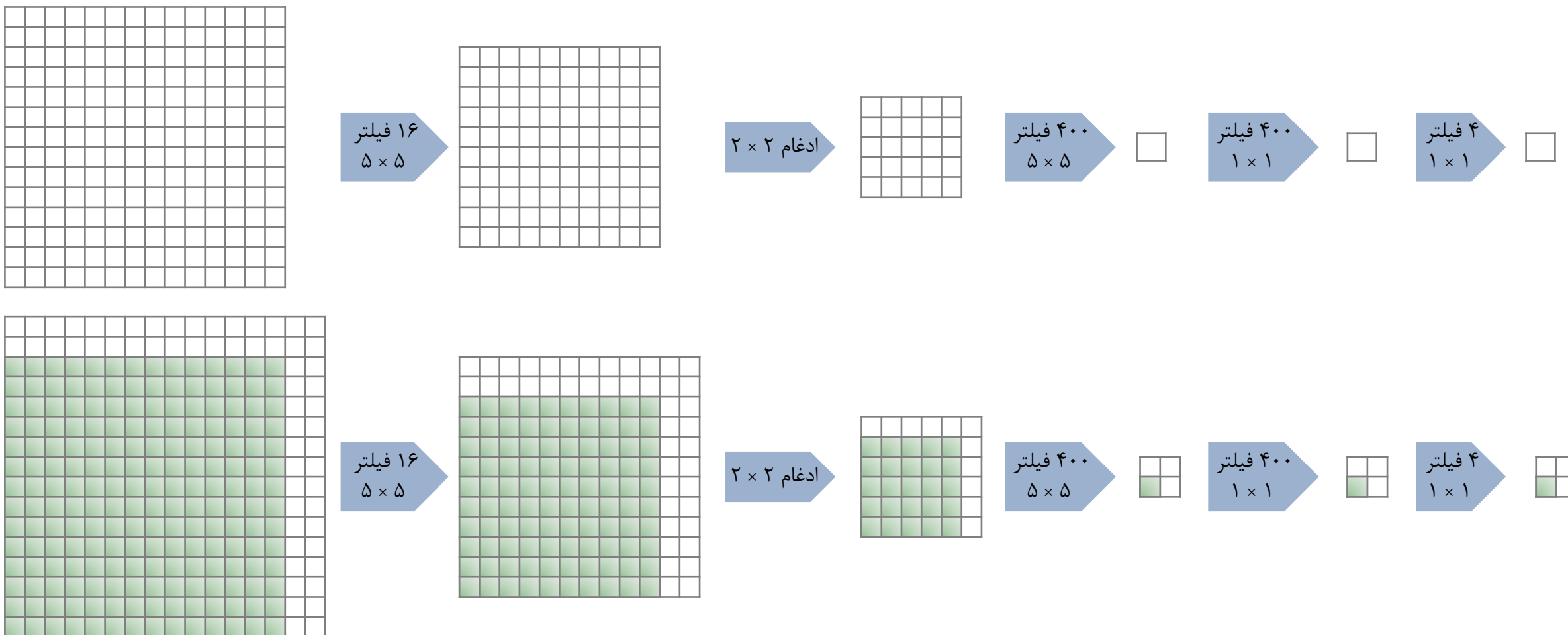
پنجره لغزان: پیاده‌سازی کانولوشنی

۲۶



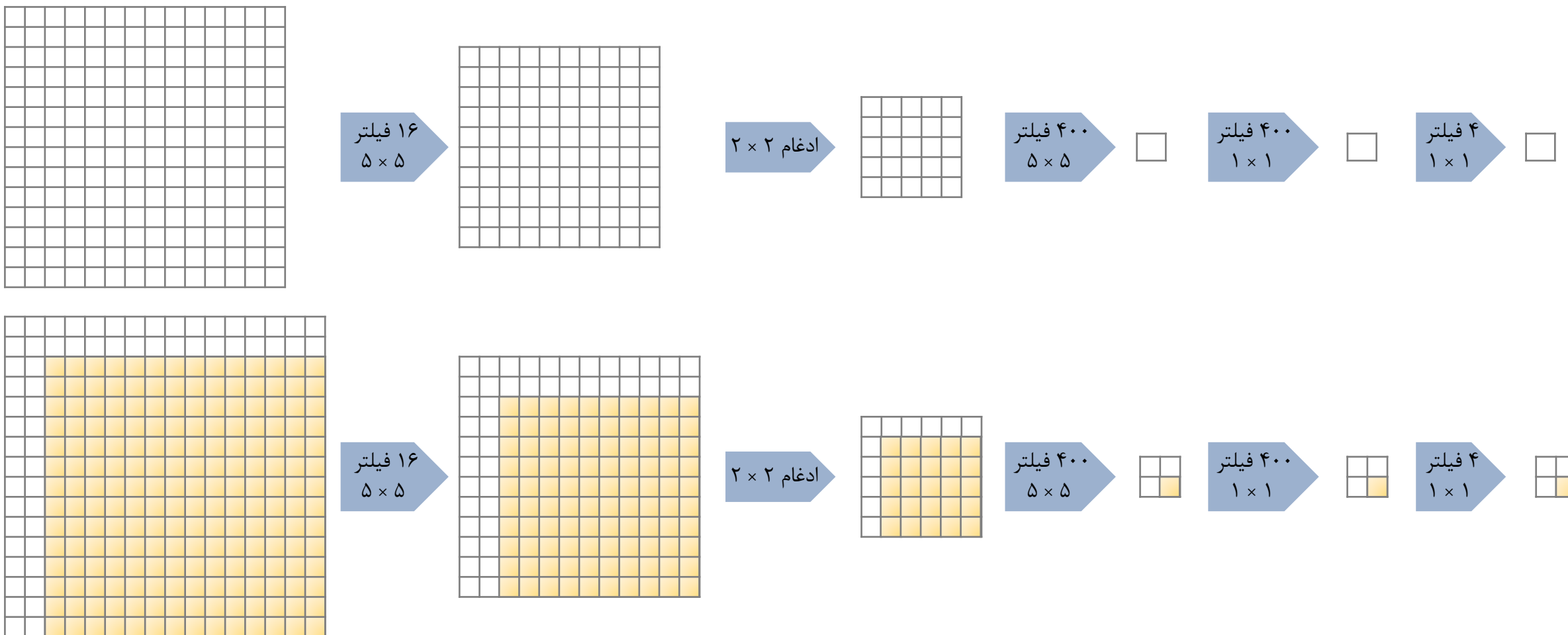
پنجره لغزان: پیاده‌سازی کانولوشنی

۲۷



پنجره لغزان: پیاده‌سازی کانولوشنی

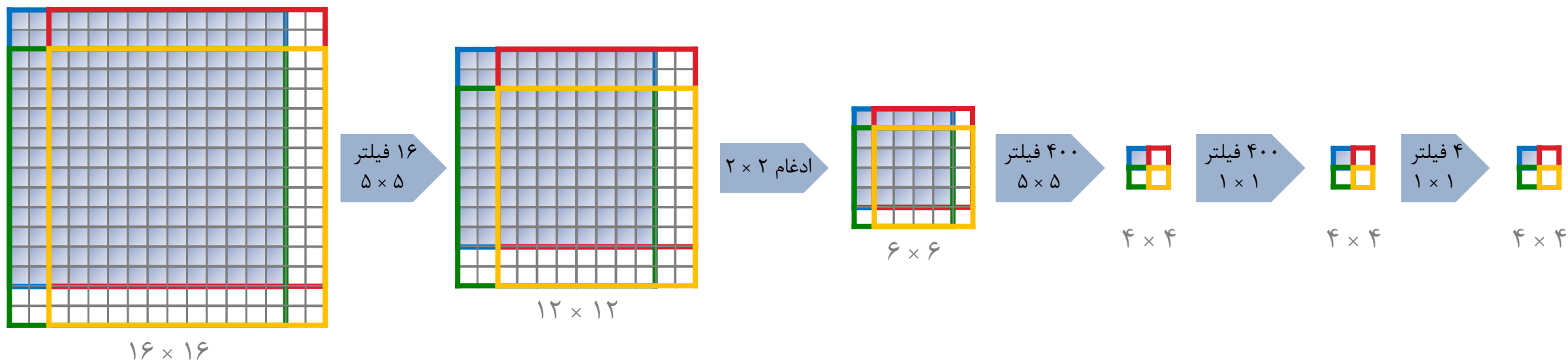
۲۸



پنجره لغزان: پیاده‌سازی کانولوشنی

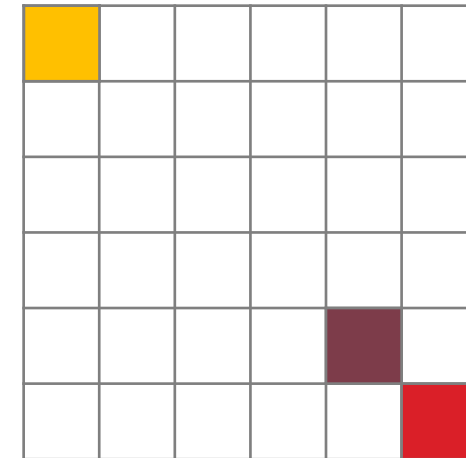
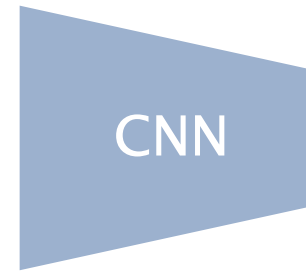
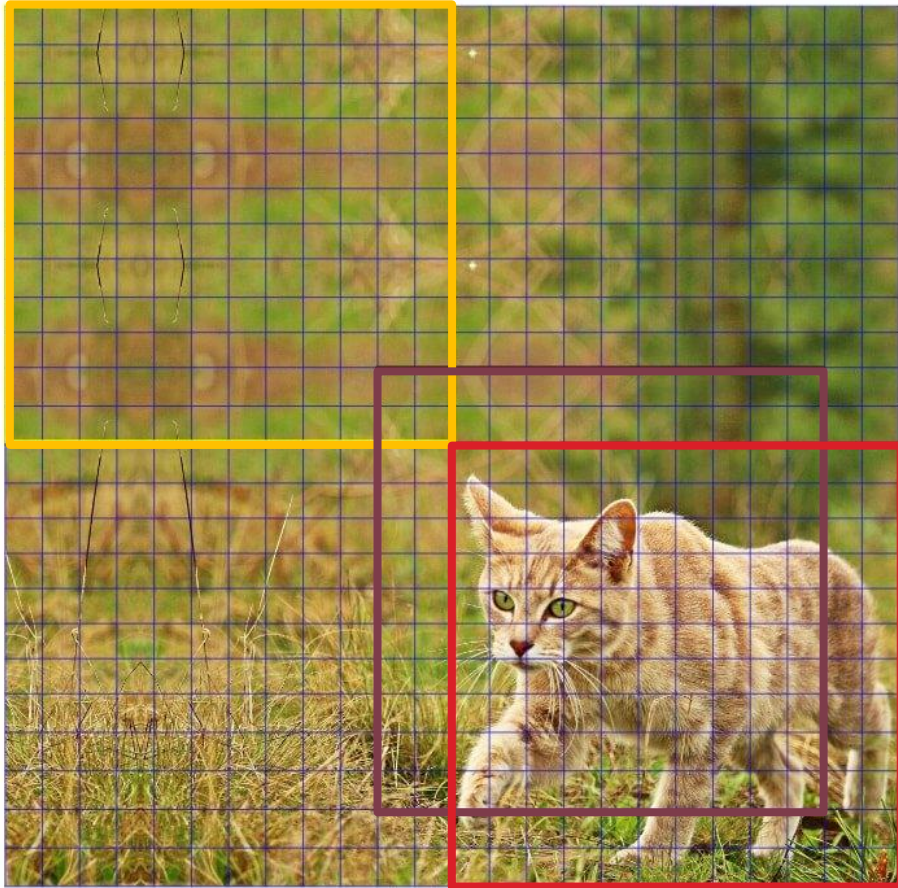
۲۹

انجام محاسبات مربوط به پنجره لغزان به صورت موازی و همزمان!



پنجره لغزان: پیاده‌سازی کانولوشنی

۳۰



احتمال وجود اشیا
در هر یک از پنجره‌ها

اشتراک بر روی اجتماع

۳۱

□ معیاری به منظور محاسبه میزان همپوشانی دو جعبه محاطی.

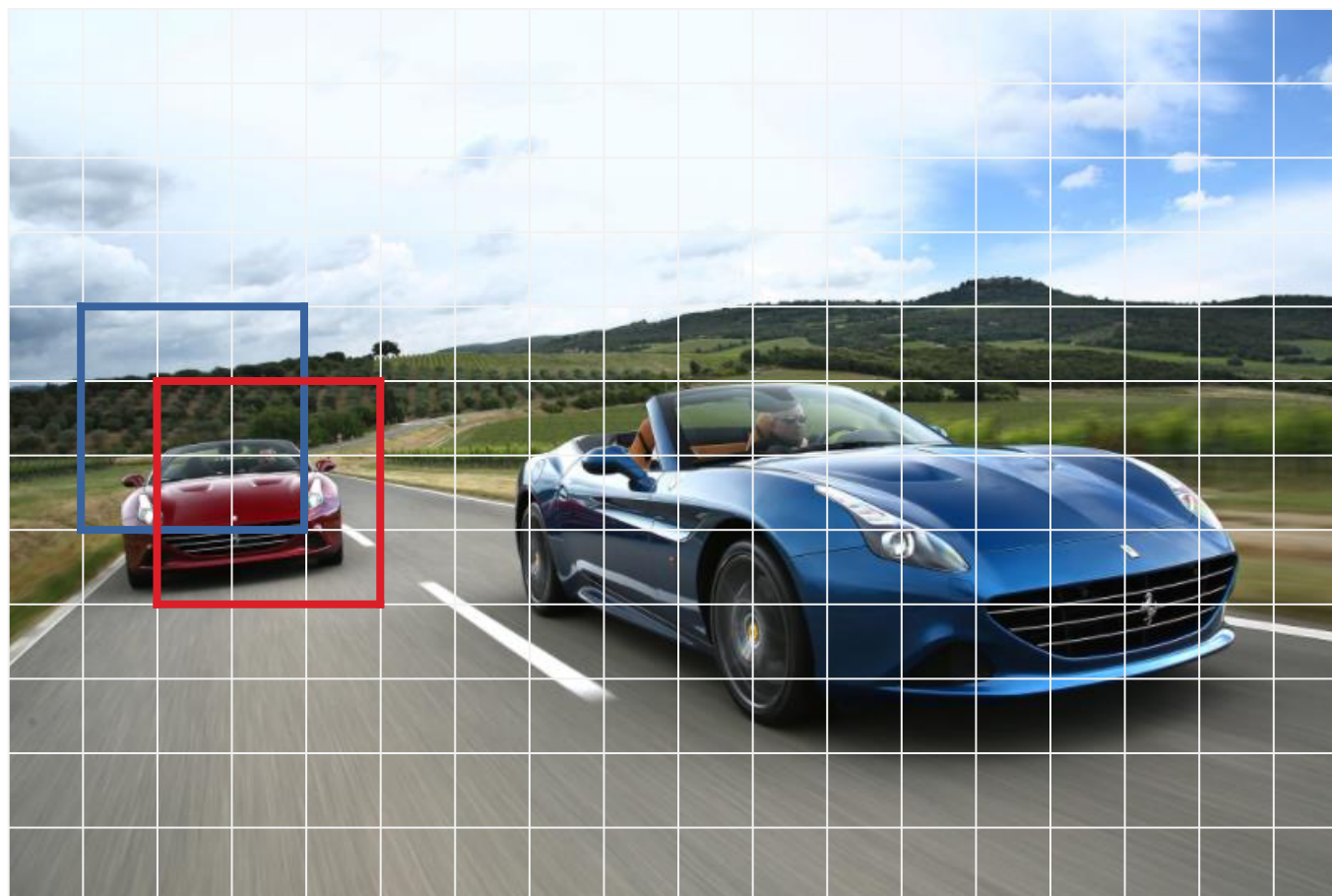


$$IoU = \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|} \geq 0.5$$

سرکوب غیر بیشینه

۳۲

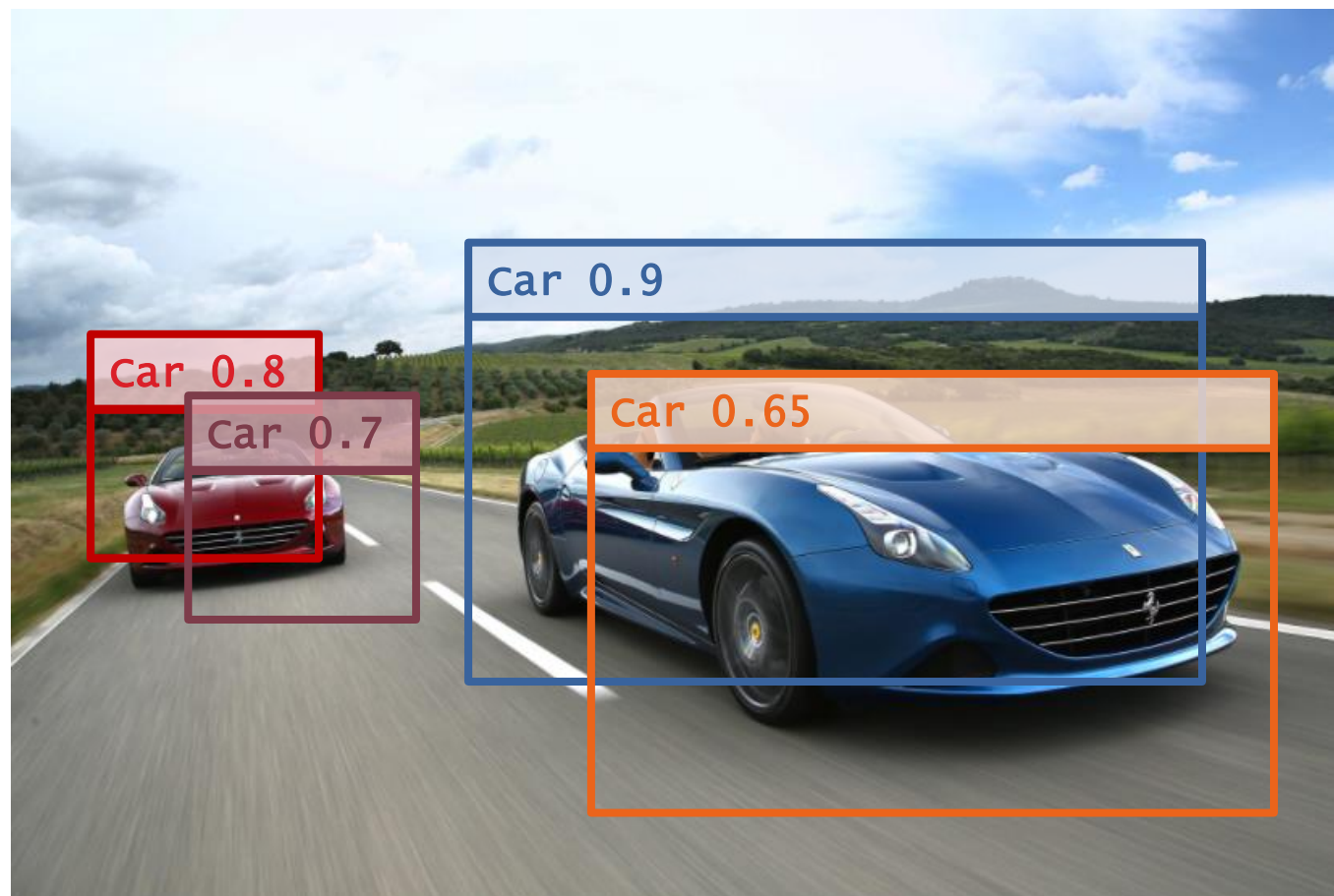
□ اطمینان از این که هر شی، بیش از یک بار شناسایی نمی شود.



سرکوب غیر بیشینه

۳۳

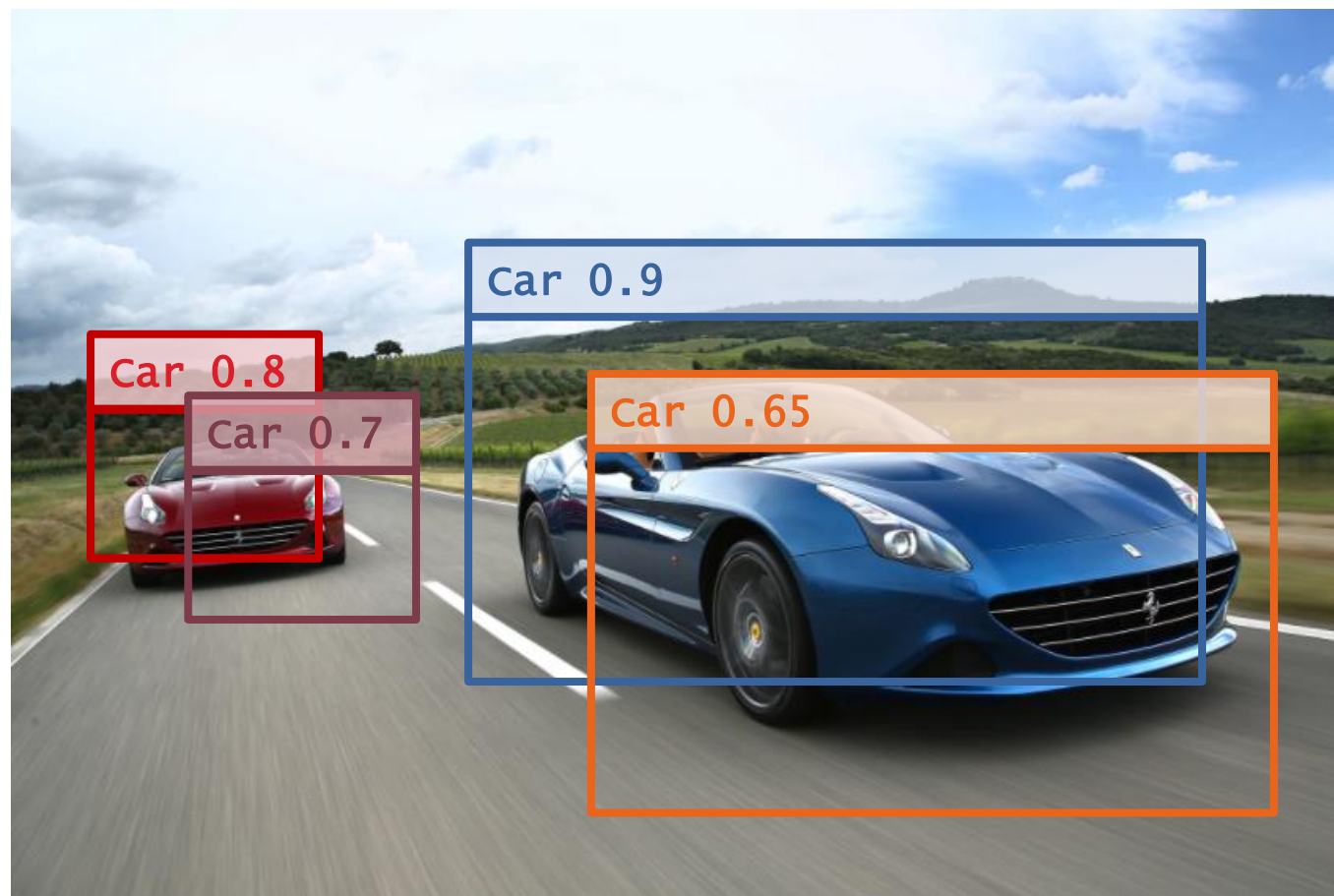
□ اطمینان از این که هر شی، بیش از یک بار شناسایی نمی شود.



سرکوب غیر بیشینه

۳۴

□ اطمینان از این که هر شی، بیش از یک بار شناسایی نمی شود.



سرکوب غیر بیشینه

۳۵

□ اطمینان از این که هر شی، بیش از یک بار شناسایی نمی شود.



سرکوب غیر بیشینه: الگوریتم

۳۶

$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \end{bmatrix}$$

□ مرحله اول: حذف تمام پیش‌بینی‌ها با احتمال کمتر از ۰/۶

□ مرحله دوم: تا زمانی که پیش‌بینی (جعبه محاطی) پردازش نشده‌ای وجود دارد:

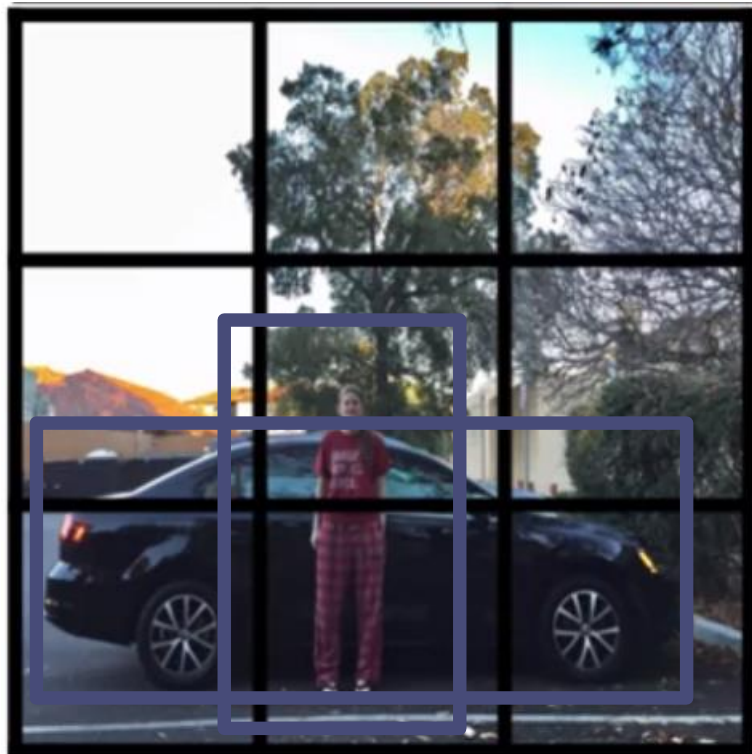
□ یک جعبه با بیشترین احتمال را انتخاب و آن را به عنوان خروجی در نظر بگیر.

□ تمامی جعبه‌هایی را که هم‌پوشانی زیادی با جعبه انتخاب شده دارند، حذف کن.


$$IoU \geq 0.5$$

همپوشانی اشیا و جعبه‌های لنگر

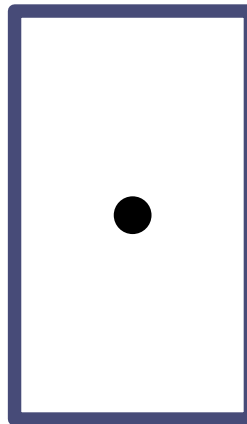
۳۷



$$y = \begin{bmatrix} y_{a1} \\ y_{a2} \end{bmatrix}$$

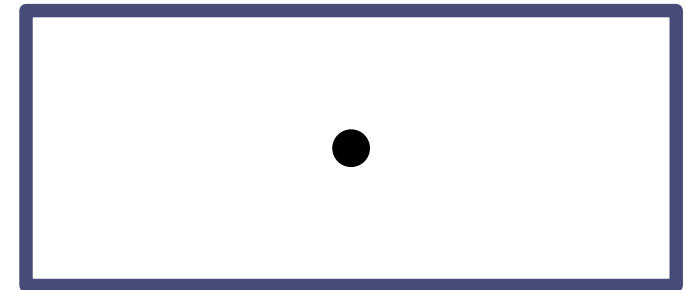
به ازای هر خانه

جعبه لنگر ۱



$$y_{a1} = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ C_1 \\ C_2 \\ C_3 \end{bmatrix}$$

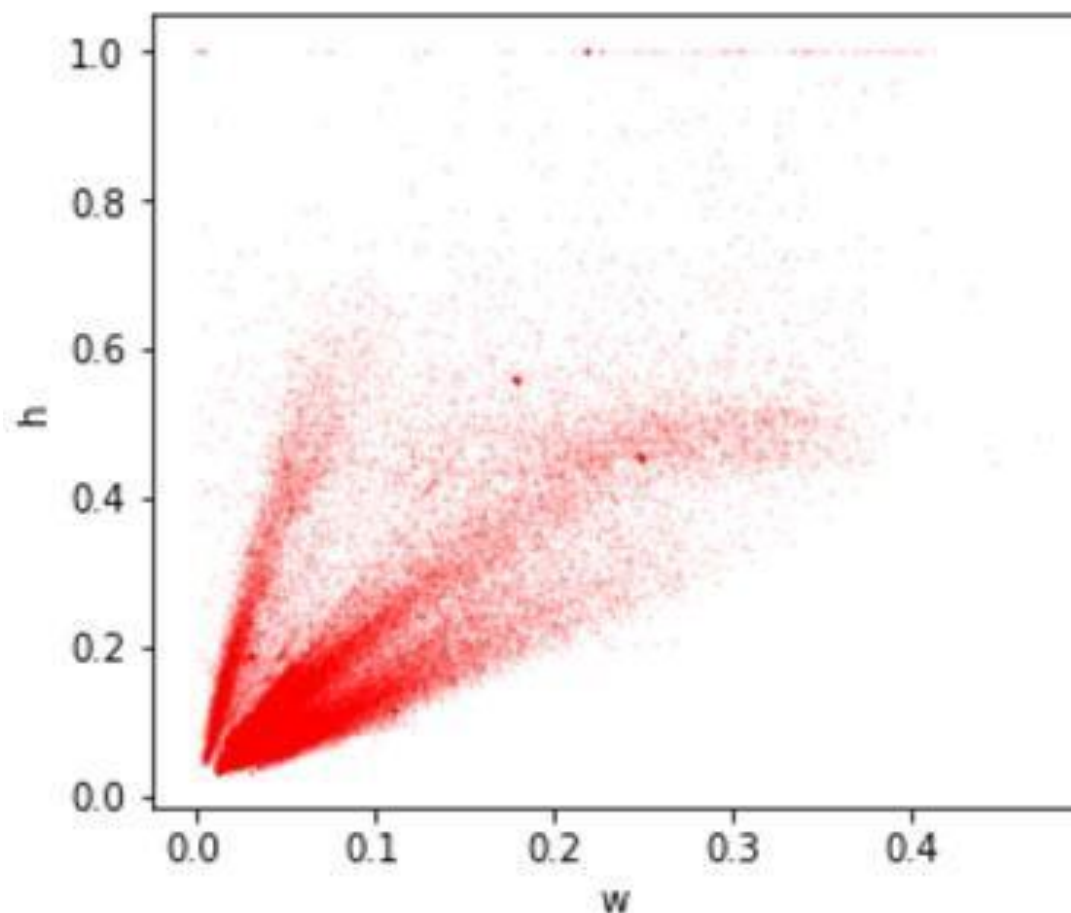
جعبه لنگر ۲



$$y_{a2} = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ C_1 \\ C_2 \\ C_3 \end{bmatrix}$$

جعبه‌های لنگر

۳۸



□ انتخاب جعبه‌های لنگر

□ طراحی جعبه‌های لنگر به صورت دستی

□ استفاده از الگوریتم خوشه‌بندی [۵ خوشه]

$$D(B_1, B_2) = 1 - IoU(B_1, B_2)$$

معیار فاصله

الگوریتم YOLO

۳۹

□ ایجاد مجموعه آموزشی.

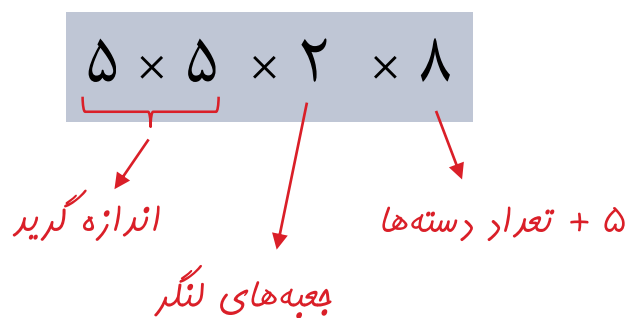
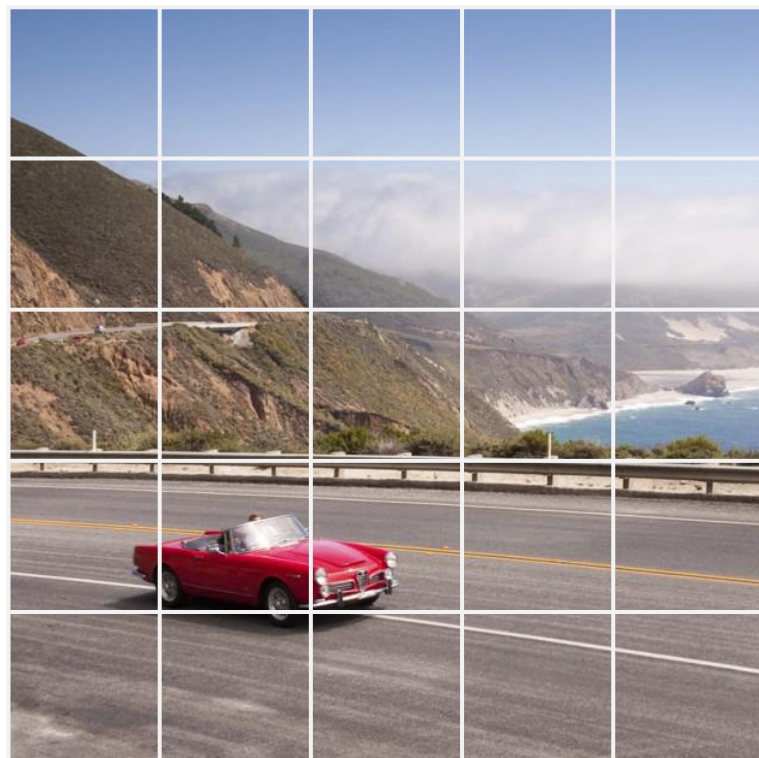
□ دسته‌ها.

۱ - رهگذر

۲ - خودرو

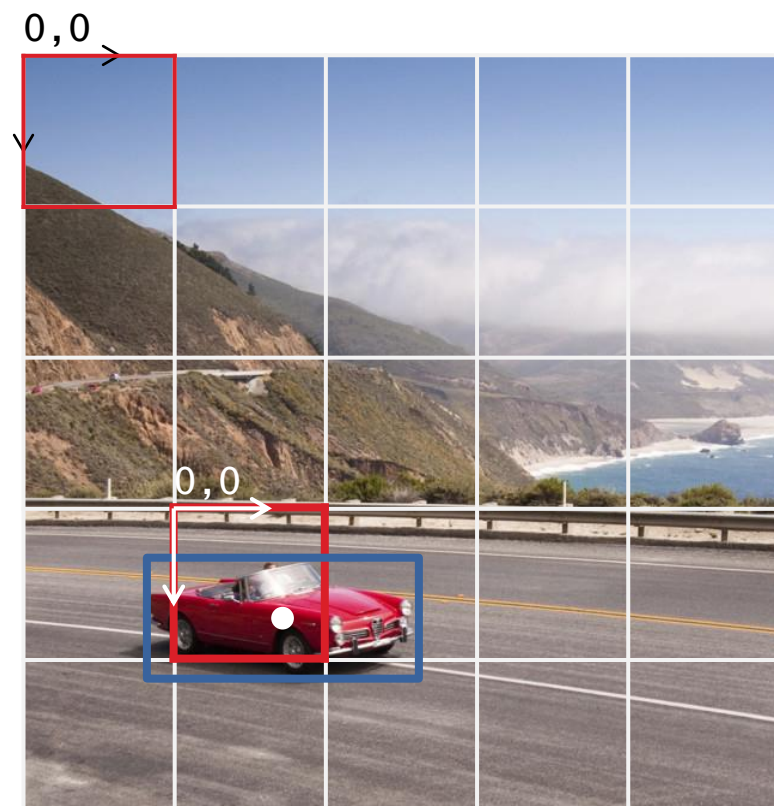
۳ - موتورسیکلت

□ خروجی.



الگوریتم YOLO

۴۰



خانه ۱

$p_c = 0$
 $b_x = 0$
 $b_y = 0$
 $b_h = 0$
 $b_w = 0$
 $C_1 = 0$
 $C_2 = 0$
 $C_3 = 0$
 $p_c = 0$
 $b_x = 0$
 $b_y = 0$
 $b_h = 0$
 $b_w = 0$
 $C_1 = 0$
 $C_2 = 0$
 $C_3 = 0$

خانه ۱۷

$p_c = 0$
 $b_x = 0$
 $b_y = 0$
 $b_h = 0$
 $b_w = 0$
 $C_1 = 0$
 $C_2 = 0$
 $C_3 = 0$
 $p_c = 1$
 $b_x = 0.7$
 $b_y = 0.7$
 $b_h = 0.8$
 $b_w = 1.9$
 $C_1 = 0$
 $C_2 = 1$
 $C_3 = 0$

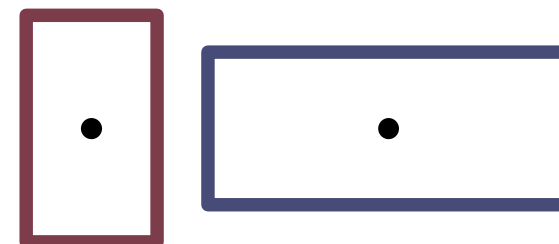
□ ایجاد مجموعه آموزشی.

□ دسته‌ها.

۱ - رهگذر

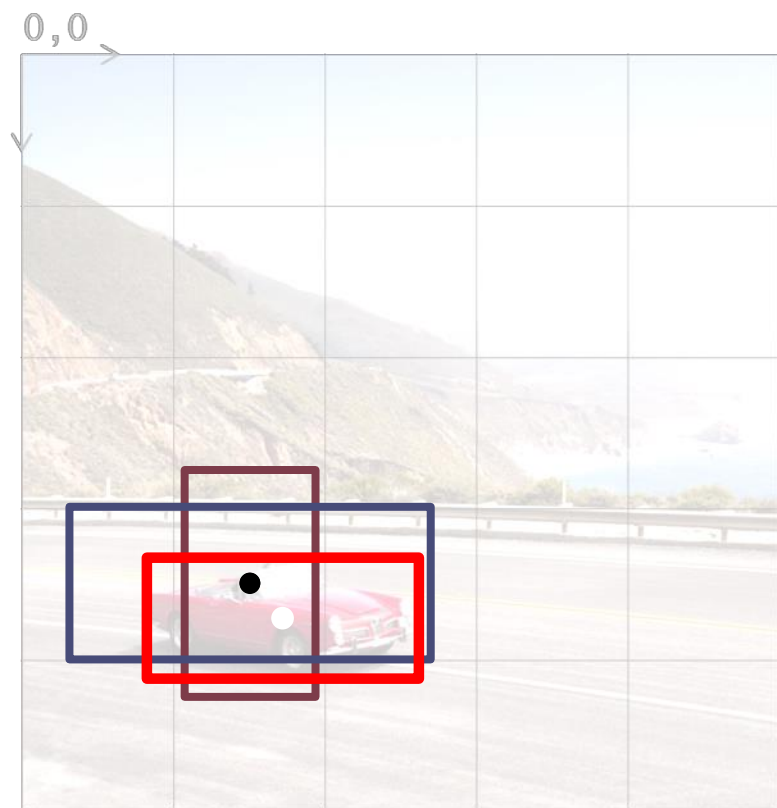
۲ - خودرو

۳ - موتورسیکلت



الگوریتم YOLO

۴۱



۱۹ × ۱۹ × ۵ × ۱

خانه ۱

$p_c = 0$
 $b_x = 0$
 $b_y = 0$
 $b_h = 0$
 $b_w = 0$
 $C_1 = 0$
 $C_2 = 0$
 $C_3 = 0$
 $p_c = 0$
 $b_x = 0$
 $b_y = 0$
 $b_h = 0$
 $b_w = 0$
 $C_1 = 0$
 $C_2 = 0$
 $C_3 = 0$

خانه ۱۷

$p_c = 0$
 $b_x = 0$
 $b_y = 0$
 $b_h = 0$
 $b_w = 0$
 $C_1 = 0$
 $C_2 = 0$
 $C_3 = 0$
 $p_c = 1$
 $b_x = 0.7$
 $b_y = 0.7$
 $b_h = 0.8$
 $b_w = 1.9$
 $C_1 = 0$
 $C_2 = 1$
 $C_3 = 0$

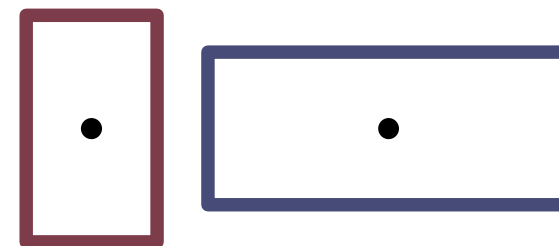
□ ایجاد مجموعه آموزشی.

□ دسته‌ها.

۱ - رهگذر

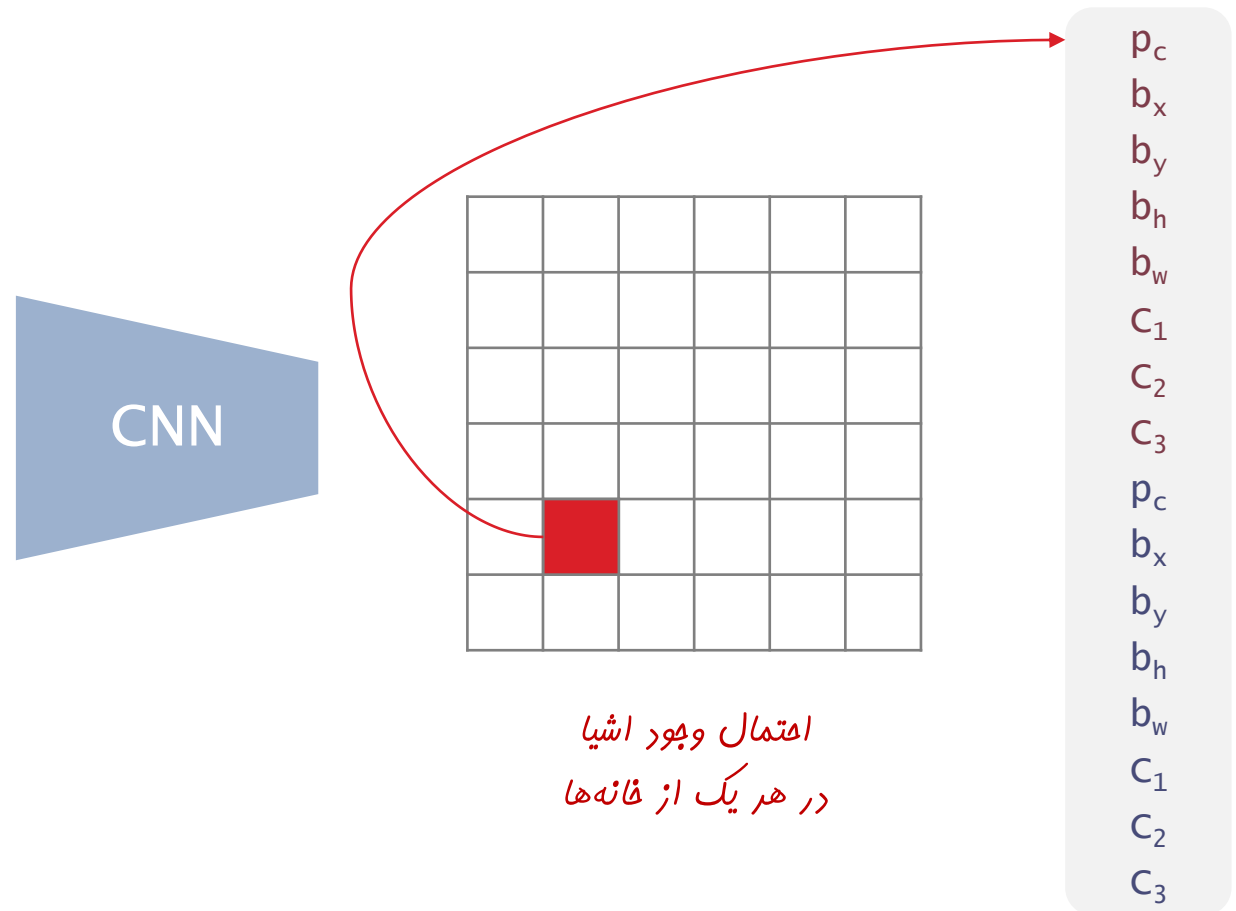
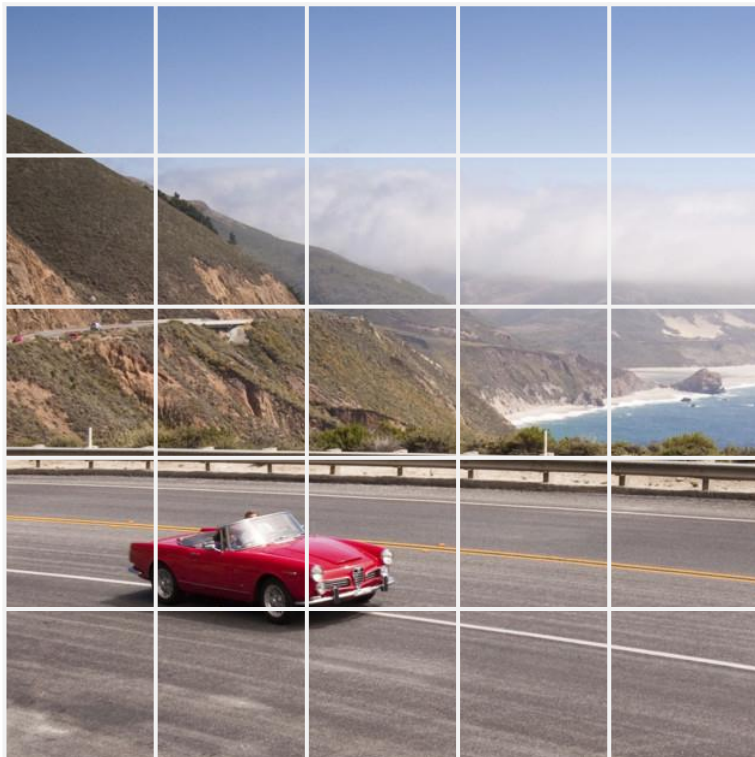
۲ - خودرو

۳ - موتورسیکلت



الگوریتم YOLO: پیش‌بینی

۴۲



الگوریتم YOLO: سرکوب غیر بیشینه

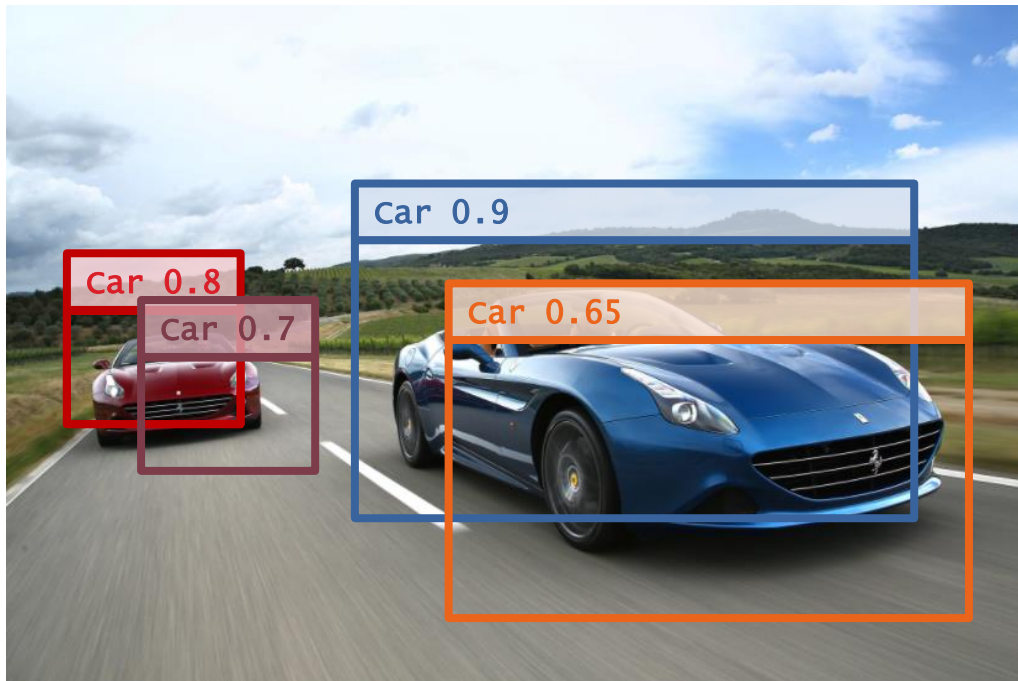
۴۳

□ پیش‌بینی‌هایی را که احتمال کمی دارند، حذف کن.

□ برای هر دسته [ره‌گذر، خودرو، موتورسیکلت]:

□ تا زمانی که پیش‌بینی دیگری وجود دارد:

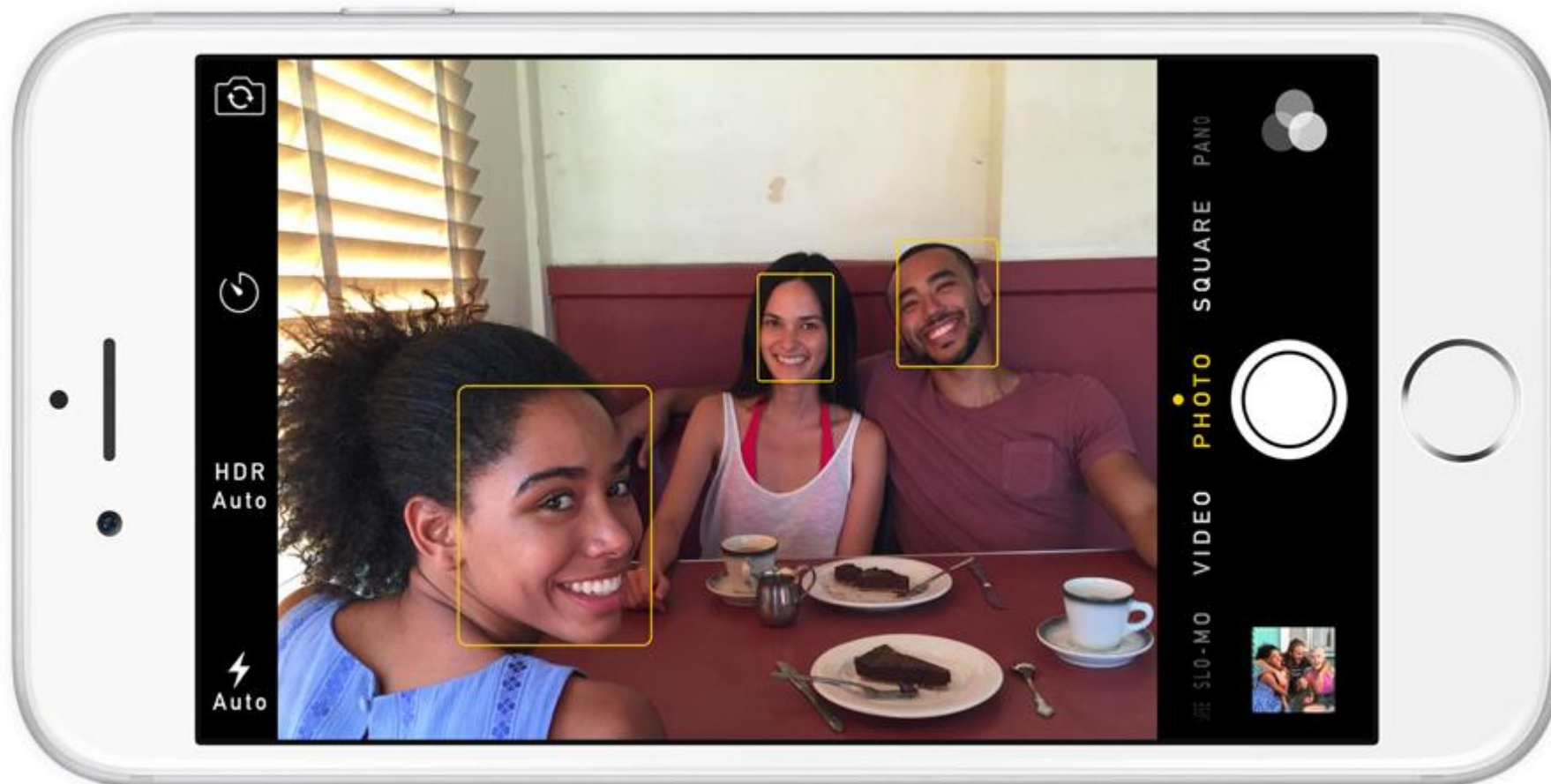
- پیش‌بینی با بیشترین احتمال را انتخاب کن.
- این پیش‌بینی را به عنوان خروجی در نظر بگیر.
- تمام پیش‌بینی‌هایی را که هم‌پوشانی بالایی با پیش‌بینی انتخاب شده دارند، حذف کن.



پیشنهادهای نامیه: الگوریتم R-CNN

تشخیص چهره

۴۵



یادگیری تک شات

۴۶

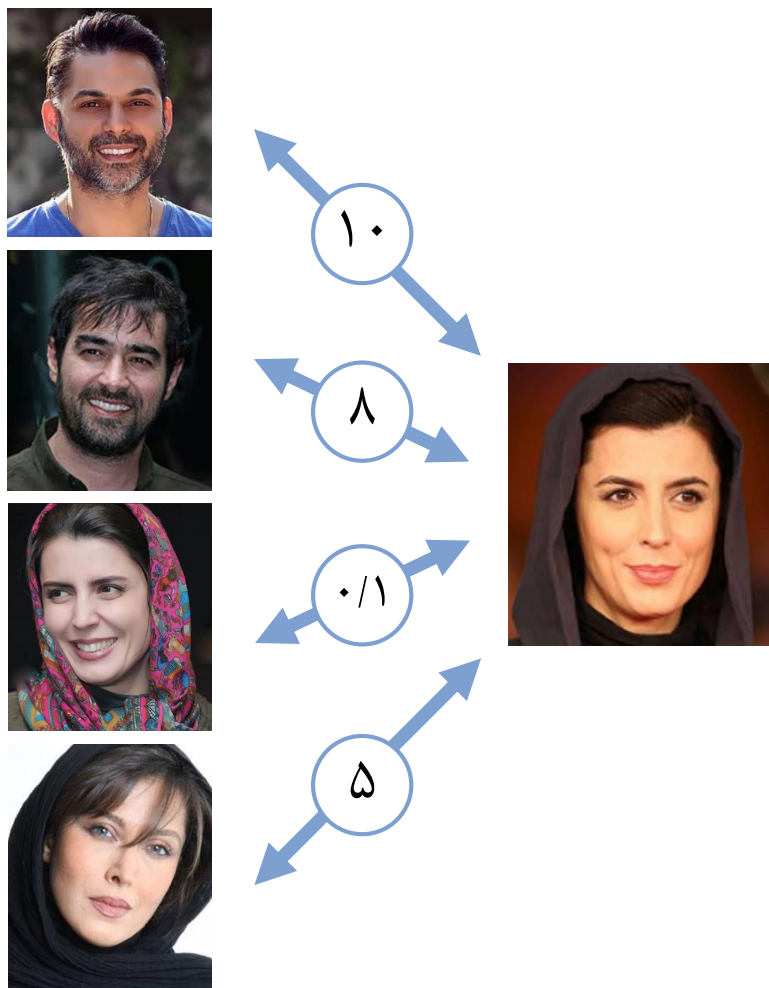
□ یادگیری از روی یک نمونه آموزشی [یا یک تعداد بسیار کم از نمونه‌های آموزشی].



یادگیری معیار شباهت

۴۷

□ تابع فاصله.



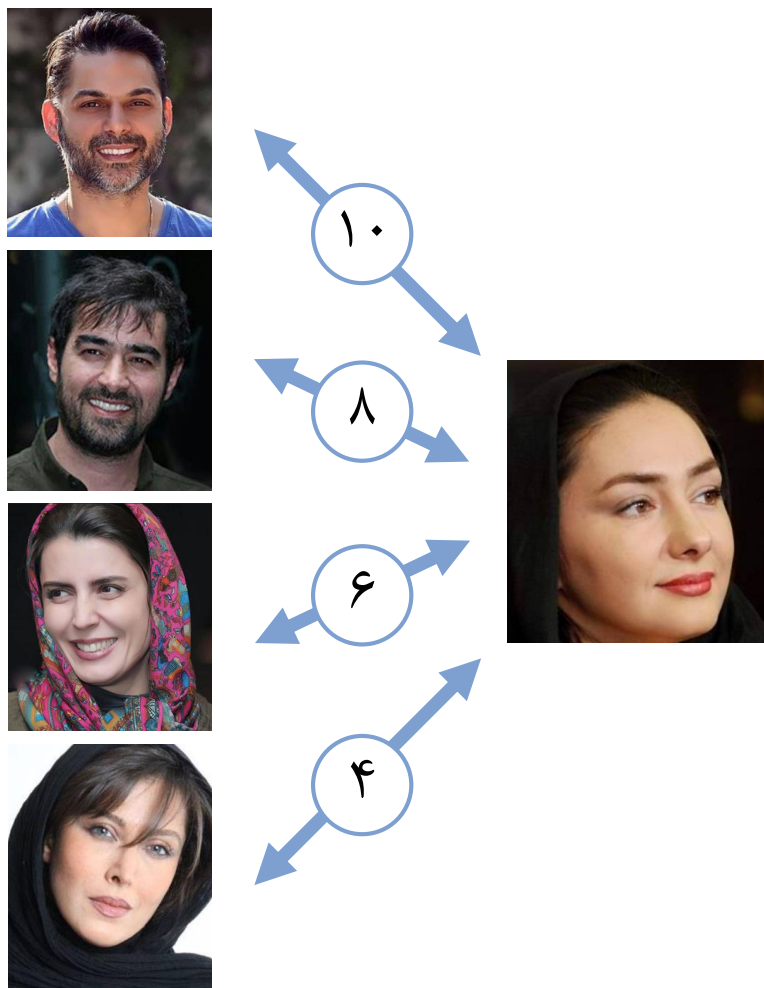
$$d(\text{Image 1}, \text{Image 2}) \leq \tau$$

$$d(\text{Image 1}, \text{Image 3}) > \tau$$

یادگیری معیار شباهت

۴۸

□ تابع فاصله.



$$d(\text{Image 1}, \text{Image 2}) \leq \tau$$

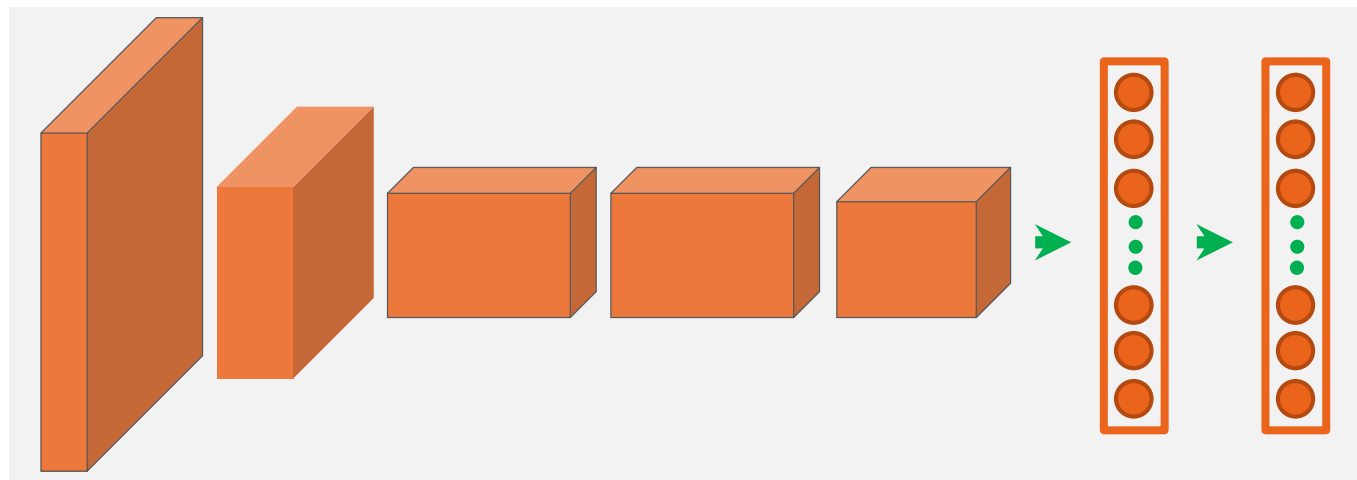
$$d(\text{Image 1}, \text{Image 3}) > \tau$$

شبکه سیامی

۴۹



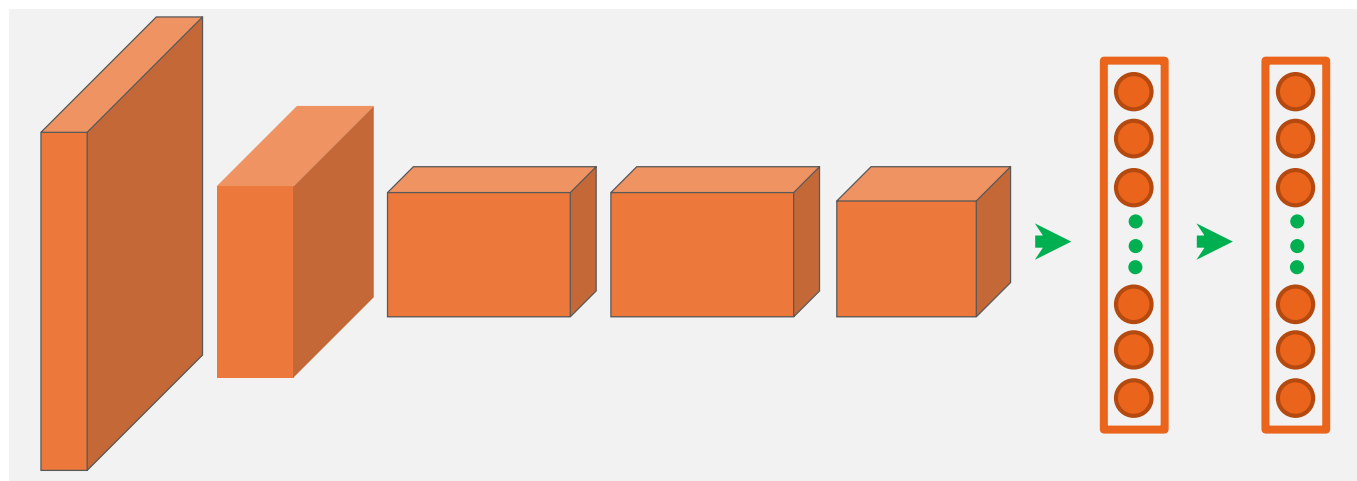
$x^{(1)}$



بدرار ویژگی [۱۲۸ بعد]
 $\rightarrow f(x^{(1)})$



$x^{(2)}$



$\rightarrow f(x^{(2)})$

شبکه سیامی: تابع فاصله

۵۰

□ محاسبه فاصله دو تصویر $x^{(i)}$ و $x^{(j)}$

$$d(x^{(i)}, x^{(j)}) = \|f(x^{(i)}) - f(x^{(j)})\|_2^2$$

□ آموزش پارامترهای شبکه کانولوشنی.

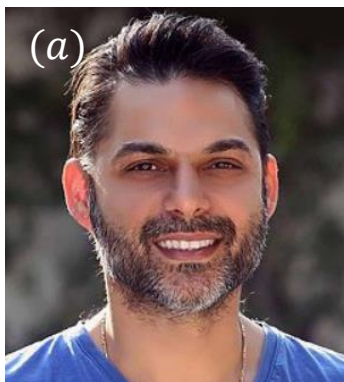
اگر دو تصویر $x^{(i)}$ و $x^{(j)}$ متعلق به یک شخص باشند $\|f(x^{(i)}) - f(x^{(j)})\|^2 \rightarrow 0$

اگر دو تصویر $x^{(i)}$ و $x^{(j)}$ متعلق به یک شخص نباشند $\|f(x^{(i)}) - f(x^{(j)})\|^2 \rightarrow \infty$

شبکه سیامی: آموزش

۵۱

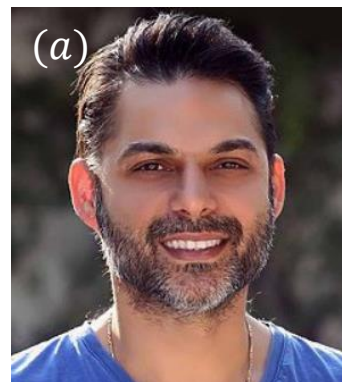
□ تابع هزینه سه گانه.



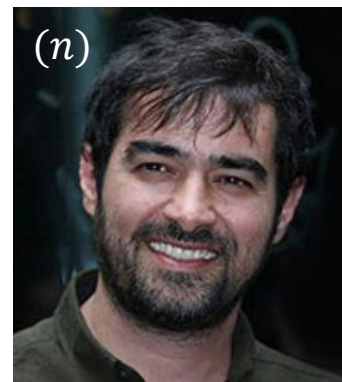
لنگر



مثبت



لنگر



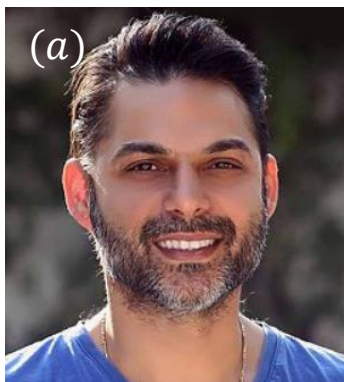
منفی

$$\|f(a) - f(p)\|^2 \leq \|f(a) - f(n)\|^2$$

شبکه سیامی: آموزش

۵۲

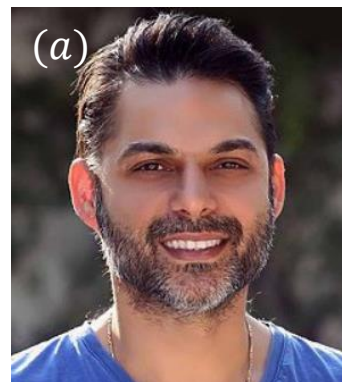
□ تابع هزینه سه گانه.



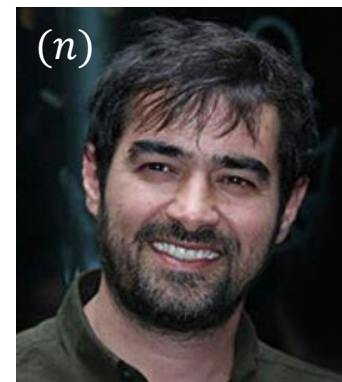
لنگر



مثبت



لنگر



منفی

$$\|f(a) - f(p)\|^2 - \|f(a) - f(n)\|^2 + \alpha \leq 0$$

اجتناب از پاسخ های بدیهی!

شبکه سیامی: آموزش

۵۳

□ تابع هزینه سه گانه.

$$L(a, p, n) = \max(\|f(a) - f(p)\|^2 - \|f(a) - f(n)\|^2 + \alpha, 0)$$

□ انتخاب یک مجموعه از سه تایی های مناسب برای مجموعه آموزشی.

□ انتخاب سه تایی هایی که یادگیری آنها برای شبکه عصبی **دشوار** است!

$$d(a, p) \cong d(a, n)$$

شبکه سیامی: آموزش

۵۴

لنجر



مثبت



منفی



⋮



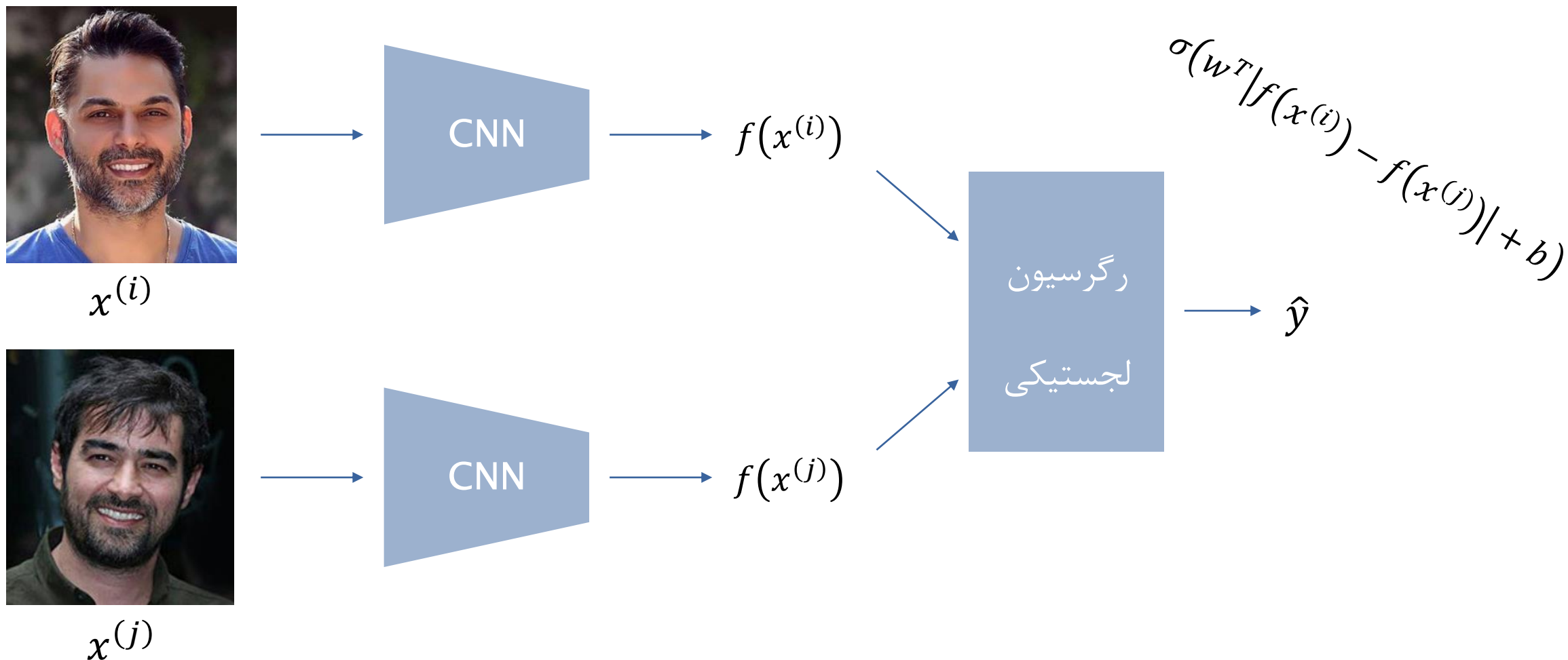
⋮



⋮

تابع هزینه دودویی

۵۵



چالش شناسایی نهنگ (کگل)

۵۶



- تعداد ۹۸۵۰ تصویر
- بیش از ۳۰۰۰ نهنگ



انتقال سبک

۵۷

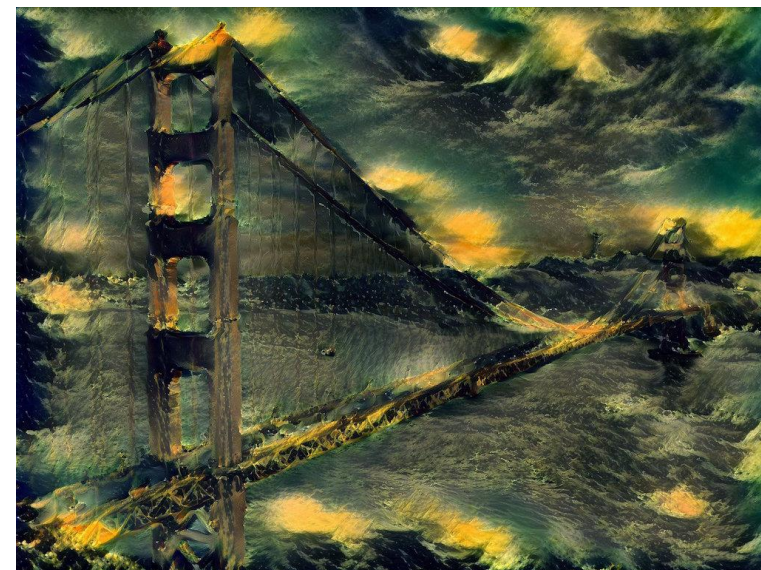
محتوا



سبک



تصویر تولید شده



تصویرسازی یادگیری شبکه‌های عمیق

۵۸

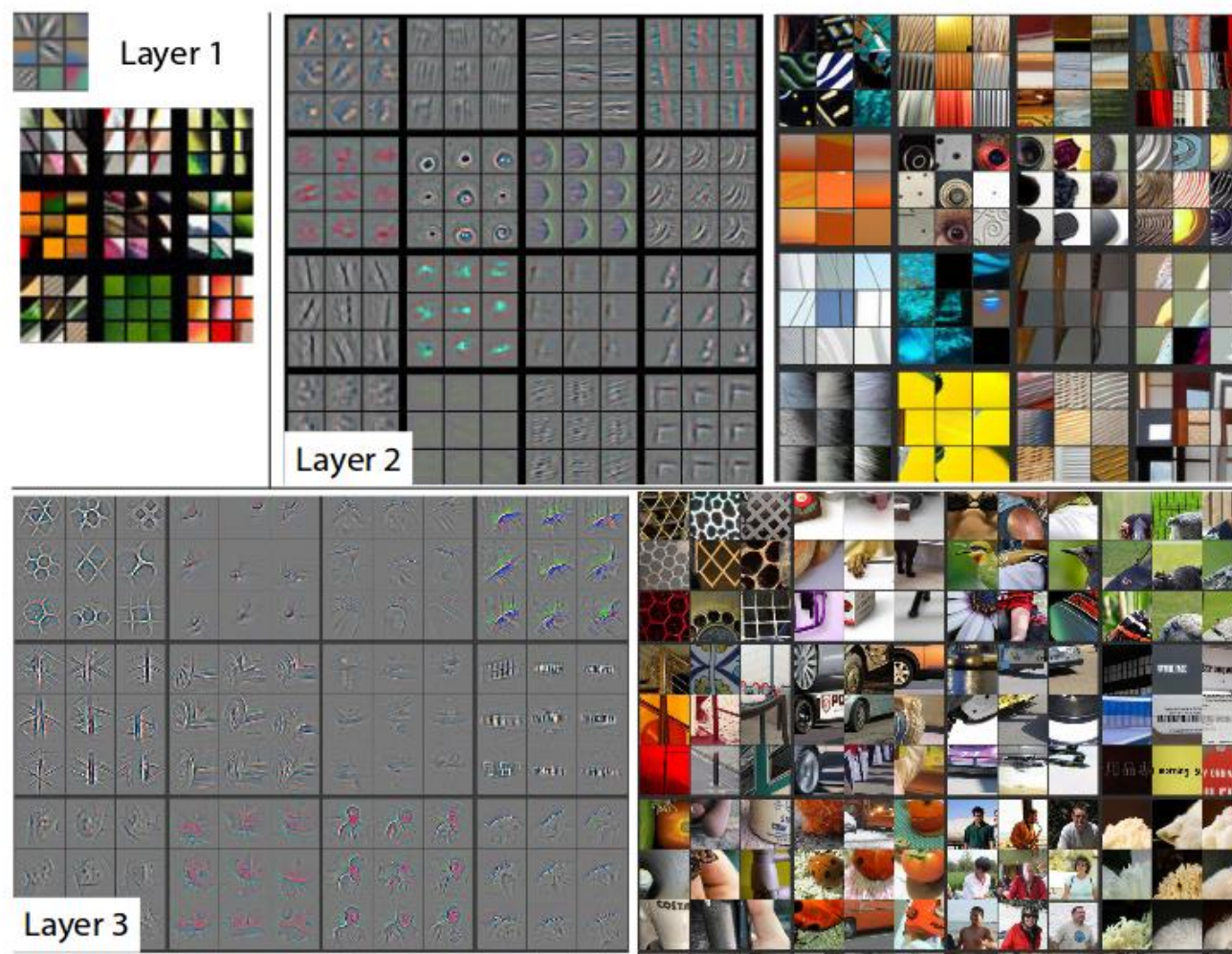
□ لایه کانولوشن اول.

۱. انتخاب یک نورون در لایه اول

۲. یافتن بخش‌هایی از تصویر که فعالیت این نورون را بیشینه می‌سازد.

تصویرسازی یادگیری شبکه‌های عمیق

۵۹



تابع هزینه

۶۰

- تابع هزینه محتوا
- تابع هزینه سبک

$$L(C, S, G) = \alpha L_{content}(C, G) + \beta L_{style}(S, G)$$

انتقال سبک: الگوریتم

۶۱

- ایجاد G به صورت تصادفی
- استفاده از الگوریتم گرادیان کاهشی (!) برای کمینه‌سازی تابع هزینه

$$L(C, S, G) = \alpha L_{content}(C, G) + \beta L_{style}(S, G)$$

- انتخاب لایه مخفی l
- استفاده از یک شبکه کانولوشنی آموزش یافته
- محاسبه خروجی (فعالیت) لایه مخفی برای تصاویر C و G
- اگر فعالیت لایه مخفی برای این دو تصویر تقریباً برابر باشد، آنگاه محتوای دو تصویر با یکدیگر مشابه است.

$$L_{content}(C, G) = \|a^{(l)}(C) - a^{(l)}(G)\|^2$$

تابع هزینه سبک

۶۳

$$L(C, S, G) = \alpha L_{content}(C, G) + \beta L_{style}(S, G)$$

□ انتخاب لایه مخفی l

□ تعریف سبک به عنوان همبستگی میان فعالیت‌ها بر روی کانال‌ها

□ محاسبه خروجی (فعالیت) لایه مخفی برای تصاویر S و G

$$a_{i,j,k}^{(l)}$$

$$G^{(l)}(S)$$

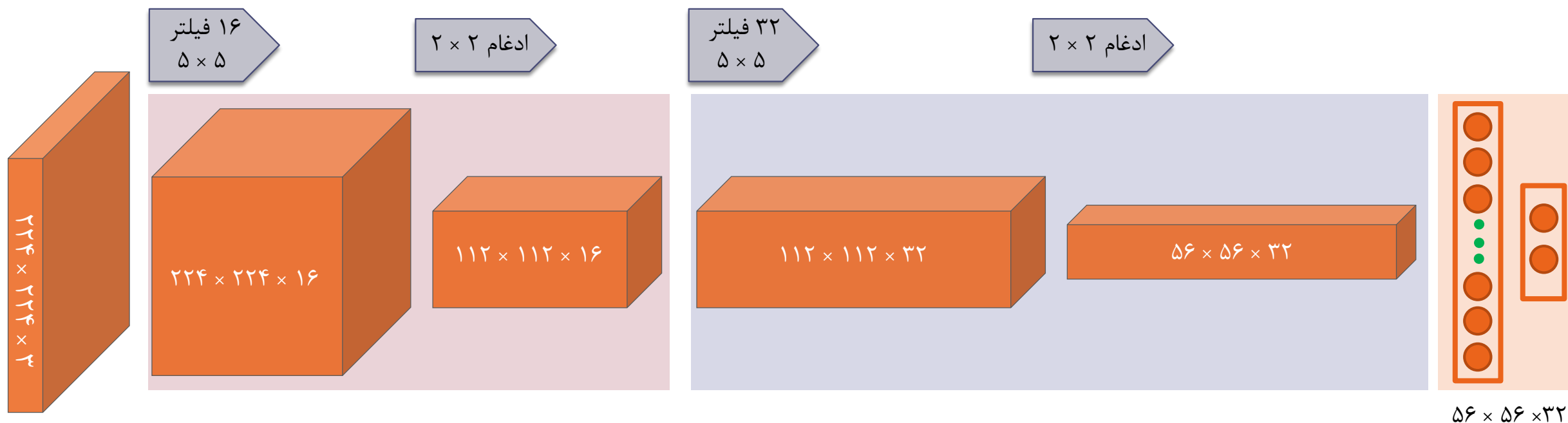
$$G_{kk'}^{(l)}(S) = \sum_{i=1}^{n_H^{(l)}} \sum_{j=1}^{n_W^{(l)}} a_{ijk}^{(l)} \cdot a_{ijk'}^{(l)}$$

$$G_{kk'}^{(l)}(G) = \sum_{i=1}^{n_H^{(l)}} \sum_{j=1}^{n_W^{(l)}} a_{ijk}^{(l)} \cdot a_{ijk'}^{(l)}$$

پروست: هفته اول

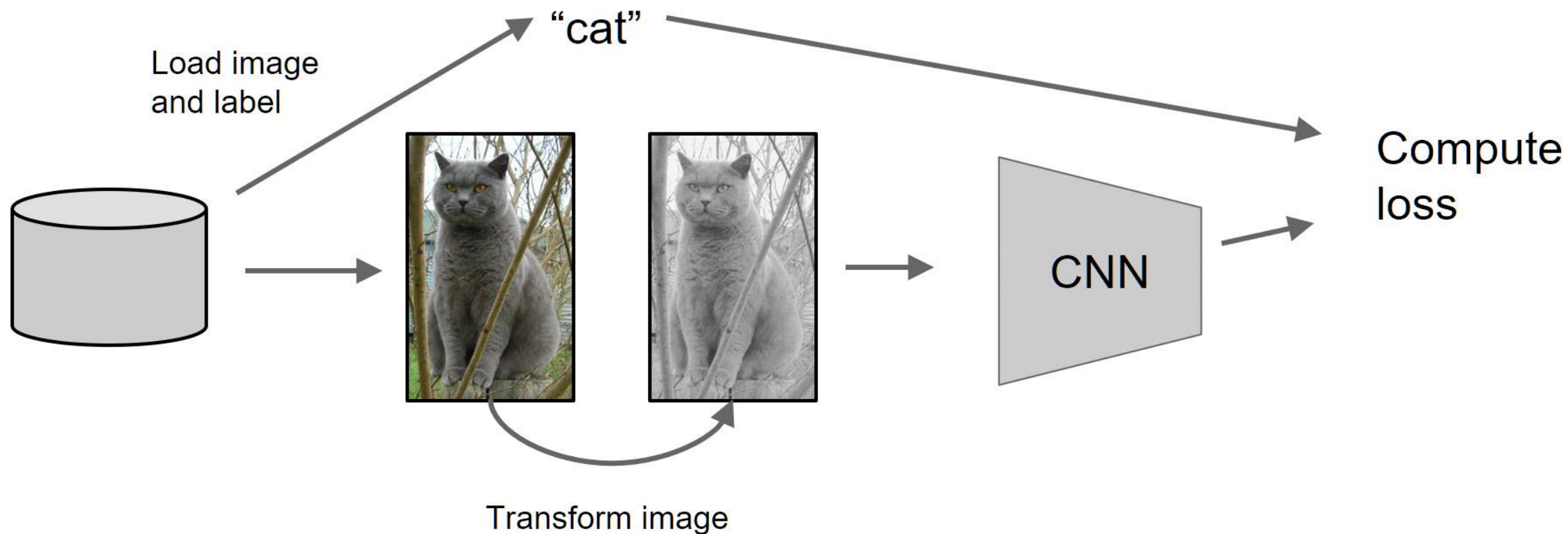
۶۴

□ شبکه کانولوشنی برای دسته‌بندی سگ و گربه



پایه: داده‌افزایی

۶۵



پیوست: داده‌افزایی

۶۶

