

شبکه‌های عصبی کانولوشن

سید ناصر رضوی
www.snrazavi.ir
۱۳۹۶

فهرست

۲

- 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

گوشهای مسائل در بینایی ماشین

۳

□ دسته‌بندی تصاویر.



هوایپما

فودرو

پرنده

گربه

گوزن

سگ

قریباغه

اسب

کشتی

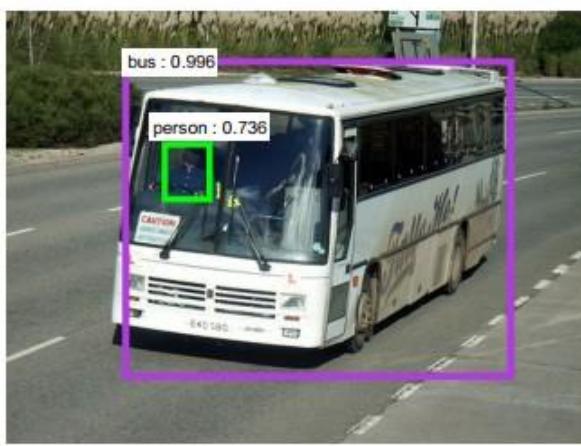
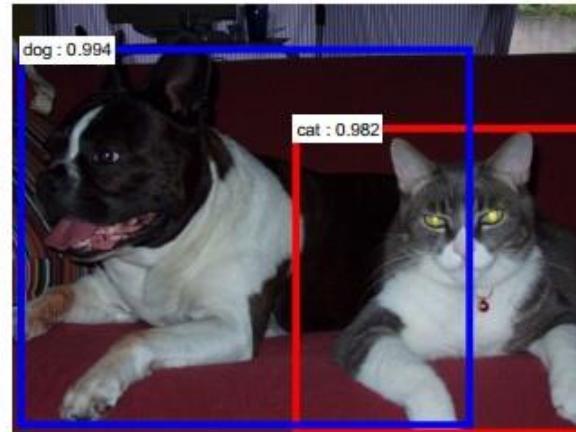
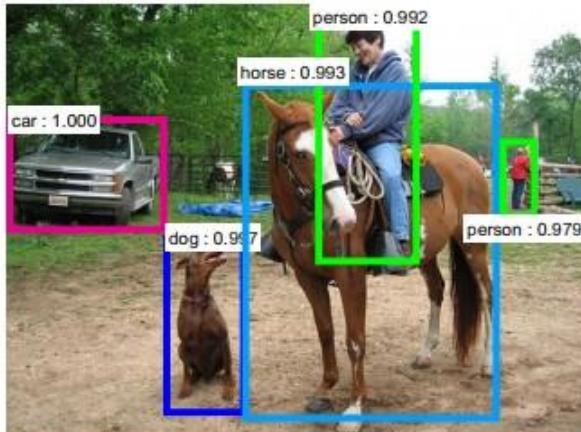
کامیون

۳۰۰ × ۱۰۰ × ۳

گونه‌های مسائل در بینایی ماشین

۴

□ شناسایی اشیا.



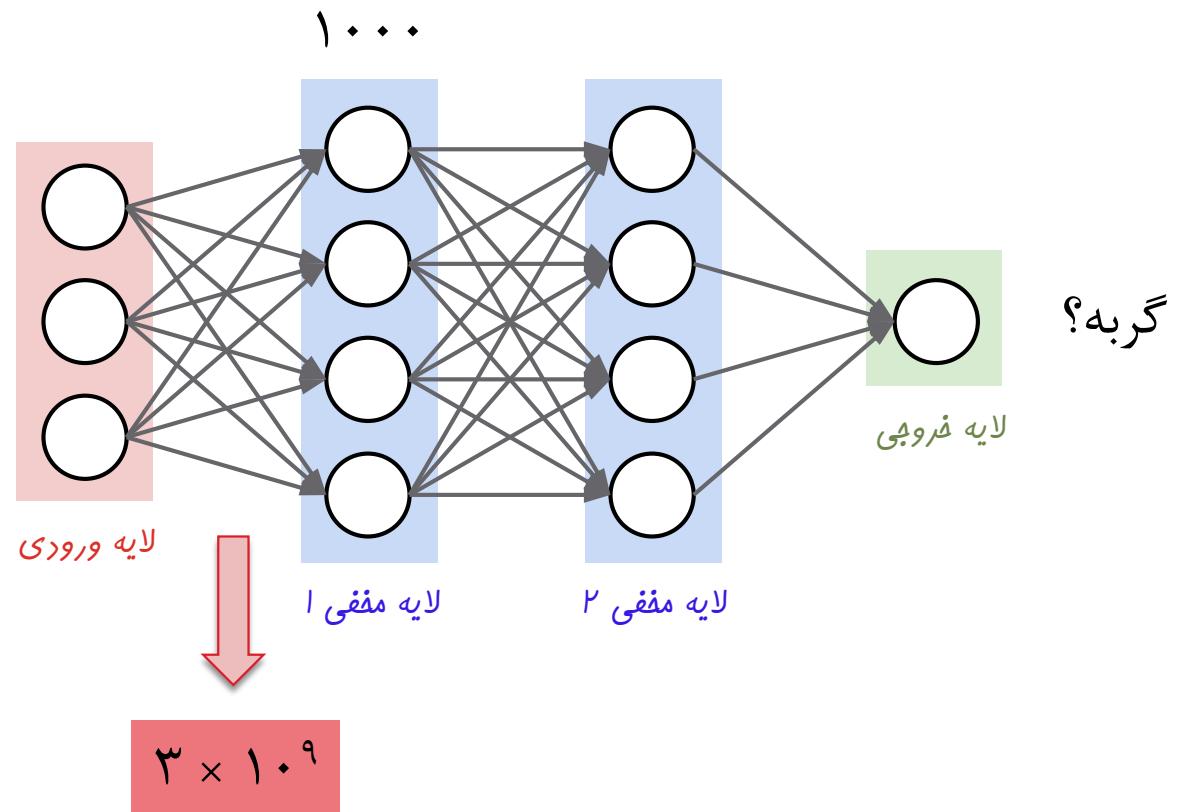
تصاویر بزرگ

۵

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

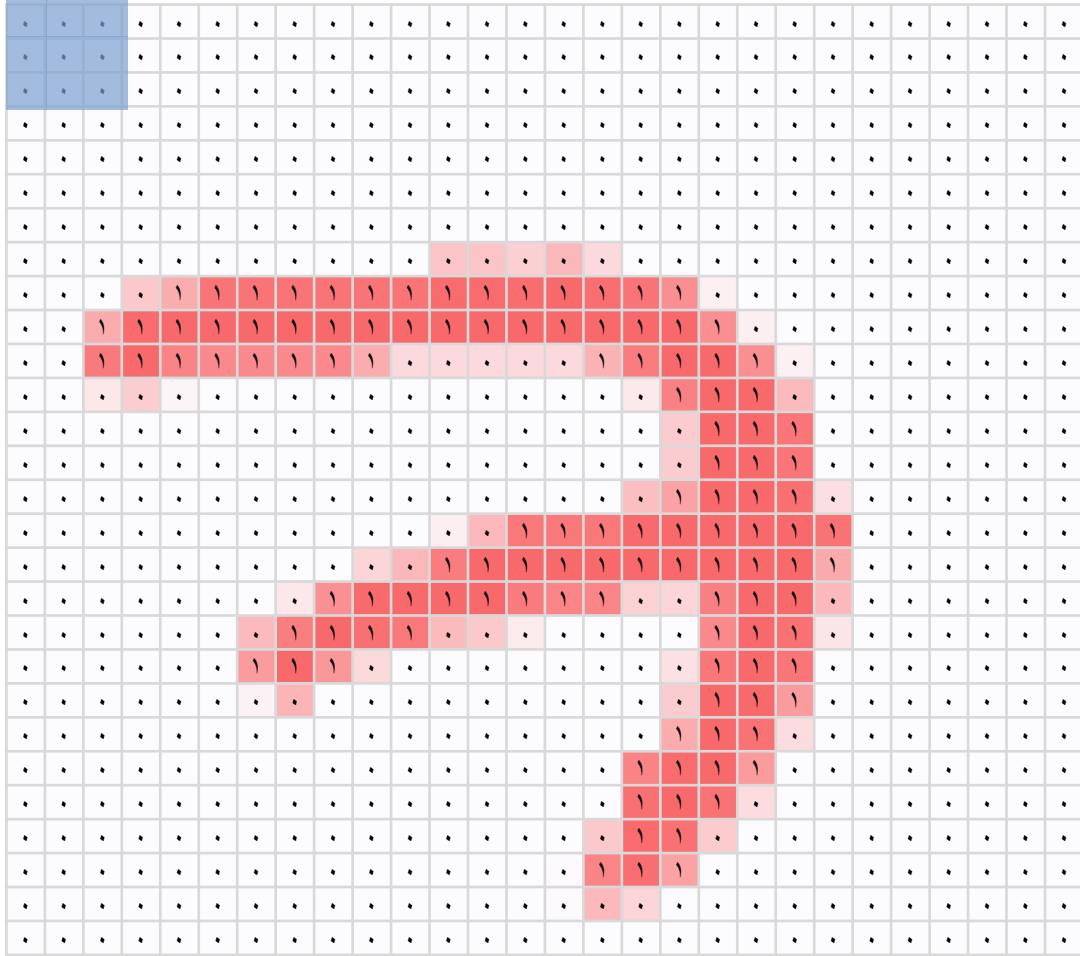


$$1000 \times 1000 \times 3$$

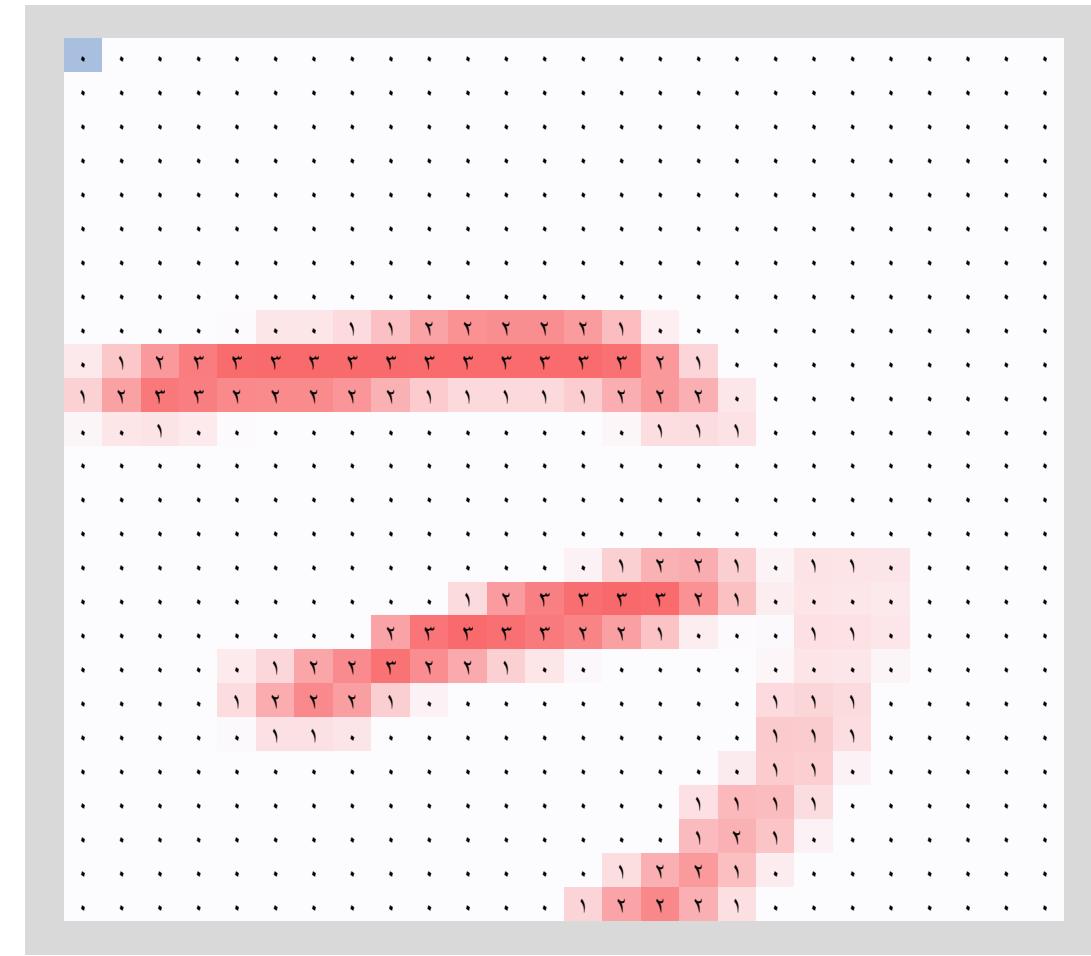


کانولوشن: فیلتر تشفیص لبه

۶

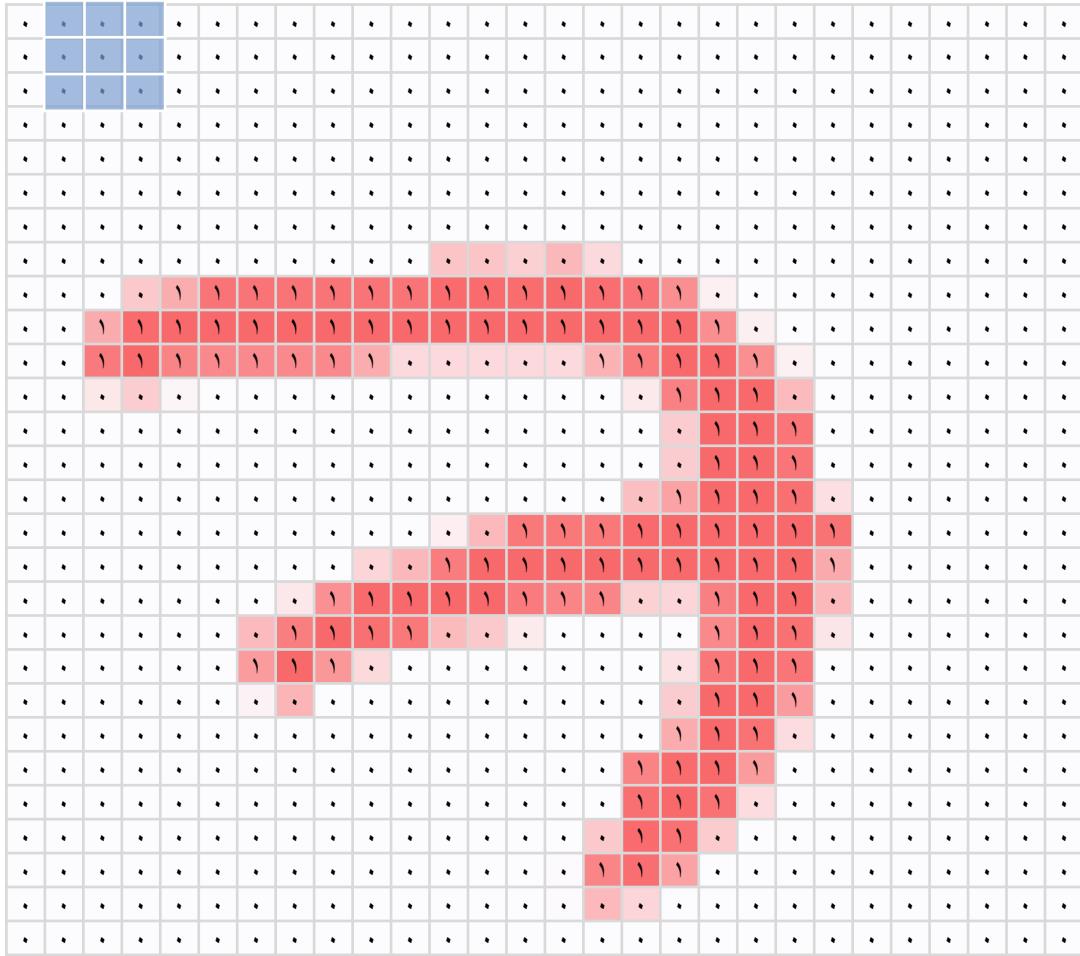


$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

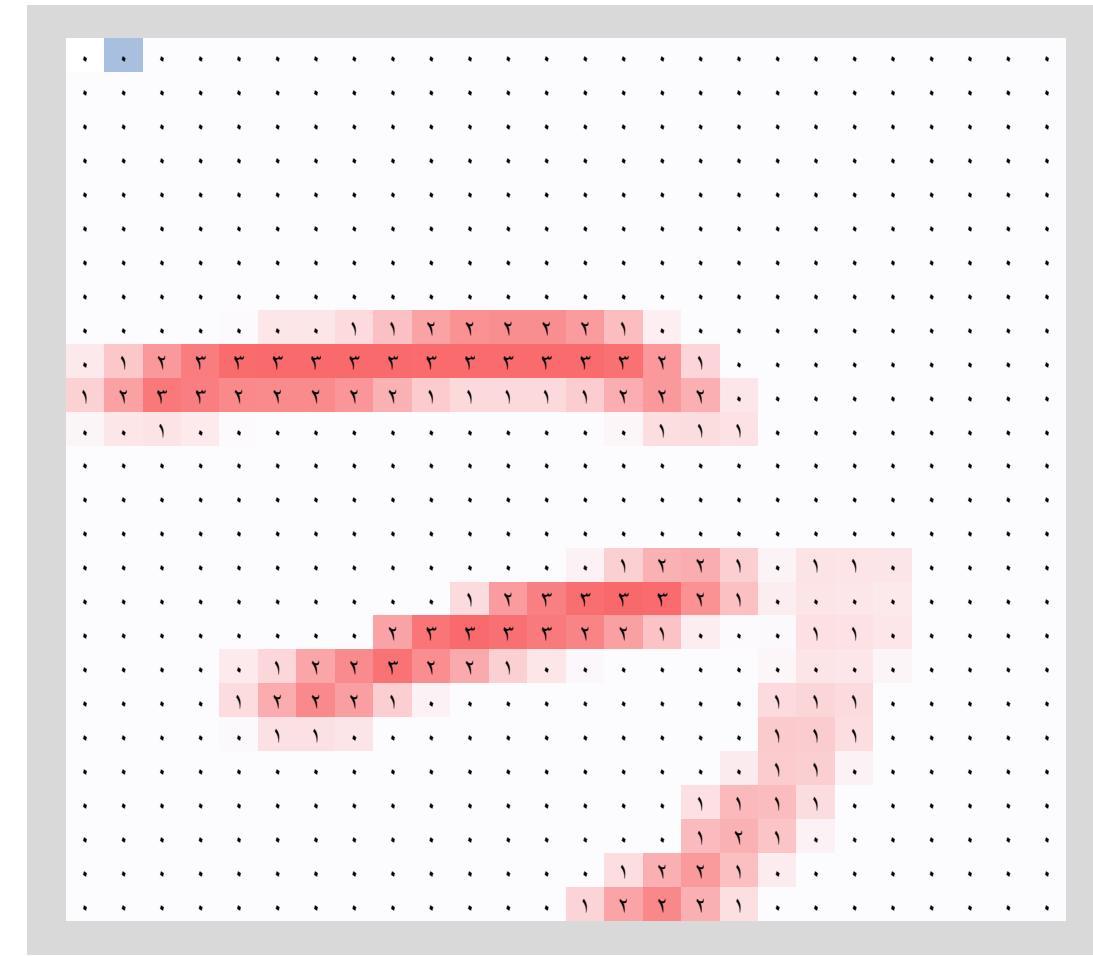


کانولوشن: فیلتر تشفیص لبه

۷

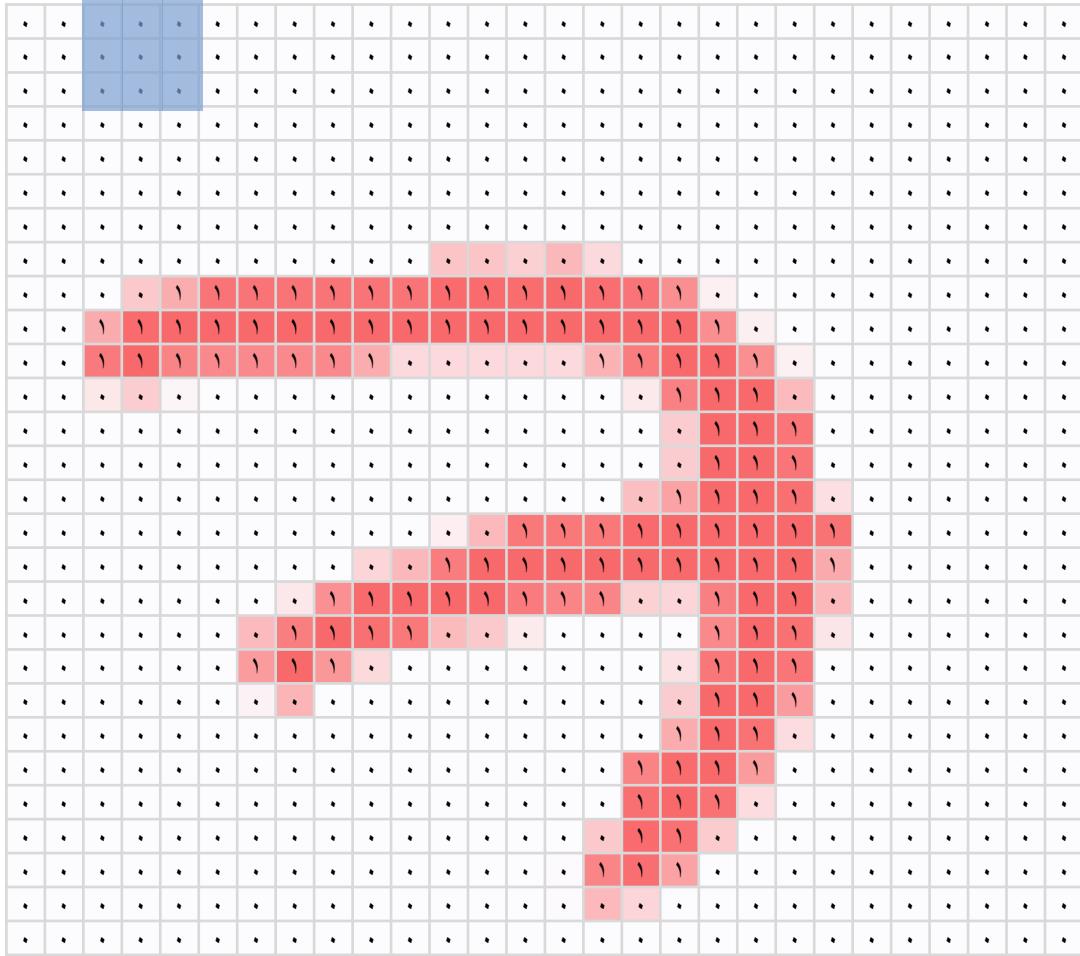


$$\begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{matrix}$$

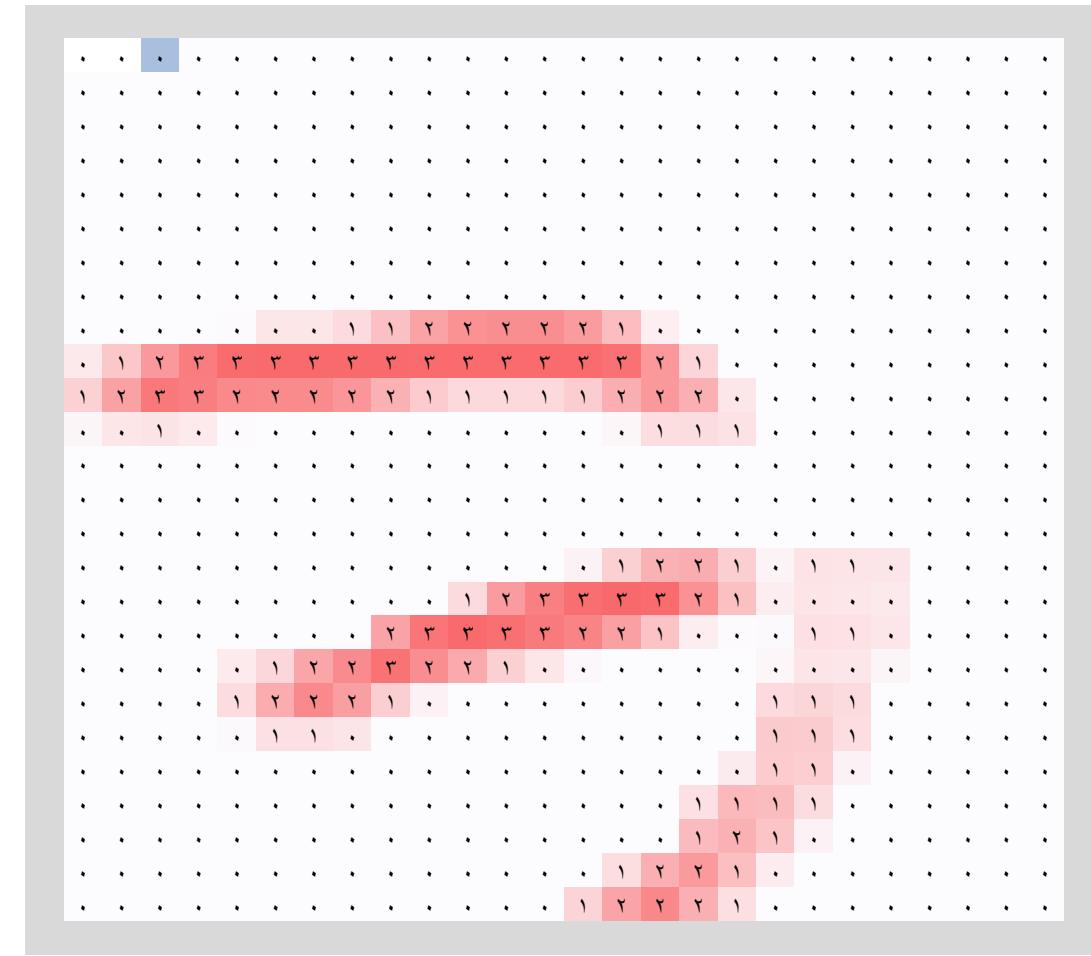


کانولوشن: فیلتر تشفیص لبه

۸

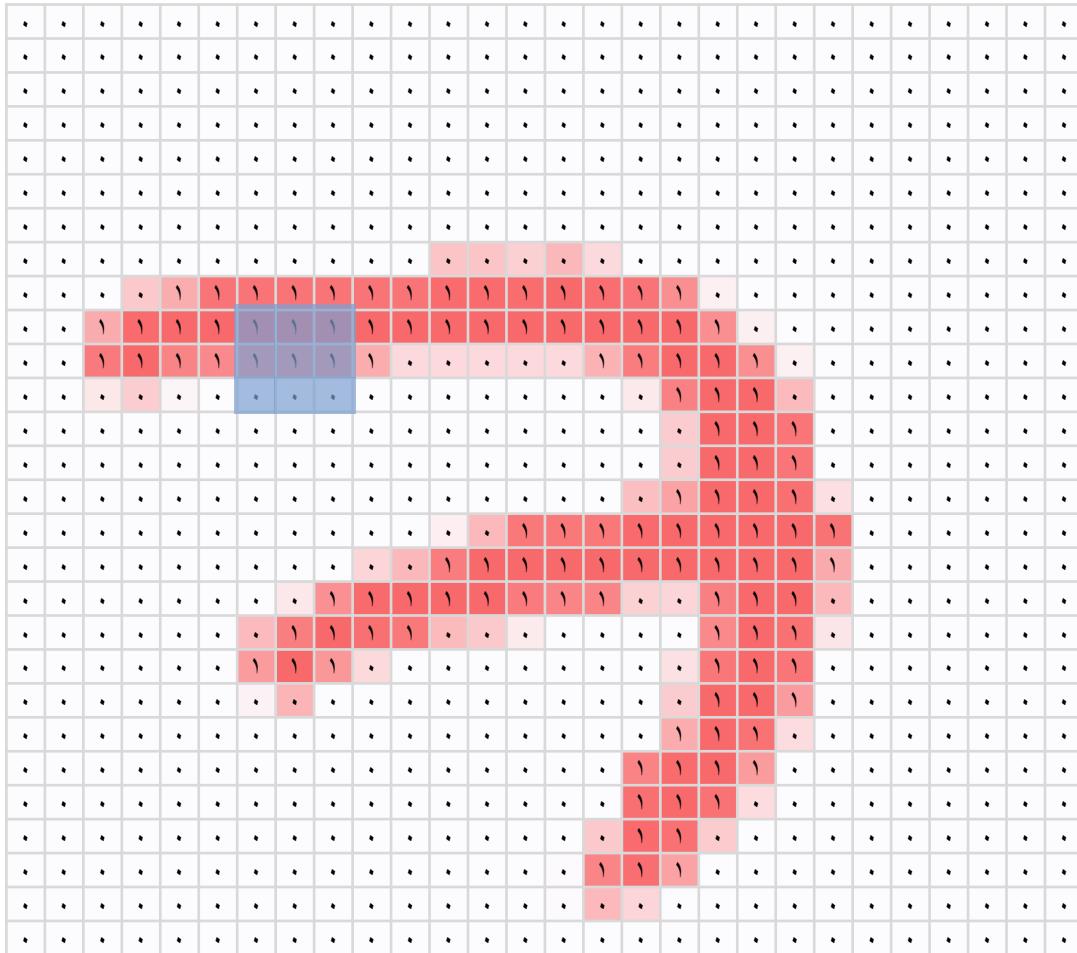


$$\begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{matrix}$$

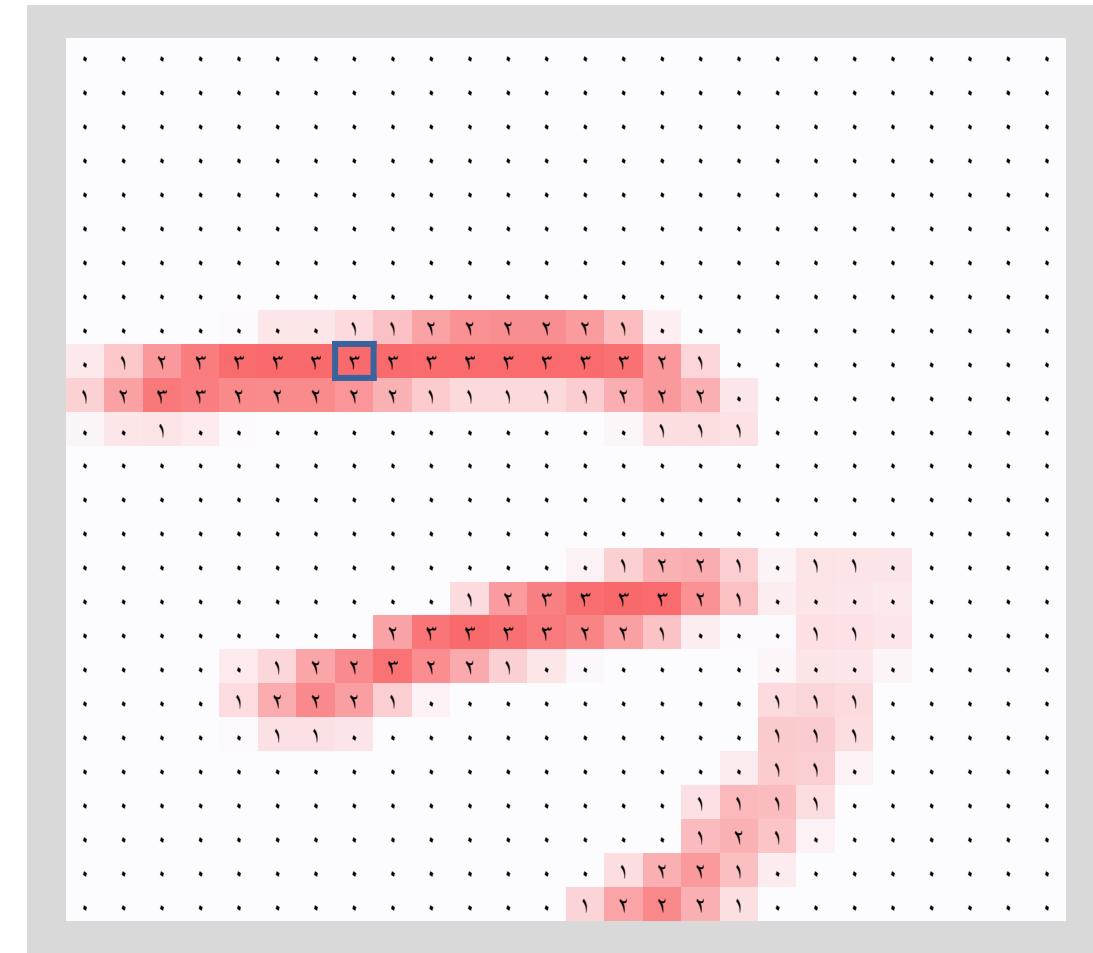


کانولوشن: فیلتر تشفیص لبه

۹

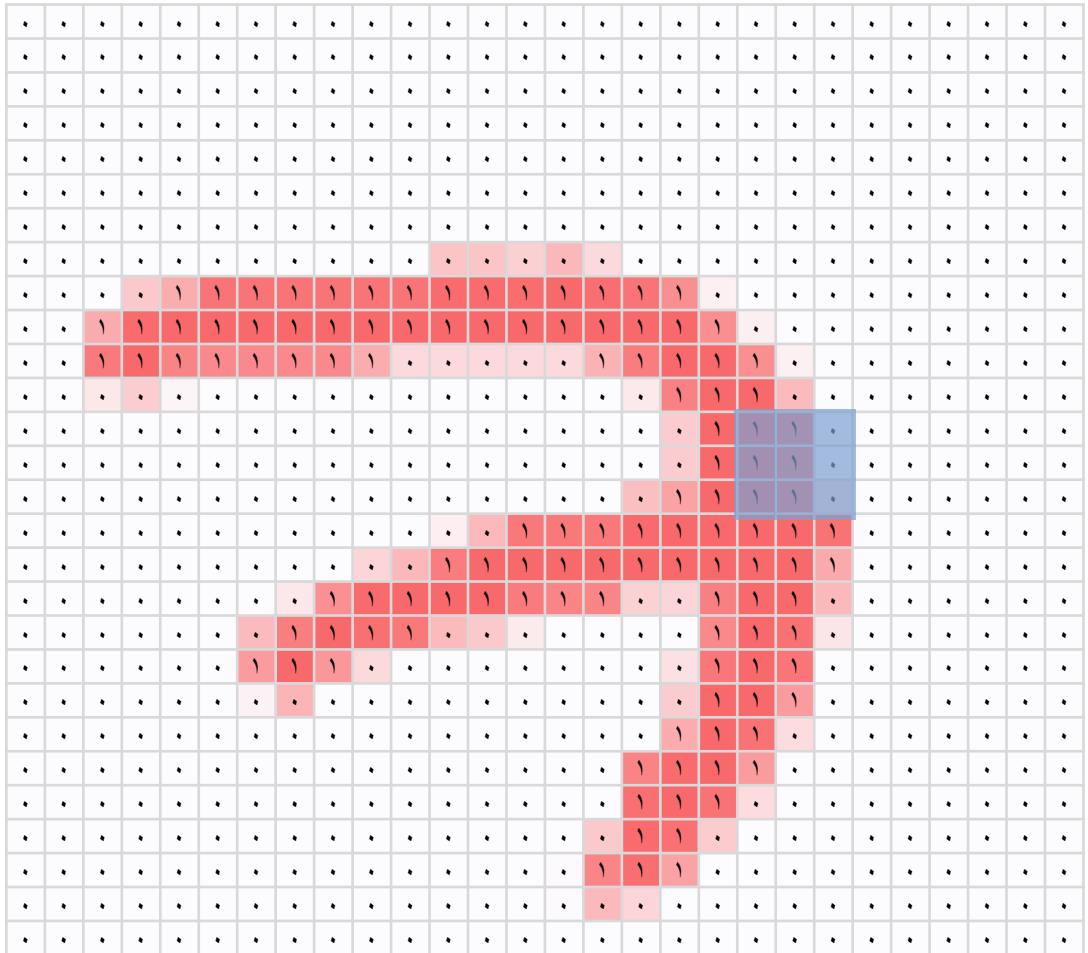


$$\begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{matrix}$$

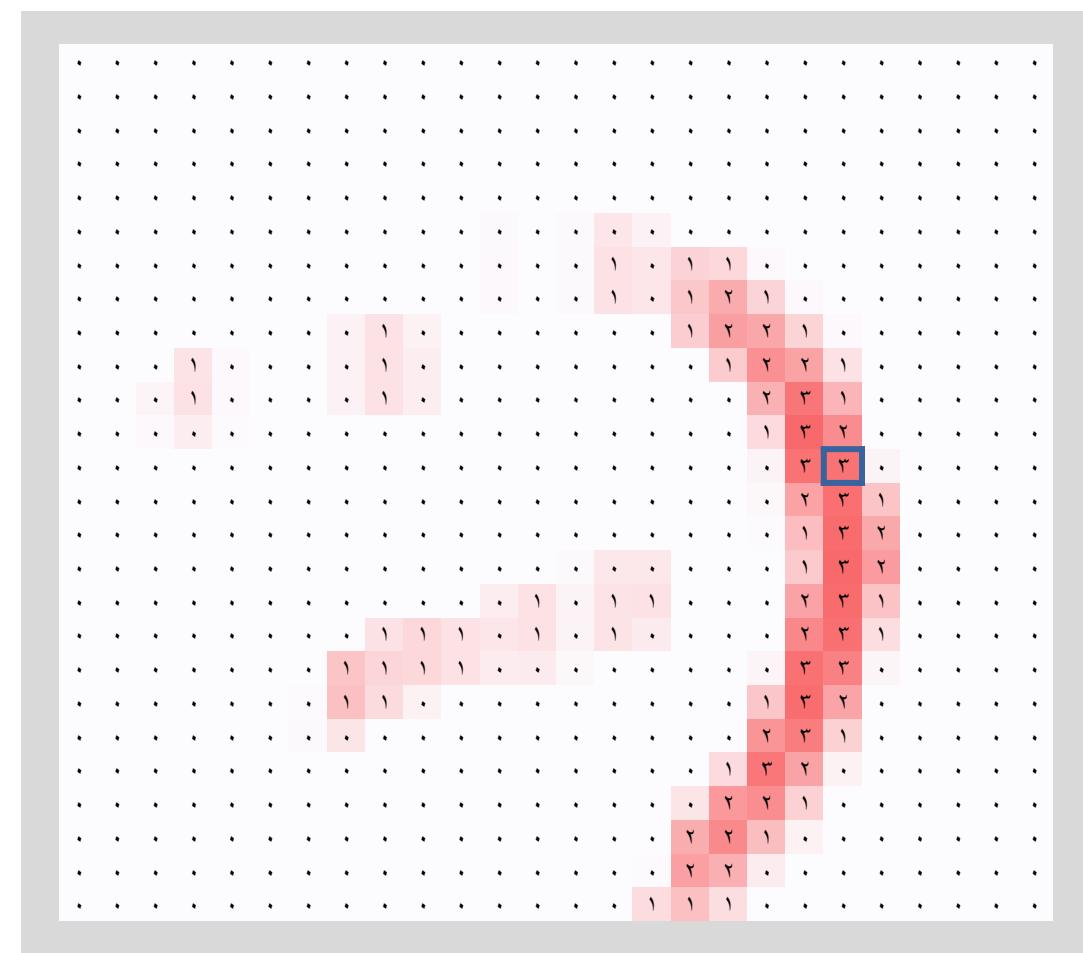


کانولوشن: تشخیص لبه‌های عمودی

۱۰

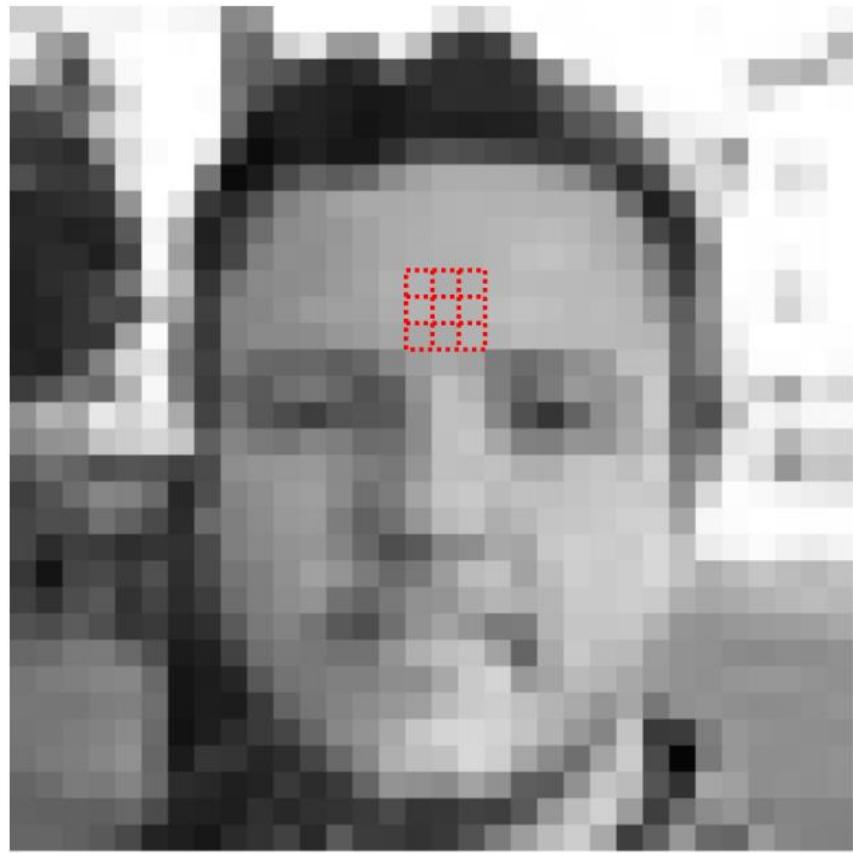


$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$



لایه کانولوشن: یک نگاه دقیق‌تر

۱۱

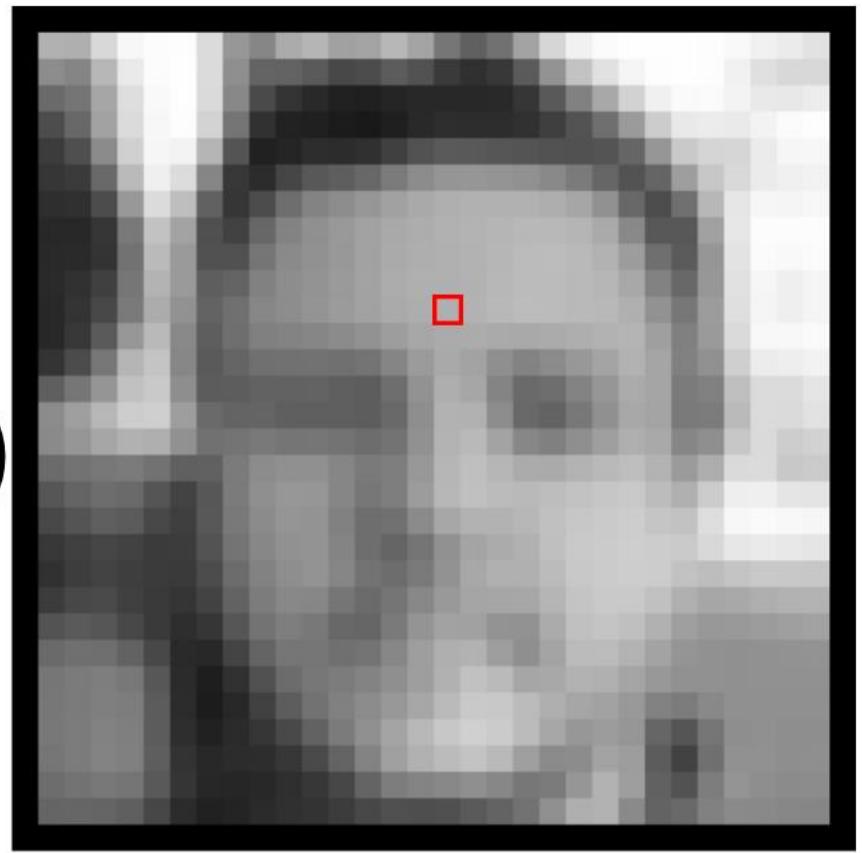


input image

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

kernel:

blur



output image

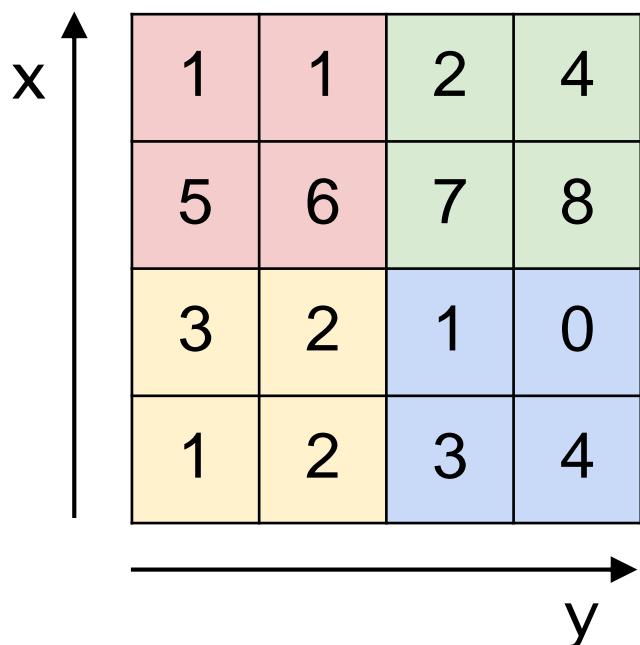
ادغام بیشینه

۱۲

□ لایه ادغام. کوچک‌تر کردن بازنمایی!

□ بر روی هر نقشه فعالیت به صورت جداگانه عمل می‌کند.

یک برش با عمق ۱



ادغام بیشینه با فیلتر 2×2
و اندازه گام ۲

6	8
3	4

تئشیفیص اشیا

۱۳

□ دسته‌بندی تصویر و شناسایی اشیا.



دسته‌بندی تصویر



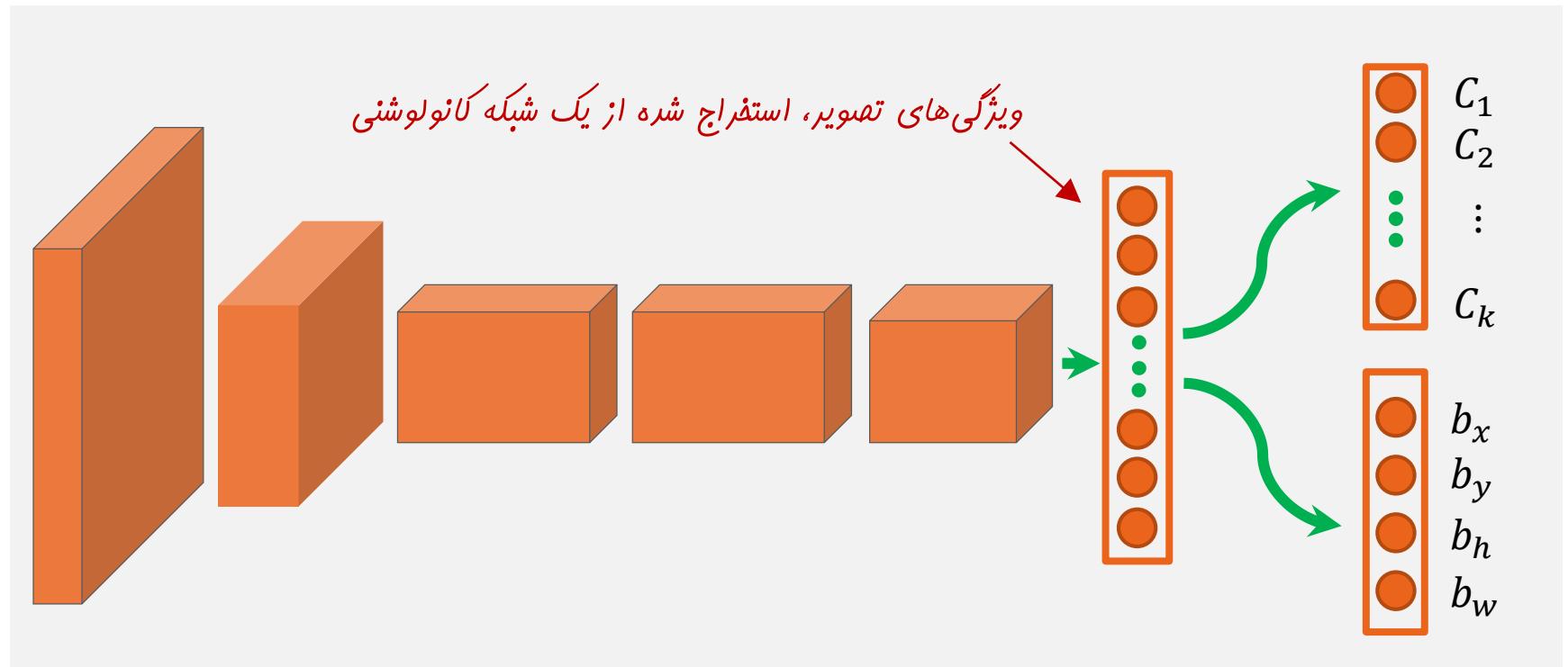
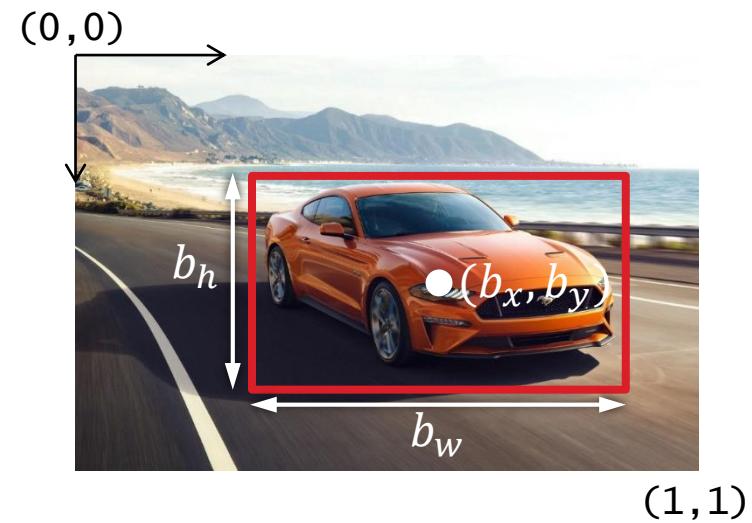
دسته‌بندی و مکان‌یابی



شناسایی اشیا

دسته‌بندی و مکان‌یابی

۱۴

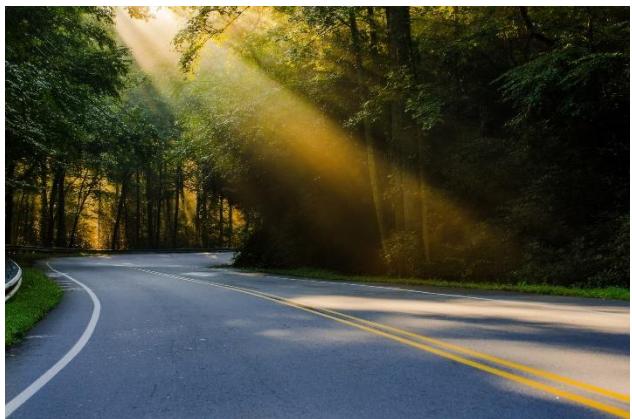


تابع هزینه: مجموع مربعات خطا

۱۵



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



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

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

کلاس‌ها □

- ۱ - رهگذر
- ۲ - خودرو
- ۳ - موتورسیکلت
- ۴ - پس‌زمینه

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

تَشخِيص نقاط عطف: تشخیص احساسات

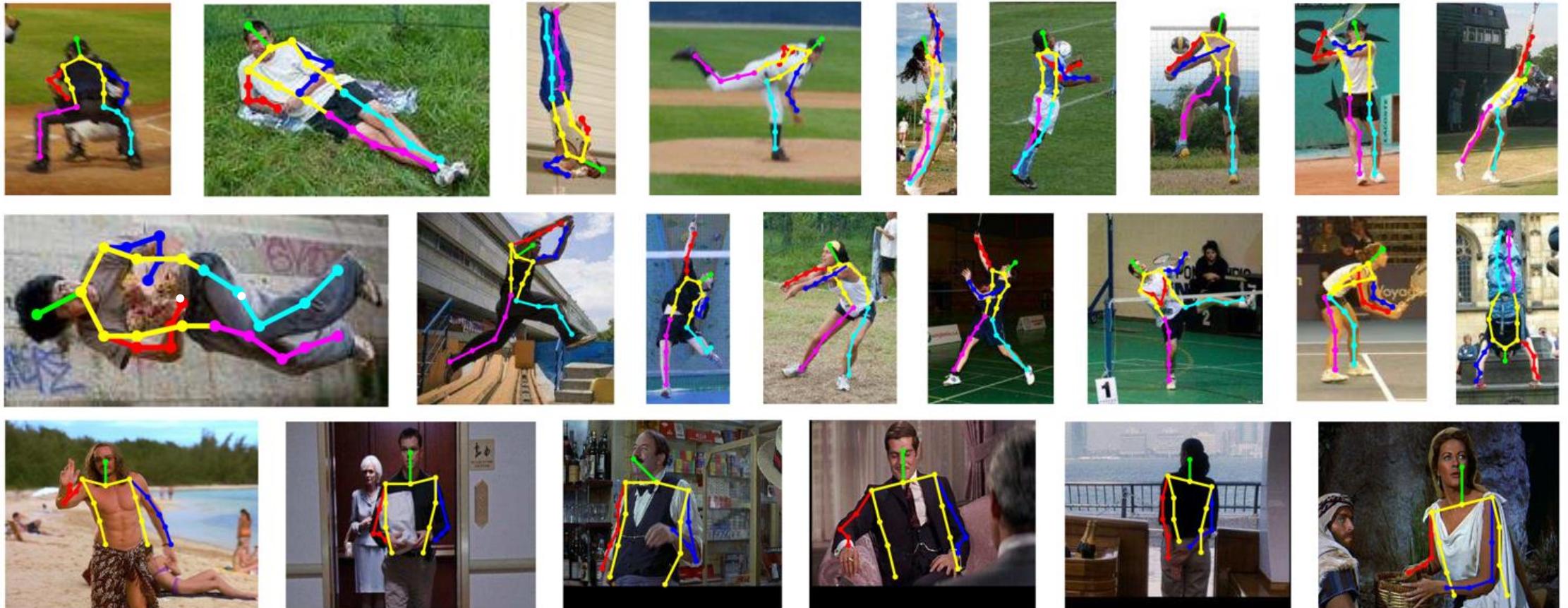
۱۶



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

تَسْخِيص نَقَاط عَطْف: تَسْخِيص وَضْعَيْت

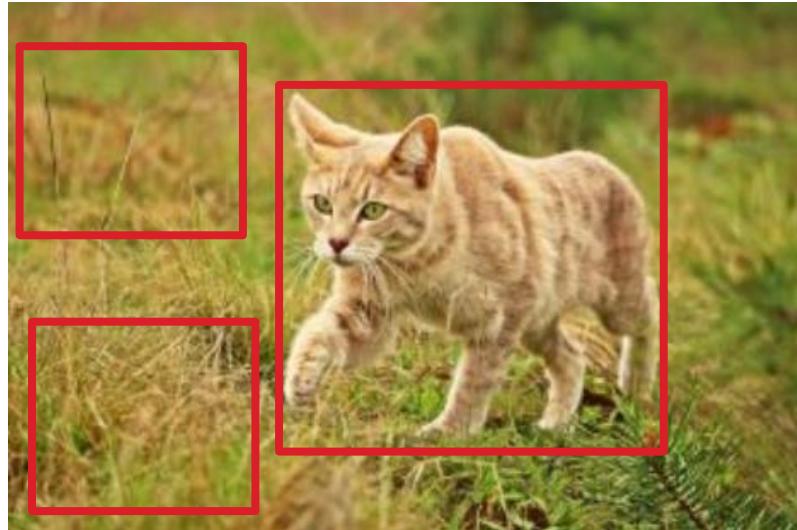
۱۷



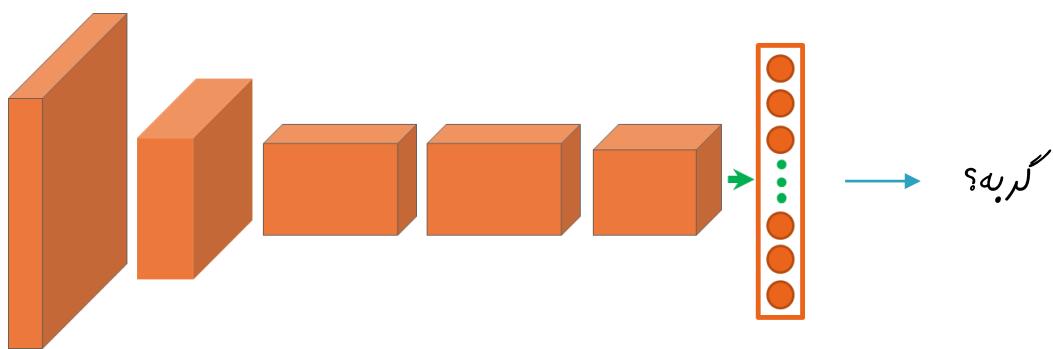
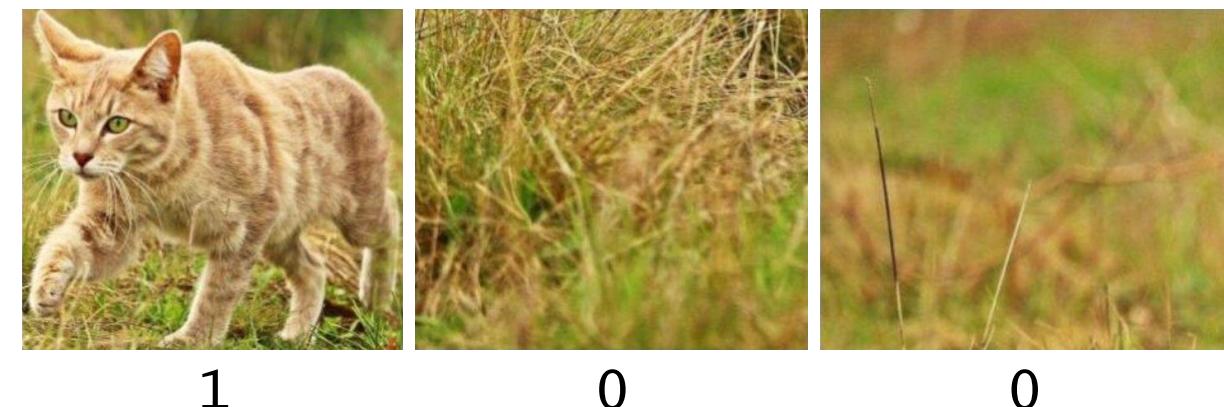
Xianjie Chen and Alan Yuille, Articulated Pose Estimation by a Graphical Model with Image Dependent Pairwise, NIPS, 2014.

شناختی اشیا: پنجمین لغزان

۱۸



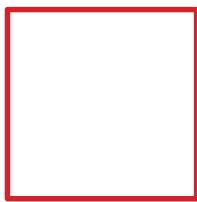
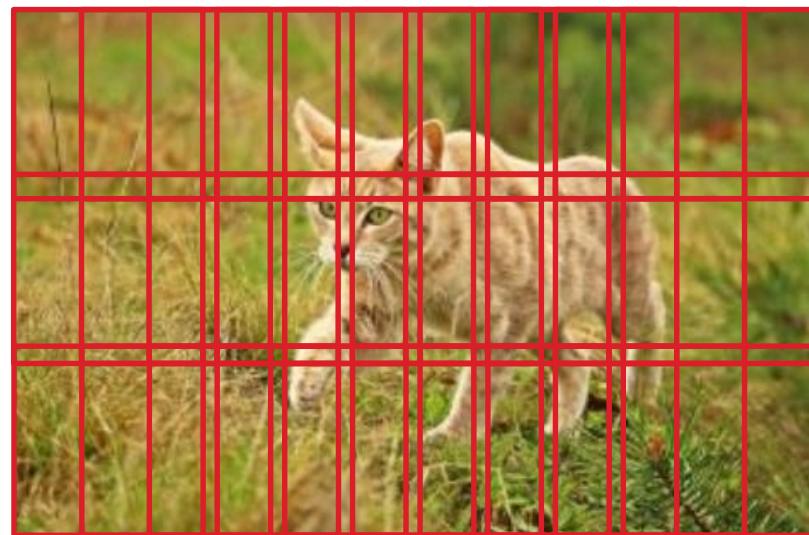
ایجاد مجموعه آموزشی



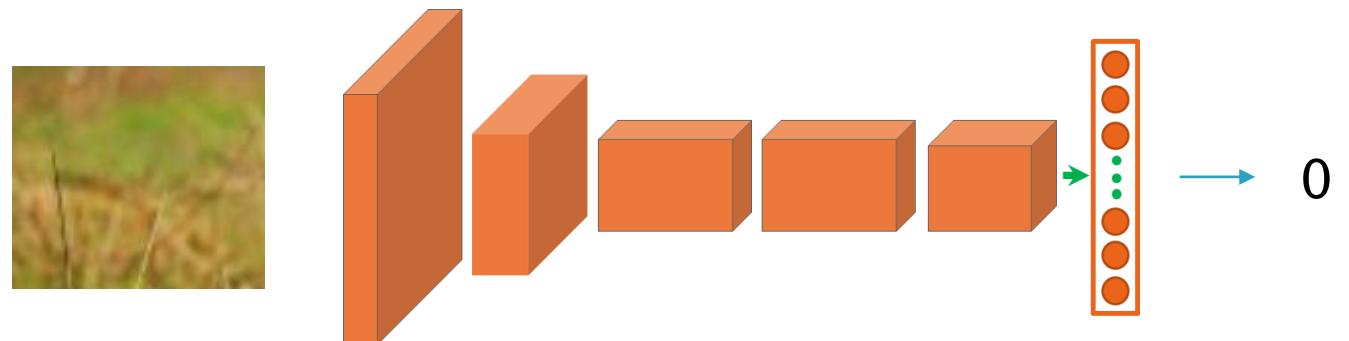
آموزش شبکه کانولوشنی

شناسایی اشیا: پنجره لغزان

۱۹

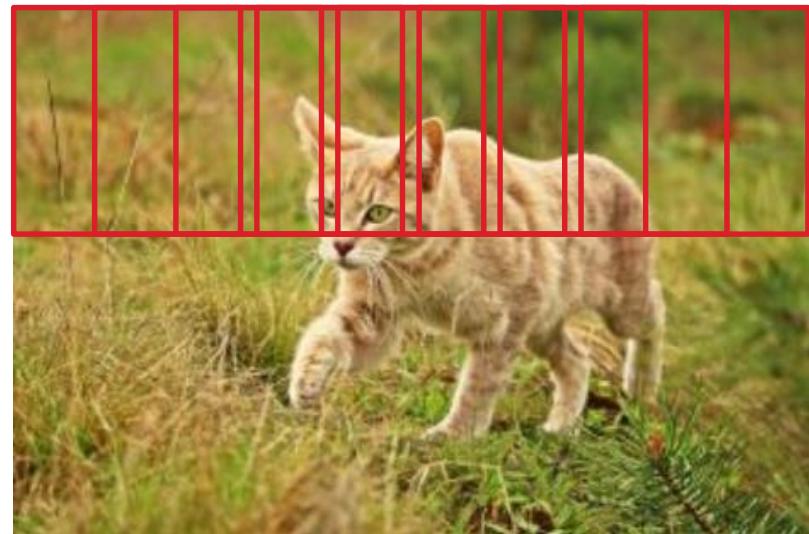


شناسایی اشیا با پنجره لغزان

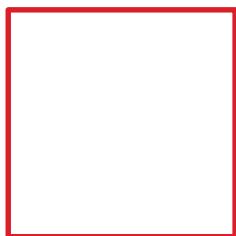
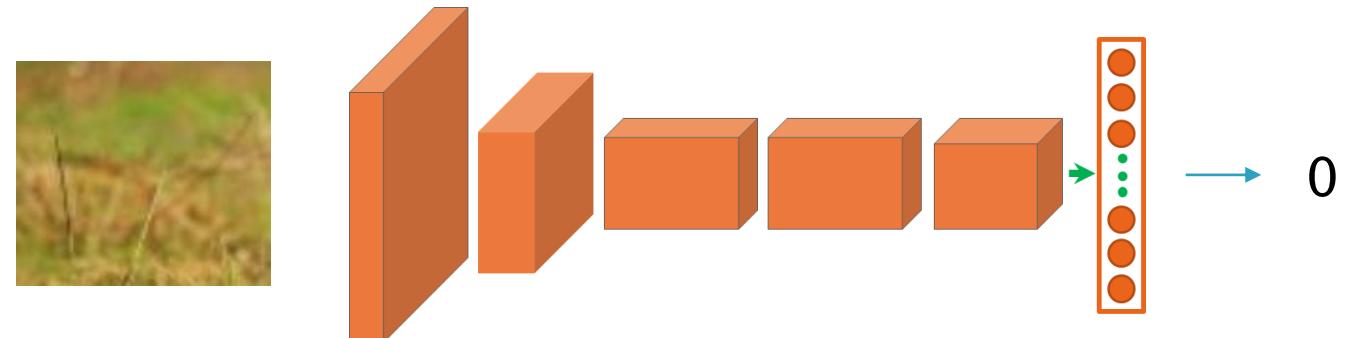


شناسایی اشیا: پنجره لغزان

۲۰

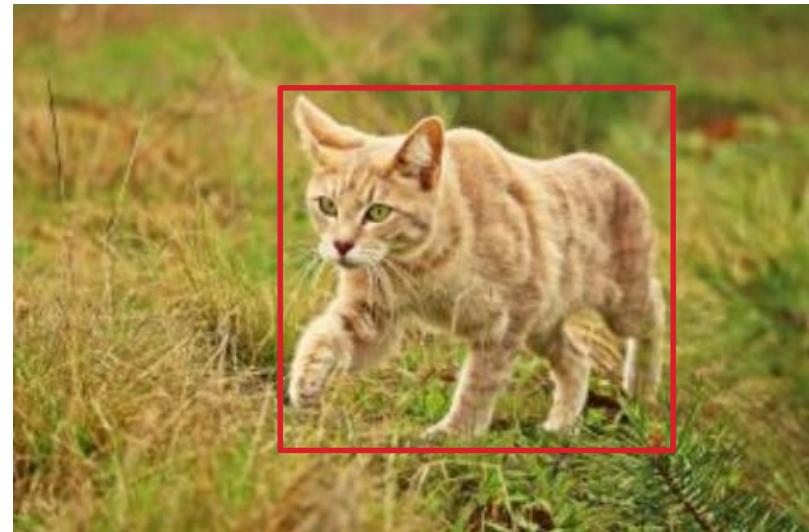


شناسایی اشیا با پنجره لغزان

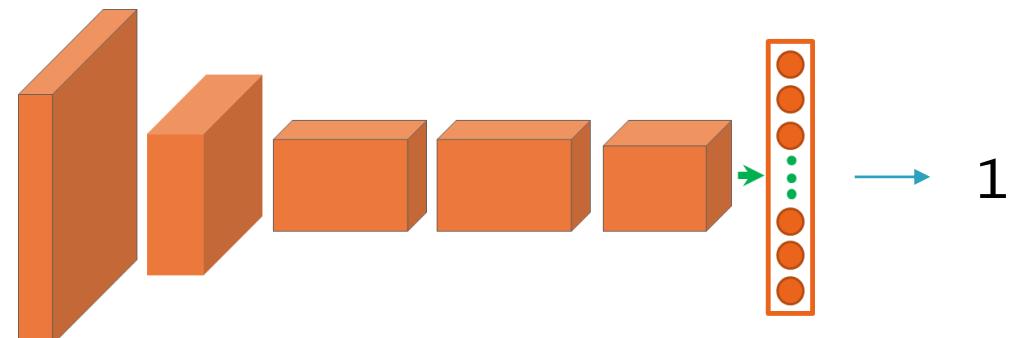


شناسایی اشیا: پنجره لغزان

۲۱



شناسایی اشیا با پنجره لغزان

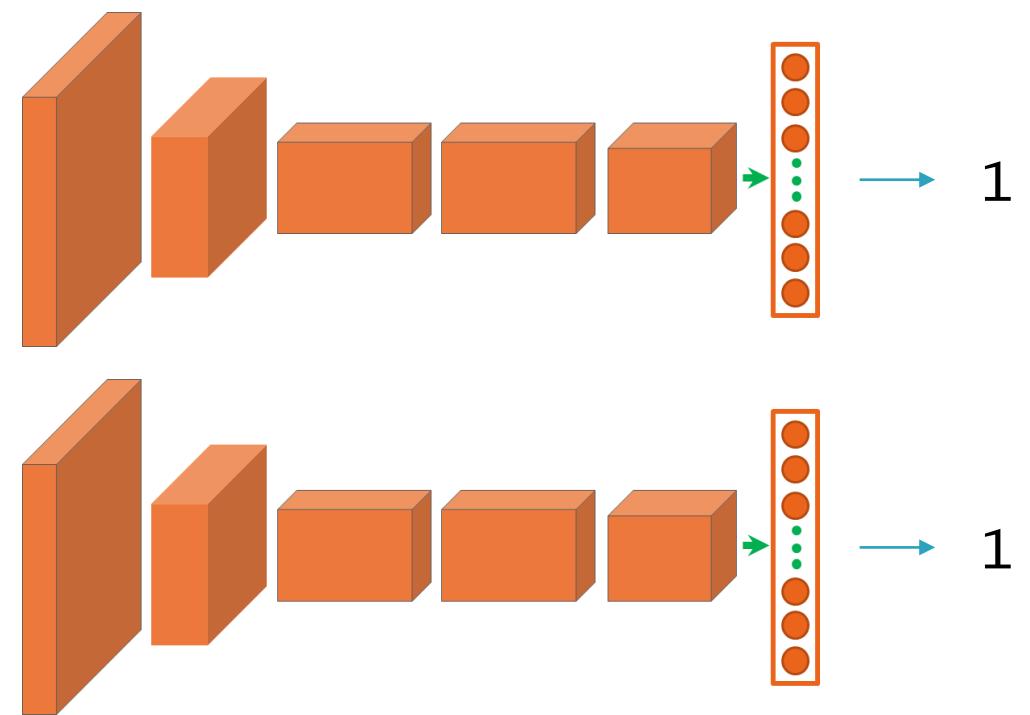
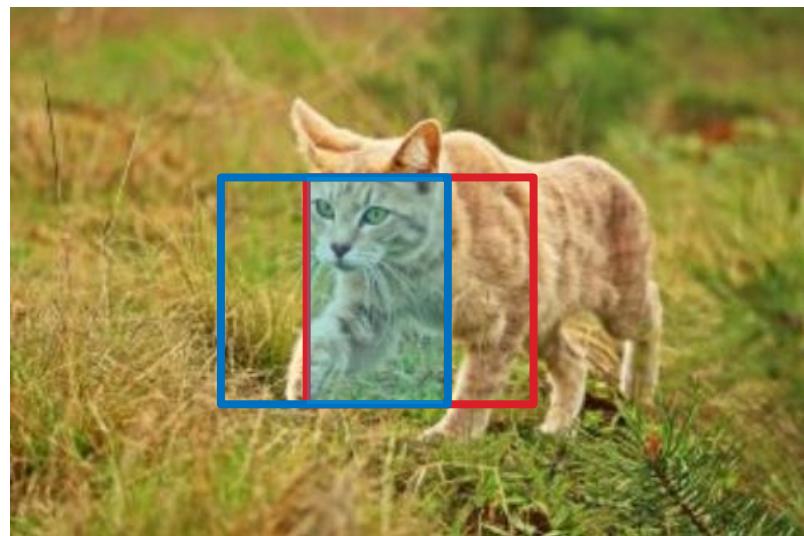


هزینه محاسباتی بسیار بالا!

پنجه لغزان: کاهش هزینه‌های محاسباتی

۲۲

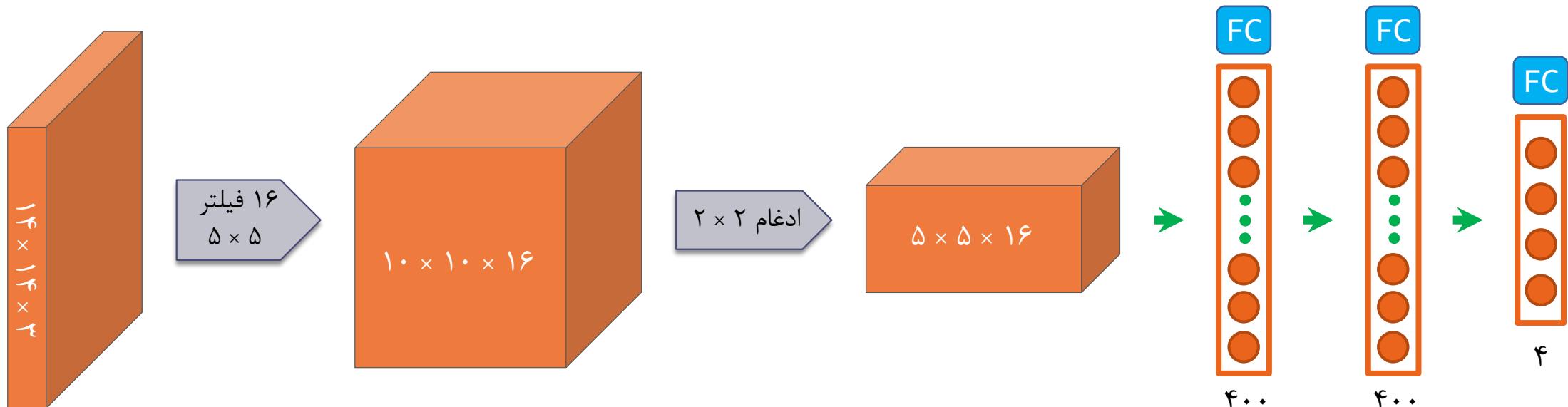
شناسایی اشیا با پنجه لغزان



پنجه لغزان: پیاده‌سازی کانولوشنی

۲۳

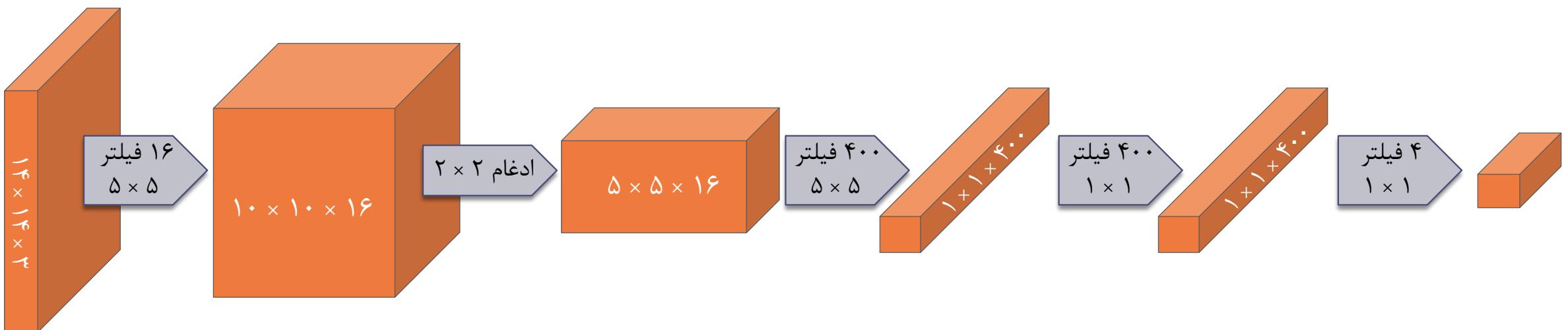
□ تبدیل لایه‌های کاملاً متصل به لایه‌های کانولوشنی.



پنجه لغزان: پیاده‌سازی کانولوشنی

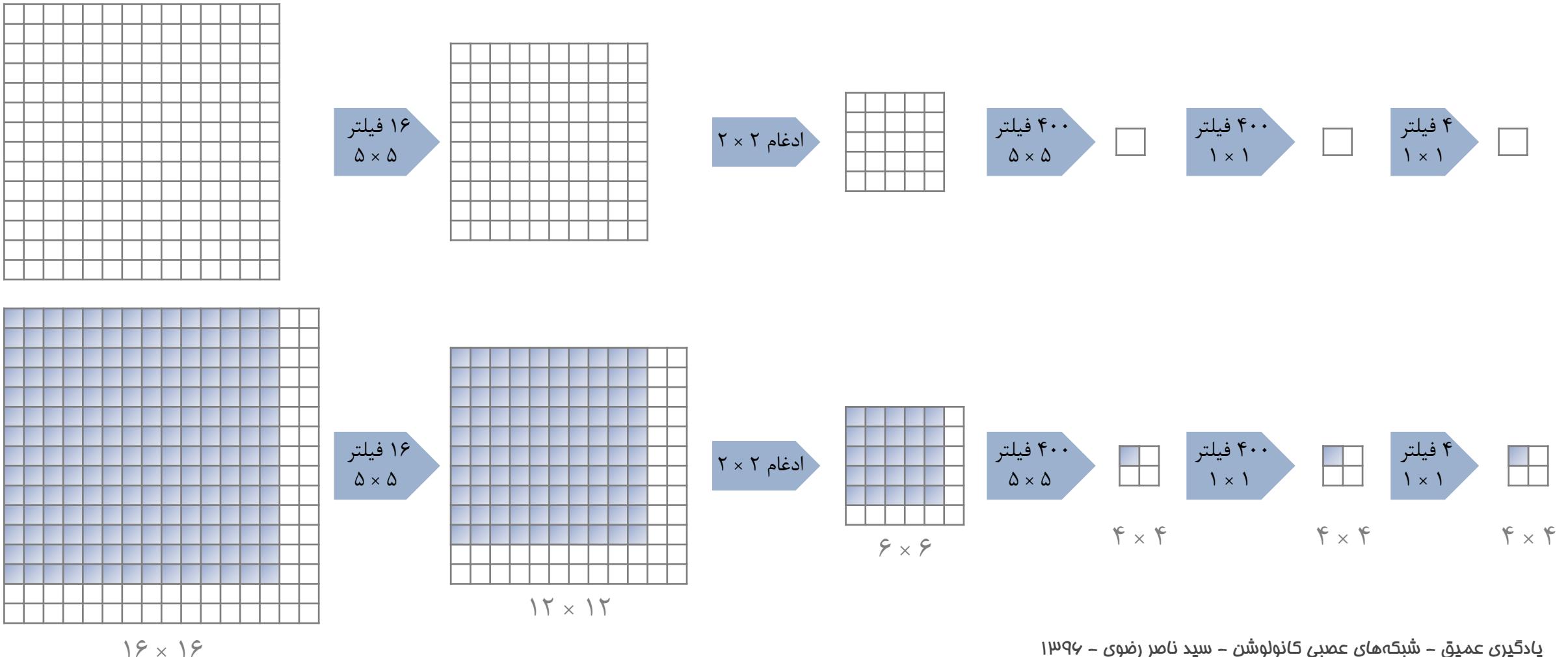
۲۴

□ تبدیل لایه‌های کاملاً متصل به لایه‌های کانولوشنی.



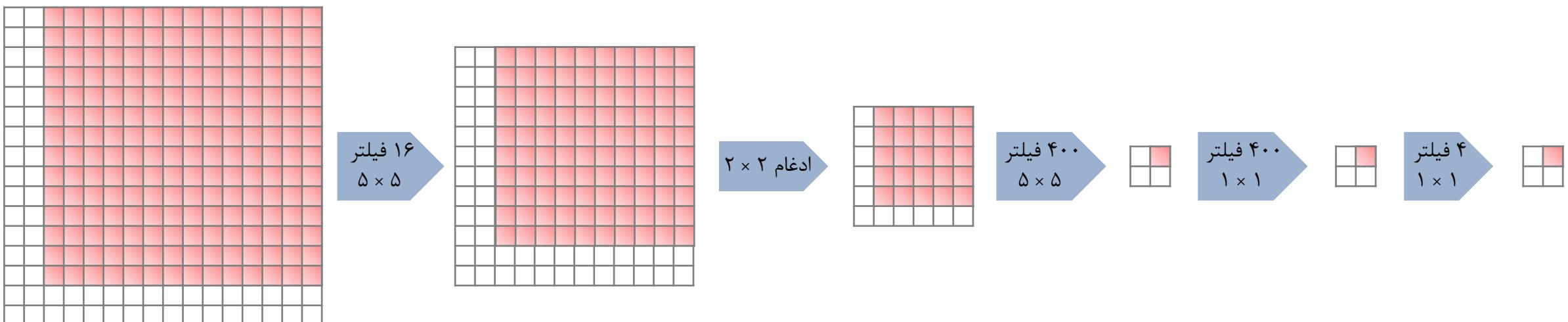
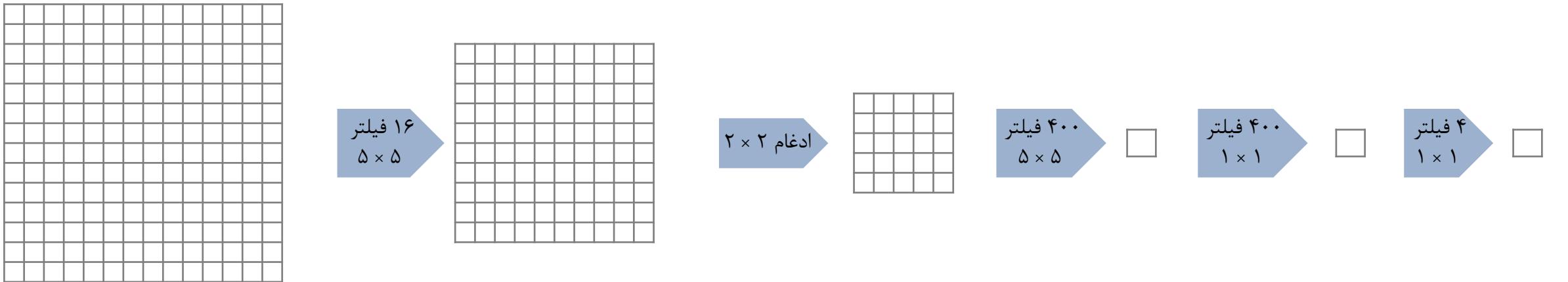
پنجه لغزان: پیاده‌سازی کانولوشن

۲۵



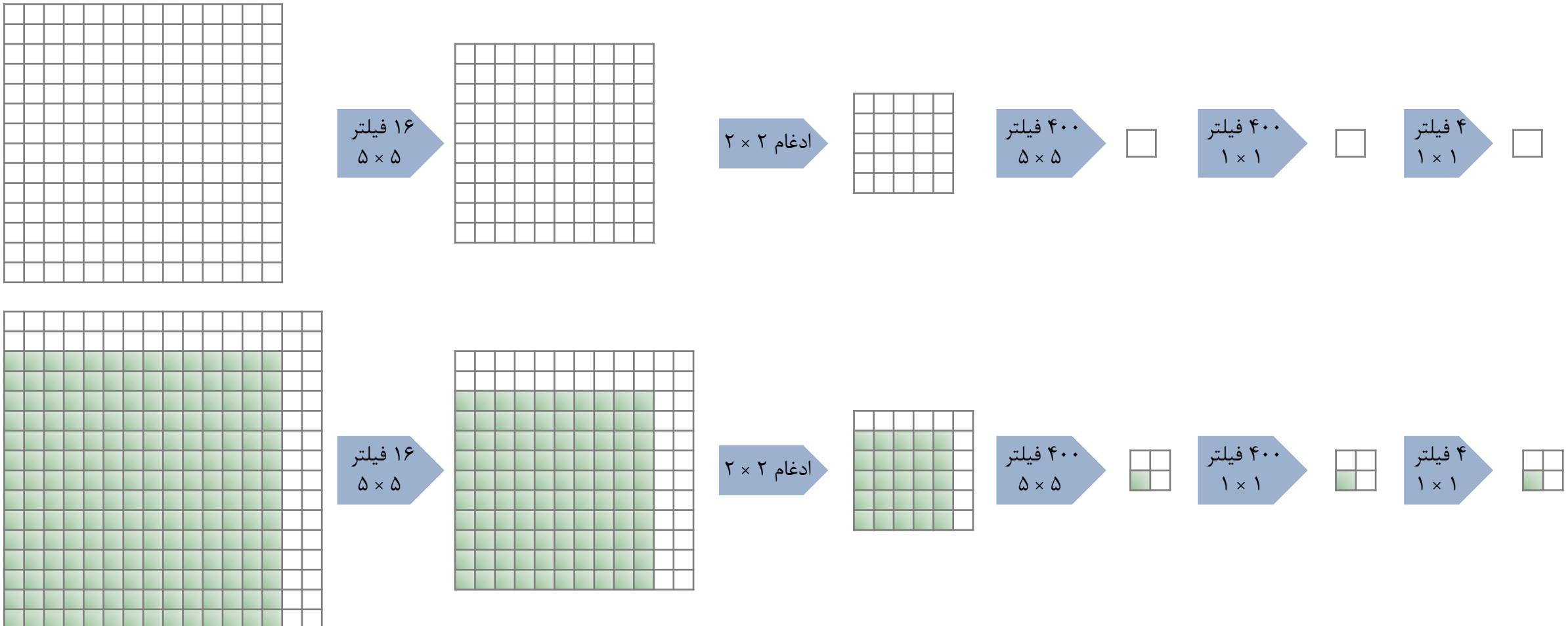
پنجه لغزان: پیاده‌سازی کانولوشن

۲۶



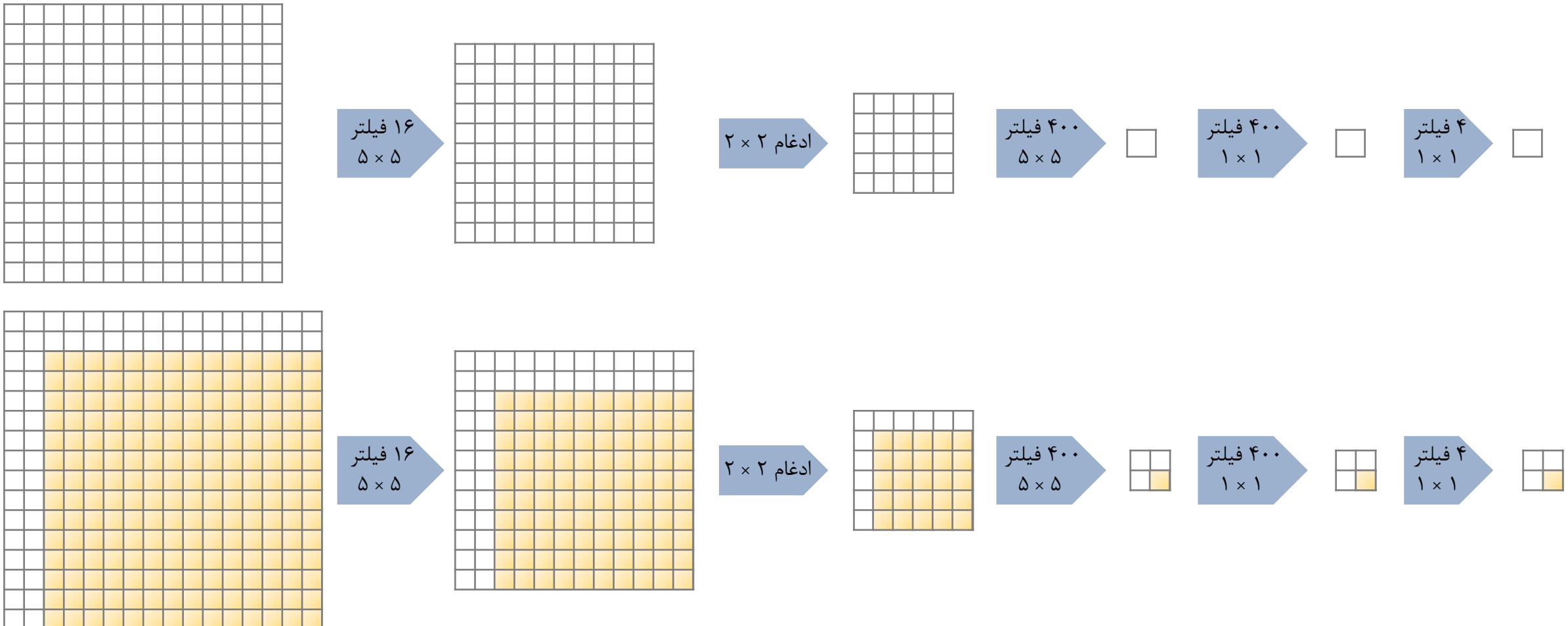
پنجه لغزان: پیاده‌سازی کانولوشن

۲۷



پنجه لغزان: پیاده‌سازی کانولوشن

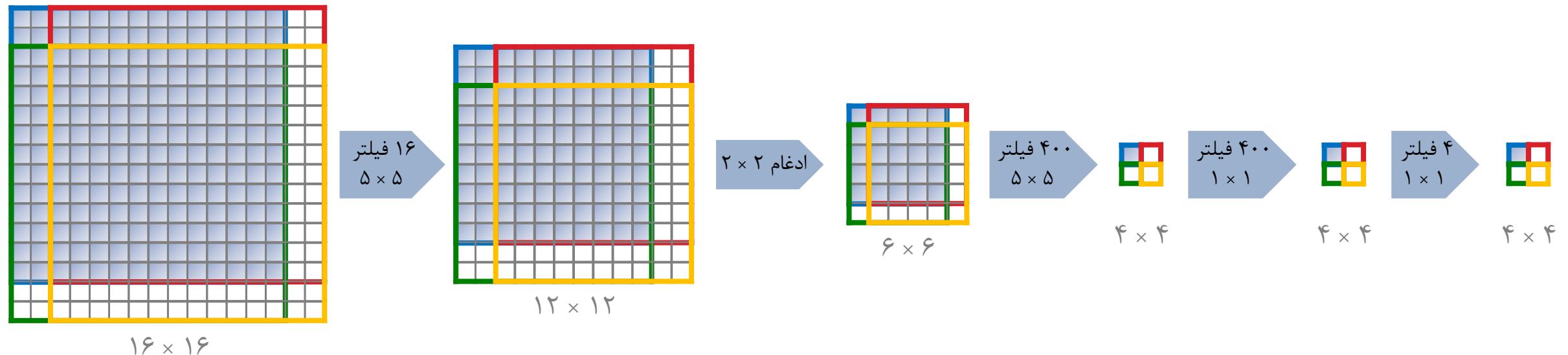
۲۸



پنجه لغزان: پیاده‌سازی کانولوشن

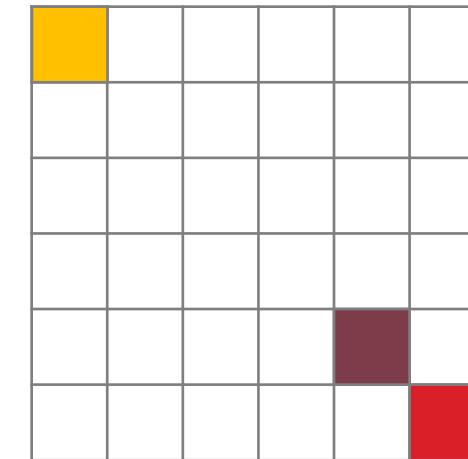
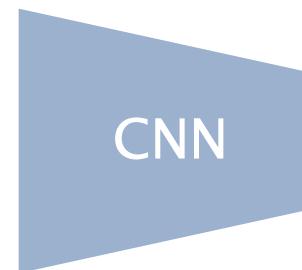
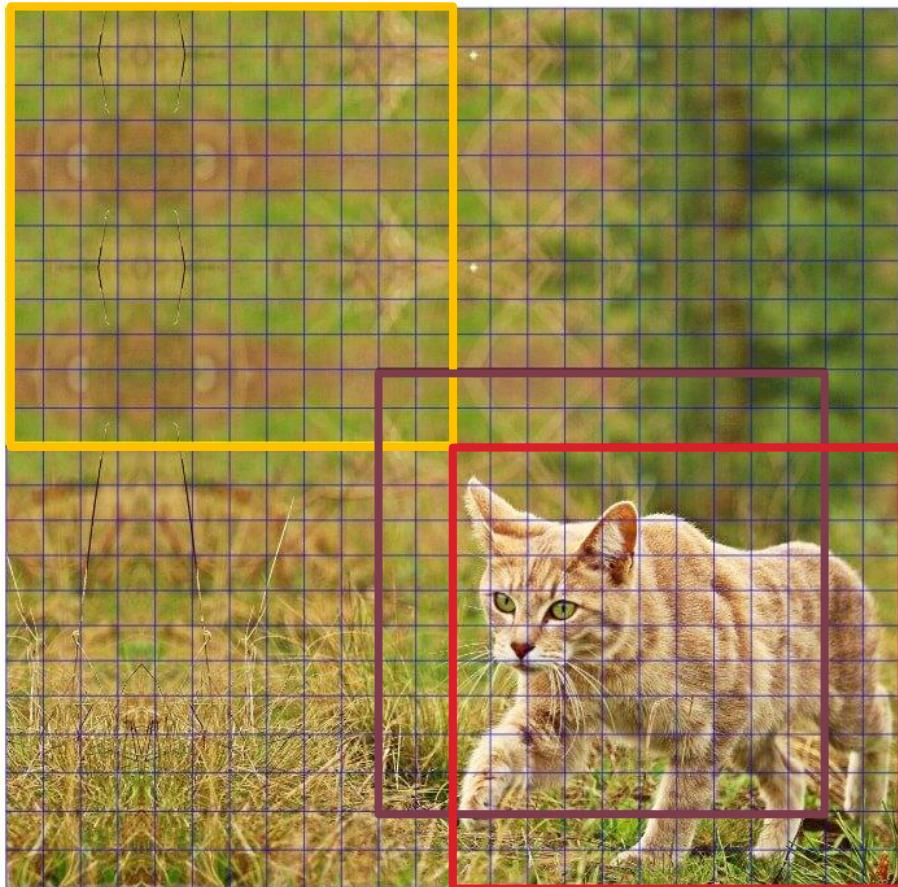
۲۹

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



پنجه لغزان: پیاده‌سازی کانولوشنی

۳۰



احتمال وجود اشیا
در هر یک از پنجره‌ها

اشتراك بر اروى اجتماع

۳۱

□ معیاری به منظور محاسبه میزان همپوشانی دو جعبه محاطی.

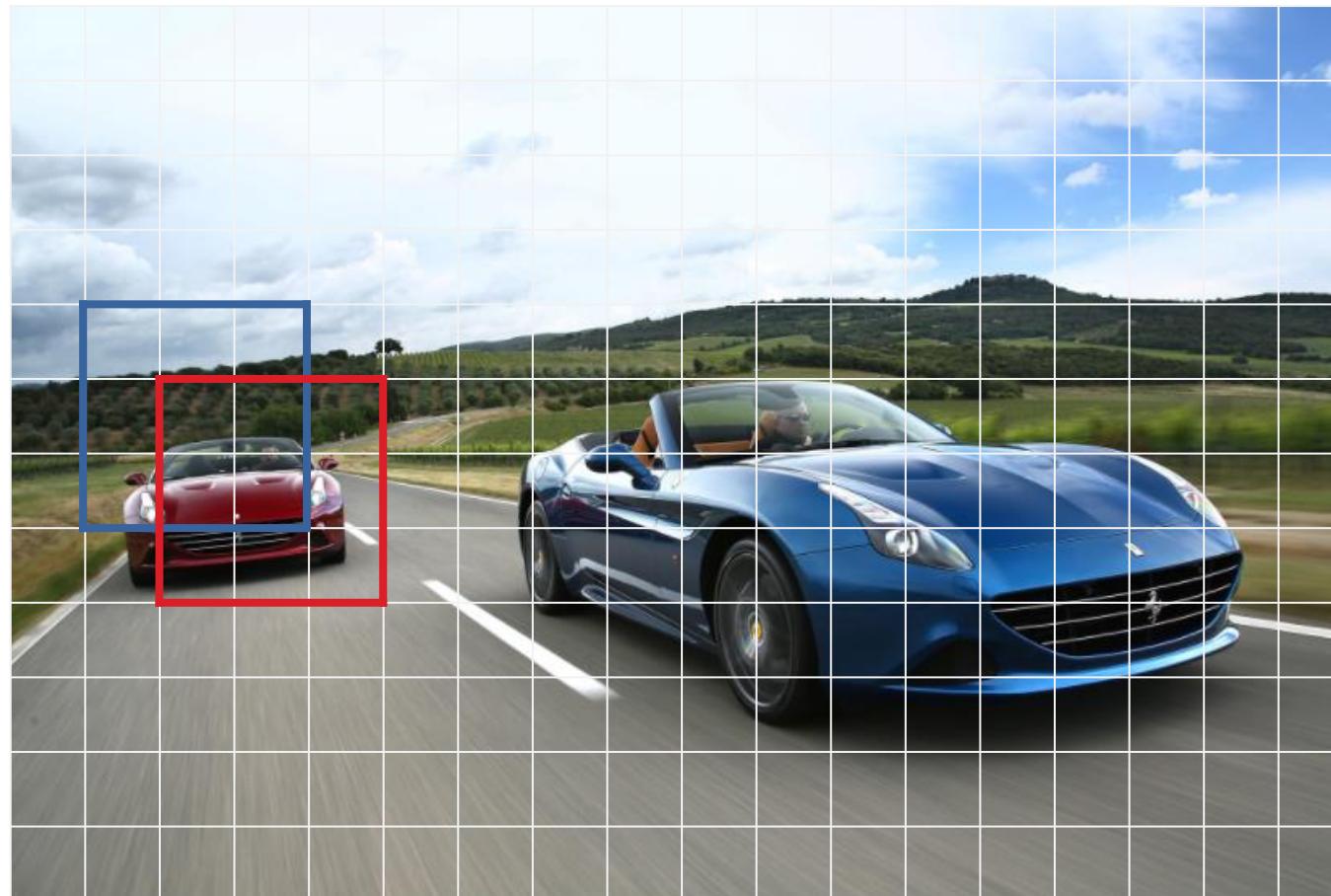


$$IoU = \frac{|B_1 \cap B_2|}{|B_1 \cup B_2|} \geq 0.5$$

سرکوب غیر بیشینه

۳۲

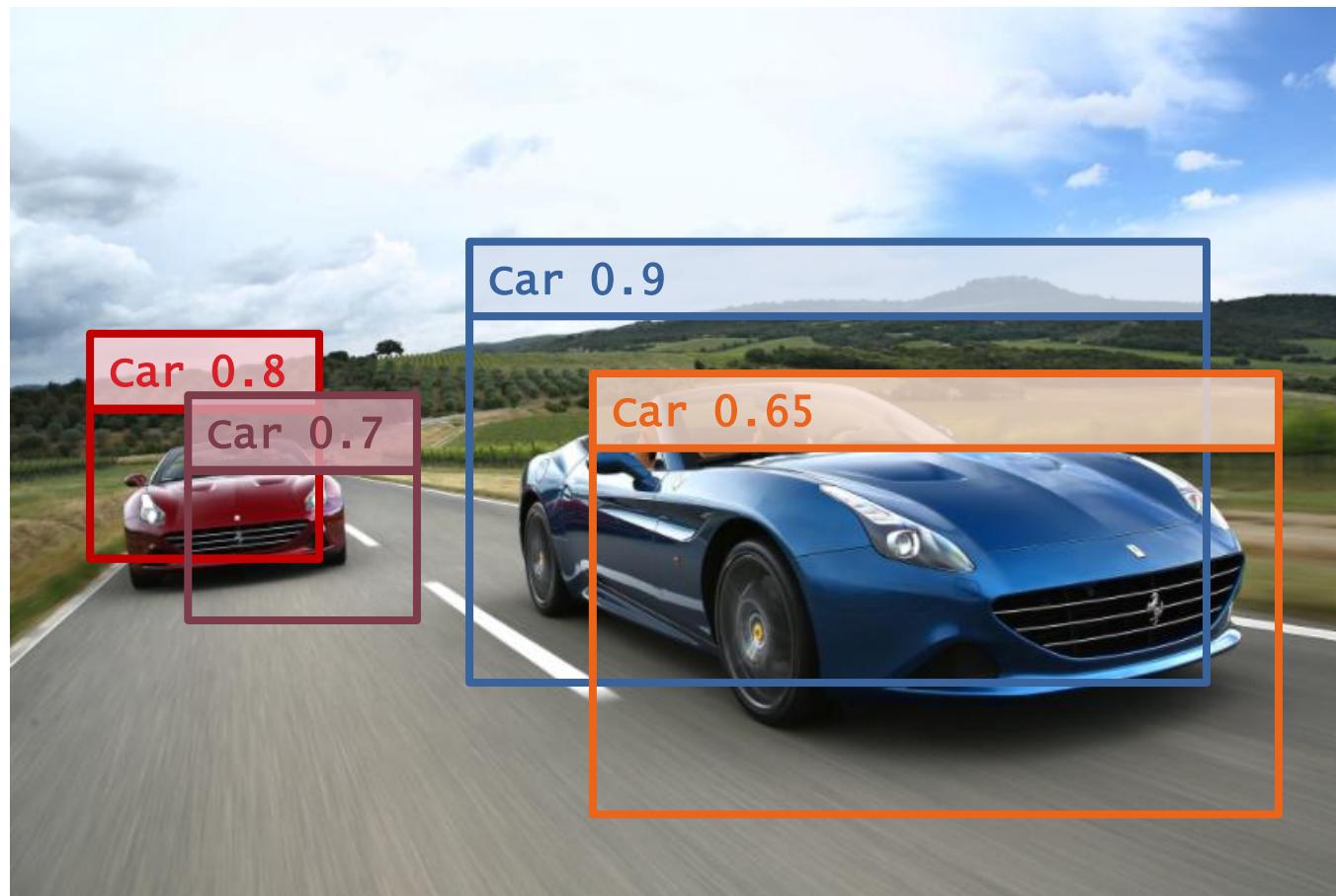
اطمینان از این که هر شی، بیش از یک بار شناسایی نمی‌شود.



سرگوب غیر بیشینه

۳۳

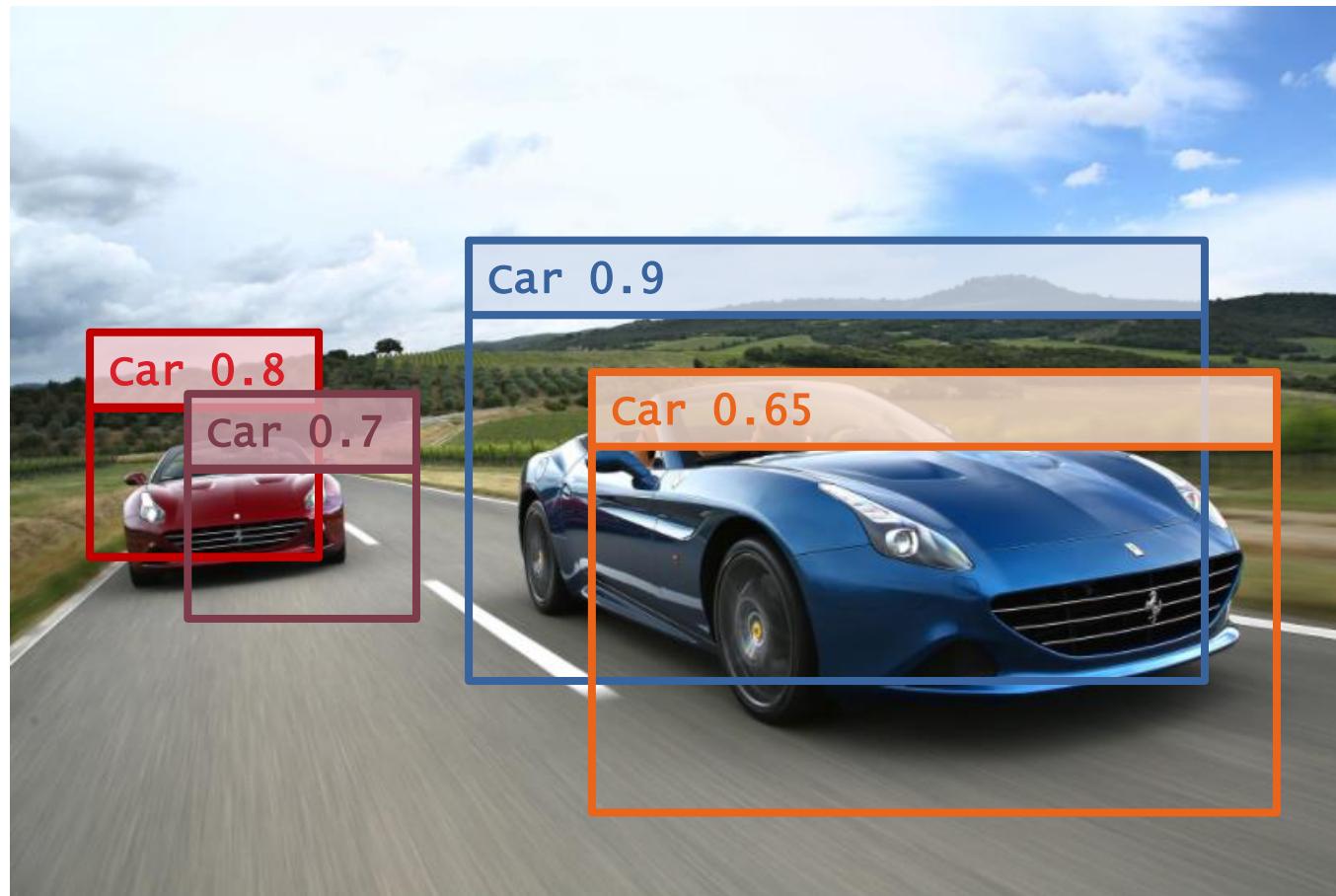
اطمینان از این که هر شی، بیش از یک بار شناسایی نمی‌شود. □



سرگوب غیر بیشینه

۳۴

اطمینان از این که هر شی، بیش از یک بار شناسایی نمی‌شود.



سرکوب غیر بیشینه

۳۵

اطمینان از این که هر شی، بیش از یک بار شناسایی نمی‌شود.



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

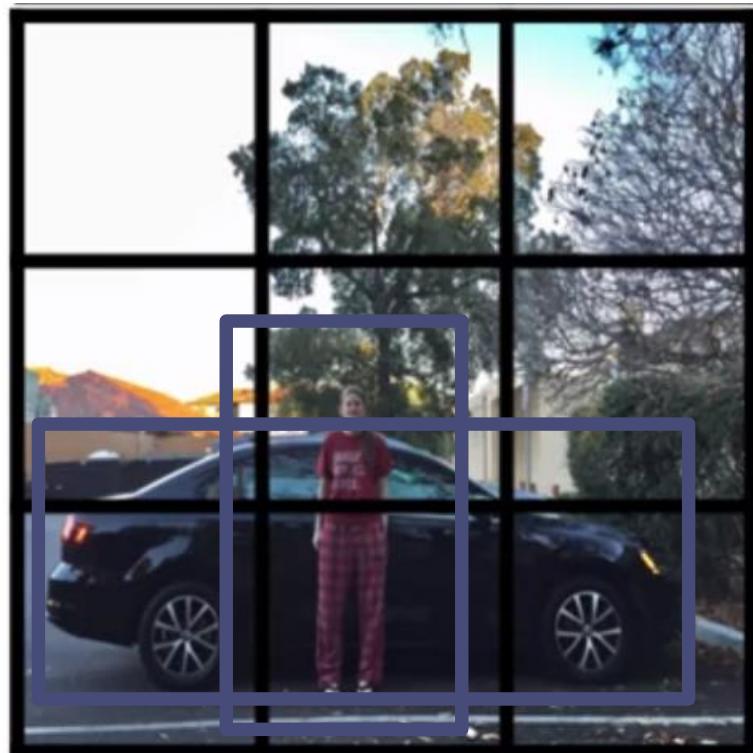
- مرحله دوم: تا زمانی که پیش‌بینی (جعبه محاطی) پردازش نشده‌ای وجود دارد:
- یک جعبه با بیشترین احتمال را انتخاب و آن را به عنوان خروجی در نظر بگیر.
 - تمامی جعبه‌هایی را که همپوشانی زیادی با جعبه انتخاب شده دارند، حذف کن.



$$IoU >= 0.5$$

همپوشانی اشیا و جعبه‌های لنگر

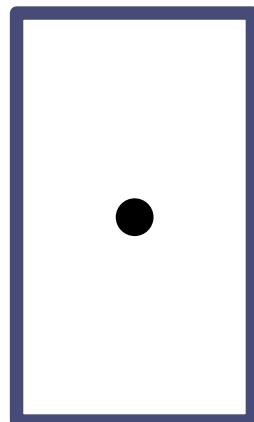
۳۷



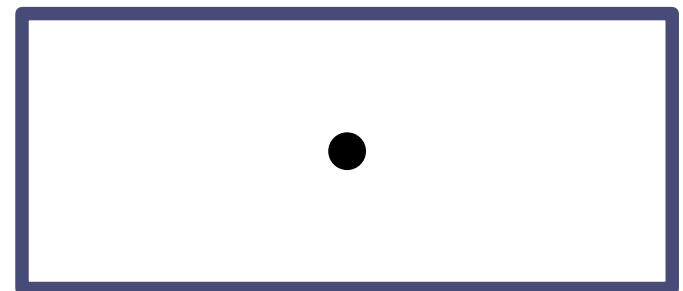
$$y = \begin{bmatrix} y_{a1} \\ y_{a2} \end{bmatrix}$$

به ازای هر خانه

جعبه لنگر ۱



جعبه لنگر ۲

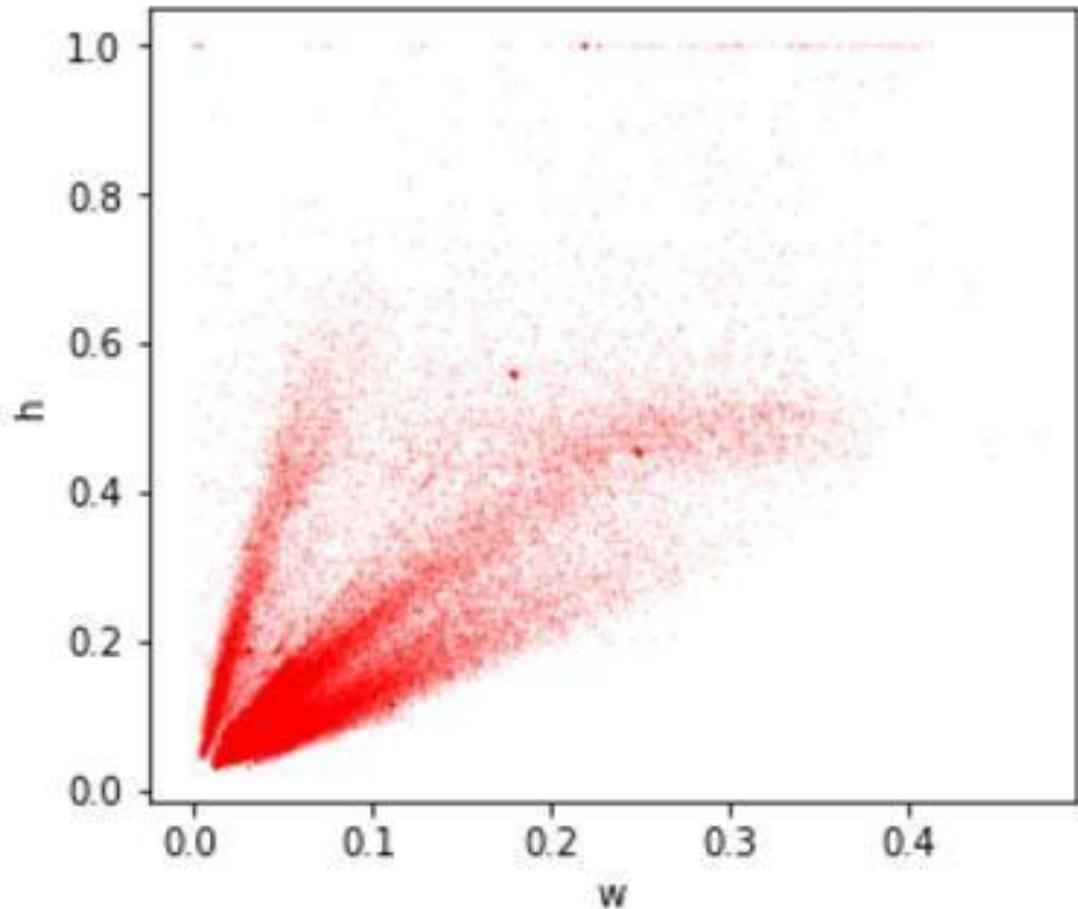


$$y_{a1} = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ C_1 \\ C_2 \\ C_3 \end{bmatrix}$$

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

جعبه‌های لنگر

۳۸



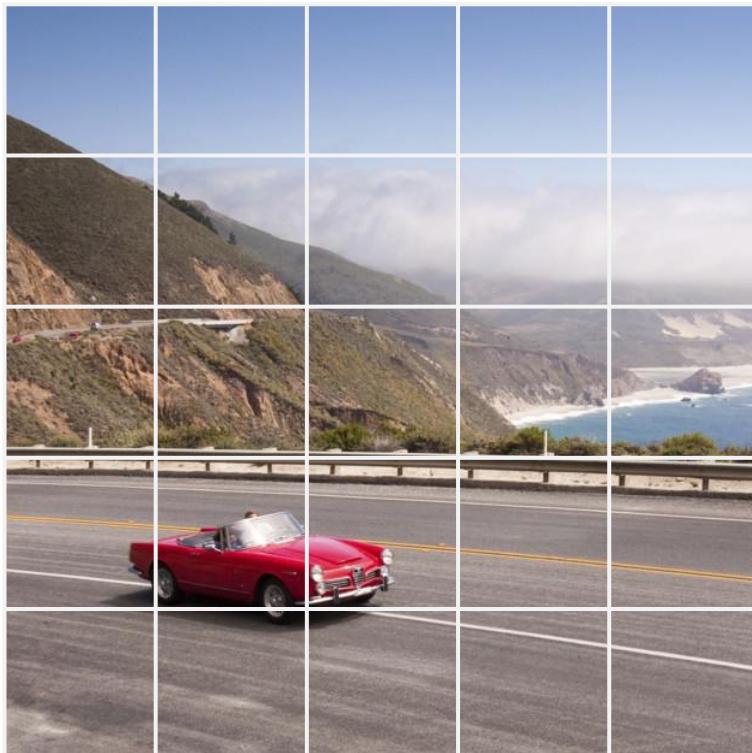
- انتخاب جعبه‌های لنگر
- طراحی جعبه‌های لنگر به صورت دستی
- استفاده از الگوریتم خوشبندی [۵ خوشه]

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

معیار فاصله

الکوئیت YOLO

۳۹



□ ایجاد مجموعه آموزشی.

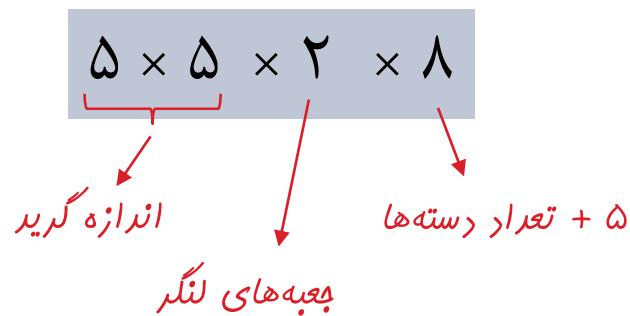
□ دسته‌ها.

۱ - رهگذر

۲ - خودرو

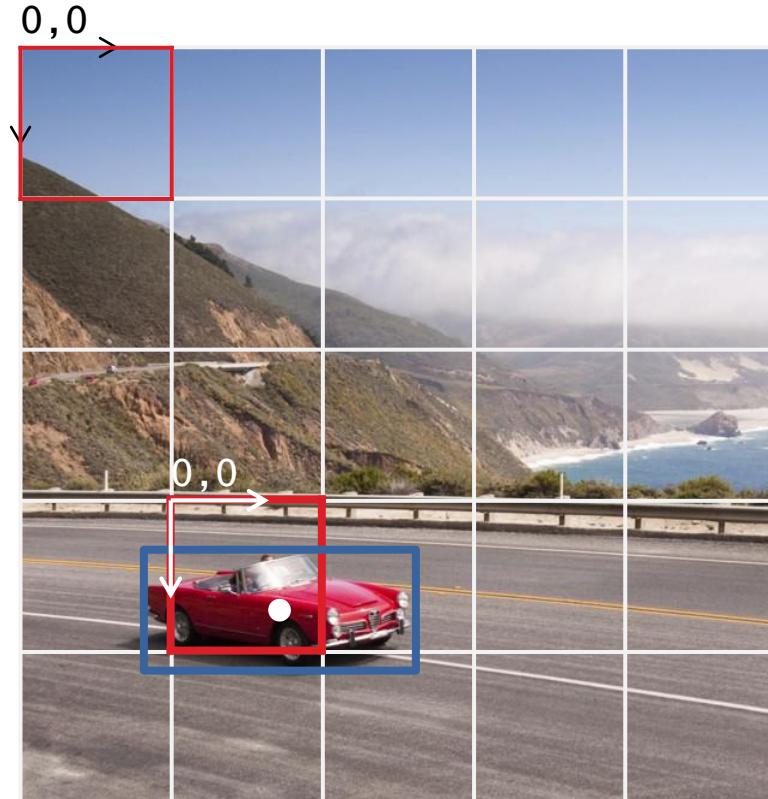
۳ - موتورسیکلت

□ خروجی.



الکوریتم YOLO

۴۰



فانه ۱

$$\begin{aligned} p_c &= 0 \\ b_x &= 0 \\ b_y &= 0 \\ b_h &= 0 \\ b_w &= 0 \\ C_1 &= 0 \\ C_2 &= 0 \\ C_3 &= 0 \\ p_c &= 0 \\ b_x &= 0 \\ b_y &= 0 \\ b_h &= 0 \\ b_w &= 0 \\ C_1 &= 0 \\ C_2 &= 0 \\ C_3 &= 0 \end{aligned}$$

فانه ۱۷

$$\begin{aligned} p_c &= 0 \\ b_x &= 0 \\ b_y &= 0 \\ b_h &= 0 \\ b_w &= 0 \\ C_1 &= 0 \\ C_2 &= 0 \\ C_3 &= 0 \\ p_c &= 1 \\ b_x &= 0.7 \\ b_y &= 0.7 \\ b_h &= 0.8 \\ b_w &= 1.9 \\ C_1 &= 0 \\ C_2 &= 1 \\ C_3 &= 0 \end{aligned}$$

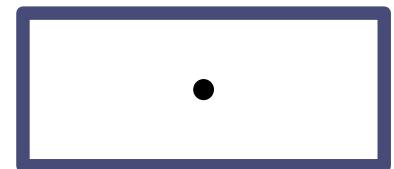
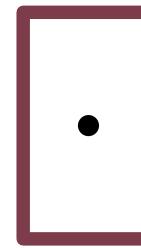
□ ایجاد مجموعه آموزشی.

□ دسته‌ها.

۱ - رهگذر

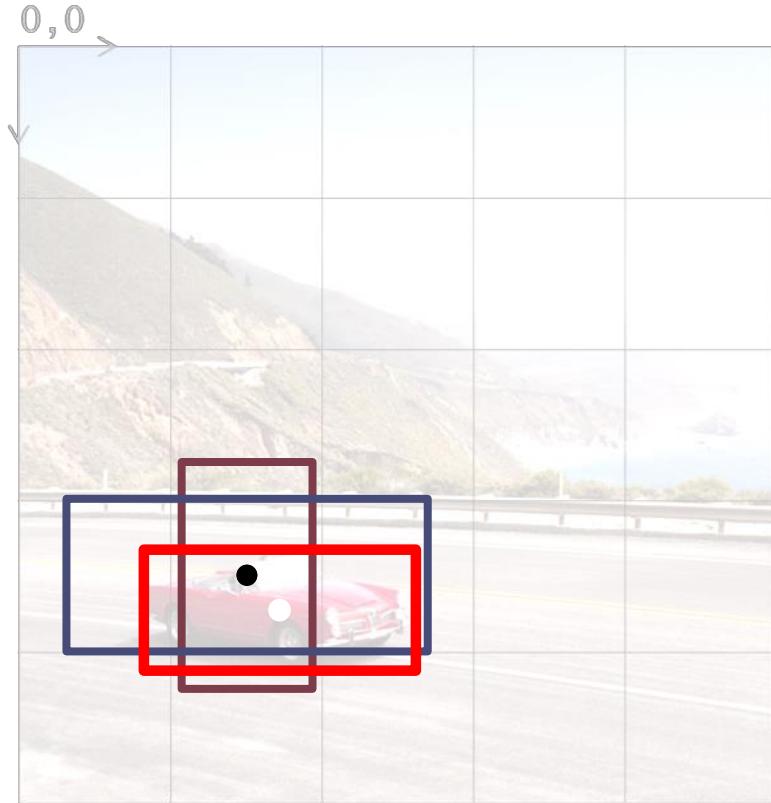
۲ - خودرو

۳ - موتورسیکلت



الگوریتم YOLO

۴۱



$19 \times 19 \times 5 \times \lambda$

فانه ۱

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

فانه ۱۷

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

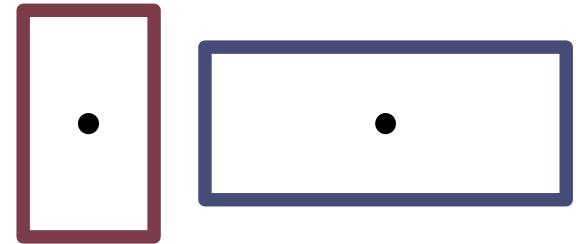
□ ایجاد مجموعه آموزشی.

□ دسته‌ها.

۱ - رهگذر

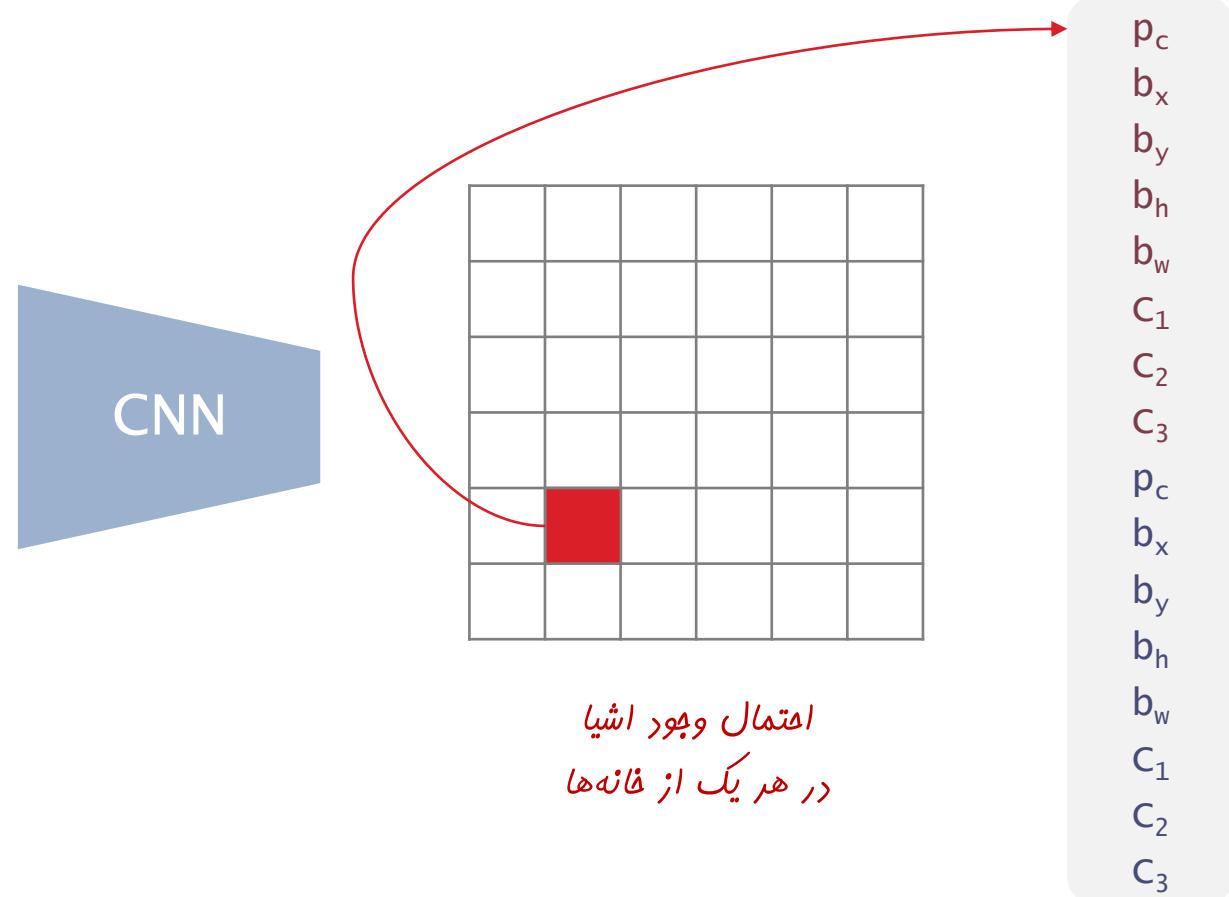
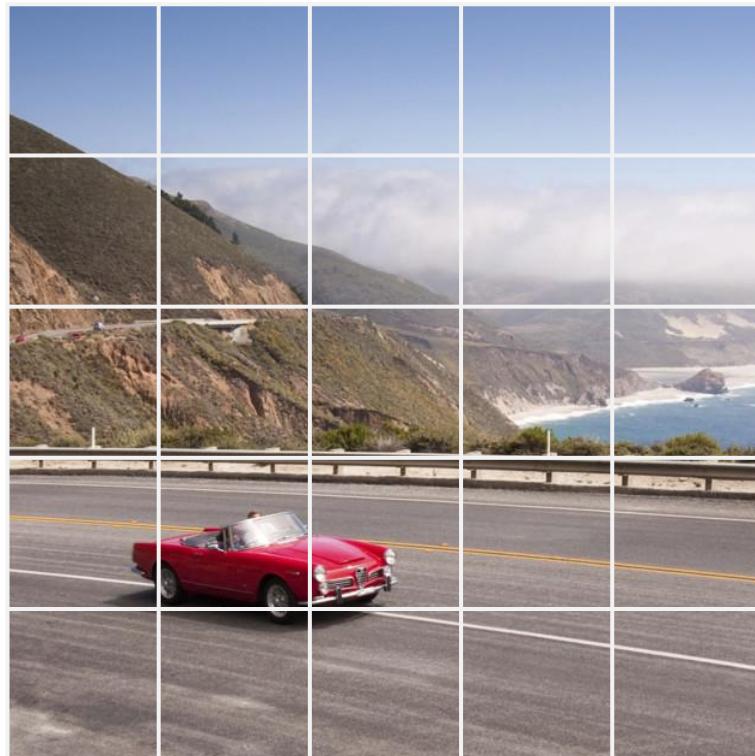
۲ - خودرو

۳ - موتورسیکلت



الگوریتم YOLO: پیش‌بینی

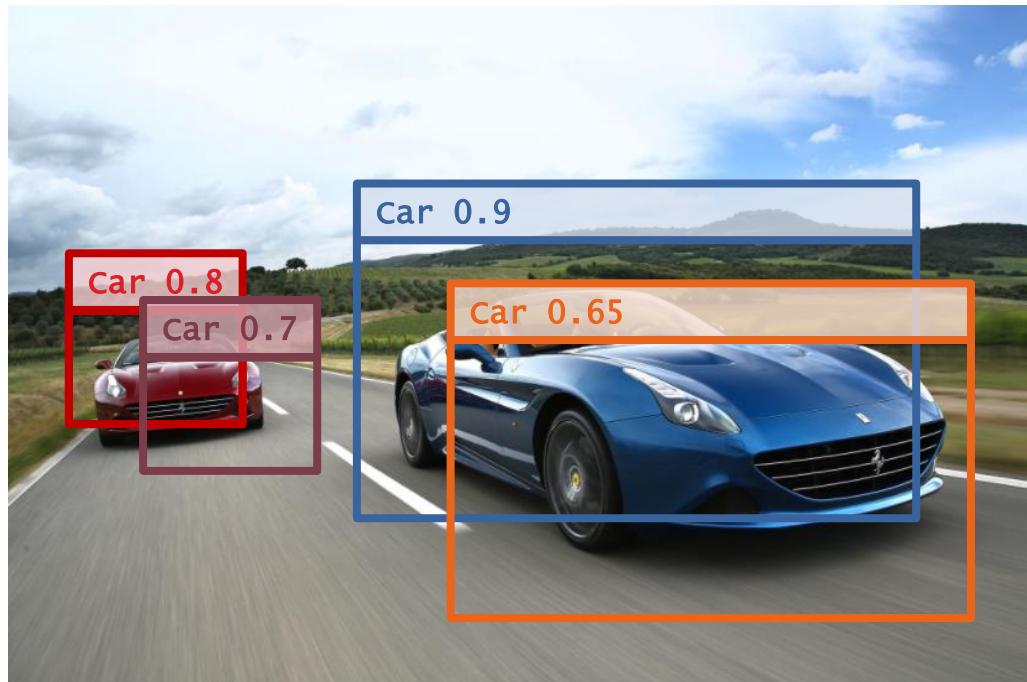
۴۲



الکوئیتم YOLO: سرکوب غیر بیشینه

۴۳

□ پیش‌بینی‌هایی را که احتمال کمی دارند، حذف کن.



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

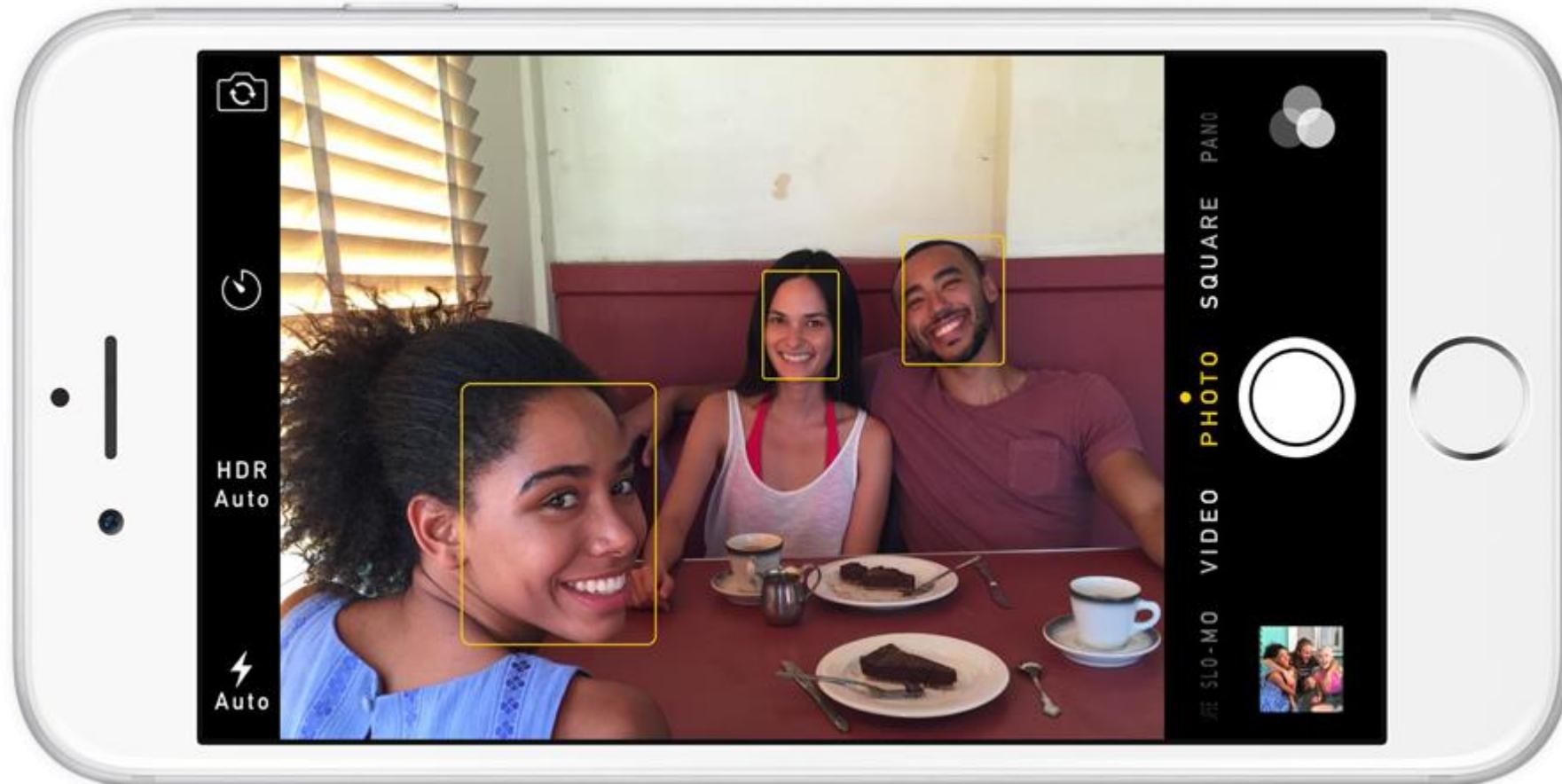
- تا زمانی که پیش‌بینی دیگری وجود دارد:
- پیش‌بینی با بیشترین احتمال را انتخاب کن.
- این پیش‌بینی را به عنوان خروجی در نظر بگیر.
- تمام پیش‌بینی‌هایی را که همپوشانی بالایی با پیش‌بینی انتخاب شده دارند، حذف کن.

پیشنهاد نامیه: الگوریتم R-CNN

۴۴

تَسْخِيص چهاره

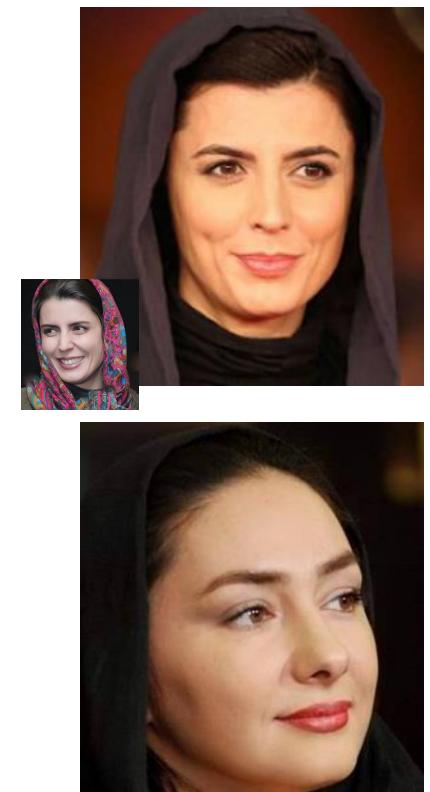
۴۵



یادگیری تک شات

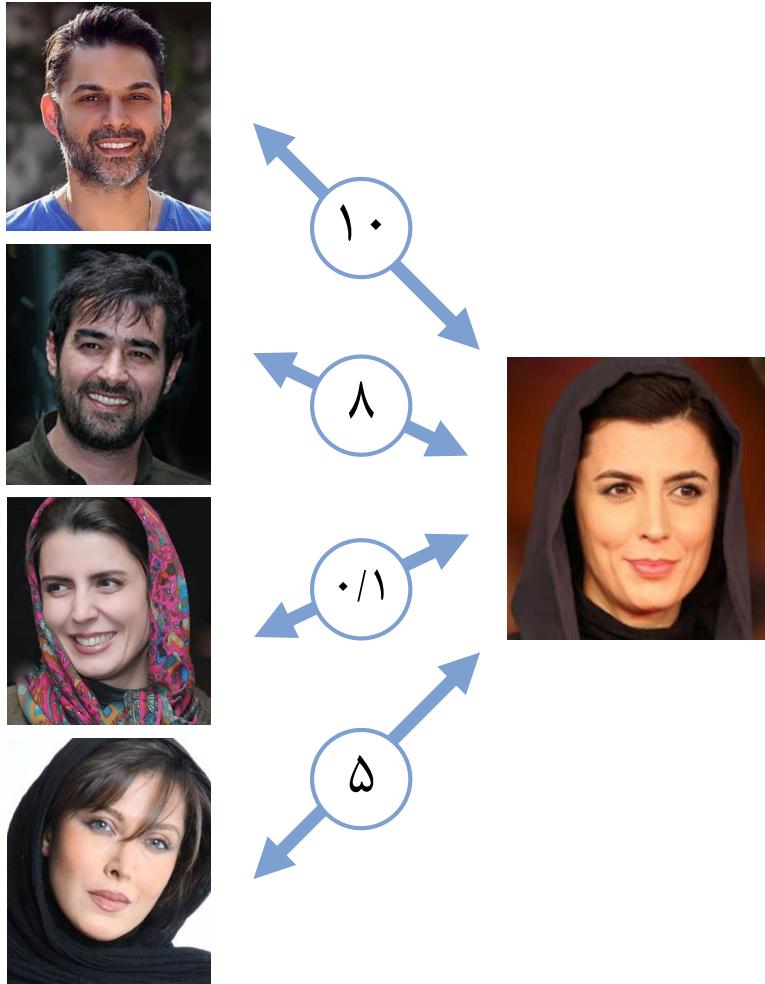
۴۶

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



یادگیری معیار شباهت

۴۷



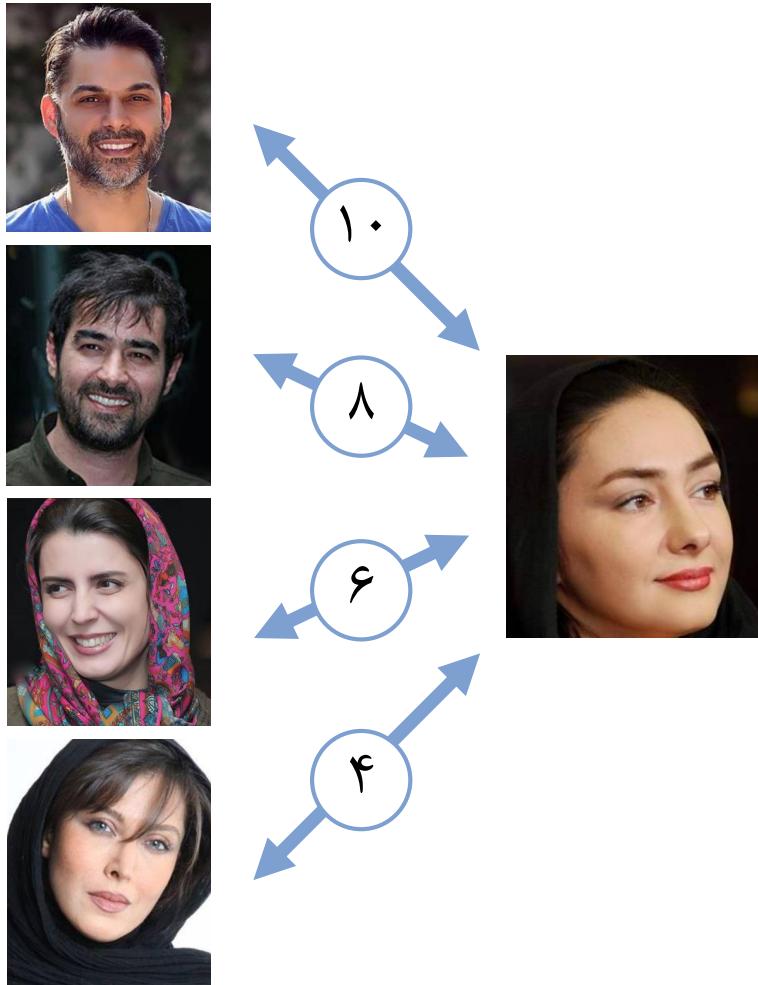
□ تابع فاصله.

$$d(\text{}) \leq \tau$$

$$d(\text{}) > \tau$$

یادگیری معیار شباهت

۴۸



□ تابع فاصله.

$$d(\text{}, \text{}) \leq \tau$$

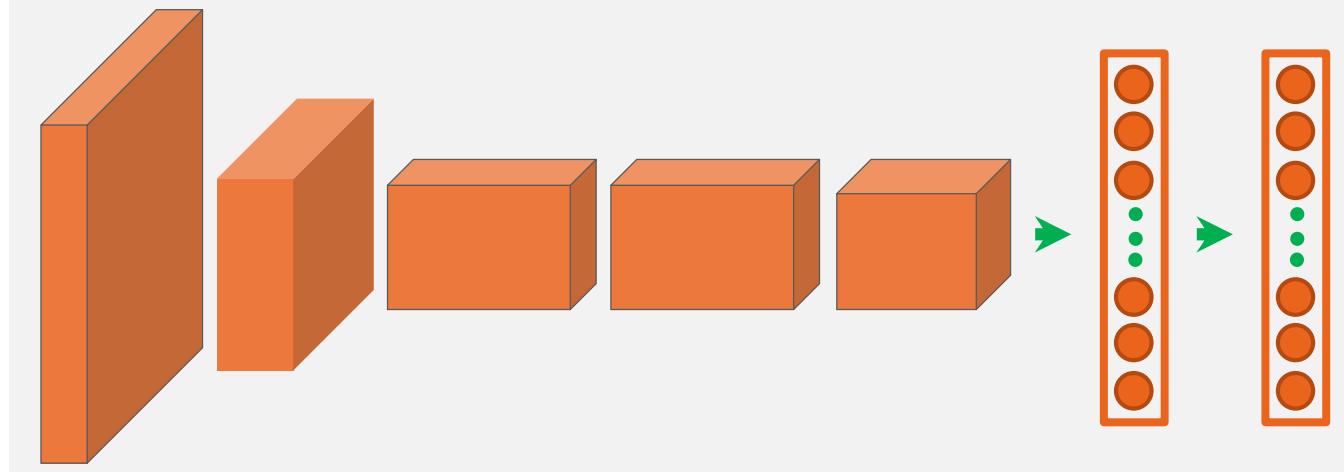
$$d(\text{}, \text{}) > \tau$$

شبکه سیاهی

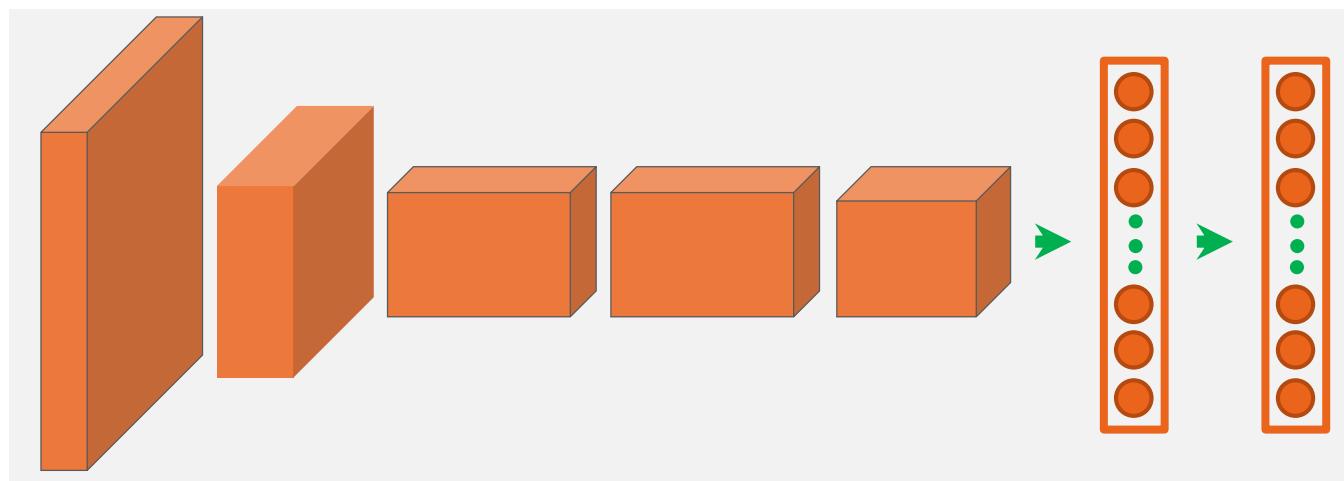
۴۹



$x^{(1)}$



$x^{(2)}$



شبکه سیاهی: تابع فاصله

۵۰

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

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

□ آموزش پارامترهای شبکه کانولوشنی.

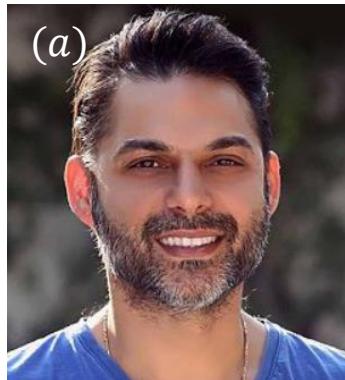
$\|f(x^{(i)}) - f(x^{(j)})\|^2 \rightarrow 0$ اگر دو تصویر $x^{(j)}$ و $x^{(i)}$ متعلق به یک شخص باشند

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

شبکه سیاهی: آموزش

۵۱

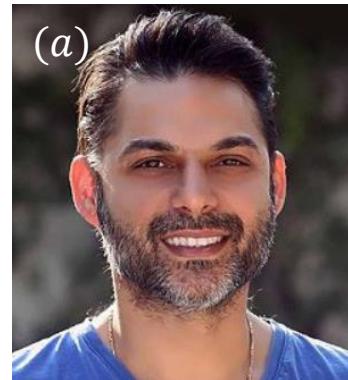
□ تابع هزینه سه‌گانه.



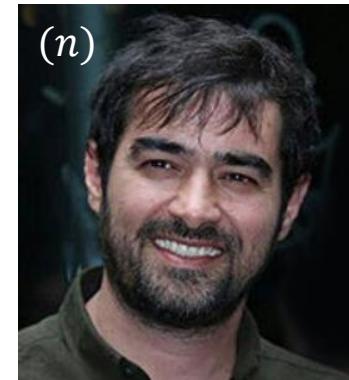
نگر



مثبت



نگر



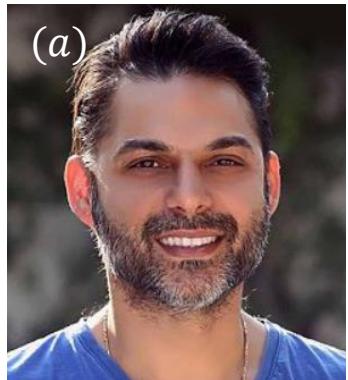
منفی

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

شبکه سیاهی: آموزش

۵۲

□ تابع هزینه سه‌گانه.



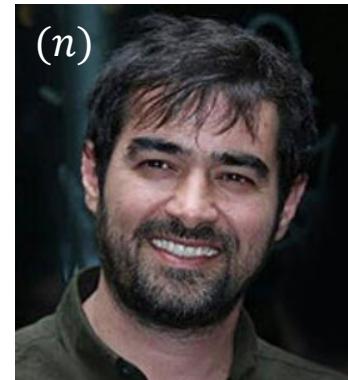
نگر



مثبت



نگر



منفی

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

اجتناب از پاسخ‌های بدیعی!

شبکه سیاهی: آموزش

۵۳

- تابع هزینه سه‌گانه.

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

- انتخاب یک مجموعه از سه‌تایی‌های مناسب برای مجموعه آموزشی.
- انتخاب سه‌تایی‌هایی که یادگیری آنها برای شبکه عصبی دشوار است!

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

شبکه سیاهی: آموژش

۵۴

لنگر



مثبت



منفی



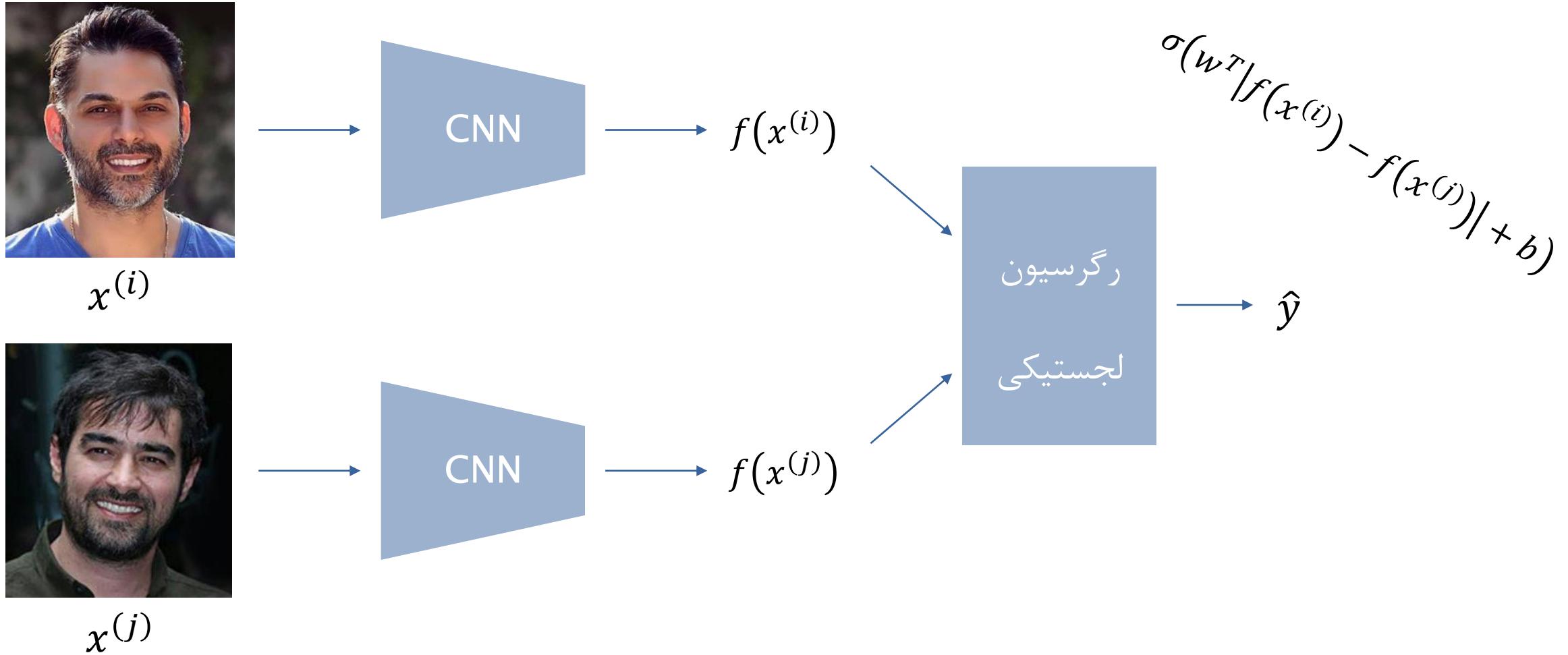
:

:

:

تابع هزینه دودویی

۵۵



چالش شناسایی نهنگ (کل)

۵۶



تعداد ۹۸۵۰ تصویر

بیش از ۳۰۰۰ نهنگ



انتقال سبک

۵۷

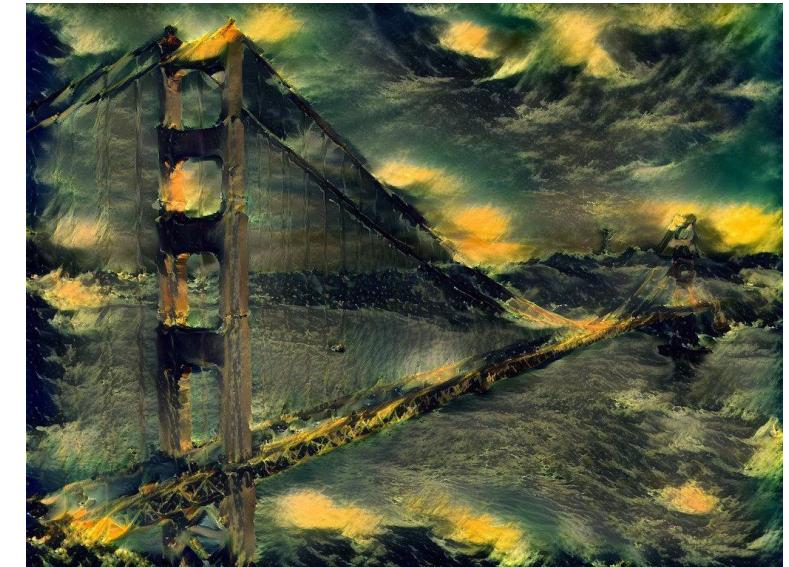
محتوا



سبک



تصویر تولید شده



تصویرسازی یادگیری شبکه‌های عمیق

۵۸

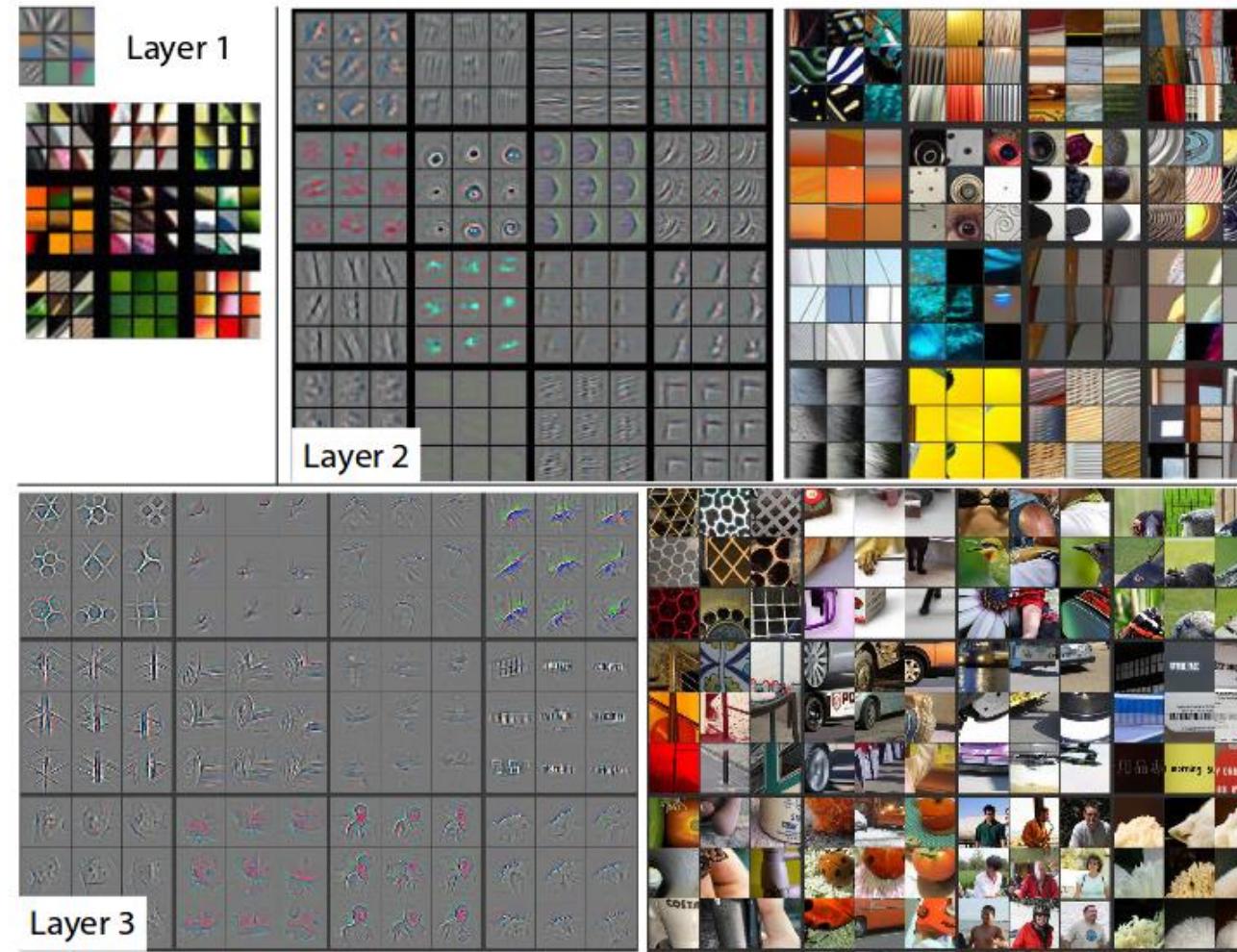
□ لایه کانولوشن اول.

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 کاوشی (!) برای کمینه‌سازی تابع هزینه

تابع هزینه محتوا

۶۲

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

- 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

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

$$a_{i,j,k}^{(l)} \quad G^{(l)}(S)$$

□ انتخاب لایه مخفی l
□ تعریف سبک به عنوان **همبستگی** میان فعالیت‌ها بر روی کانال‌ها
□ محاسبه خروجی (فعالیت) لایه مخفی برای تصاویر S و G

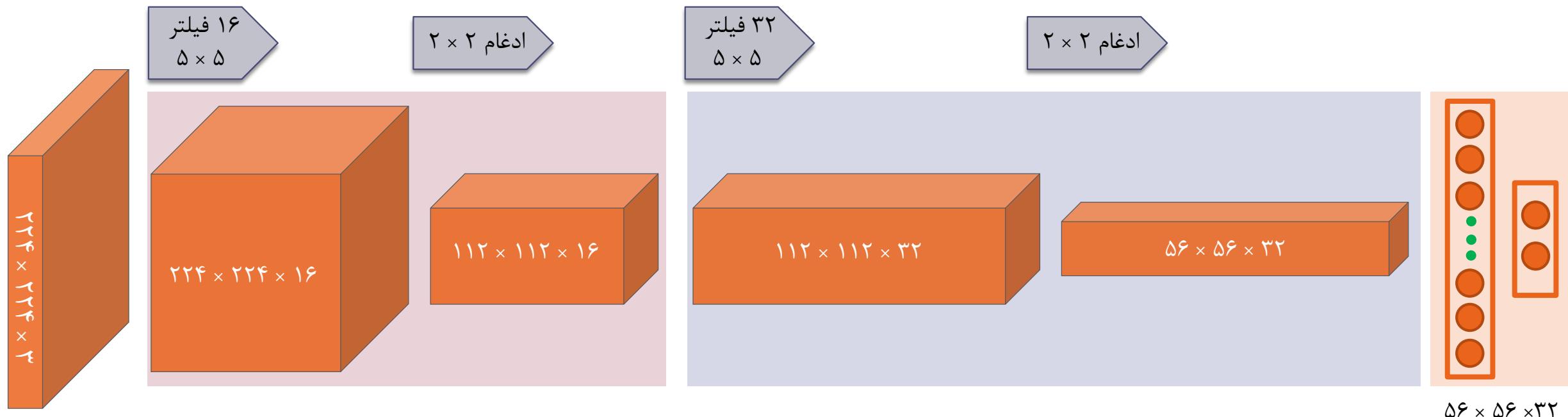
$$G_{kk'}^{(l)}(S) = \sum_{i=1}^{n_H^{(l)}} \sum_{j=1}^{n_W^{(l)}} a_{ijk}^{(l)} \cdot a_{ijk'}^{(l)}$$

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

پیوست: هفته اول

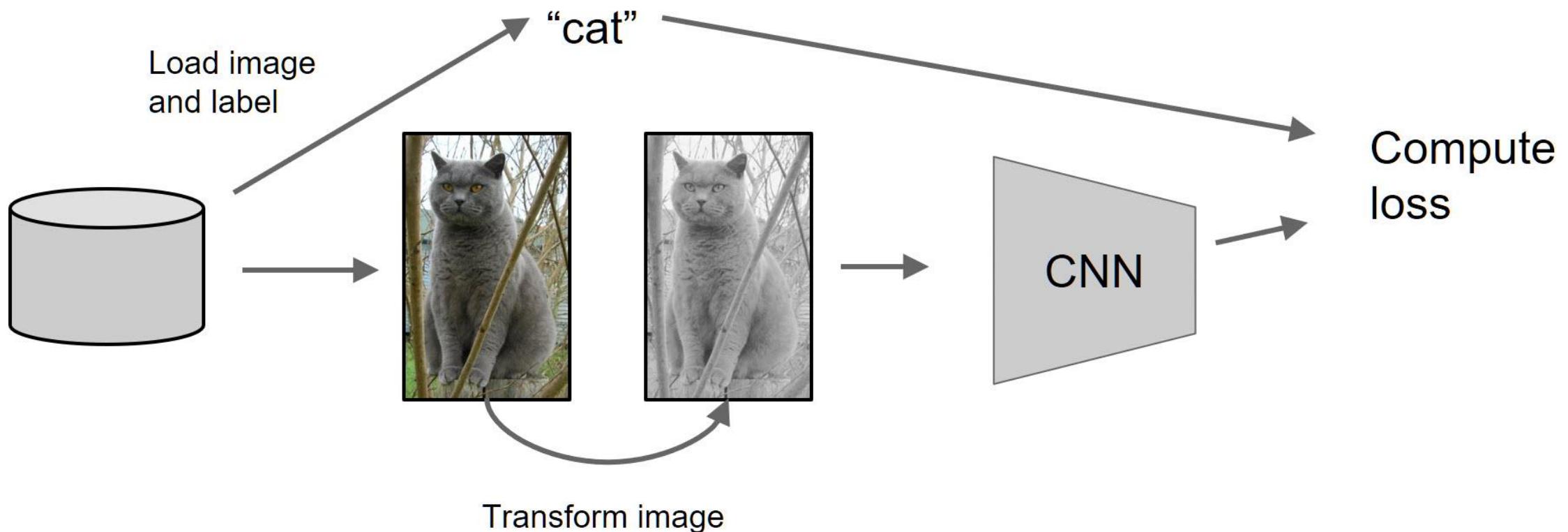
۶۴

□ شبکه کانولوشنی برای دسته‌بندی سگ و گربه



پیوست: داده افزایی

۶۵



پیوست: داده افزایی

۶۶

