

رسالة محمد



# یادگیری عمیق

مدرس: محمدرضا محمدی

بهار ۱۴۰۲

روش‌شناسی کاربردی

Practical Methodology

# جستجوی معماری مشتق پذیر (DARTS)

- فضای جستجو با قرار دادن وزن برای هر حالت به صورت پیوسته مدل می شود
- اتصالات ضعیف حذف می شوند!

$$\bar{o}^{(i,j)}(x) = \sum_{o \in \mathcal{O}} \frac{\exp(\alpha_o^{(i,j)})}{\sum_{o' \in \mathcal{O}} \exp(\alpha_{o'}^{(i,j)})} o(x)$$

---

**Algorithm 1:** DARTS – Differentiable Architecture Search

---

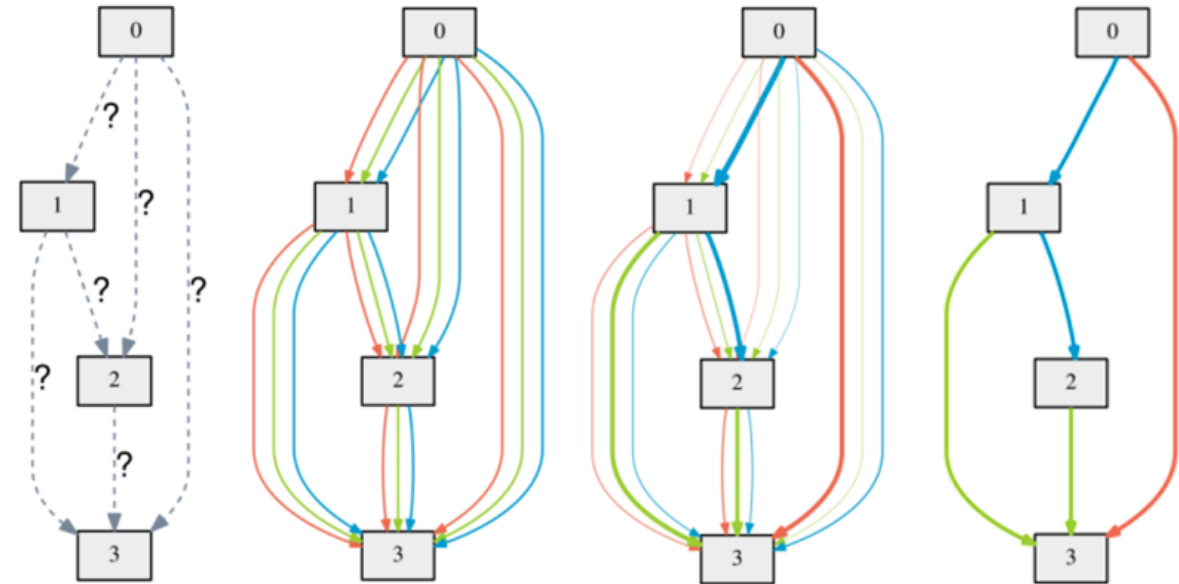
Create a mixed operation  $\bar{o}^{(i,j)}$  parametrized by  $\alpha^{(i,j)}$  for each edge  $(i, j)$

**while** not converged **do**

1. Update architecture  $\alpha$  by descending  $\nabla_{\alpha} \mathcal{L}_{val}(w - \xi \nabla_w \mathcal{L}_{train}(w, \alpha), \alpha)$   
( $\xi = 0$  if using first-order approximation)
2. Update weights  $w$  by descending  $\nabla_w \mathcal{L}_{train}(w, \alpha)$

Derive the final architecture based on the learned  $\alpha$ .

---



# جستجوی معماری مشتق پذیر (DARTS)

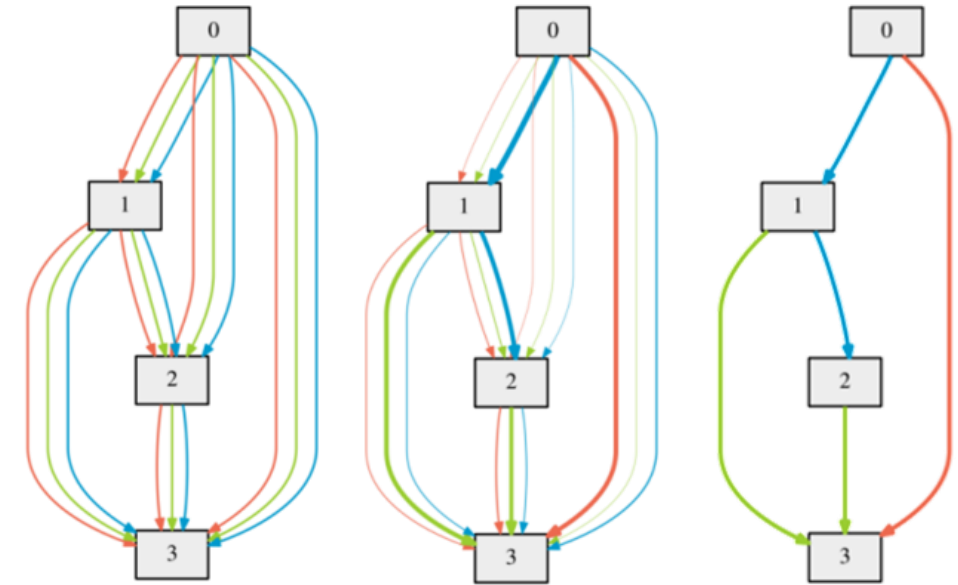
Table 1: Comparison with state-of-the-art image classifiers on CIFAR-10 (lower error rate is better). Note the search cost for DARTS does not include the selection cost (1 GPU day) or the final evaluation cost by training the selected architecture from scratch (1.5 GPU days).

Architecture	Test Error (%)	Params (M)	Search Cost (GPU days)	#ops	Search Method
DenseNet-BC (Huang et al., 2017)	3.46	25.6	–	–	manual
NASNet-A + cutout (Zoph et al., 2018)	2.65	3.3	2000	13	RL
NASNet-A + cutout (Zoph et al., 2018) <sup>†</sup>	2.83	3.1	2000	13	RL
BlockQNN (Zhong et al., 2018)	3.54	39.8	96	8	RL
AmoebaNet-A (Real et al., 2018)	3.34 ± 0.06	3.2	3150	19	evolution
AmoebaNet-A + cutout (Real et al., 2018) <sup>†</sup>	3.12	3.1	3150	19	evolution
AmoebaNet-B + cutout (Real et al., 2018)	2.55 ± 0.05	2.8	3150	19	evolution
Hierarchical evolution (Liu et al., 2018b)	3.75 ± 0.12	15.7	300	6	evolution
PNAS (Liu et al., 2018a)	3.41 ± 0.09	3.2	225	8	SMBO
ENAS + cutout (Pham et al., 2018b)	2.89	4.6	0.5	6	RL
ENAS + cutout (Pham et al., 2018b) <sup>*</sup>	2.91	4.2	4	6	RL
Random search baseline <sup>‡</sup> + cutout	3.29 ± 0.15	3.2	4	7	random
DARTS (first order) + cutout	3.00 ± 0.14	3.3	1.5	7	gradient-based
DARTS (second order) + cutout	2.76 ± 0.09	3.3	4	7	gradient-based

<sup>\*</sup> Obtained by repeating ENAS for 8 times using the code publicly released by the authors. The cell for final evaluation is chosen according to the same selection protocol as for DARTS.

<sup>†</sup> Obtained by training the corresponding architectures using our setup.

<sup>‡</sup> Best architecture among 24 samples according to the validation error after 100 training epochs.

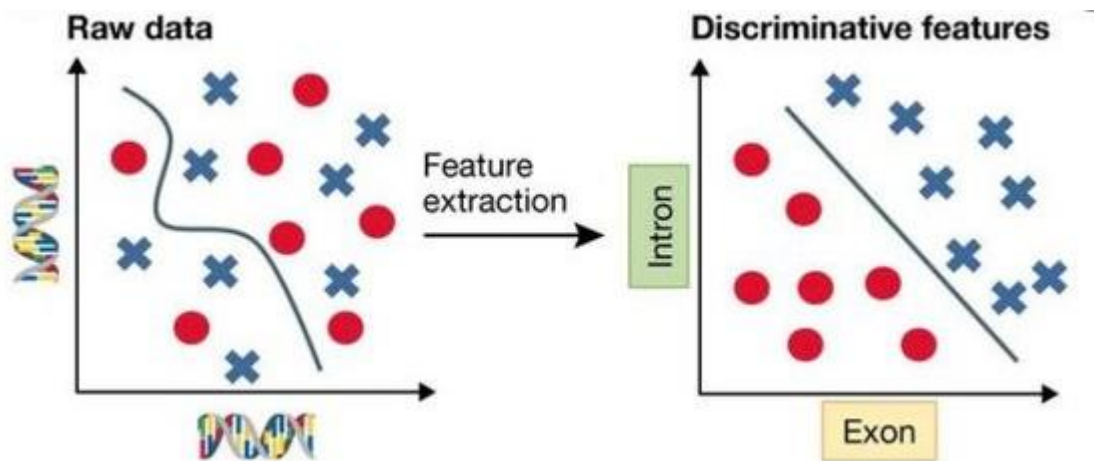


یادگیری بازنمایی

Representation Learning

# یادگیری بازنمایی

- بازنمایی مناسب داده‌های ورودی اثر بسیار زیادی در عملکرد الگوریتم‌های یادگیری ماشین دارد
- به خصوص برای داده‌های با ابعاد بالا (مانند تصاویر) بسیار مهم است
  - برای مجموعه داده‌های کوچک با چالش جدی روبرو است
- یادگیری بازنمایی می‌تواند به صورت با ناظر یا بدون ناظر انجام شود



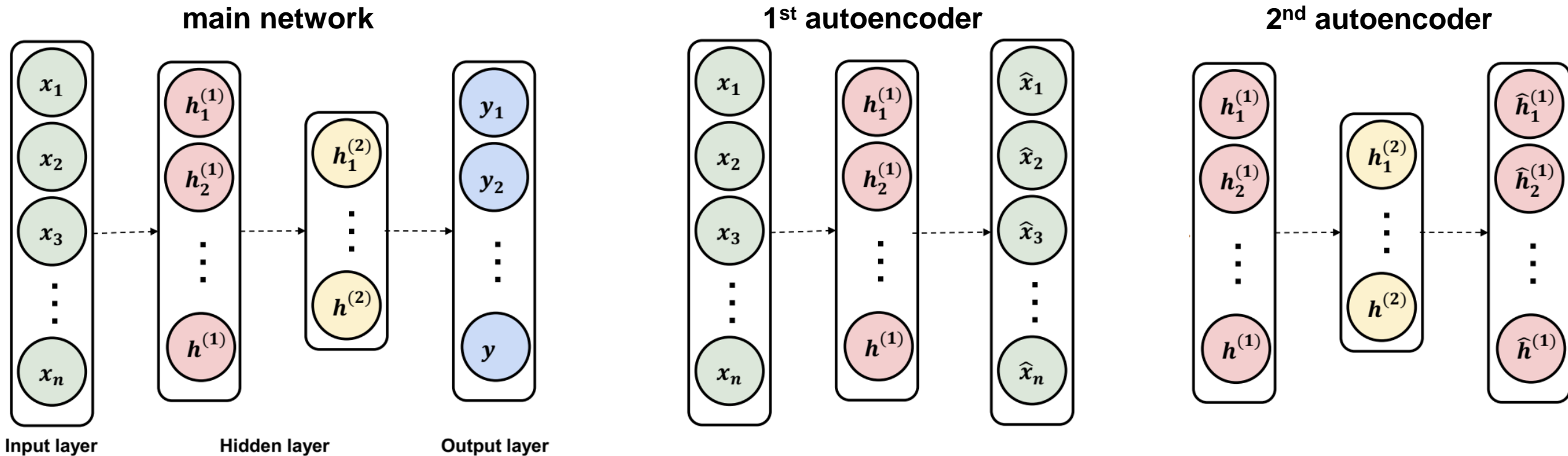
# پیش‌آموزش حریصانه لایه‌ها

- مقداردهی اولیه وزن‌های شبکه یک گام مهم در پیاده‌سازی شبکه‌های عصبی عمیق است
  - روش‌هایی مانند Xavier برای وزن‌دهی تصادفی مناسب پیشنهاد شده‌اند
- آموزش همزمان تعداد بسیار زیادی لایه متوالی با دشواری‌های زیادی همراه بوده است
  - یکی از ایده‌ها این بوده است که پیش از آموزش همزمان تمام لایه‌ها برای مسئله اصلی، وزن‌های لایه‌ها یکی یکی تنظیم شوند
  - این ایده در حدود سال ۲۰۰۶ خیلی مورد توجه بود



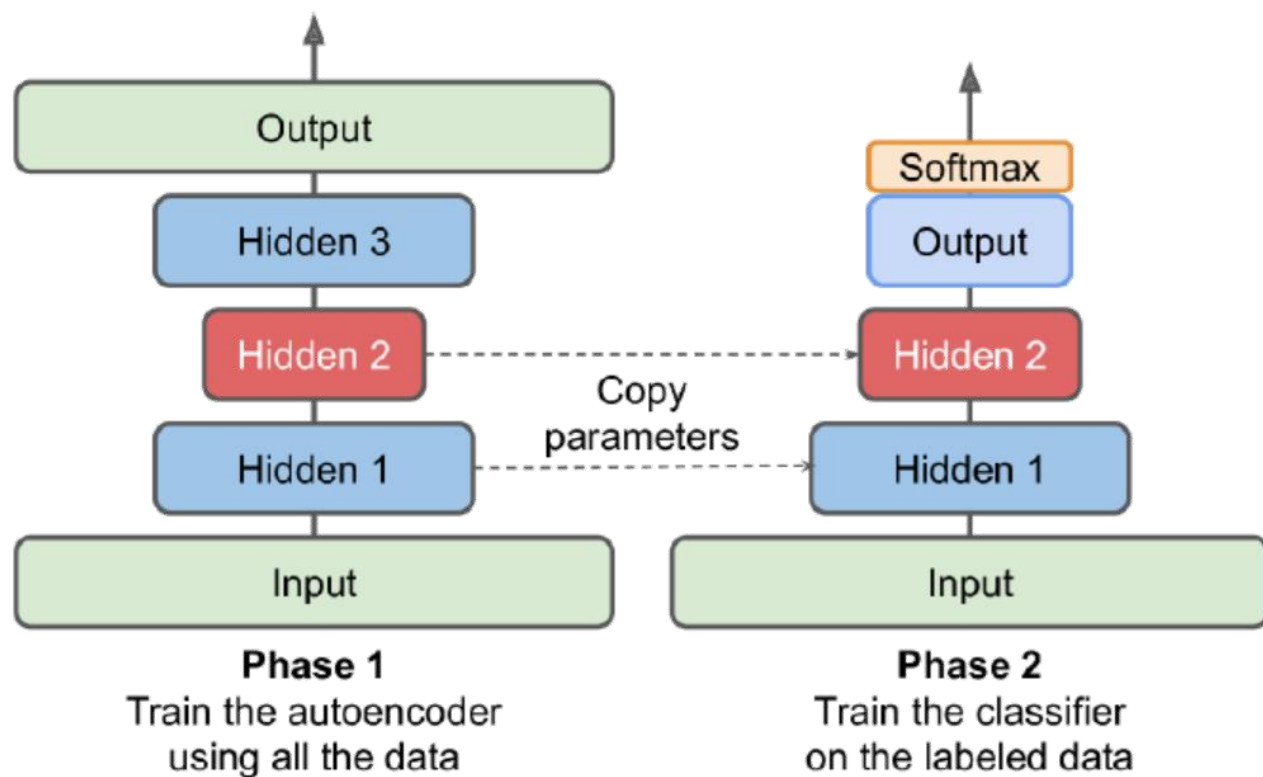
# پیش‌آموزش حریصانه لایه‌ها

- در هر گام از پیش‌آموزش، تنها وزن‌های یک لایه با هدف بازسازی داده بهینه می‌شوند
- سپس، می‌توان تنظیم دقیق با ناظر را انجام داد



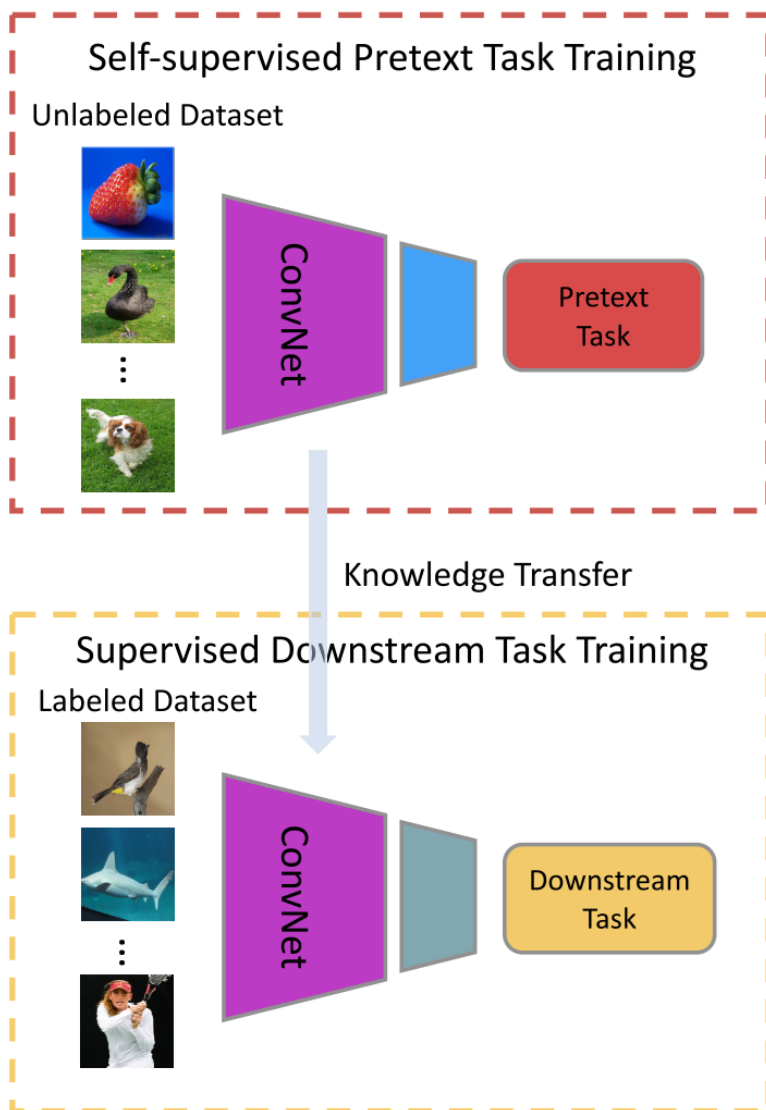
# پیش‌آموزش حریصانه لایه‌ها

- این رویکرد قبل از توسعه تکنیک‌های مدرن برای آموزش شبکه‌های بسیار عمیق (ReLU، بهینه‌سازهای بهتر، معماری‌های بهتر، نرمال‌سازی و ...) انجام می‌شد



- یادگیری بدون ناظر برای بهبود عملکرد شبکه‌های عمیق همچنان پر استفاده است
- به خصوص زمانیکه تعداد داده‌های برچسب‌خورده کم است
- در رویکردهای مدرن، استفاده از داده‌های بدون ناظر تنها برای پیش‌آموزش نیست

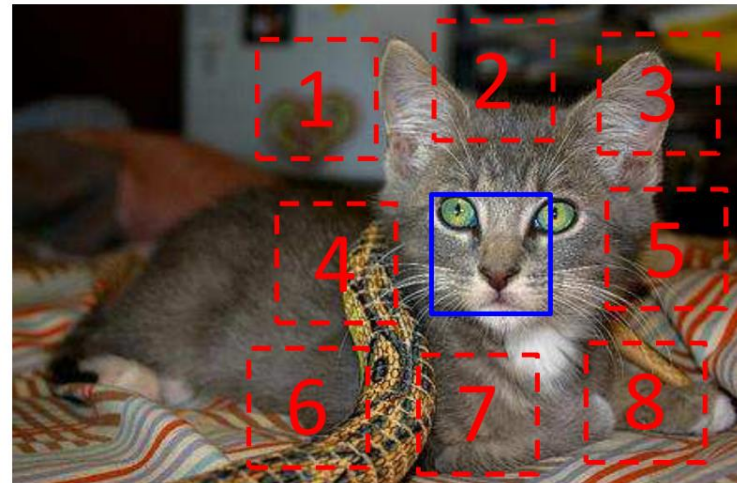
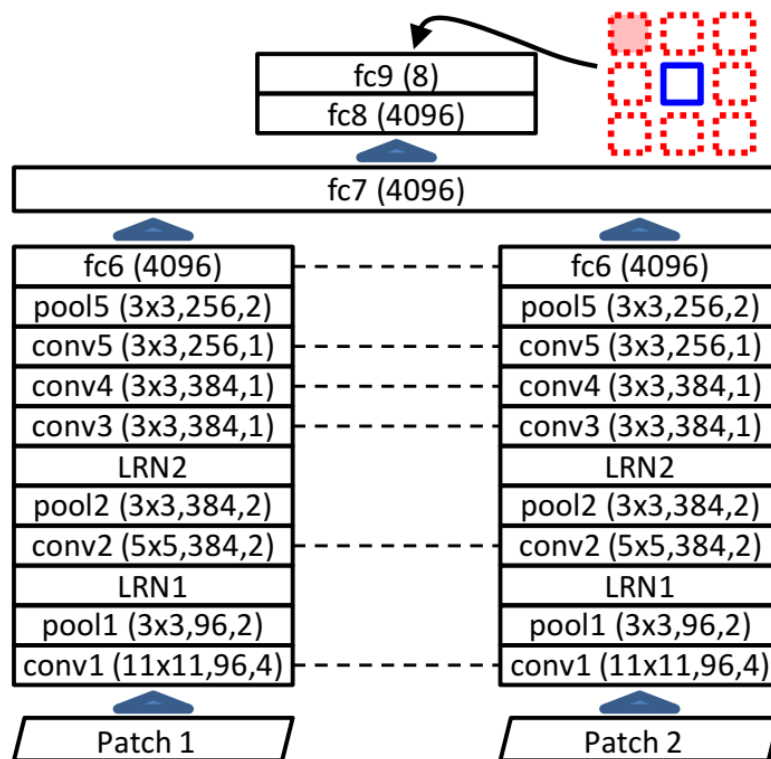
# یادگیری خودنظارتی (Self-Supervised)



- روش‌های یادگیری خود نظارتی برای یادگیری ویژگی‌های عمومی از داده‌های بدون برچسب در مقیاس بزرگ پیشنهاد شده‌اند
- مسئله‌های کمکی دو ویژگی مشترک دارند:
  - برای حل مسئله کمکی لازم است تا ویژگی‌های بصری توسط شبکه‌های کانولوشنی استخراج شوند
  - شبه‌برچسب‌های مورد نیاز در مسئله کمکی باید به صورت خودکار برای داده‌های مورد نظر قابل تولید باشند
- این رویکرد برای مسئله‌های غیر از بینایی کامپیوتر مانند پردازش صوت و پردازش متن نیز قابل استفاده است

# مثال: پیش‌بینی موقعیت

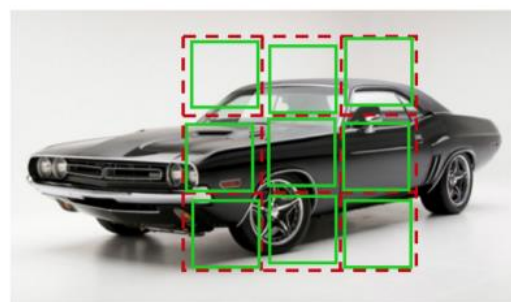
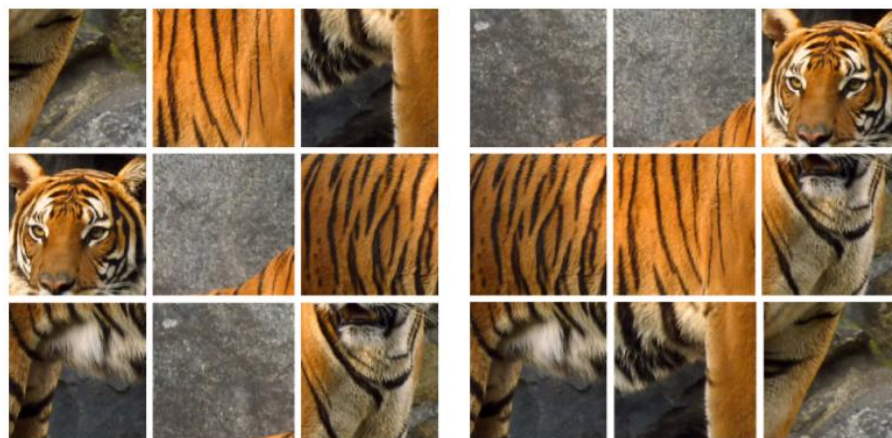
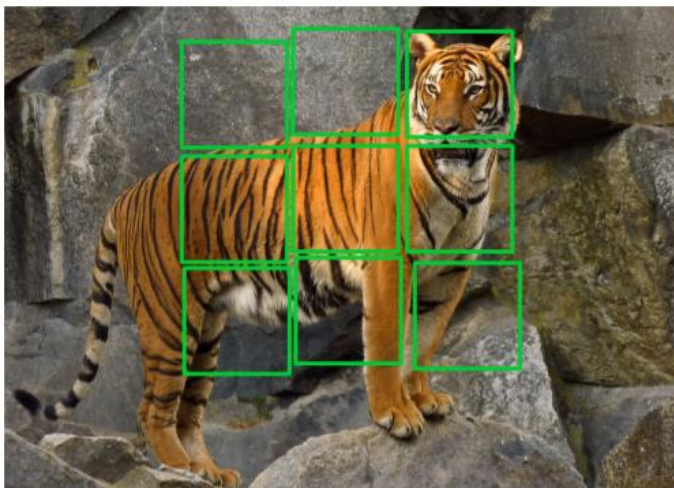
- یادگیری موقعیت نسبی اجزاء تصویر می‌تواند منجر به شناخت خوبی از محتوای تصویر شود
- برای حل این مسئله، نیاز است تا شبکه بتواند اشیاء و اجزاء آنها را یاد بگیرد



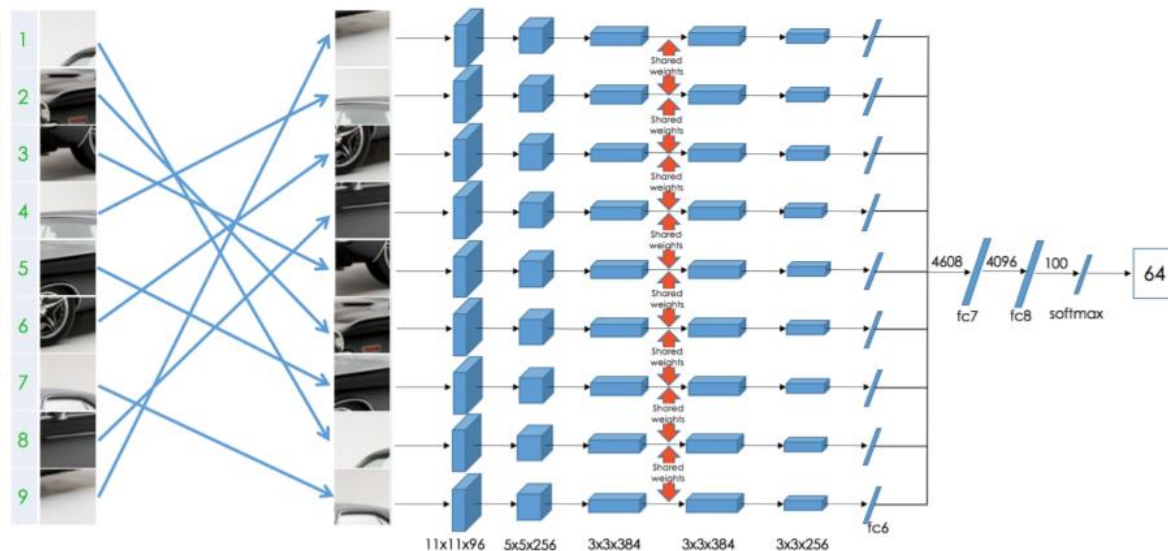
$$X = \left( \begin{array}{c} \text{cat face} \\ \text{cat ears} \end{array} \right); Y = 3$$



# مثال: حل جورچین

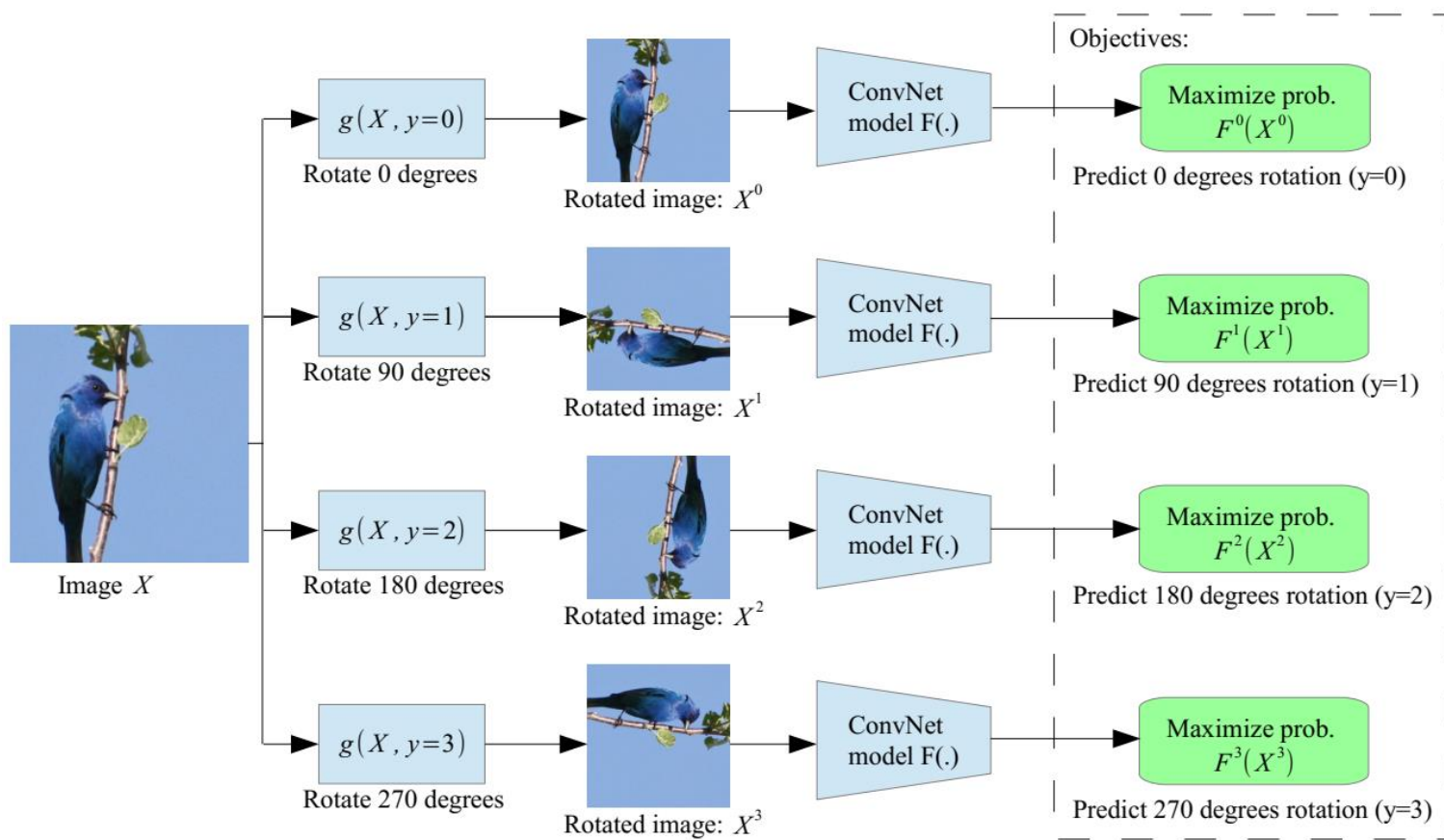


Permutation Set		
index	permutation	Reorder patches according to the selected permutation
64	9,4,6,8,3,2,5,1,7	



# مثال: تخمین چرخش

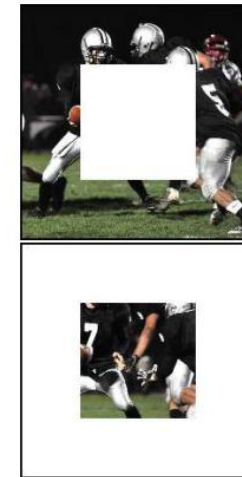
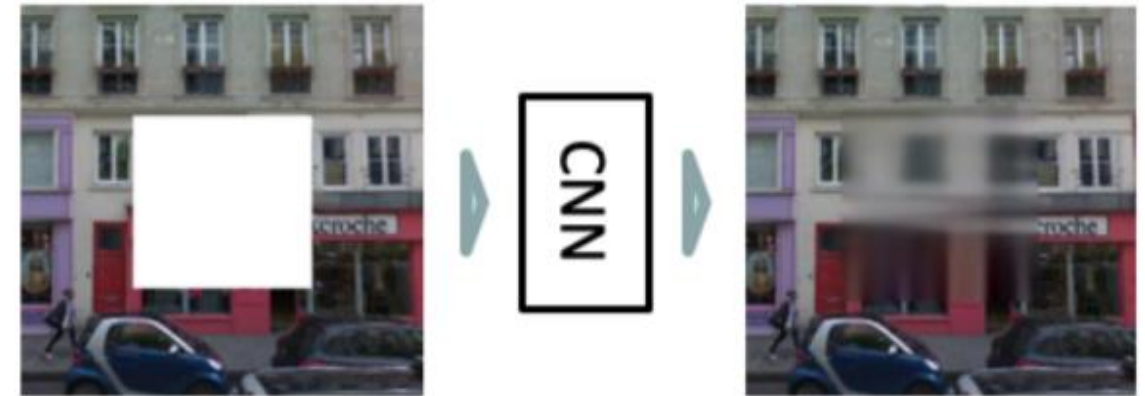
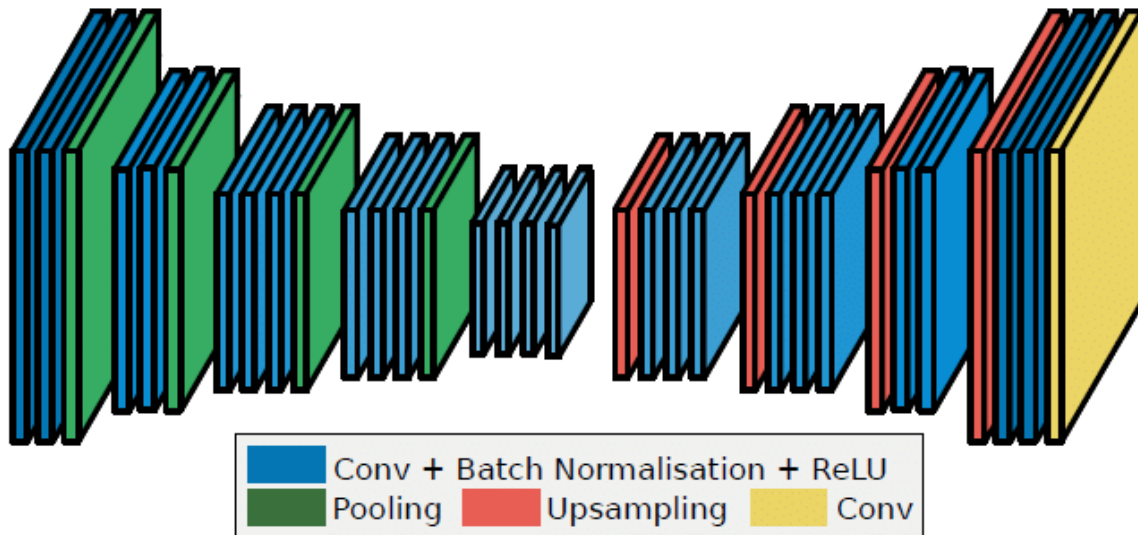
- با آموزش مدل برای تشخیص میزان چرخش ۲ بعدی، ویژگی‌های تصویر آموخته می‌شود



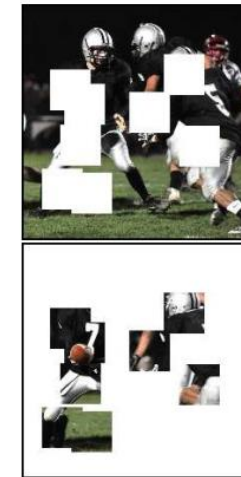
# مثال: پیش بینی محتوا

$$\mathcal{L}_{rec}(x) = \left\| \hat{M} \odot \left( x - F \left( (1 - \hat{M}) \odot x \right) \right) \right\|_2$$

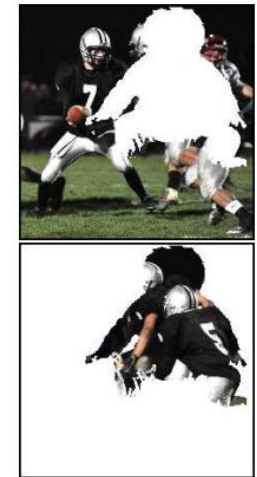
## Convolutional Encoder-Decoder



(a) Central region



(b) Random block

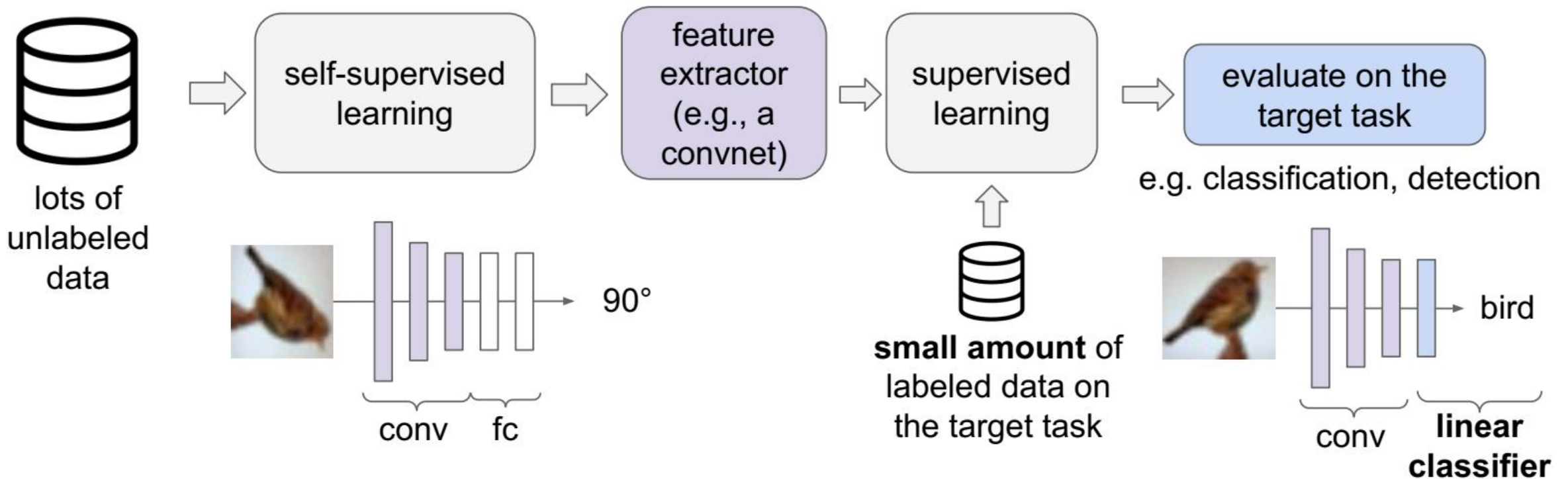


(c) Random region



# ارزیابی روش‌های SSL

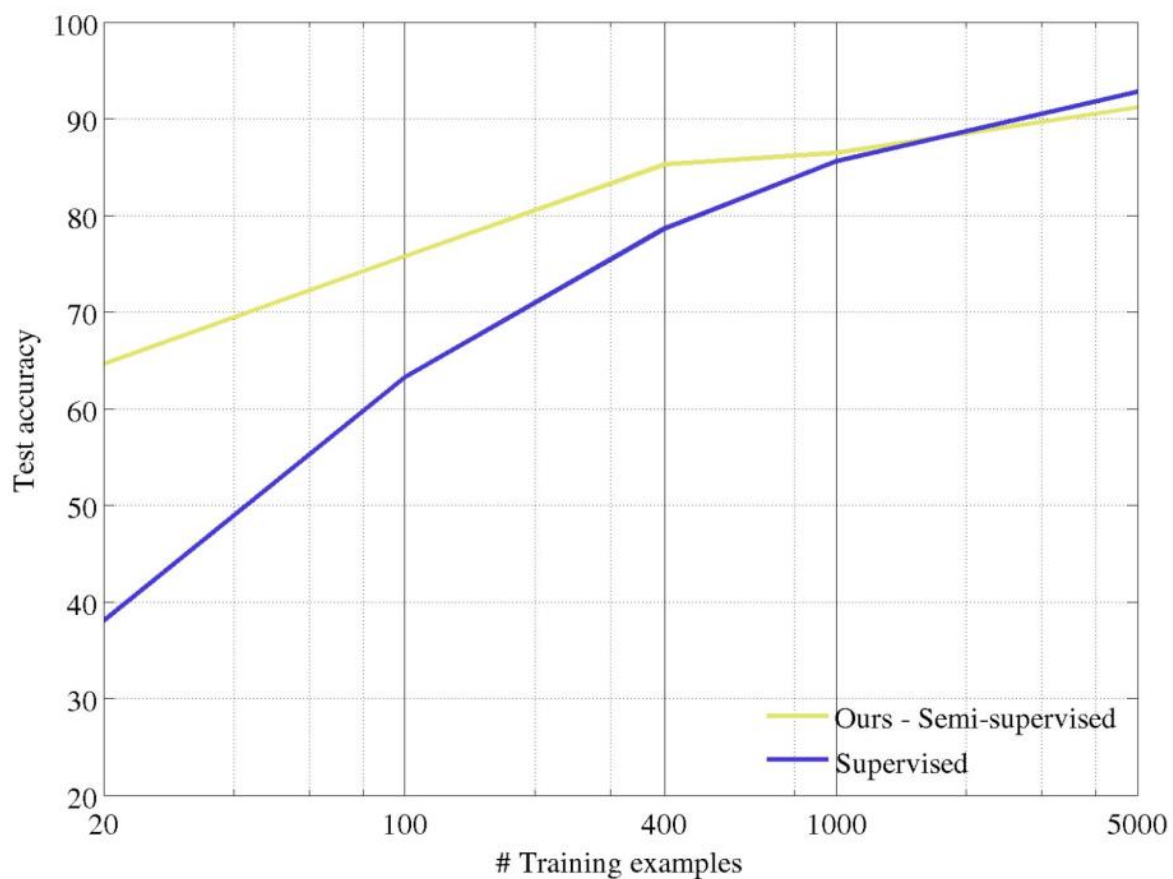
- کیفیت عملکرد مدل در حل مسئله self-supervised مورد توجه نیست بلکه نحوه عملکرد ویژگی‌های آموخته شده برای حل مسئله اصلی برای مقایسه الگوریتم‌ها استفاده می‌شود





# مثال X

- آموزش SSL بر روی تمام داده‌های CIFAR10 انجام شده است
- لایه‌های conv1 و conv2 منجمد شده‌اند و لایه‌های conv3 و linear با زیرمجموعه‌ای از داده‌های دارای برچسب آموزش داده شده است



# ارزیابی روش‌های SSL

- ارزیابی‌ها بر روی مجموعه‌داده Pascal VOC 2007 انجام شده است

- وزن‌های تصادفی عملکرد خوبی ندارند

- پیش‌آموزش روی مجموعه‌داده ImageNet بهبود بسیار خوبی ایجاد کرده است

- روش‌های مختلف یادگیری خودنظارتی تلاش کرده‌اند بدون استفاده از برچسب‌های ImageNet، ویژگی‌های خوبی را بیاموزند

	Classification (%mAP)		Detection (%mAP)	Segmentation (%mIoU)
Trained layers	fc6-8	all	all	all
ImageNet labels	78.9	79.9	56.8	48.0
Random		53.3	43.4	19.8
Random rescaled Krähenbühl et al. (2015)	39.2	56.6	45.6	32.6
Egomotion (Agrawal et al., 2015)	31.0	54.2	43.9	
Context Encoders (Pathak et al., 2016b)	34.6	56.5	44.5	29.7
Tracking (Wang & Gupta, 2015)	55.6	63.1	47.4	
Context (Doersch et al., 2015)	55.1	65.3	51.1	
Colorization (Zhang et al., 2016a)	61.5	65.6	46.9	35.6
BIGAN (Donahue et al., 2016)	52.3	60.1	46.9	34.9
Jigsaw Puzzles (Noroozi & Favaro, 2016)	-	67.6	53.2	37.6
NAT (Bojanowski & Joulin, 2017)	56.7	65.3	49.4	
Split-Brain (Zhang et al., 2016b)	63.0	67.1	46.7	36.0
ColorProxy (Larsson et al., 2017)		65.9		38.4
Counting (Noroozi et al., 2017)	-	67.7	51.4	36.6
(Ours) RotNet	<b>70.87</b>	<b>72.97</b>	<b>54.4</b>	<b>39.1</b>

# یادگیری چندوظیفه (Multi-Task Learning)

- نتایج نشان می‌دهد که ترکیب وظایف (حتی از طریق یک معماری ساده چند سر) عملکرد را بهبود می‌بخشد

