清華大學

Tsinghua University

i-VisionGroup

Towards Annotation-Efficient Learning: Few-Shot, Self-Supervised, and Incremental Learning Approaches

CVPR 2020 Tutorial

目录

- ☐ 自监督的定义

- ☐ 自监督的实现
  - 借口任务的构造
  - 应用时的改动

- ☐ 应用和评价

目录

- 自监督的定义

- 自监督的实现
  - 借口任务的构造
  - 应用时的改动

- 应用和评价

自监督的动机

- 有监督的深度学习常常表现出很好的性能
  当任务和数据集合适的时候

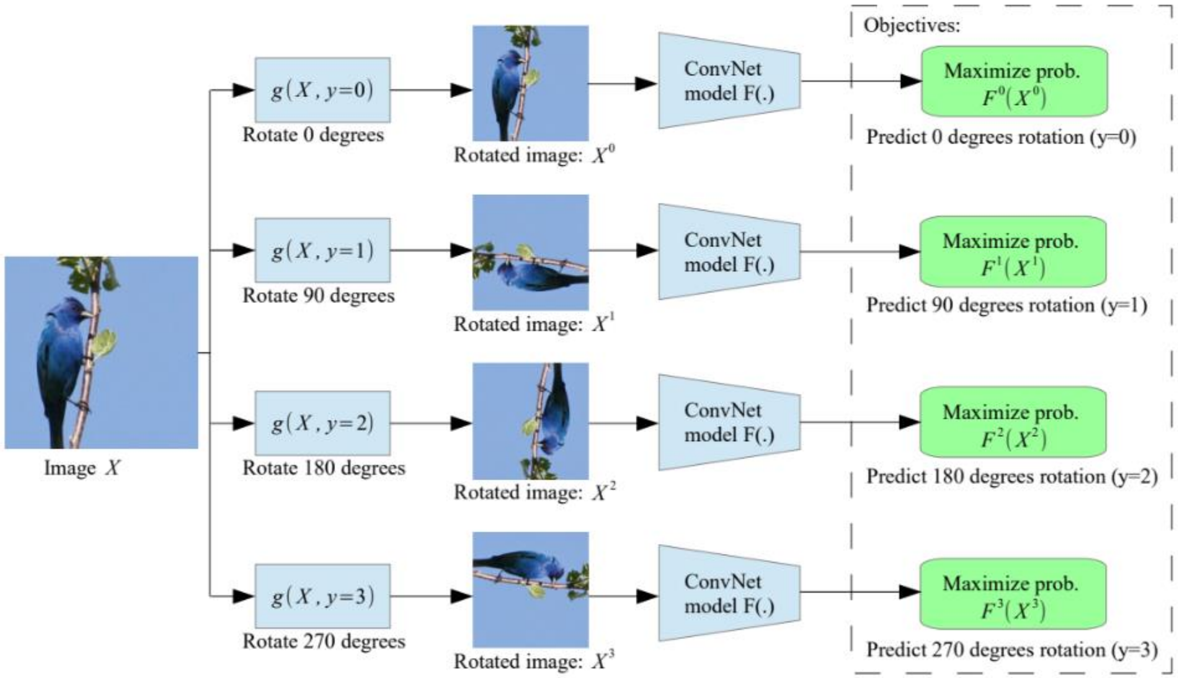
- 更希望通过表示学习获得通用的特征
  不受训练目标的影响

- 自监督模块只需要数据本身提供监督信号
  不需要额外人工标注的信号

自监督的定义

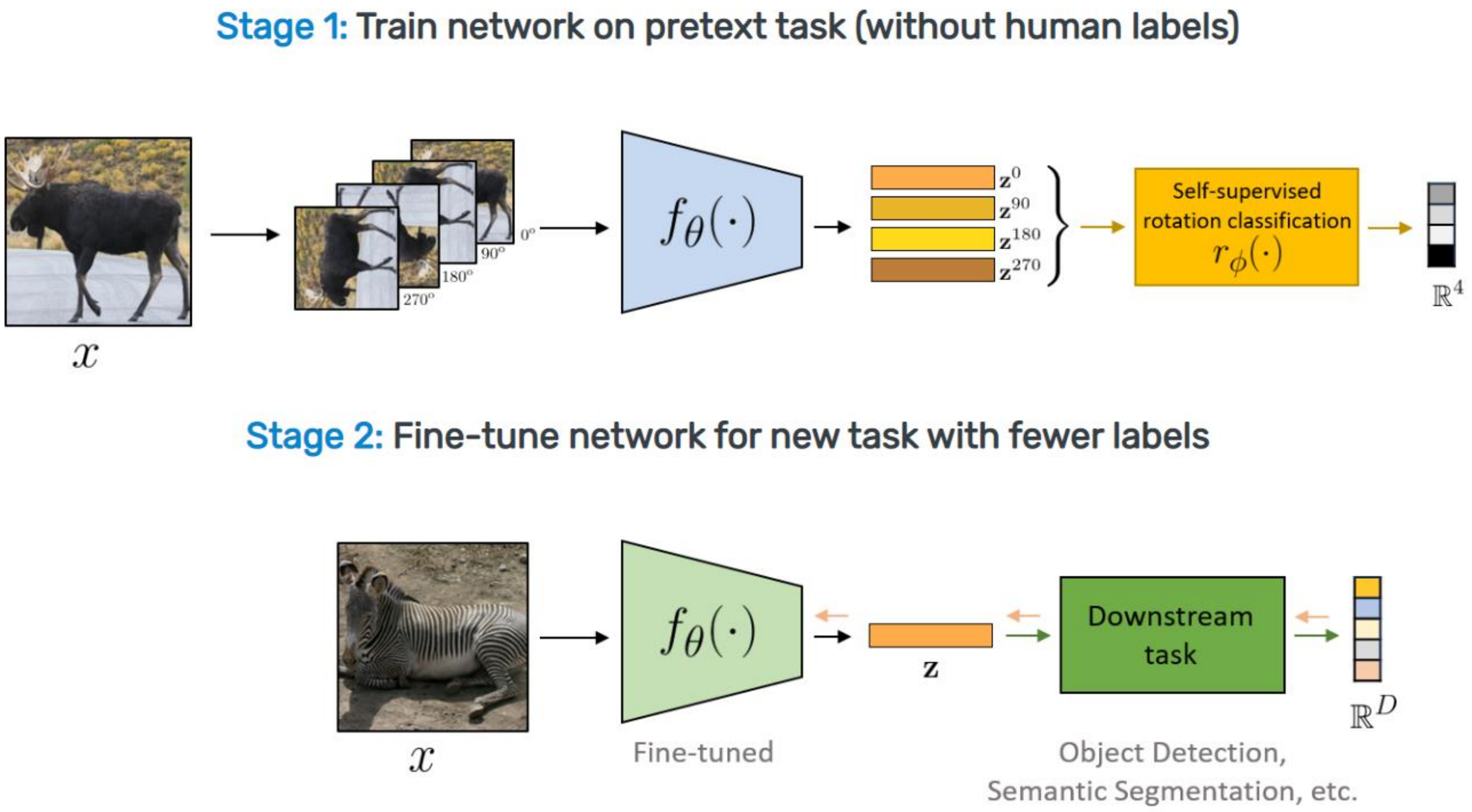
- 数据本身提供监督信号
- 定义一个pretext任务作为训练目标
- 通常辅助任务目标是数据的一部分
- 学习到的特征可以用在不同的下游任务

Example: Rotation prediction



Predict the orientation of the image

自监督流程



Stage 1: Train network on pretext task (without human labels)
x
0°
90°
180°
270°
$f_\theta(\cdot)$
$\mathbf{z}^0$
$\mathbf{z}^{90}$
$\mathbf{z}^{180}$
$\mathbf{z}^{270}$
Self-supervised rotation classification
$r_\phi(\cdot)$
$\mathbb{R}^4$
Stage 2: Fine-tune network for new task with fewer labels
x
$f_\theta(\cdot)$
Fine-tuned
$\mathbf{z}$
Downstream task
$\mathbb{R}^D$
Object Detection, Semantic Segmentation, etc.

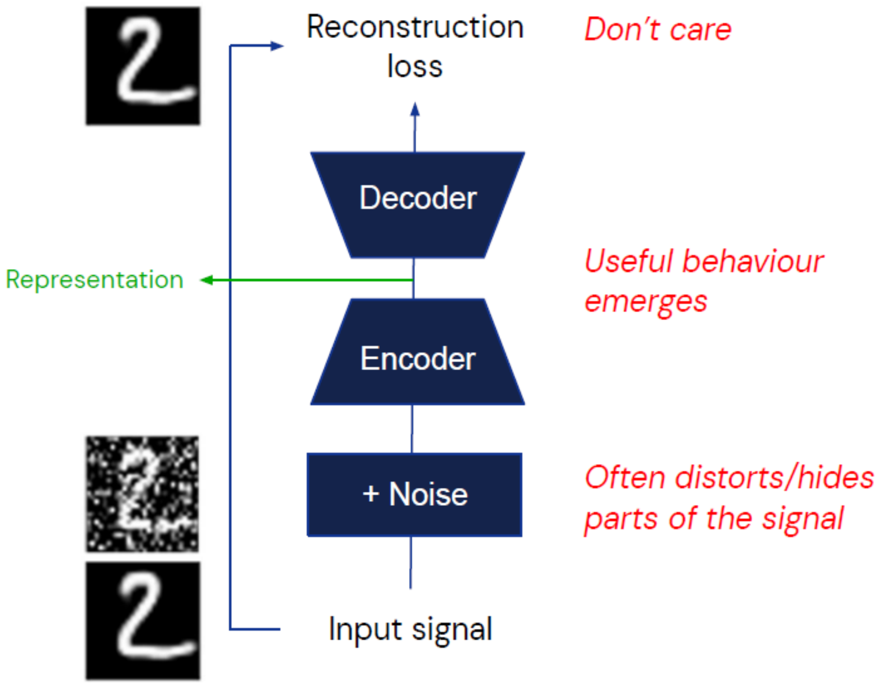
目录

- 自监督的定义
- 自监督的实现
  - 借口任务的构造
  - 应用时的改动
- 应用和评价

借口任务

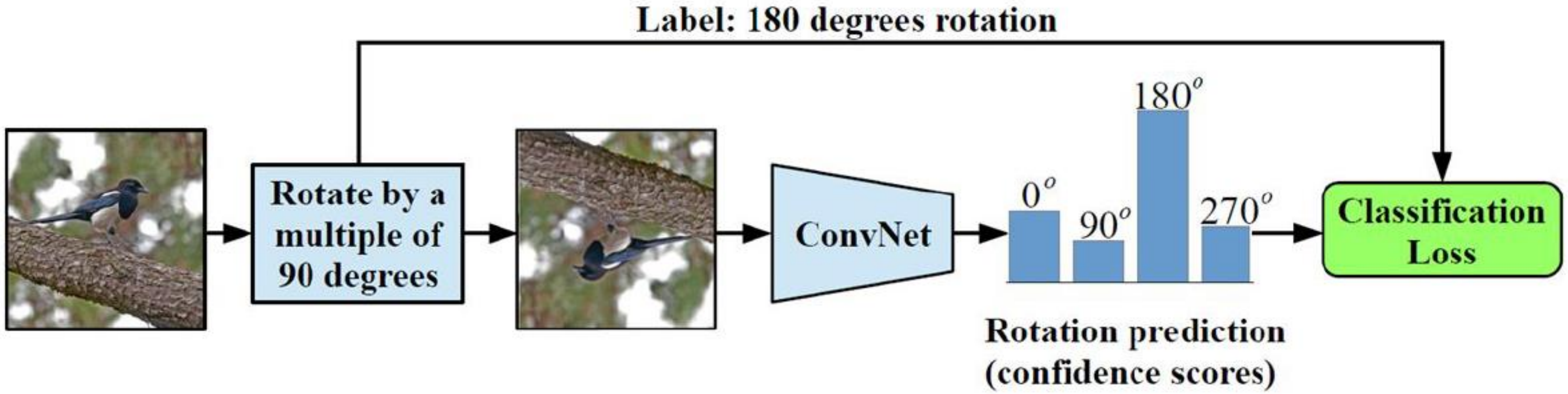
- 目的：学习一个好的特征表示
  要求：不依赖数据集标注

- 直观的辅助任务构造：
  - 旋转
  - 重建
  - 结构推断
  - 实例分类

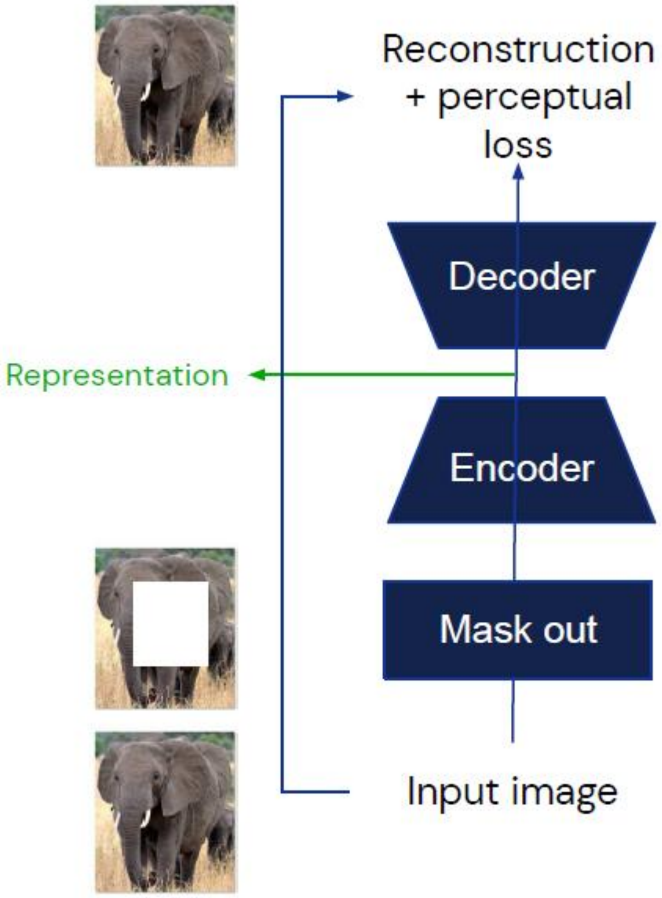


旋转

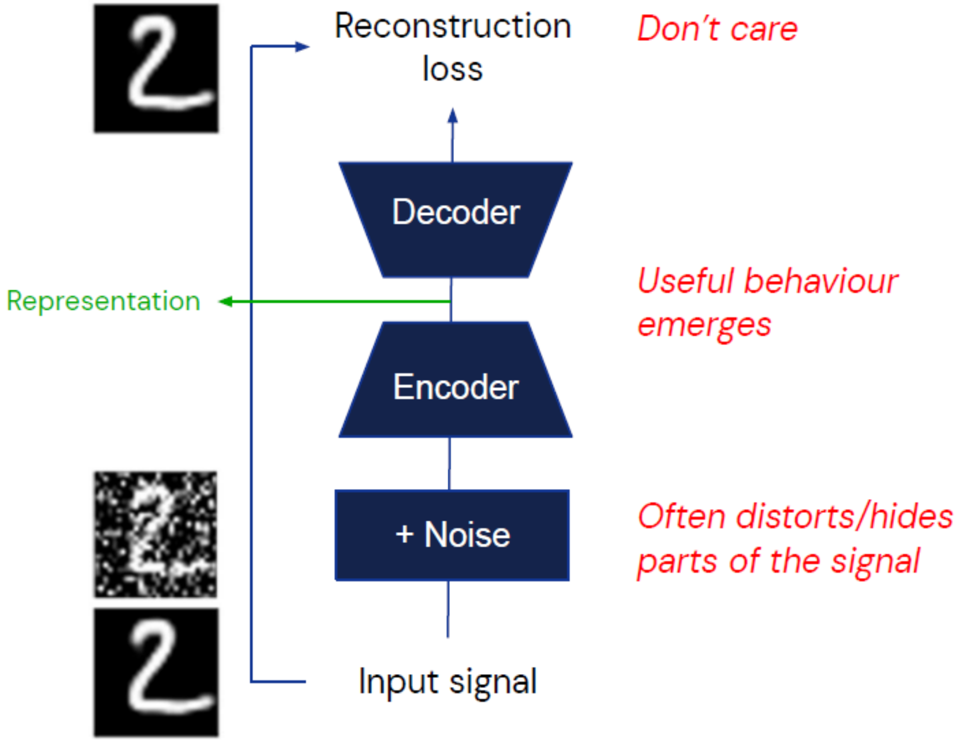
Rotation prediction



重建



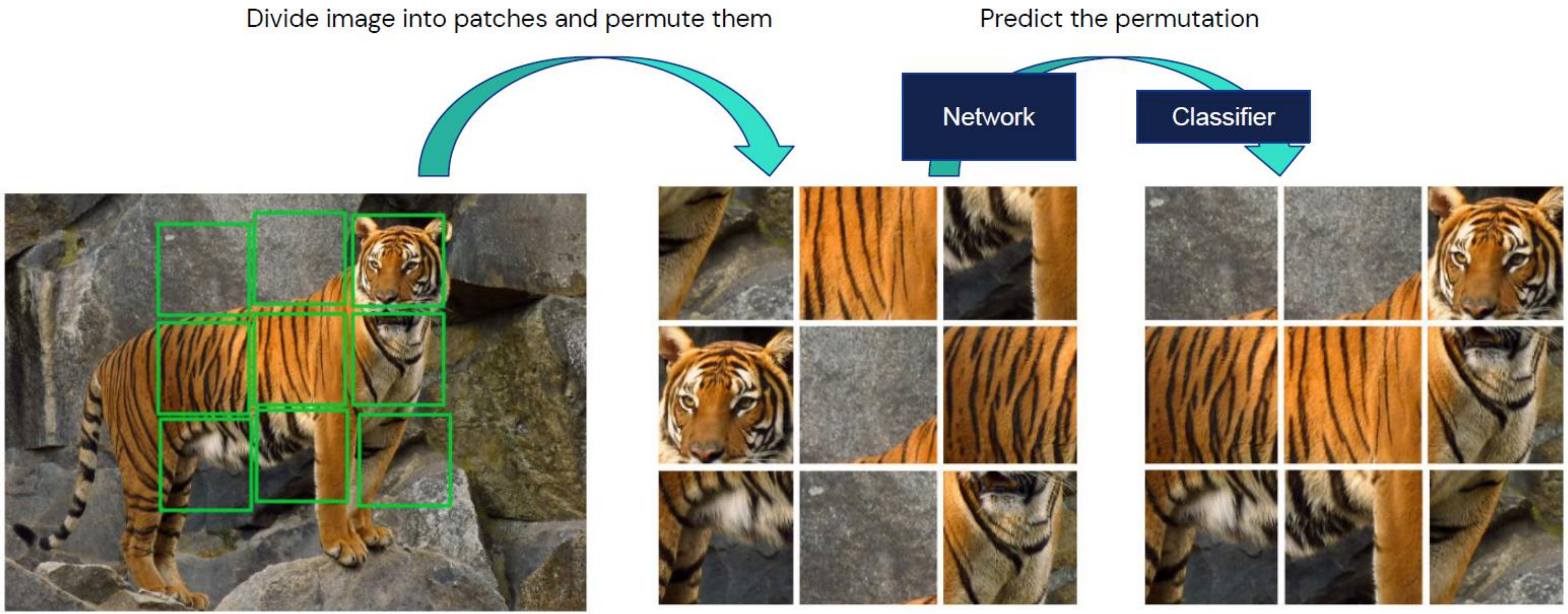
Reconstruction
+ perceptual
loss
Decoder
Representation
Encoder
Mask out
Input image



Reconstruction
loss
Don't care
Decoder
Representation
Useful behaviour
emerges
Encoder
+ Noise
Often distorts/hides
parts of the signal
Input signal

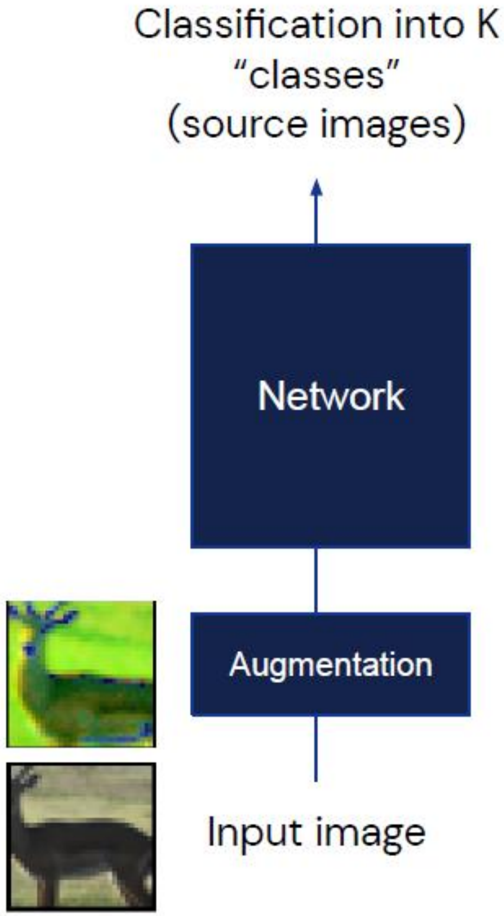
结构推断

Jigsaw puzzles



实例分类


This
is a distorted crop extracted from an image, which of these crops has the same source image?


Classification into K
"classes"
(source images)
Network
Augmentation
Input image

目录

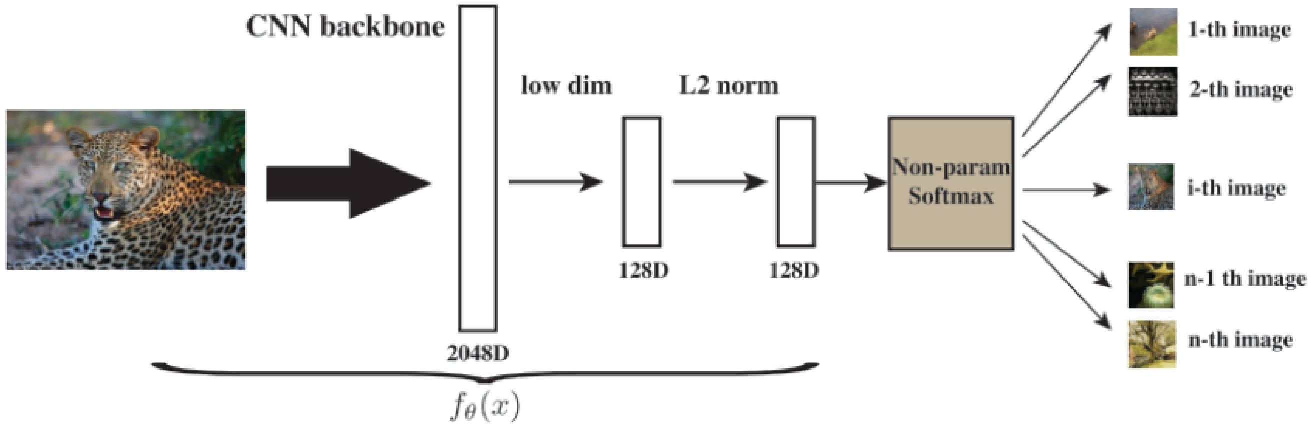
- 自监督的定义

- 自监督的实现
  - 辅助任务的构造
  - 应用时的改动

- 应用和评价

实际应用中的考虑

- 非参数化分类器



原softmax：和数据无关
非参数化：跟数据高度相关

Self-supervised learning as image instance-level discrimination

$$\mathcal{L}_{\text{non-param-softmax}}(q) = -\log \frac{\exp(q^\top k_q)}{\sum_{i\in N}\exp(q^\top k_i)} \qquad \mathcal{L}_{\text{softmax}}(q, c(q)) = -\log \frac{\exp(q^\top w_{c(q)})}{\sum_{c\in C}\exp(q^\top w_c)}$$

Z. Wu et al., Unsupervised Feature Learning via Non-Parametric Instance Discrimination, CVPR 2018

Noise-Contrastive Estimation (NCE)

- NCE：多分类变成二分类（只区分正样本和负样本）

$$\mathcal{L}_{\text{NCE}}(q) = -\log \frac{\exp(q^\top k_q)}{Z_q}$$

$$Z_q = \sum_{i \in N} \exp(q^\top k_i) \simeq \frac{N}{K} \sum_{j \in K} \exp(q^\top k_j)$$

$$\mathcal{L}_{\text{non-param-softmax}}(q) = -\log \frac{\exp(q^\top k_q)}{\sum_{i \in N} \exp(q^\top k_i)}$$

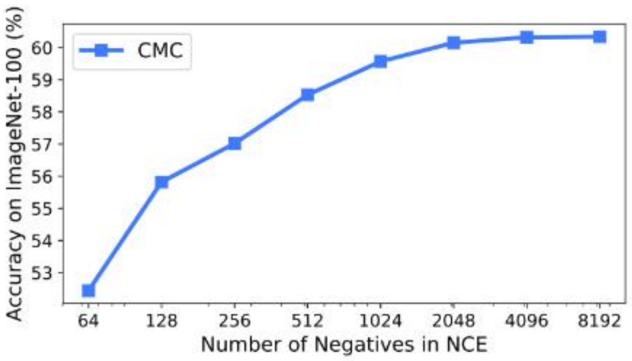
- 接着变形成infoNCE

$$\mathcal{L}_{\text{InfoNCE}}(q) = -\log \frac{\exp(q^\top k_q)}{\sum_{i \in K} \exp(q^\top k_i)}$$

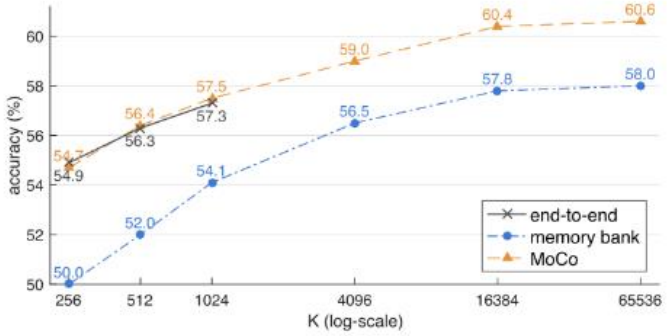
InfoNCE

- infoNCE是近些年来在对比自监督学习任务中表现最好的loss函数
  其性能和负样本的数量有关，总的趋势是负样本越多，性能越好

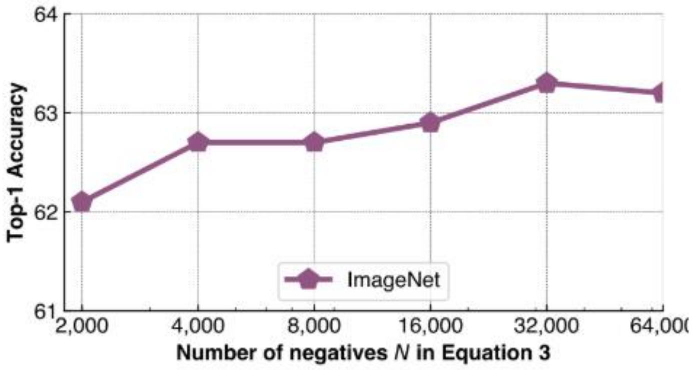
CMC [Tian et al. (2019)]



MoCo [He et al. (2020)]

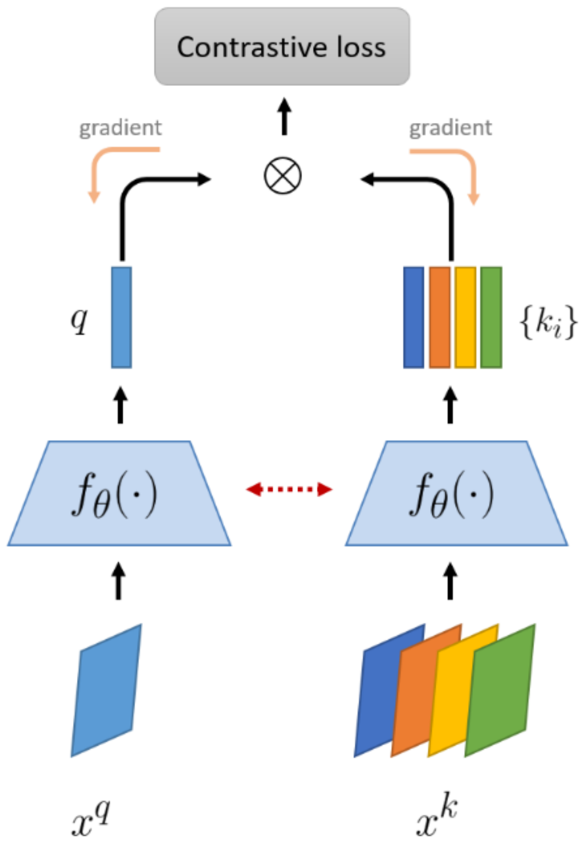


PiRL [Misra and van der Maaten (2020)]

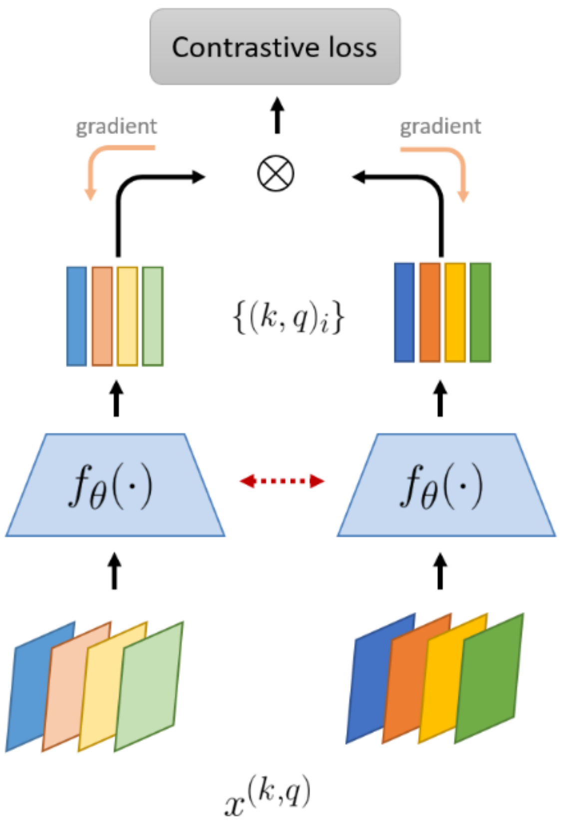


Y. Tian et al., Contrastive Multiview Coding, ArXiv 2019
K. He et al., Momentum Contrast for Unsupervised Visual Representation Learning, CVPR 2020
I. Misra and L. van der Maaten, Self-Supervised Learning of Pretext-Invariant Representations, CVPR 2020

训练方案

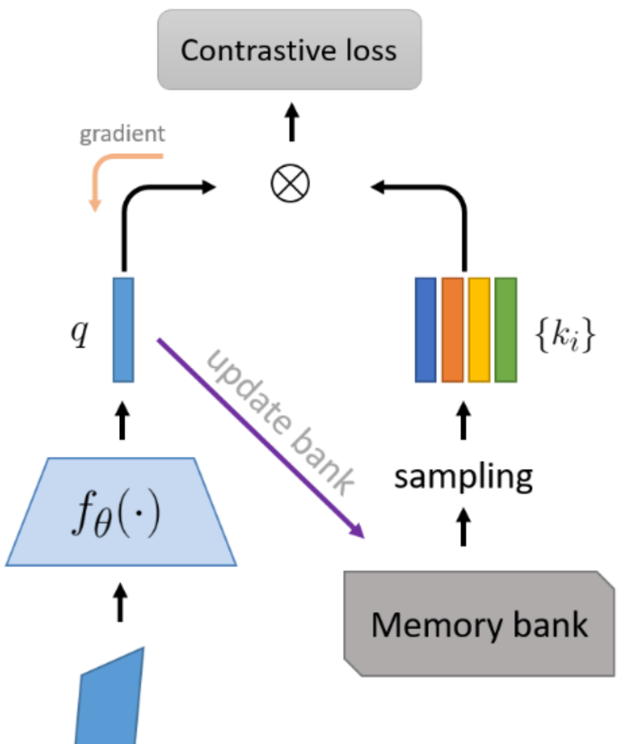


只计算一个loss



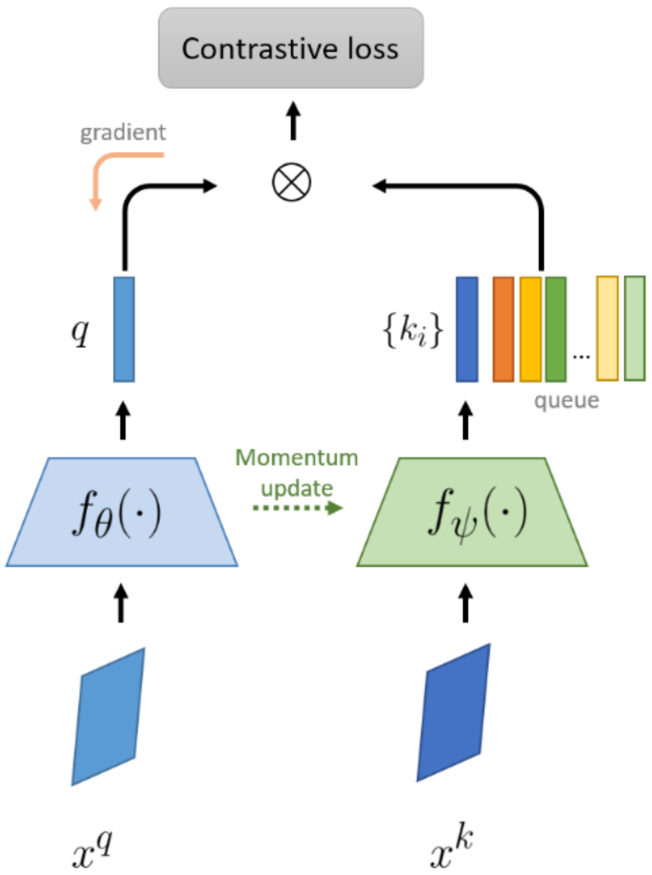
同时计算N对loss

训练方案



$$-\log \frac{\exp(q^\top k_q)}{\sum_{i \in N} \exp(q^\top k_i)}$$

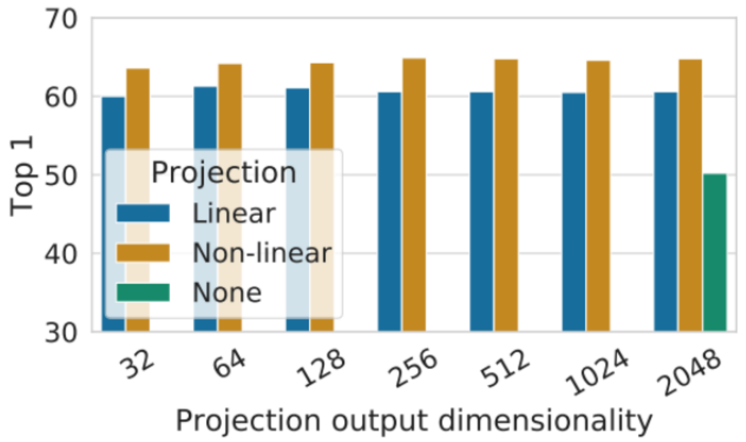
存储查询字典



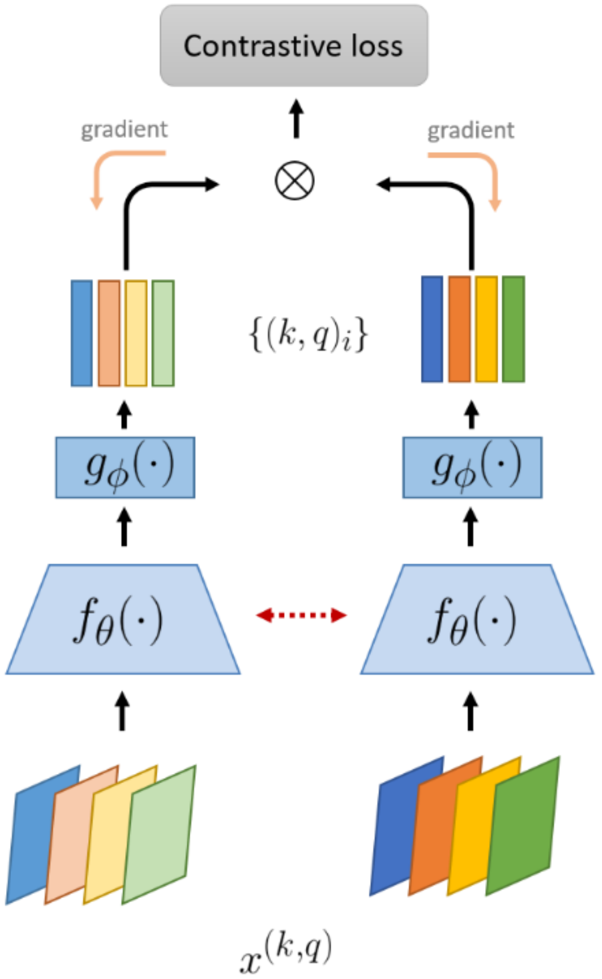
存储网络参数

网络结构

- 输出投影（output projection）
  - 增加投影层g
  - g在下游任务中被移除
  - 可以提升性能6-10%

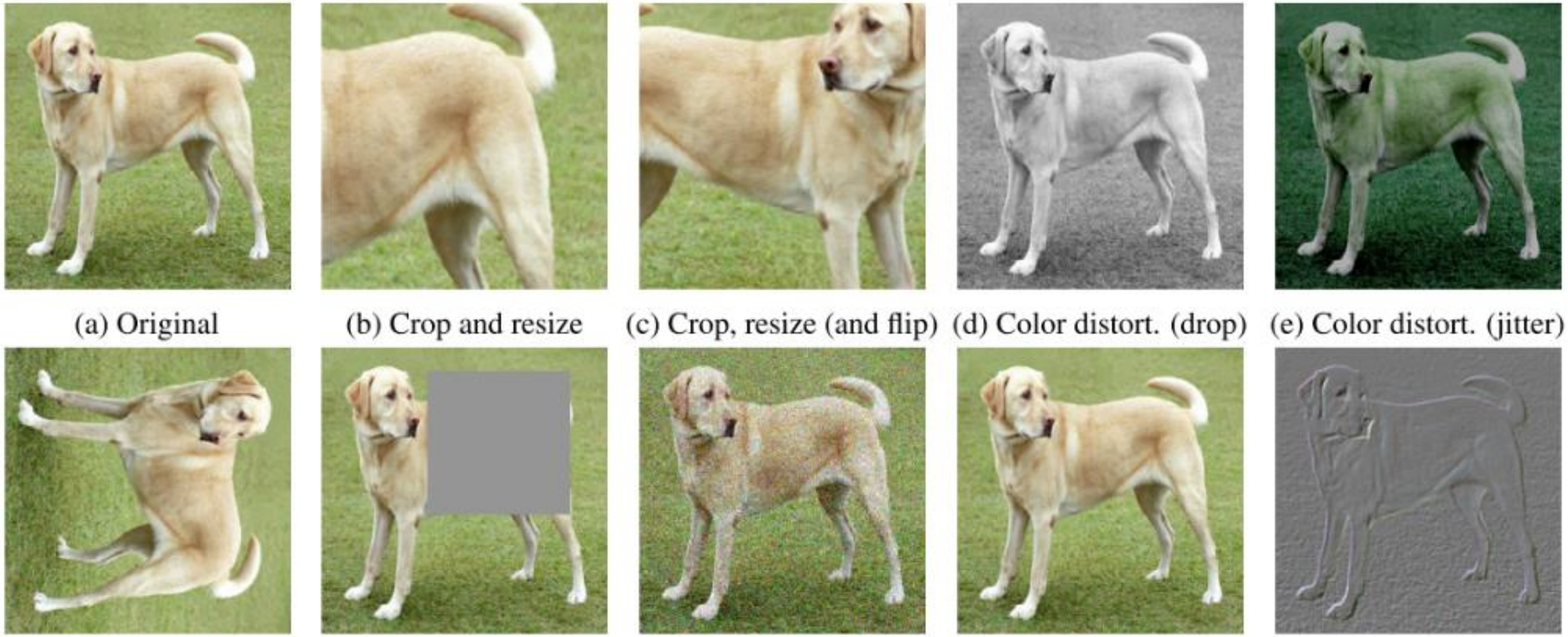


Linear evaluation of representations with different projection heads $g_\psi(\cdot)$ and various output dimensions.



T. Chen et al., A Simple Framework for Contrastive Learning of Visual Representations, ArXiv 2020
K. He et al., Improved Baselines with Momentum Contrastive Learning, ArXiv 2020

数据增广



(a) Original (b) Crop and resize (c) Crop, resize (and flip) (d) Color distort. (drop) (e) Color distort. (jitter)

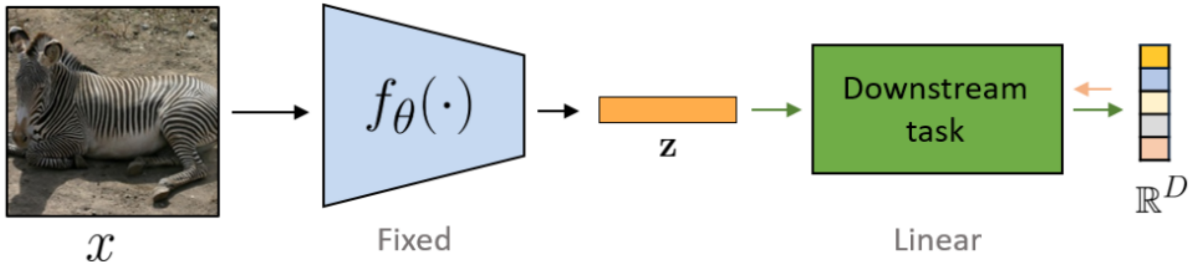
(f) Rotate {90°, 180°, 270°} (g) Cutout (h) Gaussian noise (i) Gaussian blur (j) Sobel filtering

Typical data augmentation operators used for visual representation learning

目录

评价

- 线性分类任务
  后面接全连接或者线性SVM

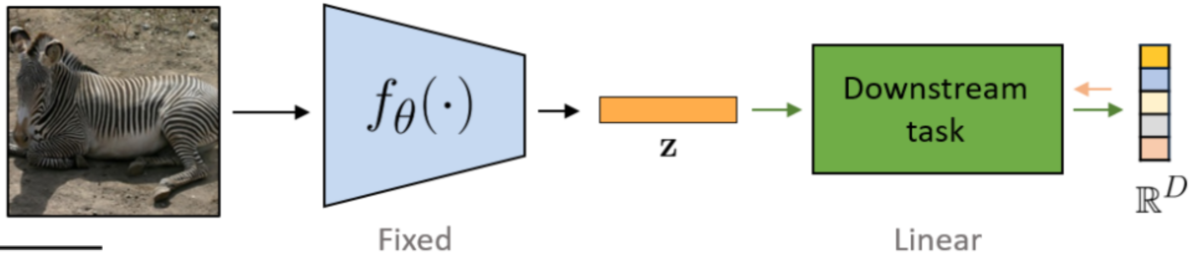


Method	Parameters	Transfer Dataset			
		ImageNet	VOC07	Places205	iNat.
ResNet-50 using evaluation setup of [19]					
Supervised	25.6M	75.9	87.5	51.5	45.4
Colorization [19]	25.6M	39.6	55.6	37.5	–
Rotation [18]	25.6M	48.9	63.9	41.4	23.0
NPID++ [72]	25.6M	59.0	76.6	46.4	32.4
MoCo [24]	25.6M	60.6	–	–	–
Jigsaw [19]	25.6M	45.7	64.5	41.2	21.3
PIRL (**ours**)	25.6M	**63.6**	**81.1**	**49.8**	**34.1**

在性能上可以接近
全监督算法

评价

- Few-shot learning
  每一类样本只有少量数据



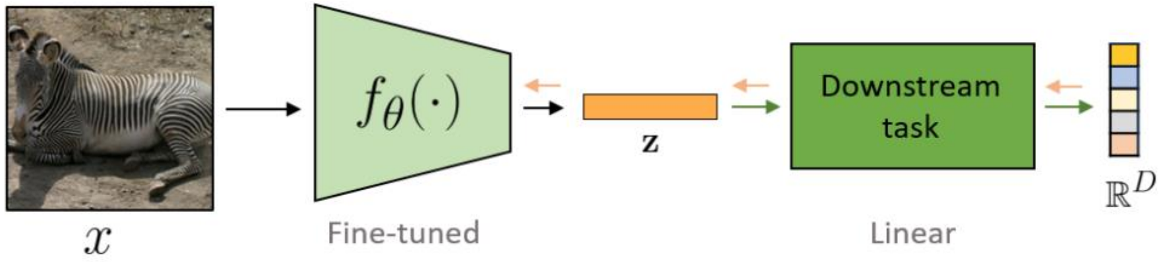
Method \ Classes	Novel n =1	5	10	50	Base Linear
Supervised CC [20]	56.8	74.1	78.1	82.7	73.7
RotNet	40.8	56.9	61.8	68.1	52.3
RelLoc [15]	40.2	57.1	62.6	68.8	50.4
Deeper Clustering	47.8	66.6	72.1	78.4	60.3
BoWNet	48.7	67.9	74.0	79.9	65.0
BoWNet ×2	**49.1**	67.6	73.6	79.9	65.6
BoWNet ×3	48.6	**68.9**	**75.3**	**82.5**	**66.0**

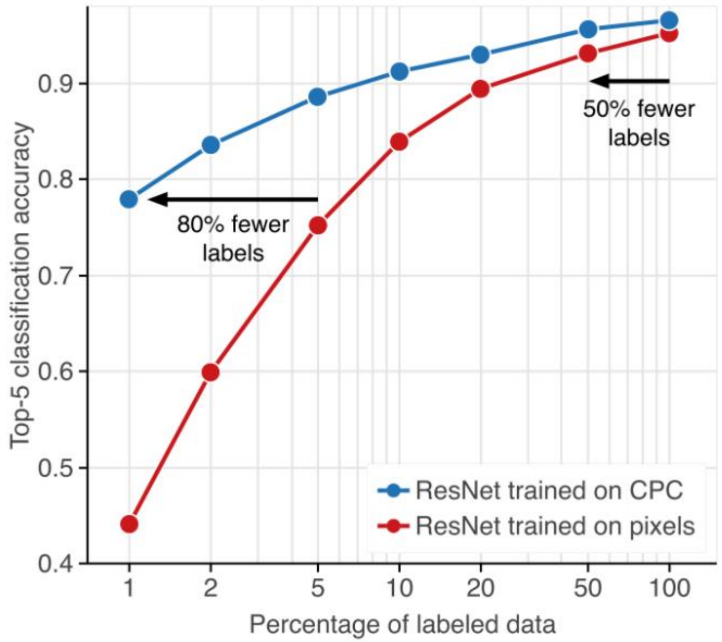
Average 5-way classification accuracies on the test set of MiniImageNet.

在性能上甚至可以
和监督算法竞争

评价

- 高效分类
  特征提取的网络参数可以微调



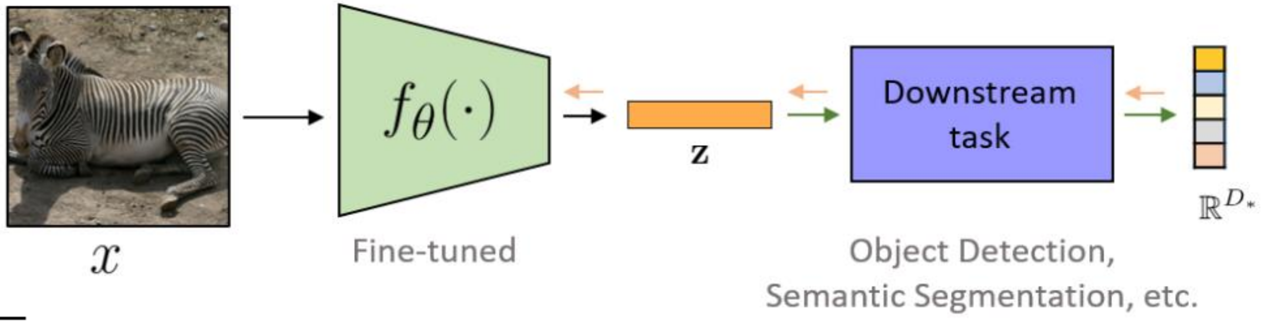


ImageNet accuracy of models trained with few labels: CPCv2 vs. supervised

在带标注数据量极少
的情况下有优势

评价

- 迁移学习
  在其他任务中的表现，如目标检



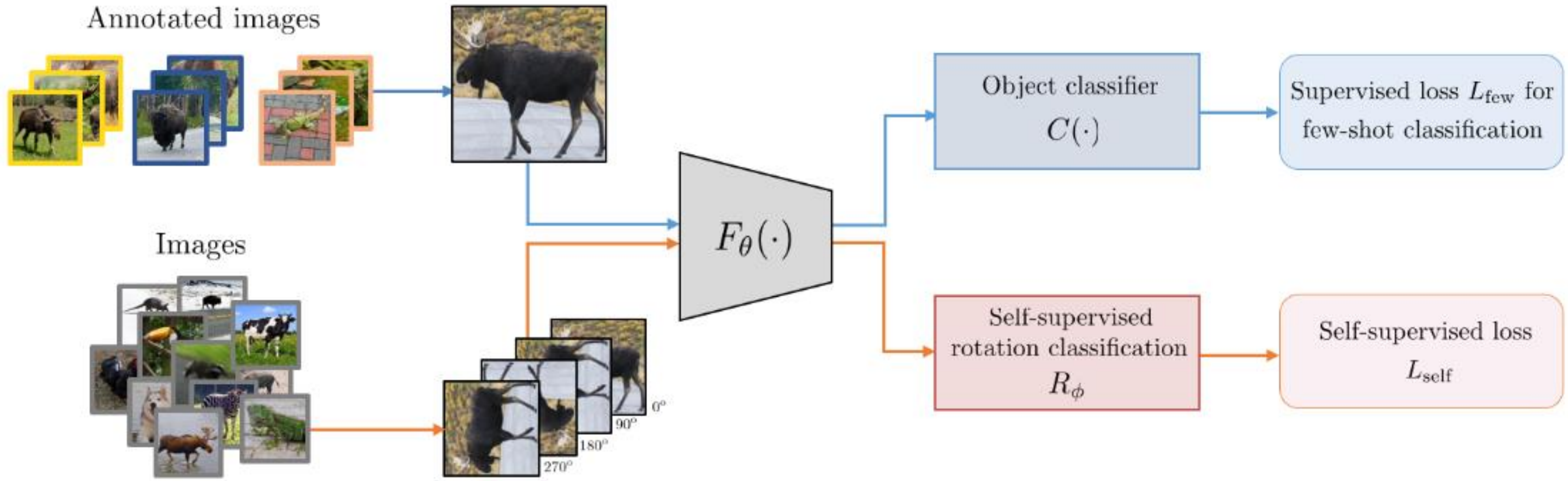
Method	Network	AP^{50}	AP^{75}	AP^{all}
Supervised*	R-50	80.8	58.5	53.2
Rotation†	R-50	72.5	49.3	46.3
Jigsaw†	R-50	75.1	52.9	48.9
NPID++†	R-50	79.1	56.9	52.3
PIRL†	R-50	80.7	59.7	54.0
MoCo v1	R-50	<u>81.4</u>	<u>61.2</u>	55.2
MoCo v2	R-50	**82.5**	**64.0**	**57.4**
BoWNet	R-50	<u>81.3</u>	<u>61.1</u>	<u>55.8</u>

Object detection with Faster R-CNN fine-tuned on VOC `trainval07+12` *and evaluated on* `test07`.
Networks are pre-trained with self-supervision on ImageNet.

自监督算法的加入
被证明是有效的
优于直接的监督算法

应用场景

- 小样本学习

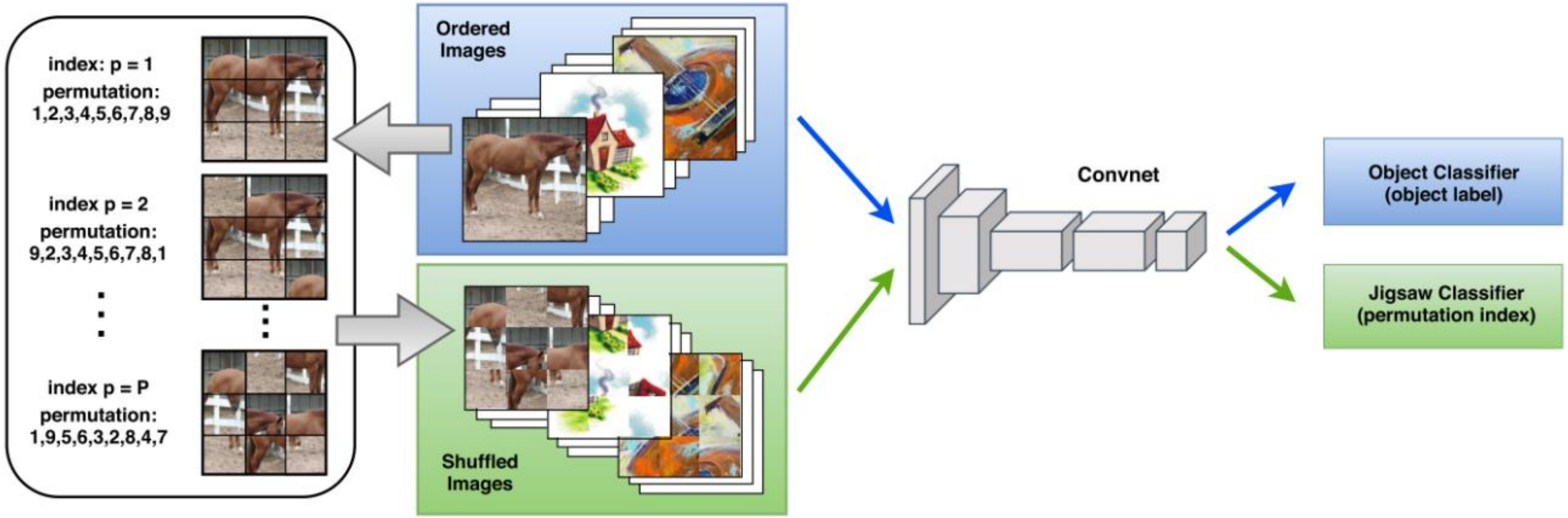


- 自监督学习作为辅助任务
- 目的是使得提取的特征更具有推广性

S. Gidaris et al., Boosting Few-Shot Learning with Self-Supervision, ICCV 2019; J.C. Su et al., When Does Self-supervision Improve Few-shot Learning?, ArXiv 2019

应用场景

- 领域泛化适应
- 数据集内部差异很大



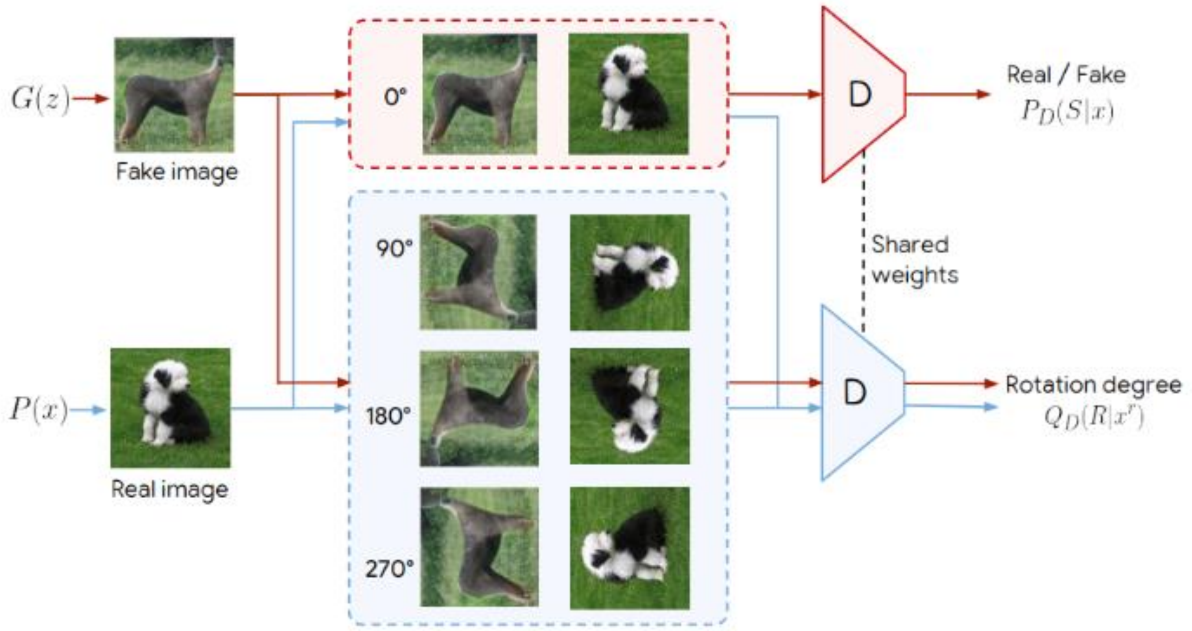
◆ 构造自监督任务可以获得更可靠的特征

F.M. Carlucci et al., Domain Generalization by Solving Jigsaw Puzzles, CVPR 2019

应用场景

- GAN网络的训练

- 自监督辅助任务的加入
- 网络倾向于学习稳定的特征
- 类似数据增广



Discriminator with rotation-based self-supervision.

自监督方法的特性

Key findings

- Large impact of CNN architecture *[Kolesnikov et al. (2019)]*
- Increasing the size of pre-training dataset benefits higher capacity networks *[Goyal et al. (2019)]*
- Varying benefits of increasing the pretext task difficulty *[Goyal et al. (2019), Chen et al. (2020)]*
- Currently greatest benefits are in low annotated data regimes *[Henaff et al. (2019); Newell and Deng (2020)]*
- In stage 2, linear classification achieves lower performance than finetuning *[Zhai et al. (2019); Newell and Deng (2020)]*

- ☐ 网络结构
- ☐ 数据规模
- ☐ 辅助任务难度
- ☐ 在小标注样本上表现较好
- ☐ Finetuning 性能更好