Data-Efficient Image Recognition with Contrastive Predictive Coding

Пудяков Ярослав

Национальный Исследовательский Университет
Высшая Школа Экономики

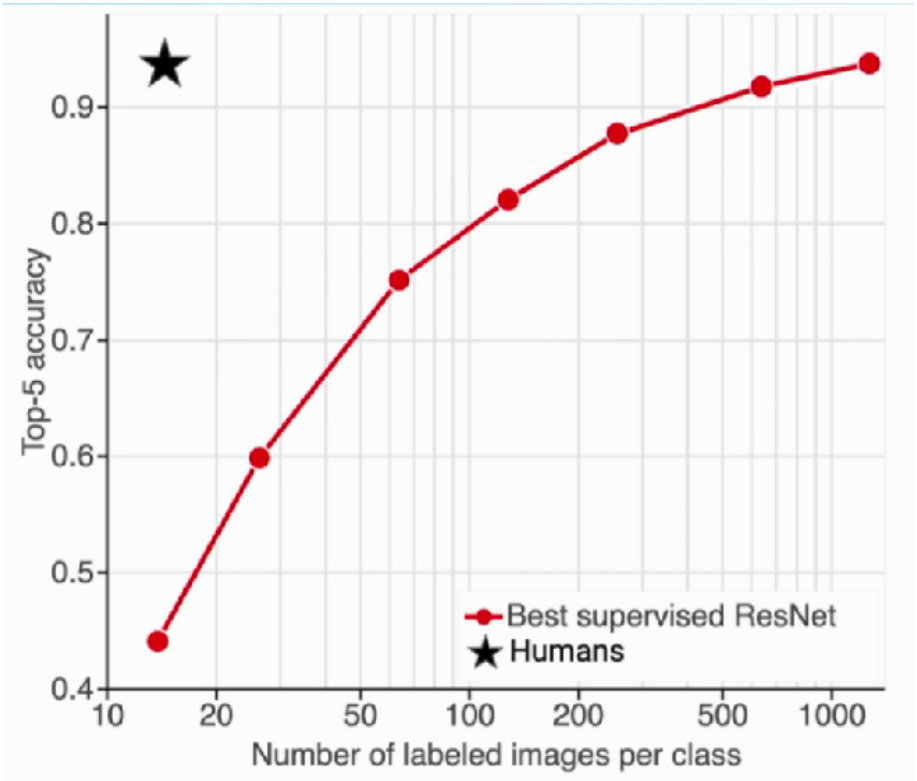
yaapudyakov@edu.hse.ru

28 ноября 2019 г.

План доклада

1. Motivation
2. Contrastive Predictive Coding
3. Prediction Task
4. Contrastive Loss
5. Unsupervised learning with CPC
6. Semi-supervised learning with CPC
7. Results: ImageNet classification. Linear separability
8. Low-data classification: fully-supervised
9. Low-data classification: semi-supervised
10. Transfer to PASCAL detection
11. Learning dynamics
12. Conclusion

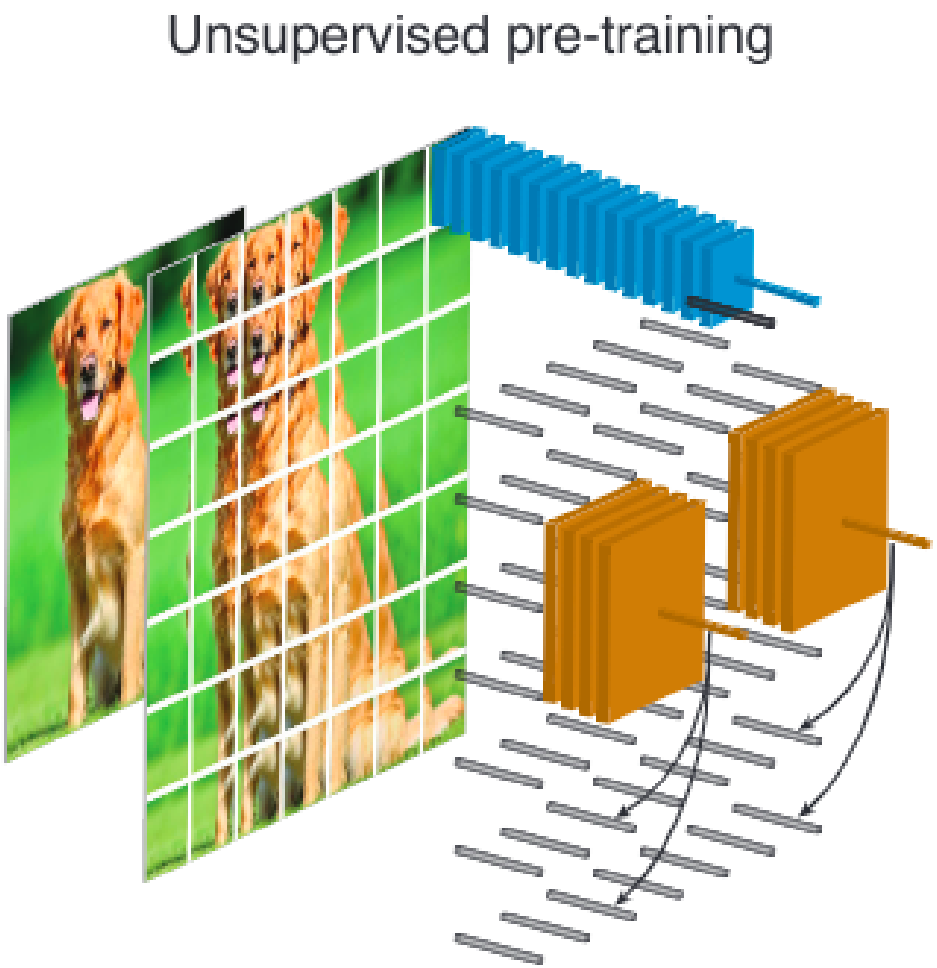
Motivation


Top-5 accuracy
0.9
0.8
0.7
0.6
0.5
0.4
10
20
50
100
200
500
1000
Number of labeled images per class
Best supervised ResNet
Humans

Contrastive Predictive Coding

- Contrastive Predictive Coding (CPC) - self-supervised метод обучения представлений, который обучается при помощи данных представленных последовательностями. Метод предсказывает признаки при помощи уже полученных представлений.
- CPC поощряет стабильные представления
- Хотя обучение CPC не контролируется, его представления имеют тенденцию быть лин. разделяемыми по классам

Prediction Task



$$z_{i,j} = f_\theta(x_{i,j})$$
$$c_{i,j} = g_{\text{context}}\left(z_{i,j}\right)$$
$$\hat{z}_{i+k,j} = W_k c_{i,j}$$

Contrastive Loss

$$\mathcal{L}_{\text{CPC}} = -\sum_{i,j,k} \log p(z_{i+k,j} | \hat{z}_{i+k,j}, \{z_l\})$$

$$= -\sum_{i,j,k} \log \frac{\exp\left(\hat{z}_{i+k,j}^T z_{i+k,j}\right)}{\exp\left(\hat{z}_{i+k,j}^T z_{i+k,j}\right) + \sum_l \exp\left(\hat{z}_{i+k,j}^T z_l'\right)}$$

$\{z_l\}$ - negative samples

Цель - корректно распознавать таргет среди множества случайно сэмплированных таргетов патчей из датасета $\{z_l\}$.

Unsupervised learning with CPC

- Увеличивая количество слоев в сети f_θ (network capacity) улучшается качество представлений
- В статье использовалась расширенная ResNet-170 (stack of ResNet-101)
- Более сложная архитектура сложнее обучается. Экспериментально получили, что BatchNormalization ухудшает эффективность обучения, а LayerNormalization наоборот повышает

Semi-supervised learning with CPC

$$\theta^* = \arg\min_\theta \frac{1}{N} \sum_{n=1}^{N} \mathcal{L}_{\text{CPC}}[f_\theta(x_n)]$$

$$\phi^* = \arg\min_\phi \frac{1}{M} \sum_{m=1}^{M} \mathcal{L}_{\text{Sup}}[g_\phi \circ f_{\theta^*}(x_m), y_m]$$

$\{x_n\}$ - dataset of N images

$\{x_m, y_m\}$ - dataset of M labeled images.

g_ϕ - 11-block ResNet

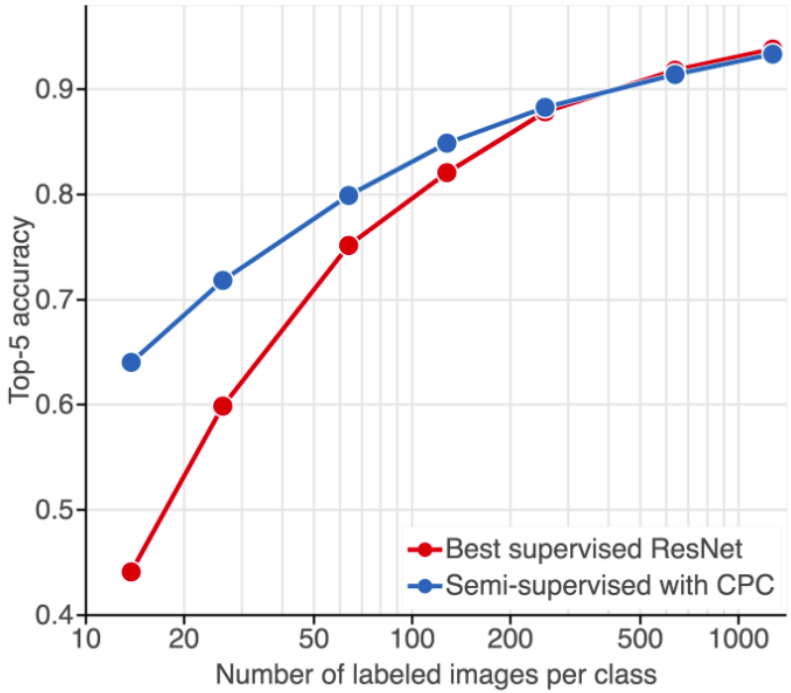
L_{Sup} - cross entropy between model predictions and image labels

Results: ImageNet classification. Linear separability

Method	Top-1	Top-5
Motion Segmentation (MS) [50]	27.6	48.3
Exemplar (Ex) [17]	31.5	53.1
Relative Position (RP) [14]	36.2	59.2
Colorization (Col) [69]	39.6	62.5
Combination of MS + Ex + RP + Col [15]	-	69.3
CPC [49]	48.7	73.6
Rotation + RevNet [36]	55.4	-
CPC (ours)	**61.0**	**83.0**

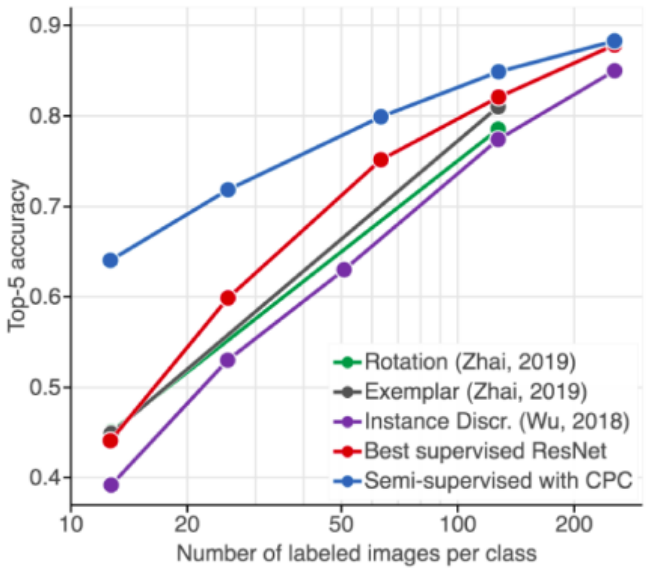
- Качество self-supervised метода оценивалось в способности линейной разделимости представлений (связь со сложностью классификации).
- Для оценки качества представлений-векторов были взяты mean-pooled CPC признаки ($z_{i,j}$) и на них обучили линейный классификатор.

Low-data classification: fully-supervised



- Чтобы исследовать эффективность архитектуры - производится обучение с варьированием числа используемых размеченных данных от 1% до 100%.
- В процессе, использовалась аугментация данных и настройка параметров по валидационной выборке.

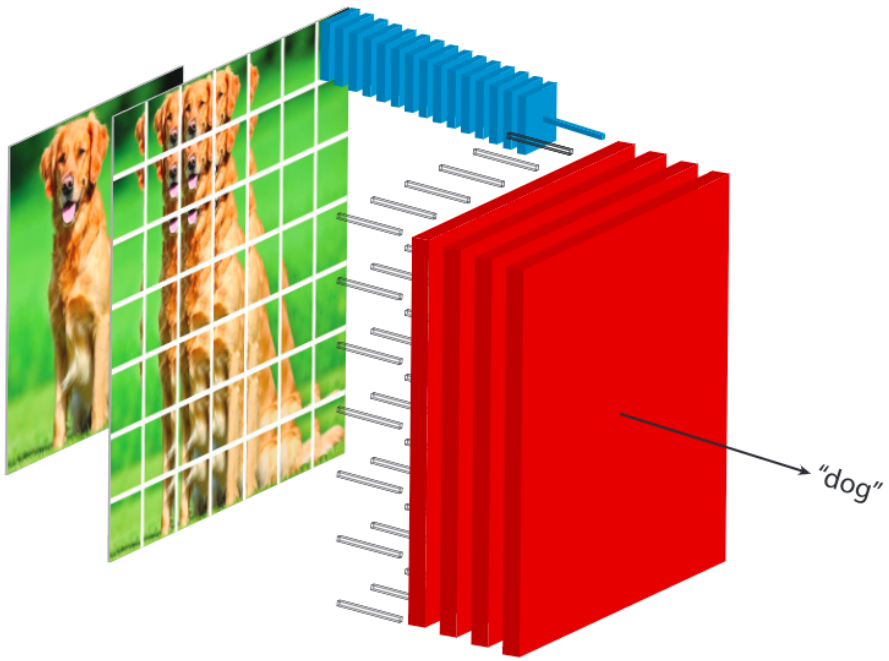
Low-data classification: semi-supervised



Labeled data **Method**	1% **Top-5 accuracy**	10%
Supervised baseline	44.10	82.08
Methods using label-propagation:		
Pseudolabeling [68]	51.56	82.41
VAT [68]	44.05	82.78
VAT + Entropy Minimization [68]	46.96	83.39
Unsup. Data Augmentation [65]	-	88.52
Rotation + VAT + Ent. Min. [68]	-	**91.23**
Methods only using representation learning:		
Instance Discrimination [64]	39.20	77.40
Exemplar [68]	44.90	81.01
Exemplar (joint training) [68]	47.02	83.72
Rotation [68]	45.11	78.53
Rotation (joint training) [68]	53.37	83.82
CPC (ours)	**64.03**	**84.88**

- Предобучаем наш feature extractor на всем датасете неразмеченных данных.
- Обучаем и настраиваем классификатор по размеченной части датасета.

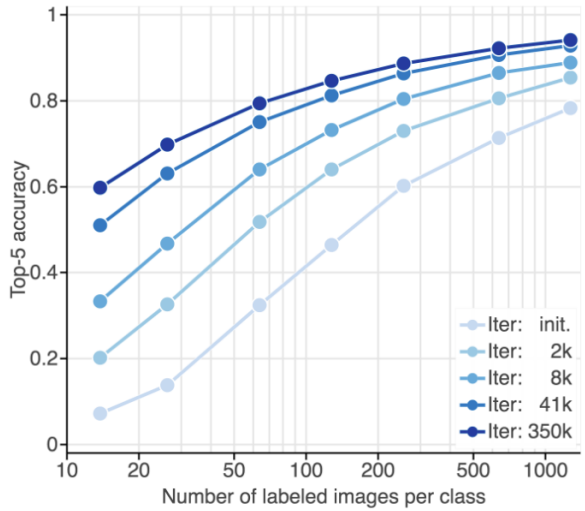
Transfer to PASCAL detection



Method	mAP
Transfer from labeled ImageNet:	
Supervised - ResNet-152	74.7
Transfer from unlabeled ImageNet:	
Exemplar (Ex) [17]	60.9
Motion Segmentation (MS) [50]	61.1
Colorization (Col) [69]	65.5
Relative Position (RP) [14]	66.8
Combination of Ex + MS + Col + RP [15]	70.5
Deep Cluster [8]	65.9
Deeper Cluster [9]	67.8
CPC - ResNet-101	70.6
CPC - ResNet-170	**72.1**

- Наши представления будут информативными, если при помощи них можно будет решать другие задачи, например object classification.
- Для этого обучили CPC представления на ImageNet. Эти признаки уже подавали на вход в Faster-RCNN
- Тестировали модель на датасете PASCAL

Learning dynamics



- В начале случайно инициализированный ResNet не дает никакой выгоды. Точность меньше 10%.
- Однако в процессе обучения CPC эти результаты быстро улучшаются; 40000 итераций достаточно, чтобы обойти контролируемые методы в режиме с низким уровнем данных, даже если многие параметры не настроены точным образом
- После 350k итераций мы получаем предельную точность, это говорит о том, что CPC действительно играет решающую роль в результатах.

Conclusion

- Semi-supervised CPC - работает!.
- State of the art классификации изображений с малым числом размеченных данных (+20% supervised, +10% semi-supervised)
- State of the art unsupervised перенос на задачу детекции изображений

References

- Data-Efficient Image Recognition with Contrastive Predictive Coding, DeepMind, 2019
- International Conference on Machine Learning Live (Grand Ballroom A), Facebook research, 2019