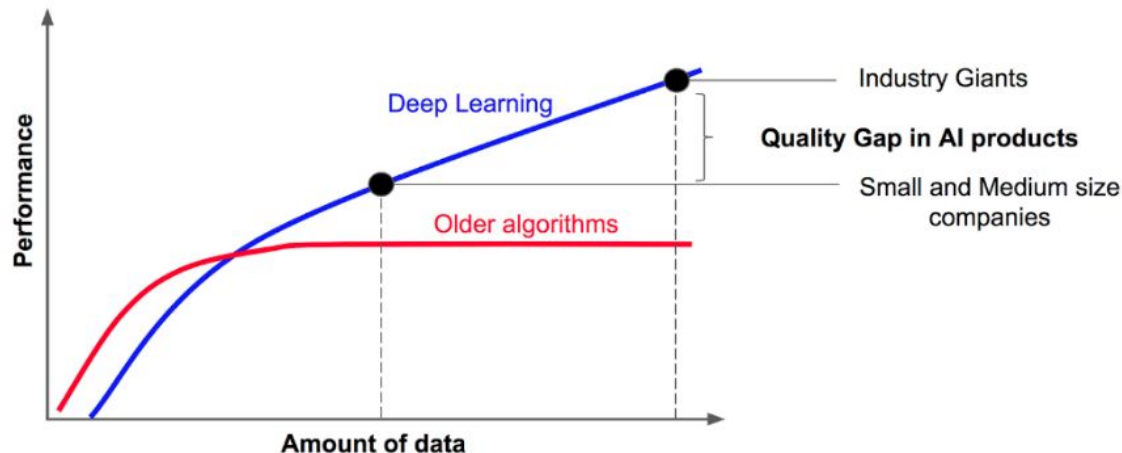


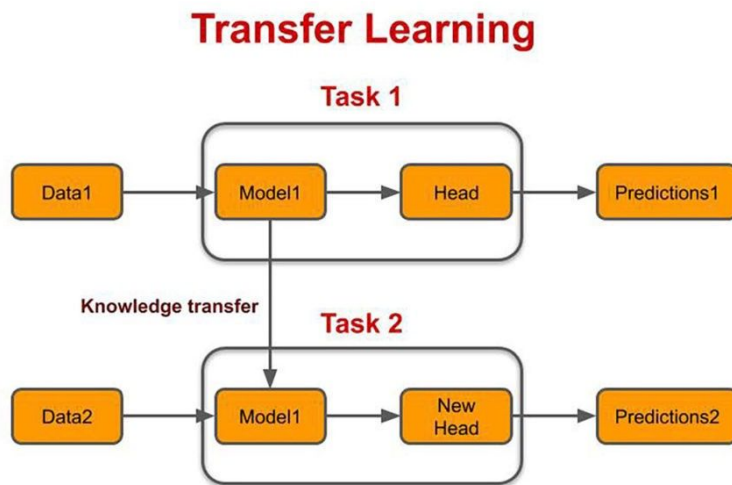
Big Self-Supervised Models are Strong
Semi-Supervised Learners

Размеченные данные



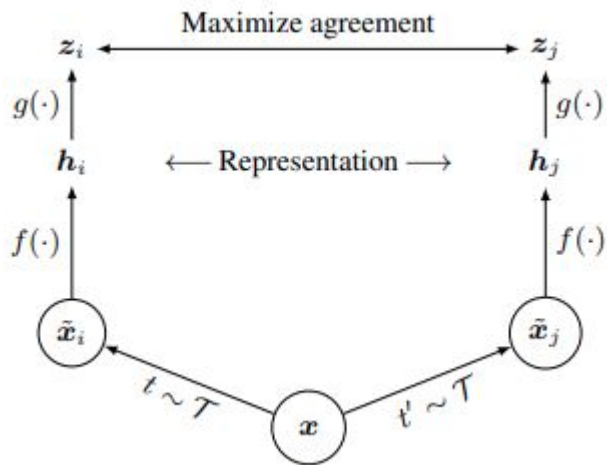
- Очень дорого размечать большие датасеты
- Специфичные области поставленных задач

Transfer learning



- Тело модели отдаем другой задаче
- Данные для двух разных задач могут очень сильно отличаться и тогда

Simple Framework for Contrastive Learning of Visual Representations



- Хотим получить представление картинок
- Для одного изображения применяются два разных типа аугментации
- Получаем положительные пары (две версии картинки после разных аугментаций/или оригинал) и отрицательные (оригинальная картинка остальные в батче)

Особенности

$$\begin{aligned}\text{sim}(\mathbf{u}, \mathbf{v}) &= \frac{\mathbf{u}^T \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} \\ &= \hat{\mathbf{u}}^T \hat{\mathbf{v}}\end{aligned}$$

- **Normalized temperature-scaled cross-entropy loss**

минимизируем его для положительных пар

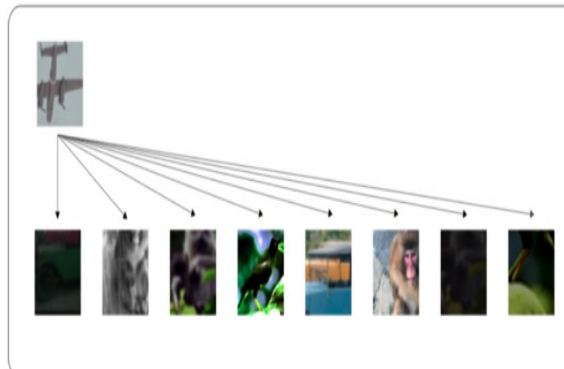
$$l_{i,j} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)}$$

- По умолчанию настройки сети: ResNet-50 для энкодера $f()$. Для декодера $g()$ - MLP 2 слоя ReLU

Работы с данными

- Случайно выбирается тип аугментации
- Композиция аугментаций улучшает работу

Batch of N examples



$2*(N-1)$ negative pairs



(a) Original



(b) Crop and resize



(c) Crop, resize (and flip)



(d) Color distort. (drop)



(e) Color distort. (jitter)



(f) Rotate $\{90^\circ, 180^\circ, 270^\circ\}$



(g) Cutout



(h) Gaussian noise



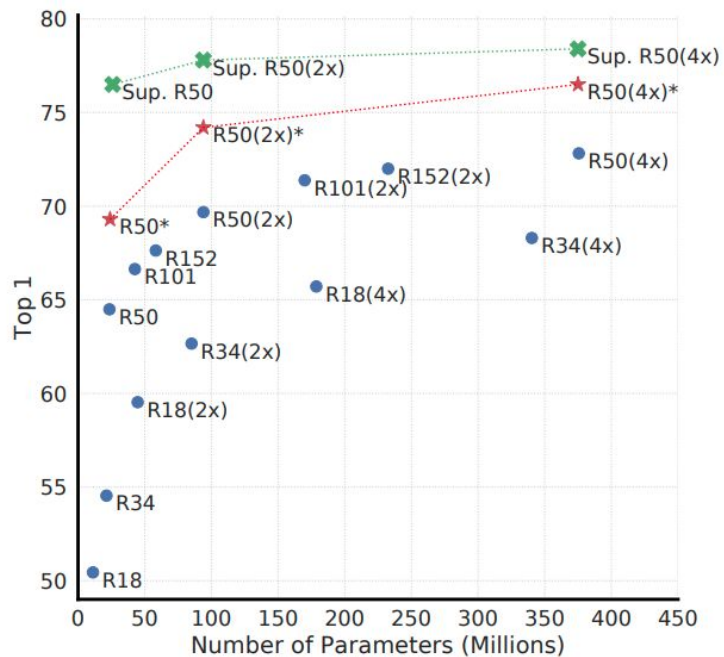
(i) Gaussian blur



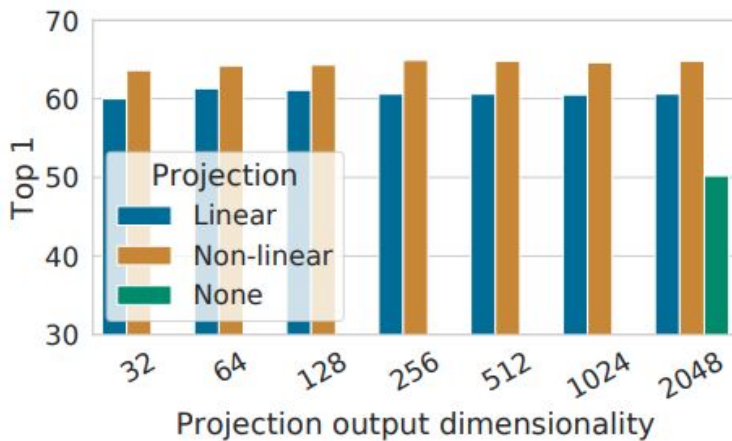
(j) Sobel filtering

Архитектура

Больше модели (для энкодера) - лучше

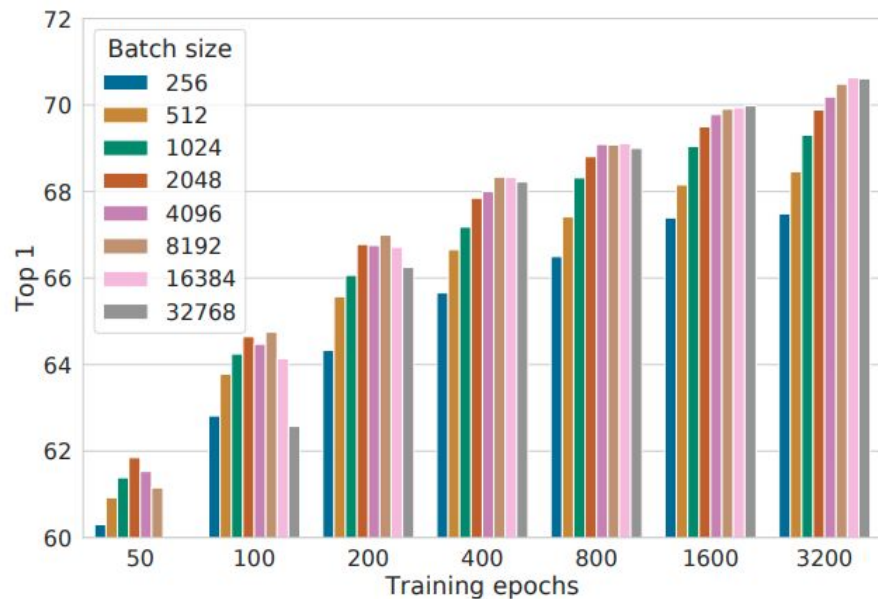
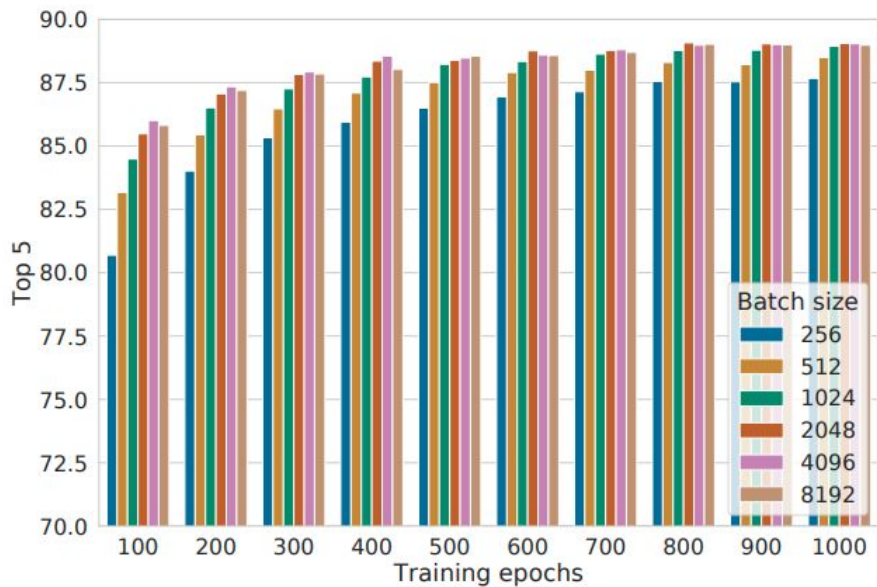


Нелинейность улучшает репрезентативные качества



Обучение

Больше batch size. Дольше обучение



Результаты

- При semi-supervised подходе на ImageNet
- При fine-tuning на 12-ти разных датасетах

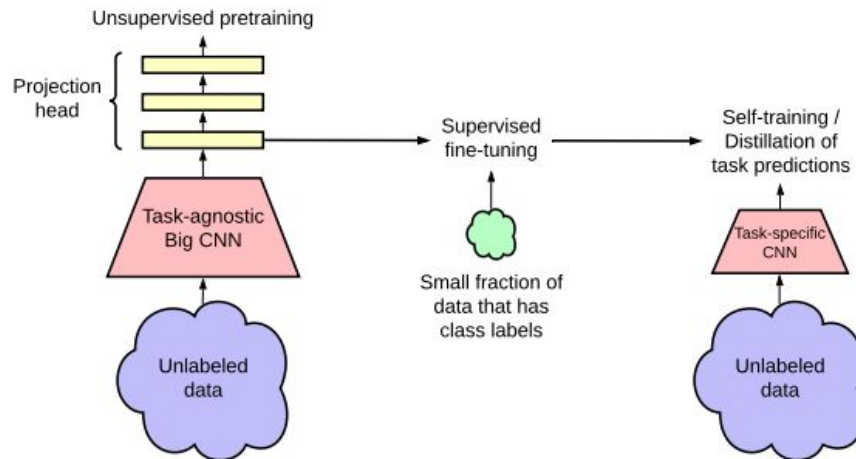
Method	Architecture	Label fraction	
		1%	10%
Top 5			
Supervised baseline	ResNet-50	48.4	80.4
<i>Methods using other label-propagation:</i>			
Pseudo-label	ResNet-50	51.6	82.4
VAT+Entropy Min.	ResNet-50	47.0	83.4
UDA (w. RandAug)	ResNet-50	-	88.5
FixMatch (w. RandAug)	ResNet-50	-	89.1
S4L (Rot+VAT+En. M.)	ResNet-50 (4×)	-	91.2
<i>Methods using representation learning only:</i>			
InstDisc	ResNet-50	39.2	77.4
BigBiGAN	RevNet-50 (4×)	55.2	78.8
PIRL	ResNet-50	57.2	83.8
CPC v2	ResNet-161(*)	77.9	91.2
SimCLR (ours)	ResNet-50	75.5	87.8
SimCLR (ours)	ResNet-50 (2×)	83.0	91.2
SimCLR (ours)	ResNet-50 (4×)	85.8	92.6

A Simple Framework for Contrastive Learning of Visual Representations

	Food	CIFAR10	CIFAR100	Birdsnap	SUN397	Cars	Aircraft	VOC2007	DTD	Pets	Caltech-101	Flowers
<i>Linear evaluation:</i>												
SimCLR (ours)	76.9	95.3	80.2	48.4	65.9	60.0	61.2	84.2	78.9	89.2	93.9	95.0
Supervised	75.2	95.7	81.2	56.4	64.9	68.8	63.8	83.8	78.7	92.3	94.1	94.2
<i>Fine-tuned:</i>												
SimCLR (ours)	89.4	98.6	89.0	78.2	68.1	92.1	87.0	86.6	77.8	92.1	94.1	97.6
Supervised	88.7	98.3	88.7	77.8	67.0	91.4	88.0	86.5	78.8	93.2	94.2	98.0

SimCLRv2

- Модель стала глубже. Вместо ResNet-50, 152-layer ResNet.
- Projection head на один слой больше.
- Дистилляция



Итоги

- В CV решение такие решения новые. Многие подсмотрели в NLP подходах

- Статья 17-го июня 2020

Method	Architecture	Top-1 Label fraction		Top-5 Label fraction	
		1%	10%	1%	10%
Supervised baseline [30]	ResNet-50	25.4	56.4	48.4	80.4
<i>Methods using unlabeled data in a task-specific way:</i>					
Pseudo-label [11, 30]	ResNet-50	-	-	51.6	82.4
VAT+Entropy Min. [37, 38, 30]	ResNet-50	-	-	47.0	83.4
UDA (w. RandAug) [14]	ResNet-50	-	68.8	-	88.5
FixMatch (w. RandAug) [15]	ResNet-50	-	71.5	-	89.1
S4L (Rot+VAT+Entropy Min.) [30]	ResNet-50 (4×)	-	73.2	-	91.2
MPL (w. RandAug) [2]	ResNet-50	-	73.8	-	-
<i>Methods using unlabeled data in a task-agnostic way:</i>					
BigBiGAN [39]	RevNet-50 (4×)	-	-	55.2	78.8
PIRL [40]	ResNet-50	-	-	57.2	83.8
CPC v2 [19]	ResNet-161(*)	52.7	73.1	77.9	91.2
SimCLR [1]	ResNet-50	48.3	65.6	75.5	87.8
SimCLR [1]	ResNet-50 (4×)	63.0	74.4	85.8	92.6
<i>Methods using unlabeled data in both ways:</i>					
SimCLRv2 distilled (ours)	ResNet-50	73.9	77.5	91.5	93.4
SimCLRv2 distilled (ours)	ResNet-50 (2×+SK)	75.9	80.2	93.0	95.0
SimCLRv2 self-distilled (ours)	ResNet-152 (3×+SK)	76.6	80.9	93.4	95.5

Вопросы

- Схема SimCLR
- Какие подходы аугментации используются для обучения?
- Лосс для contrastive learning