

# Re-labeling ImageNet: from Single to Multi-Labels, from Global to Localized Labels

Александр Чернышёв, 517 группа

# Датасет ImageNet

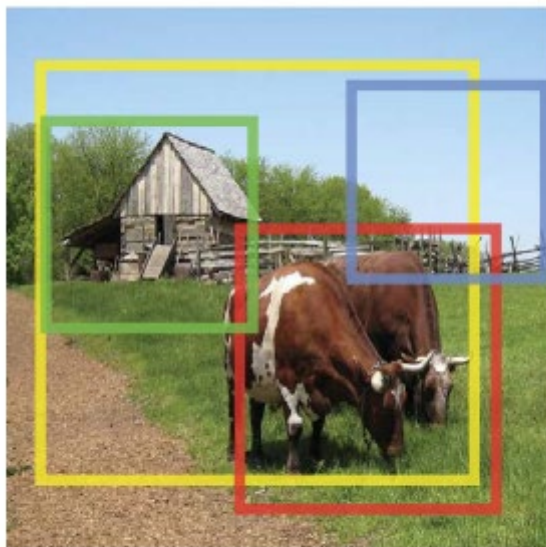
- 1.28 миллиона изображений
- 1000 классов
- Самый популярный бенчмарк для классификации изображений
- Приличный уровень шума в метках
- Переразмечать вручную — дорого



# Датасет ImageNet

- Во время обучения нейронных сетей на ImageNet часто прибегают к аугментации данных
- Используется в том числе и «случайный вырез» (random crop)
- Можем вырезать кусок картинки, на котором не будет представлена метка

⇒ проблемы как с обучением, так и с оцениванием качества моделей на ImageNet

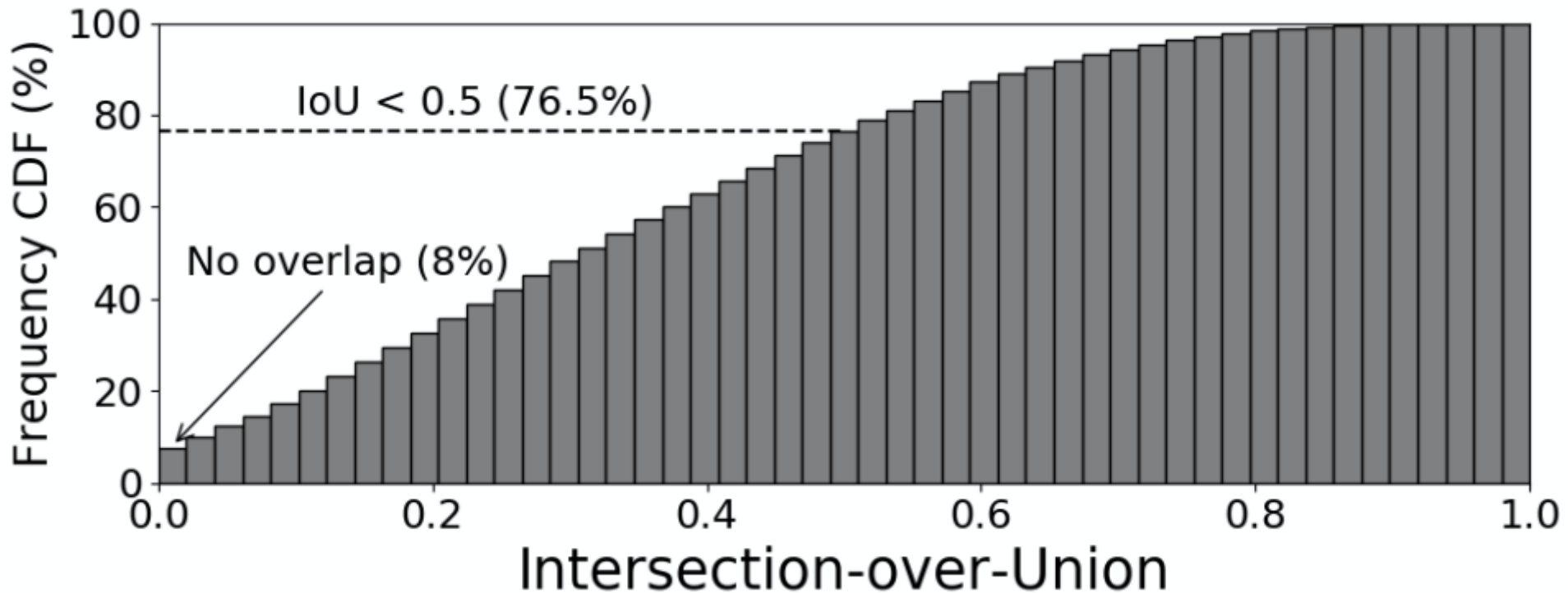


Original ImageNet label: ox 1.00



# Random crops for ImageNet

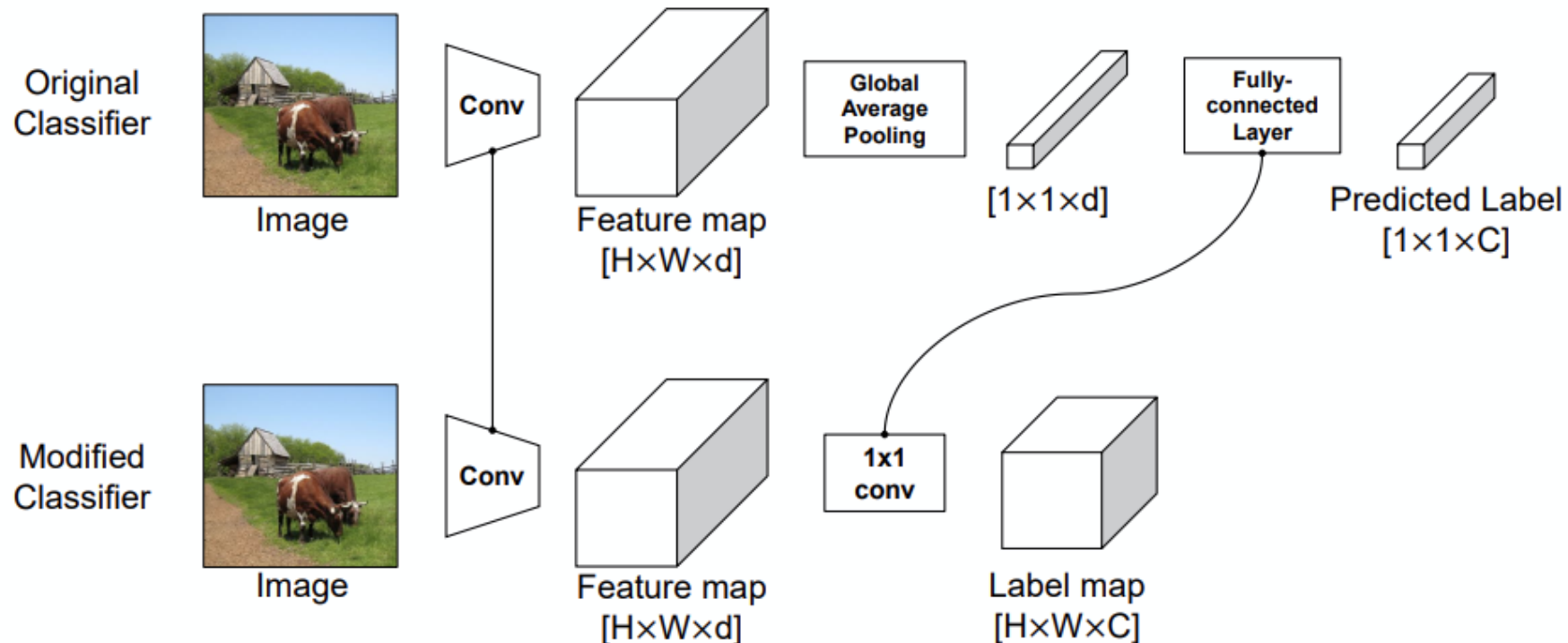
Кумулятивная гистограмма для IoU между случайными вырезами и настоящим расположением объектов. Было засэмплировано 100 случайных вырезков для каждого изображения из валидационного набора ImageNet (50 000).



⇒ у многих случайных вырезков метка будет не совсем правильной

# Алгоритм ReLabel

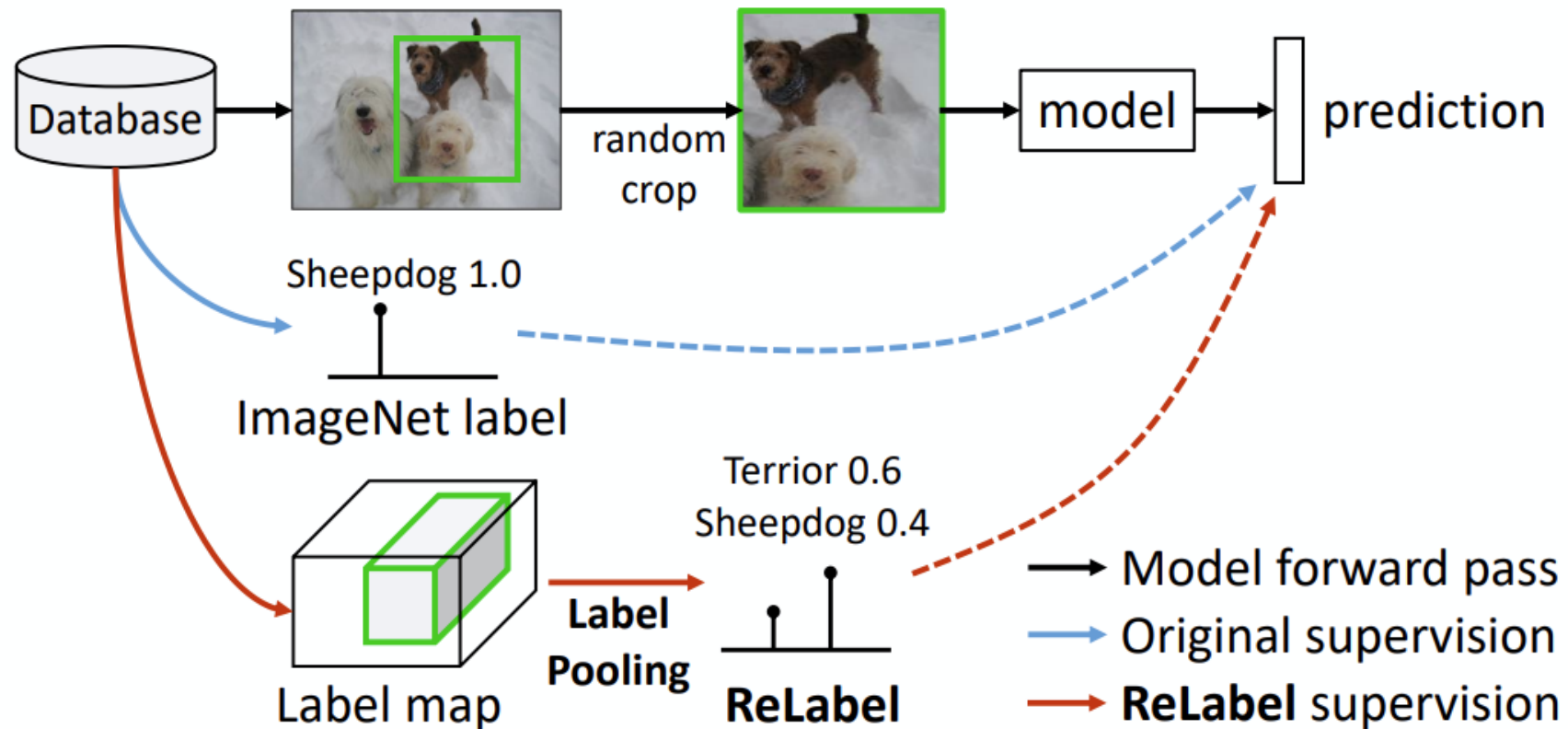
- Берем SOTA классификатор, обученный на огромном датасете (например, JFT-300M)
- Дообучаем его на ImageNet
- Заменяем в нем слой глобального пулинга на свертку 1x1
- Теперь выход такой сети  $f(x) \in \mathbb{R}^{W \times H \times C}$  для изображения  $x$  будем считать новой разметкой  $x$
- $W$  и  $H$  —размеры выхода сверточных слоев нейронной сети (не совпадают с размерами  $x$ )



# LabelPooling

LabelPooling — это подход обучения классификатора на ReLabel метках:

- Во время случайного выреза также вырезаем соответствующий кусок из ReLabel меток
- Применяем глобальный пулинг + softmax → получаем распределение классов для случайного выреза
- Используем это распределение для обучения (используется кросс энтропийная функция потерь)



# Обсуждение

В качестве SOTA модели использовалась EfficientNet-L2.

## Память:

- Размер новой разметки:  $L \in \mathbb{R}^{15 \times 15 \times 1000}$
- Нужно 1 TB памяти:  $(1.28 \times 10^6) \text{ images} \times (15 \times 15 \times 1000) \text{ dim/image} \times 4 \text{ bytes/dim} \approx 1.0 \text{ TB}$
- К счастью, обычно всего несколько классов имеют ненулевые предсказания
- Будем сохранять предсказания только для top-5 классов
- Итого всего нужно +10% дополнительной памяти от всего ImageNet.

## Время:

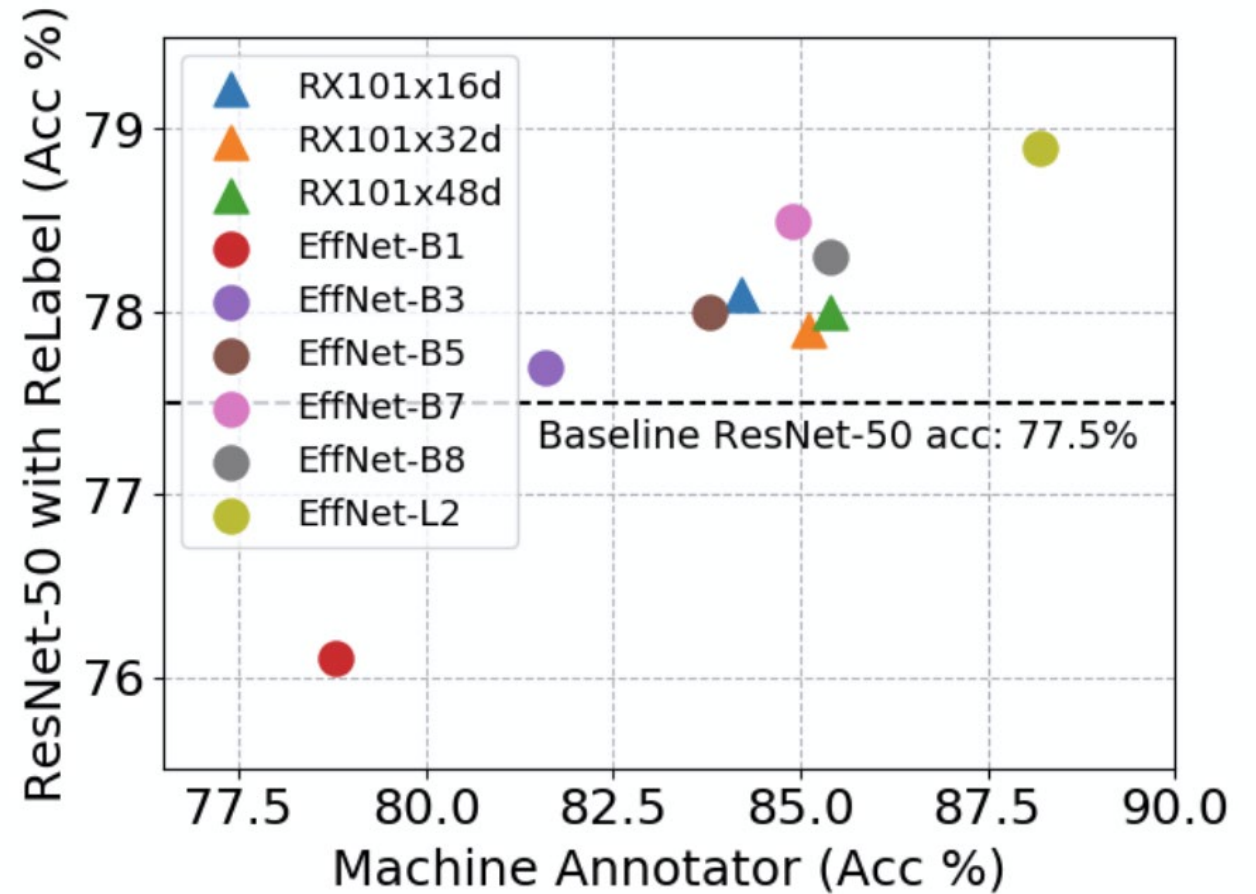
- ReLabel — всего +3.3% от всего времени обучения ResNet-50
- LabelPooling — всего +0.5% дополнительного времени нужно на обучение



# Обсуждение

## Выбор SOTA модели:

- Авторы попробовали EfficientNet-{B1, B3, B5, B7, B8}, EfficientNet-L2, ResNeXT-101 32x{32d,48d}
- Сравнивали по top-1 доли правильных ответов ResNet-50 на валидации
- Выбрали EfficientNet-L2 как наиболее точную





# Обсуждение

**Анализ факторов.** Хотим проверить, насколько нам в новых метках нужна (а) их локальность, (б) их мультиклассовость:

- Оставляем только 1 класс → localized single labels (argmax вместо softmax)
- Убираем пространственность в метках → global multi-labels
- Убираем и пространственность, и многоклассовость → global single-labels

Variants	ImageNet top-1 (%)
ReLabel (localized mutli-labels)	78.9
Localized single labels	78.4 (-0.5)
Global multi-labels	78.5 (-0.4)
Global single labels	77.5 (-1.4)
Original ImageNet labels	77.5 (-1.4)

# Эксперименты

Сравнение с другими алгоритмами переразметки:

Network	Supervision	ImageNet single-label	ImageNetV2 [40] single-label	ReaL [2] multi-label	Shankar <i>et al.</i> [43] multi-label
ResNet-50	Original	77.5	79.0	83.6	85.3
ResNet-50	Label smoothing ( $\epsilon=0.1$ ) [49]	78.0	79.5	84.0	84.7
ResNet-50	Label cleaning [2]	78.1	79.1	83.6	85.2
ResNet-50	<b>ReLabel</b>	<b>78.9</b>	<b>80.5</b>	<b>85.0</b>	<b>86.1</b>

Label smoothing — перераспределяет метки: переднему плану отдается  $(1 - \epsilon)$ , классам, отвечающим за задний план равномерно отдается  $\epsilon$ .

Label cleaning — из датасета удалены все объекты, для которых метка не согласуется с ответом от учителя (“strong teacher classifier”)

Метрика для multi-label датасетов:  $\frac{1}{N} \sum_{n=1}^N 1(\arg \max f(x_n) \in y_n)$

# Эксперименты


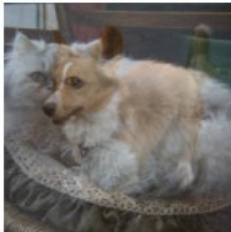


Сравнение различных CNN:

Architecture	Resources		Supervision	
	Params	Flops	Vanilla	ReLabel
ResNet-18	11.7M	1.8B	71.7	72.5 (+0.8)
ResNet-50	25.6M	3.8B	77.5	78.9 (+1.4)
ResNet-101	44.7M	7.6B	78.1	80.7 (+2.6)
EfficientNet-B0	5.3M	0.4B	77.4	78.0 (+0.6)
EfficientNet-B1	7.8M	0.7B	79.2	80.3 (+1.1)
EfficientNet-B2	9.2M	1.0B	80.3	81.0 (+0.7)
EfficientNet-B3	12.2M	1.8B	81.7	82.5 (+0.8)
ReXNet ( $\times 1.0$ )	4.8M	0.4B	77.9	78.4 (+0.5)

# Эксперименты

Выбиваем SOTA:

Model	ImageNet top1 (%)
ResNet-50	77.5
+ <b>ReLabel</b>	78.9 (+1.4)
+ <b>ReLabel</b> + CutMix	80.2 (+2.7)
+ <b>ReLabel</b> + CutMix + Extra data	81.2 (+3.7)
ResNet-101	78.1
+ <b>ReLabel</b>	80.7 (+2.6)
+ <b>ReLabel</b> + CutMix	81.6 (+3.5)

	ResNet-50	Mixup [48]
Image		
Label	Dog 1.0	Dog 0.5 Cat 0.5
	Cutout [3]	CutMix
Image		
Label	Dog 1.0	Dog 0.6 Cat 0.4

	Food-101 [3]	Stanford Cars [29]	DTD [6]	FGVC Aircraft [36]	Oxford Pets [38]
ResNet-50 (Baseline)	87.98	92.64	75.43	85.09	93.92
ResNet-50 ( <b>ReLabel</b> -trained)	88.12	92.73	75.74	88.89	94.28

# Выводы

- ReLabel помогает улучшать разметку датасетов
- Предложен алгоритм LabelPooling для обучения моделей на локализованных мультиклассовых метках
- Доказана эффективность подходов