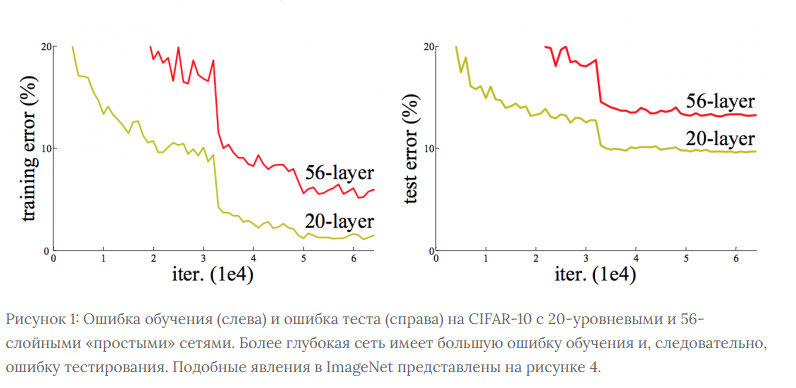
**Сеть Resnet**

**ResNet** — это сокращенное название для Residual Network (дословно  — «остаточная сеть»)

Глубокие сверточные нейронные сети превзошли человеческий уровень классификации изображений в 2015 году. Глубокие сети извлекают низко-, средне- и высокоуровневые признаки  сквозным многослойным способом, а увеличение количества stacked layers может обогатить «уровни» признаков. Stacked layer имеет решающее значение, результат ImageNet

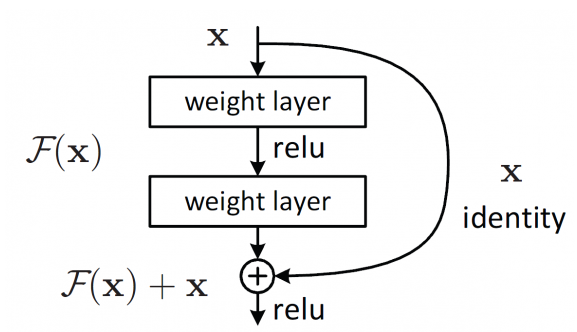


**Соединения быстрого доступа**

Чтобы преодолеть эту проблему, Microsoft ввела глубокую «остаточную» структуру обучения. Вместо того, чтобы надеяться на то, что каждые несколько stacked layers непосредственно соответствуют желаемому основному представлению, они явно позволяют этим слоям соответствовать «остаточному». Формулировка *F(x) + x* может быть реализована с помощью нейронных сетей с соединениями для быстрого доступа.

**Соединения быстрого доступа (shortcut connections)** пропускают один или несколько слоев и выполняют сопоставление идентификаторов. Их выходы добавляются к выходам stacked layers. Используя ResNet, можно решить множество проблем, таких как:

* ResNet относительно легко оптимизировать: «простые» сети (которые просто складывают слои) показывают большую ошибку обучения, когда глубина увеличивается.
* ResNet позволяет относительно легко увеличить точность благодаря увеличению глубины, чего с другими сетями добиться сложнее.

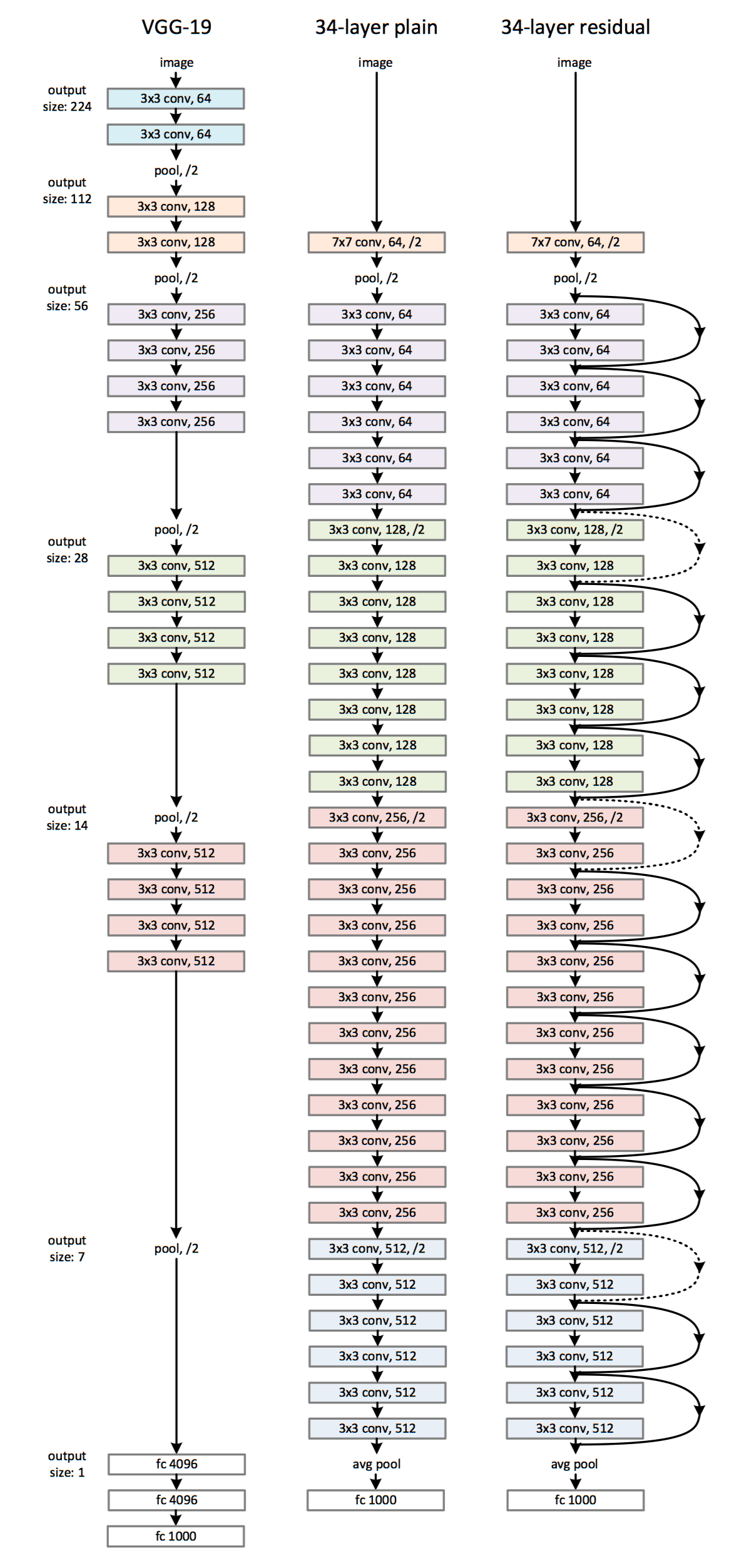


**Датасеты**

**ImageNet** — это набор данных миллионов помеченных изображений с высоким разрешением, относящихся примерно к 22 тысячам категорий. Изображения были собраны из Интернета и помечены людьми с помощью краудсорсинга. Начиная с 2010 года в рамках конкурса визуальных объектов Pascal проводится ежегодный челлендж «Крупномасштабный конкурс визуального распознавания ImageNet» (ILSVRC2013). ILSVRC использует подмножество ImageNet из примерно 1000 изображений в каждой из 1000 категорий. Существует около 1,2 миллиона обучающих образов, 50 тыс. валидаций 150 тыс. тестовых изображений.

**PASCAL VOC** предоставляет стандартизированные наборы данных изображений для распознавания классов объектов, стандартному набору инструментов для доступа к наборам данных и аннотациям, позволяет оценивать и сравнивать методы, наконец, оценивает производительность при распознавании классов объектов

**Архитектура сети**



Пример сетевой архитектуры для ImageNet. Снизу: модель VGG-19 (19,6 млрд. FLOP) в качестве эталона. Посередине: простая сеть с 34 слоями (3,6 млрд. FLOP). Сверху: ResNet с 34 слоями (3,6 миллиарда FLOP). Пунктирные быстрые соединения увеличивают размерность.

**Простая сеть**: Простые базовые линии (рис. 2, в центре) в основном вдохновлены философией сетей [VGG](https://neurohive.io/ru/vidy-nejrosetej/vgg16-model/) (рис. 2, слева). Сверточные слои в основном имеют фильтры 3×3 и следуют двум простым правилам:

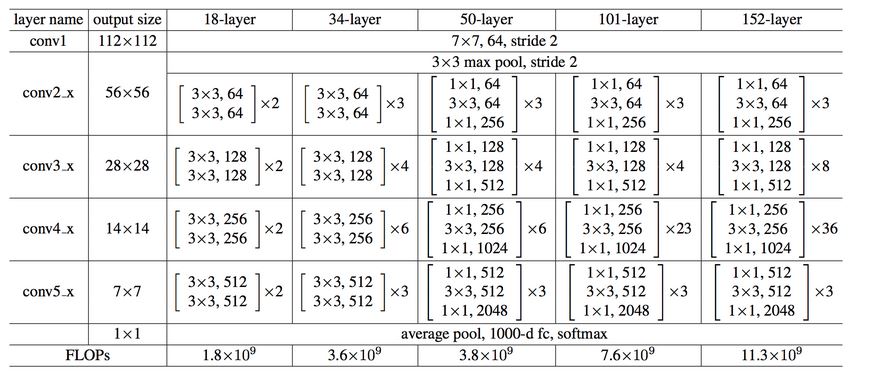
1. Для одной и той же выходной карты объектов слои имеют одинаковое количество фильтров;
2. Если размер карты объектов уменьшается вдвое, число фильтров удваивается, чтобы сохранить временную сложность каждого слоя.

Стоит отметить, что модель ResNet имеет меньше фильтров и сложность меньше, чем сети VGG.

**ResNet**: на основе описанной выше простой сети добавлено быстрое соединение (рис. 2, справа), которое превращает сеть в ее остаточную версию. Идентификационные быстрые соединения F (x {W} + x) могут использоваться непосредственно, когда вход и выход имеют одинаковые размерности (быстрые соединения сплошной линии на рис. 2). Когда размерности увеличиваются (пунктирные линии на рис. 2), он рассматривает два варианта:

1. Быстрое соединение выполняет сопоставление идентификаторов с дополнительными нулями, добавленными для увеличения размерности. Эта опция не вводит никаких дополнительных параметров.
2. Проекция быстрого соединения  в F (x {W} + x) используется для сопоставления размерностей (выполнено с помощью 1×1 сверток).

Для любой из опций, если быстрые соединения идут по картам объектов двух размерностей, они выполняются с шагом 2.



Каждый блок ResNet имеет два уровня глубины (используется в небольших сетях, таких как ResNet 18, 34) или 3 уровня (ResNet 50, 101, 152).

**50-слойная ResNet**: каждый 3-слойный блок заменяется в 34-слойной сети 3-слойным местом, в результате получается 50-слойная ResNet. Они используют вариант 2 для увеличения размерностей. Эта модель имеет 3,8 миллиарда FLOPs.

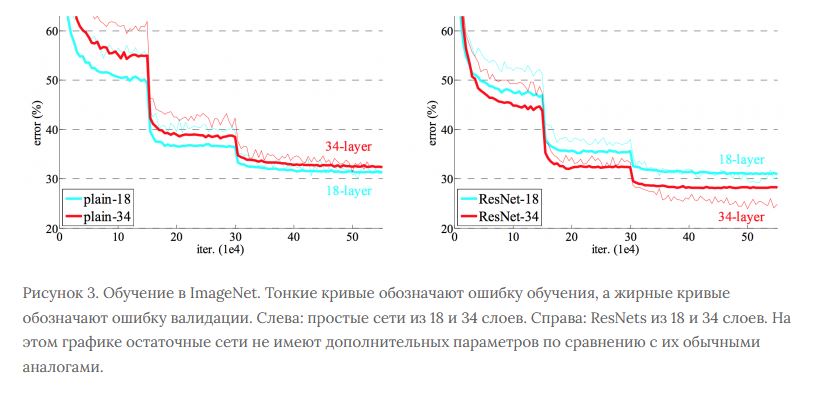
**ResNet с 101 и 152 слоями**: они создают ResNet с 101 и 152 слоями, используя больше 3-слойных блоков. Даже после увеличения глубины 152-слойная ResNet (11,3 миллиарда FLOP) имеет меньшую сложность, чем сети VGG-16/19 (15,3 / 19,6 миллиарда FLOPs).

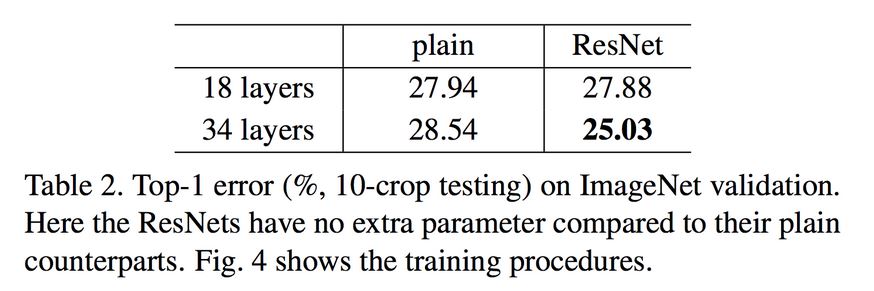
**Реализация**

Размер изображения изменяется с помощью случайной выборки его короткой стороны в [256,480] для увеличения масштаба. Кадрирование 224×224 выбирается случайным образом из изображения или его горизонтального смещения с вычитанием среднего значения для каждого пикселя. Скорость обучения стартует с 0,1 и делится на 10, когда изменение ошибок выходит на плато, модели обучаются вплоть до 60×10000 [итераций](https://neurohive.io/ru/osnovy-data-science/jepoha-razmer-batcha-iteracija/). Они используют снижение веса 0,0001 и импульс 0,9.

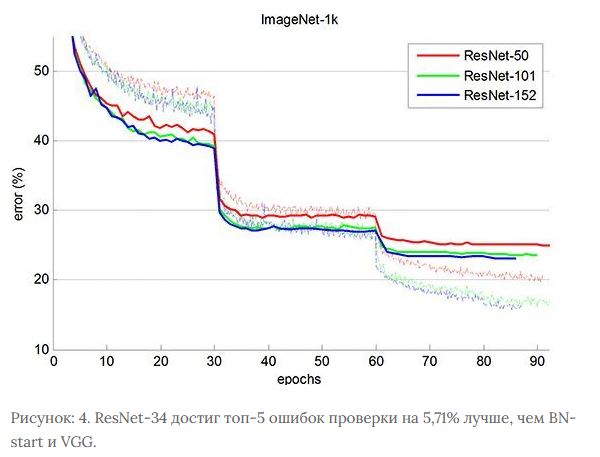
**Результат**

18-уровневая сеть — это просто подпространство в 34-уровневой сети, и она все еще работает лучше. ResNet выигрывает со значительным отрывом, если сеть глубже:





Сеть ResNet сходится быстрее, чем ее простой аналог. Рисунок 4 показывает, что более глубокие ResNet достигают лучших результатов обучения по сравнению с неглубокой сетью.



ResNet-152 достигает 4,49% в top-5 ошибок валидации. Комбинация из 6 моделей с различной глубиной достигает 3,57% в top-5 ошибок валидации.

