

# Horses or Humans Dataset with ResNet

---

ВЫПОЛНИЛ:

СТУДЕНТ ГРУППЫ М19-ИВТ-3

СУВОРКИН ПАВЕЛ

ВАРИАНТ: 73

# Horses or Humans Dataset

---



# Horses or Humans Dataset

Horses or Humans (Лошади или люди) — набор данных из 300x300 изображений, созданный Лоуренсом Морони.



# Идея создания набора данных

---

При изучении компьютерного зрения, в частности обучения DNNs (Deep Neural Networks) и CNNs (Convolutional Neural Networks), которые призваны распознавать и классифицировать изображения, есть нехватка хороших наборов данных.

Автор хотел решить эту проблему:

Автор решил создать набор данных, который можно использовать для обучения бинарных классификаторов. Он является уникальным и отличается от существующих наборов данных.

Автор хотел выяснить, возможно ли использовать Photoreal CGI для обучения нейронных сетей, которые могли бы затем распознавать и классифицировать реальные изображения той же тематики.



# Данные

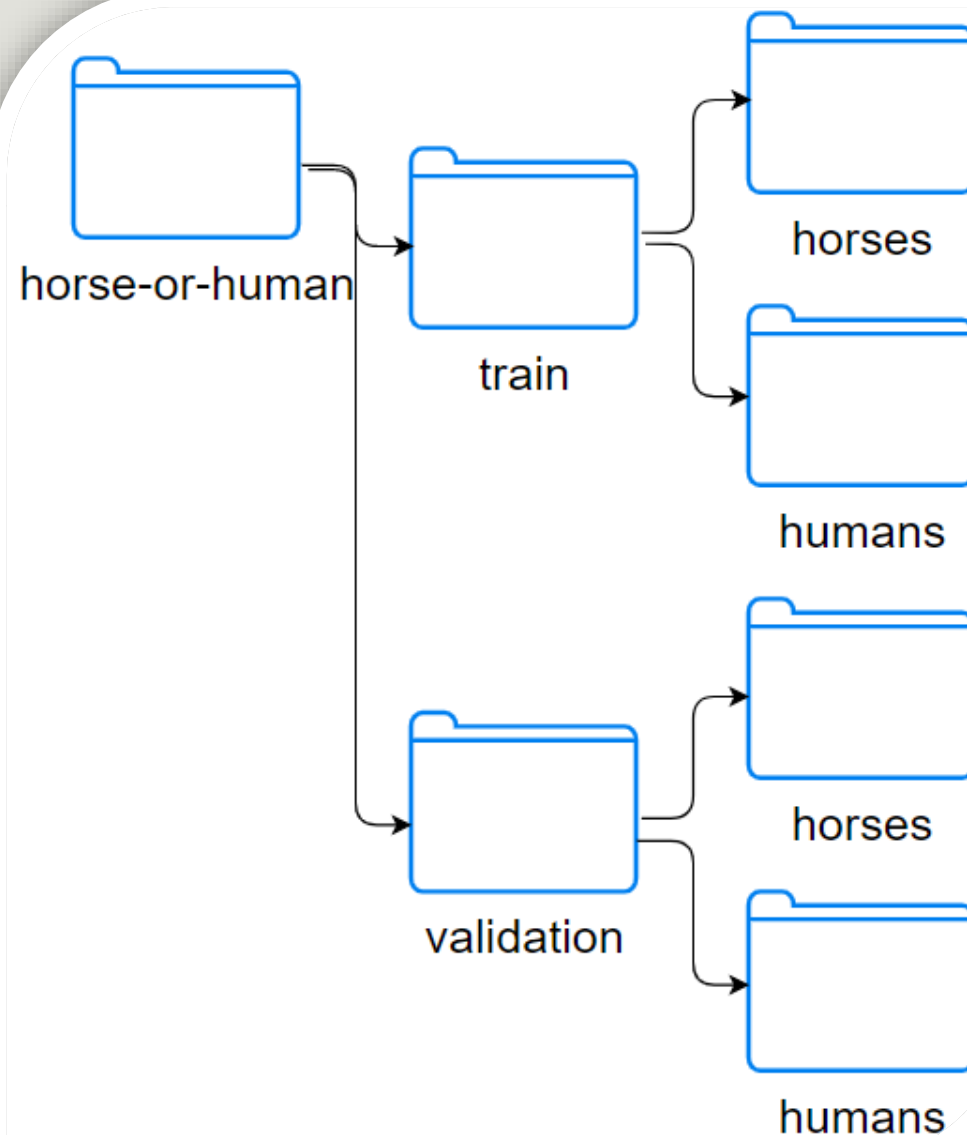
Набор содержит 500 изображений различных видов лошадей в различных позах в различных местах. Он также содержит 527 изображений людей в различных позах и местах. Особое внимание было уделено обеспечению многообразия людей, и с этой целью в учебном наборе присутствуют как мужчины, так и женщины, а также азиаты, чернокожие, южноазиаты кавказцы. Набор проверки добавляет 6 различных фигур разного пола, расы и позы для обеспечения широты данных.



Все изображения имеют 300×300 пикселей в 24-битном цвете.

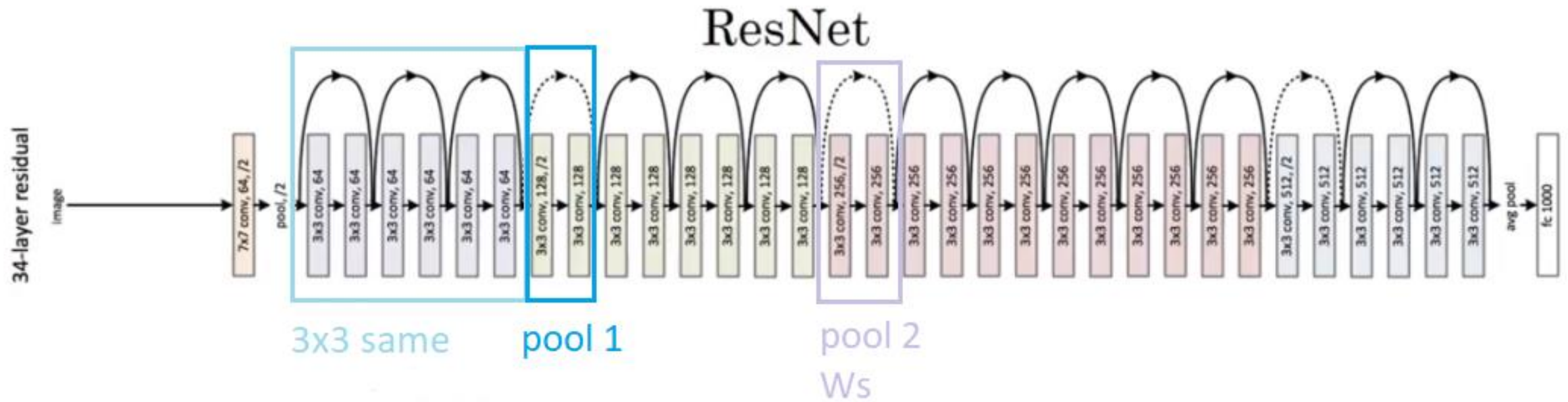
Изображения разбиты по папкам train и validation в каждой из которых есть подпапки: horses, humans. Это позволяет легко автоматически маркировать их с помощью Keras ImageGenerator.

Набор данных лицензирован как CC By 2.0, бесплатный. Доступен для коммерческого и некоммерческого использования.



# ResNet

ResNet — это сокращенное название для Residual Network, дословно — «остаточная сеть».



# Глубокие сети

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, посмотрите на результат ImageNet:

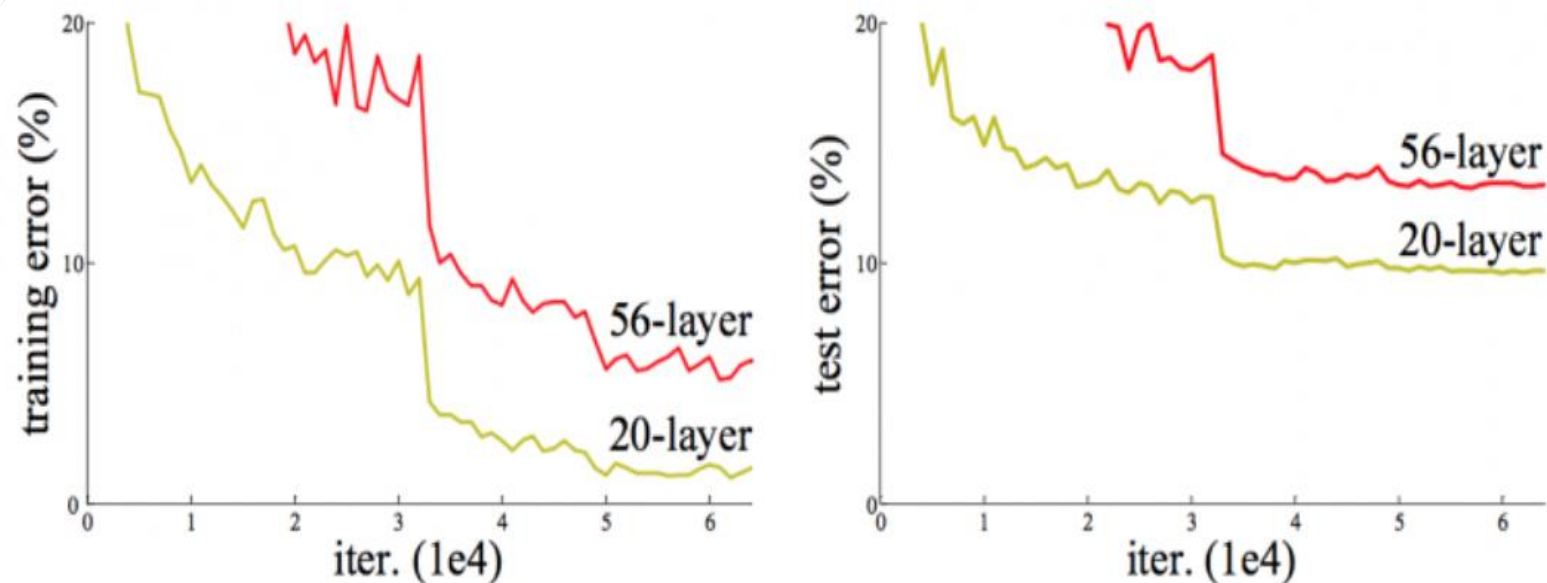


Рисунок 1: Ошибка обучения (слева) и ошибка теста (справа) на CIFAR-10 с 20-уровневыми и 56-слойными «простыми» сетями. Более глубокая сеть имеет большую ошибку обучения и, следовательно, ошибку тестирования. Подобные явления в ImageNet представлены на рисунке 4.

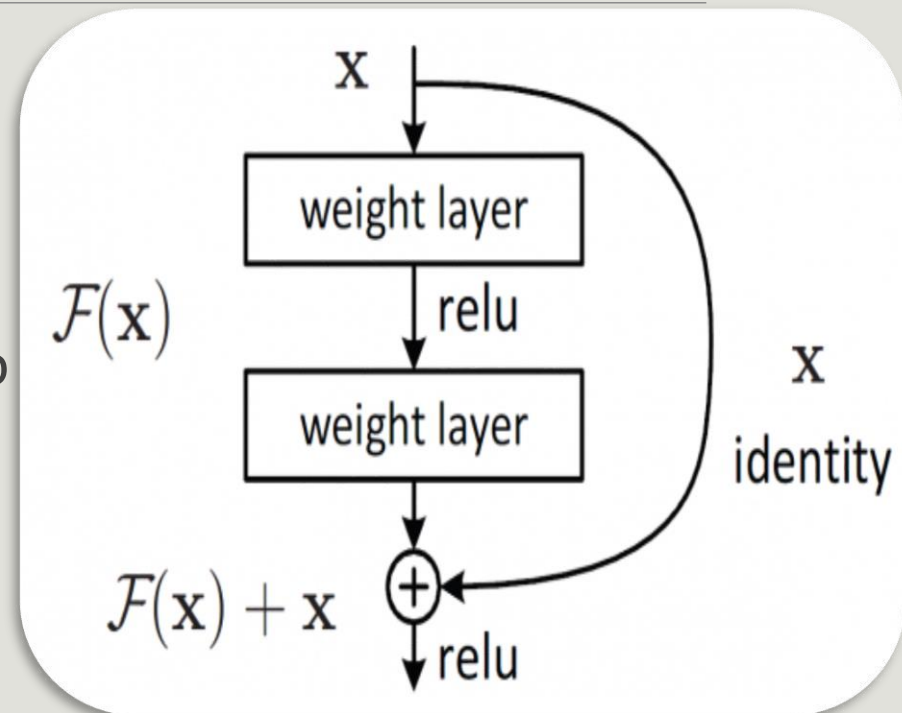


# Глубина и точность

Когда более глубокая сеть начинает сворачиваться, возникает **проблема: с увеличением глубины сети точность сначала увеличивается, а затем быстро ухудшается**.

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**.

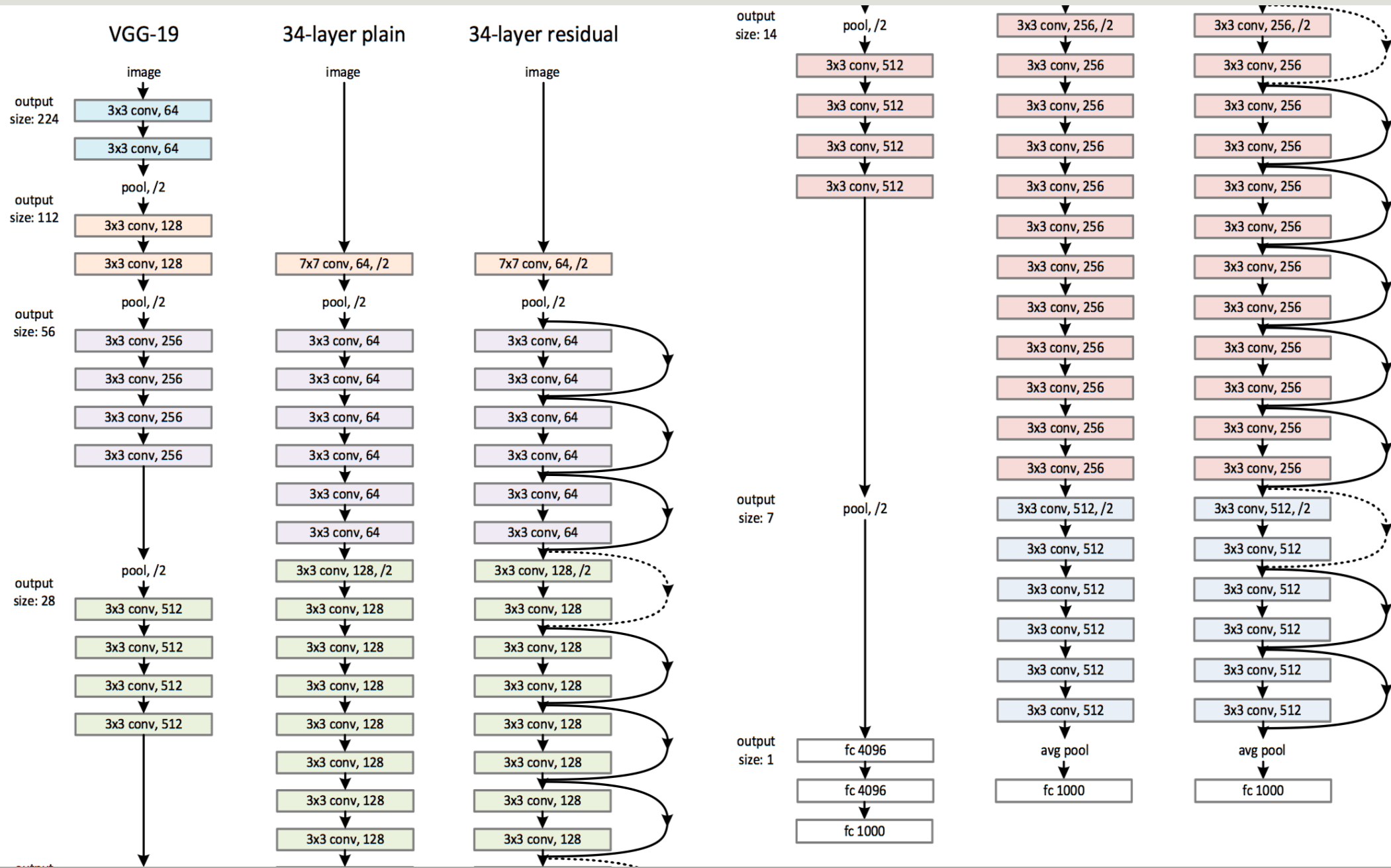


# Соединения быстрого доступа

---

**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.



Сравнение архитектур сетей VGG, Plain, Residual.

# Быстрые соединения $F(x \{W\} + x)$

---

**ResNet:** 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:

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

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 этим 3-слойным узким местом, в результате получается 50-слойная ResNet (см. Таблицу выше). 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 (см. Таблицу выше). 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).

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10 <sup>9</sup>	3.6×10 <sup>9</sup>	3.8×10 <sup>9</sup>	7.6×10 <sup>9</sup>	11.3×10 <sup>9</sup>

На сколько  
глубока ResNet



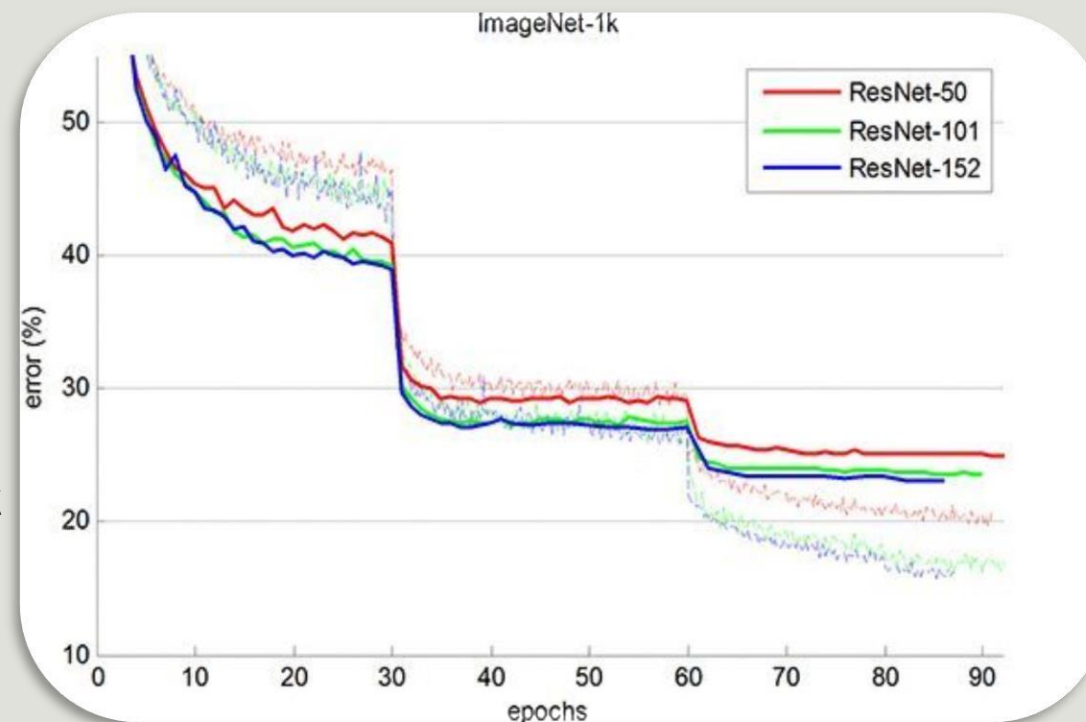
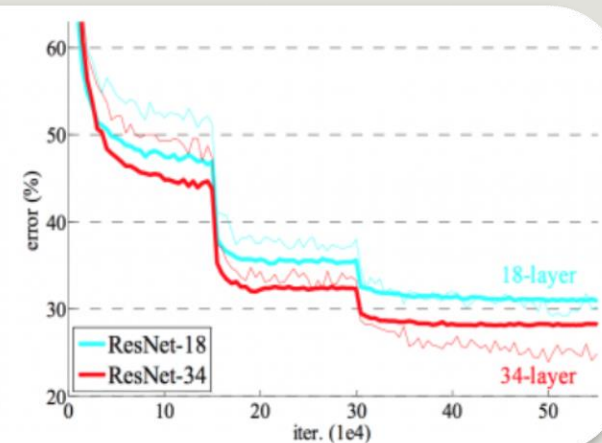
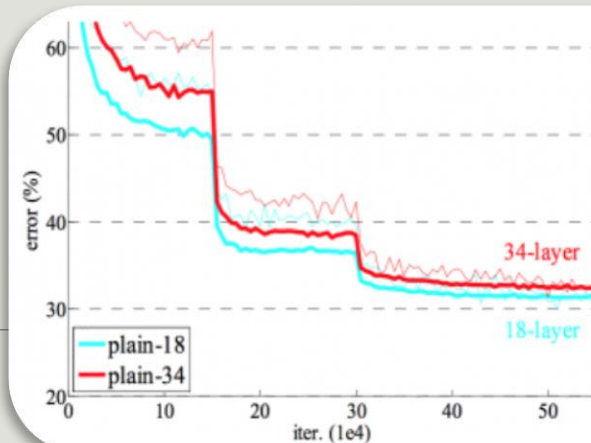


# Всё познаётся в сравнении.

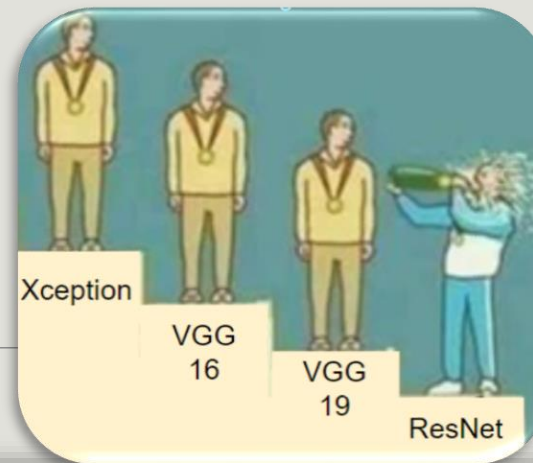
18-уровневая сеть — это просто подпространство в 34-уровневой сети, и она все еще работает лучше. ResNet выигрывает со значительным отрывом, если сеть глубже.

Обучение в ImageNet. Тонкие кривые обозначают ошибку обучения, а жирные кривые обозначают ошибку валидации. Слева: простые сети из 18 и 34 слоев. Справа: ResNets из 18 и 34 слоев. На этом графике остаточные сети не имеют дополнительных параметров по сравнению с их обычными аналогами.

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



# ТОР сетей



Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.713	0.901	138,357,544	23
VGG19	549 MB	0.713	0.900	143,667,240	26
ResNet50	99 MB	0.749	0.921	25,636,712	168
InceptionV3	92 MB	0.779	0.937	23,851,784	159
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
MobileNet	16 MB	0.704	0.895	4,253,864	88
MobileNetV2	14 MB	0.713	0.901	3,538,984	88
DenseNet121	33 MB	0.750	0.923	8,062,504	121
DenseNet169	57 MB	0.762	0.932	14,307,880	169
DenseNet201	80 MB	0.773	0.936	20,242,984	201
NASNetMobile	23 MB	0.744	0.919	5,326,716	-
NASNetLarge	343 MB	0.825	0.960	88,949,818	-

Model	Top-1 error rate	Top-5 error rate
AlexNet	37.5%	15.3%
VGG – 11	29.6%	10.4%
VGG – 15	28.1%	9.4%
VGG – 17	27%	8.8%
VGG - 19	27.3%	9.0%
GoogleNet	32%	6.67%
Inception V3	21.2%	5.6%
Inception V4	20%	5.1%
Xception	21%	5.5%
ResNet – 18	31%	11%
ResNet – 34	26.73%	8.74%
ResNet – 50	24.01%	7.02%
ResNet – 101	22.44%	6.21%
ResNet – 152	22.16%	6.16%
ResNeXt – 50 (32 × 4d)	22.2%	5.9%
DenseNet – 121	25.02%	7.71%
DenseNet – 169	23.8%	6.85%
DenseNet – 201	22.58%	6.34%
SENet	18.68%	3.79%

# Источники:

---

[Horses Or Humans Dataset – lmoroney@ \(laurencemoroney.com\)](http://laurencemoroney.com)

[ResNet: остаточная CNN для классификации изображений \(neurohive.io\)](http://neurohive.io)



Конец!

