Национальный исследовательский
Нижегородский государственный университет им. Н.И. Лобачевского
Институт информационных технологий, математики и механики

Образовательный курс
«Современные методы и технологии
глубокого обучения в компьютерном зрении»

Классификация изображений с большим числом категорий с использованием методов глубокого обучения

При поддержке компании Intel

Гетманская Александра, Кустикова Валентина

Содержание

- Цель лекции
- Постановка задачи классификации изображений
- ImageNet Large Scale Visual Recognition Challenge и набор данных ImageNet
- Обзор глубоких моделей для классификации изображений на наборе данных ImageNet
- Сравнение качества классификации и сложности глубоких моделей на наборе данных ImageNet
- Заключение

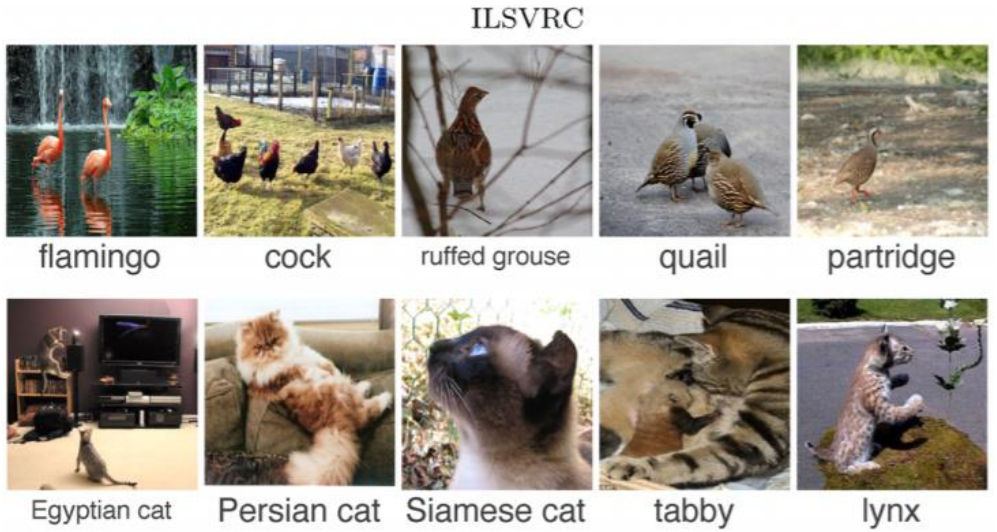
Цель лекции

- ***Цель*** – изучить глубокие нейросетевые модели для решения задачи классификации

ПОСТАНОВКА ЗАДАЧИ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ

Постановка задачи (1)

- Задача классификации изображений состоит в том, чтобы поставить в соответствие изображению класс объектов, содержащихся на этом изображении
- Примеры изображений и соответствующих им классов:



* Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A.C., Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge // International Journal of Computer Vision, 2015.

Постановка задачи (2)

- Исходное изображение представлено набором интенсивностей пикселей $I = \left(I_{ij}^{k}\right)_{\substack{0 \le i < w \\ 0 \le j < h \\ 0 \le k < 3}}$, где w и h – ширина и высота изображения, k – количество каналов
- Определено множество допустимых классов объектов на изображении $C = \{0, 1, \ldots N - 1\}$, множество идентификаторов классов однозначно соответствует множеству названий классов
- ***Задача классификации изображений*** состоит в том, чтобы каждому изображению поставить в соответствие класс, которому оно принадлежит

$$\varphi : I \to C$$

IMAGENET LARGE SCALE VISUAL RECOGNITION CHALLENGE И НАБОР ДАННЫХ IMAGENET

ImageNet Large Scale Visual Recognition Challenge

- ImageNet Large Scale Visual Recognition Challenge (ILSVRC) – конкурс по классификации изображений с большим числом категорий и детектированию объектов на изображениях
- С 2010 по 2017 годы базируется на [http://www.image-net.org], с 2017 года переехал на платформу Kaggle
- ImageNet – открытый набор данных, предоставляемый в рамках конкурса ILSVRC, содержит 14 197 122 изображений

* Russakovsky O., et al. ImageNet Large Scale Visual Recognition Challenge. – 2015. – [https://arxiv.org/pdf/1409.0575.pdf].

Набор данных ImageNet

- Состоит из 14 197 122 изображений, принадлежащих 21 841 категориям из иерархии WordNet*
- Иерархия содержит 27 категорий высокого уровня
- 1 034 908 изображений содержат разметку для задачи детектирования объектов (размечены окаймляющие прямоугольники для объектов), эти данные используются и для задачи классификации
- Разрешение изображений варьируется, среднее разрешение составляет 400x350 пикселей
- Изображения собраны из различных источников, создатели набора данных не имеют авторских прав на изображения

* Jia D., Dong W., Socher R., Li L.-J., Li K., Li F.-F. ImageNet: A large-scale hierarchical image database // In the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2009. – P. 248-255. – [https://ieeexplore.ieee.org/document/5206848].

Набор данных ImageNet для классификации изображений по данным конкурса ILSVRC 2012

- 1 000 категорий изображений
- Минимальное разрешение – 75x56 пикселей
- Максимальное разрешение – 4288x2848 пикселей
- Размер тренировочной выборки – 1 200 000 изображений
- Размер валидационной выборки – 50 000 изображений
- Размер тестовой выборки – 150 000 изображений

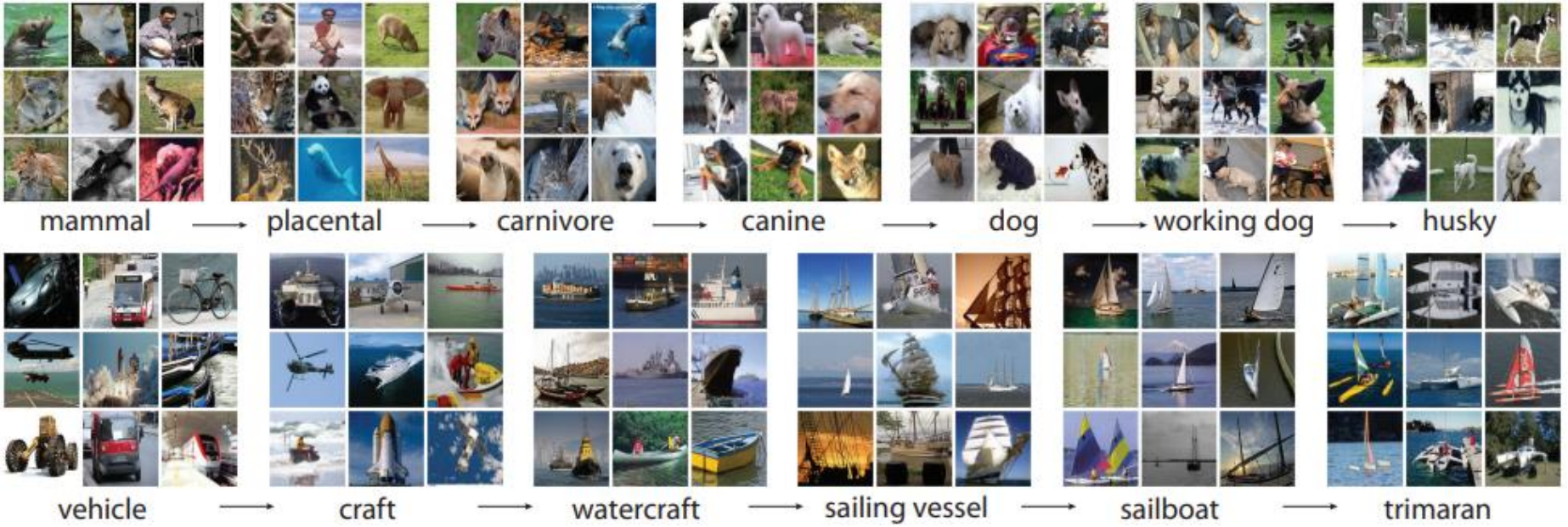
Иерархия классов WordNet (1)

- WordNet* – большая лексическая база слов английского языка
- Основным отношением между словами в WordNet является ***синонимия***
- ***Синонимы*** – слова, близкие по значению, взаимозаменяемые во многих контекстах
- Синонимы сгруппированы в ***неупорядоченные множества*** (synset)
- Группы синонимов связаны следующими отношениями:
  - ***Гиперонимия*** (гипонимия) – связь общего и частного (например, кровать – это мебель)
  - ***Меронимия*** (партонимия) – связь между объектами и их частями (например, «двигатель» – мероним по отношению к термину «автомобиль»)

* WordNet. A Lexical Database for English [https://wordnet.princeton.edu].

Иерархия классов WordNet (2)

- WordNet* содержит около 80 000 существительных
- Цель разработки набора данных ImageNet для каждого множества синонимов подобрать 500-1000 изображений



* WordNet. A Lexical Database for English [https://wordnet.princeton.edu].

** Ye T. Visual Object Detection from Lifelogs using Visual Non-lifelog Data. – 2018. – [https://www.researchgate.net/publication/324797660_Visual_Object_Detection_from_Lifelogs_using_Visual_Non-lifelog_Data].

ГЛУБОКИЕ МОДЕЛИ ДЛЯ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ НА НАБОРЕ ДАННЫХ IMAGENET

Рассматриваемые модели (1)

Наращивание глубины

- ***AlexNet (2012)***
  - Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks // Advances in neural information processing systems. – 2012. – [http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf].
- ***OverFeat (2013)***
  - Sermanet P., Eigen D., Zhang X., Mathieu M., Fergus R., LeCun Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. – 2013. – [https://arxiv.org/pdf/1312.6229.pdf].
- ***VGG-16, VGG-19, GoogLeNet (2014)***
  - Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. – 2014. – [https://arxiv.org/pdf/1409.1556.pdf].
  - Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going Deeper with Convolutions. – 2014. – [https://arxiv.org/pdf/1409.4842.pdf].

Рассматриваемые модели (2)

Решение проблемы деградации модели

- ***ResNet-*(50, 101, 152), Inception-v*(2,3) (2015)***
  - He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. – 2015. – [https://arxiv.org/pdf/1512.03385.pdf].
  - Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. – 2015. – [https://arxiv.org/pdf/1502.03167.pdf].
  - Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture for Computer Vision. – 2015. – [https://arxiv.org/pdf/1512.00567.pdf], [https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.pdf] (опубликованная версия).
- ***DenseNet-* (121, 169, 201, 264), Xception (2016)***
  - Huang G., Liu Z., Maaten L., Weinberger K.Q. Densely Connected Convolutional Networks. – 2016. – [https://arxiv.org/pdf/1608.06993.pdf].
  - Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].

Снижение количества параметров модели

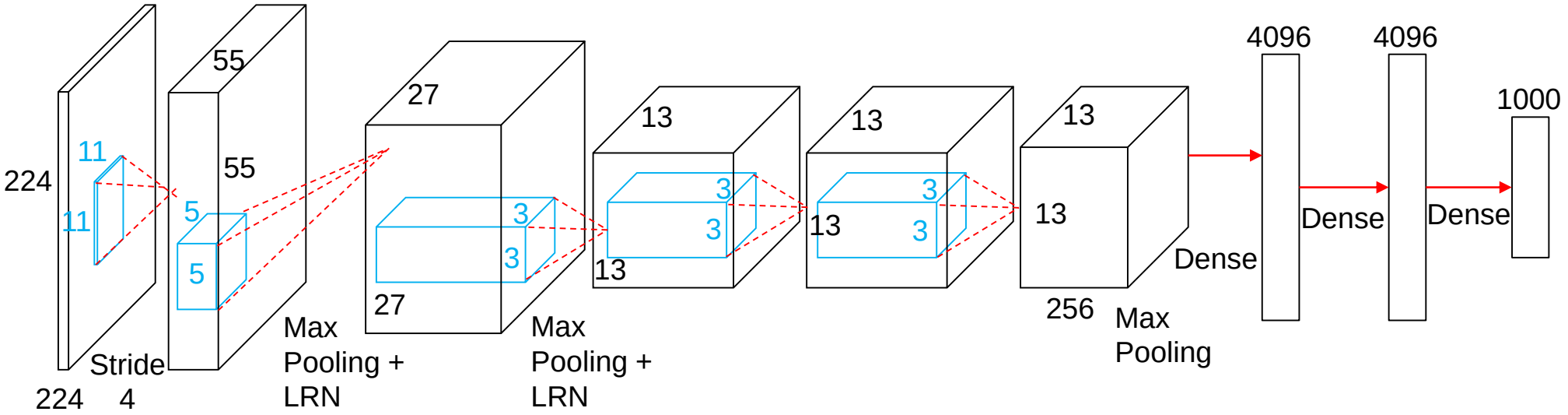
Рассматриваемые модели (3)

Снижение сложности модели

- ***MobileNet, ResNeXT-* (2017)***
  - Howard A.G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. – 2017. – [https://arxiv.org/pdf/1704.04861.pdf].
  - Xie S., Girshick R., Dollar P., Tu Z., He K. Aggregated Residual Transformations for Deep Neural Networks. – 2017. – [https://arxiv.org/pdf/1611.05431v2.pdf], [https://ieeexplore.ieee.org/document/8100117] (опубликованная версия).
- ***MobileNetV2 (2018)***
  - Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. – 2018. – [https://arxiv.org/pdf/1801.04381.pdf], [https://ieeexplore.ieee.org/document/8578572] (опубликованная версия).
- ***EfficientNet-* (B0,…,B7) (2019)***
  - Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. – 2019. – [https://arxiv.org/pdf/1905.11946.pdf].

AlexNet (1)

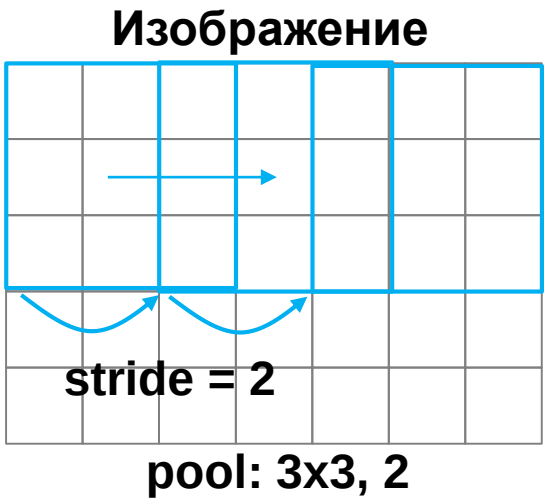
- AlexNet – первая глубокая сверточная нейронная сеть
- Разработчики сети выиграли конкурс по классификации изображений LSVRC-2012 на наборе данных ImageNet
- Ошибка классификации составила 15.3% по сравнению с ошибкой в 25.7%, полученной годом ранее



* Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks // Advances in neural information processing systems. – 2012. – [http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf].

AlexNet (2)

- Особенности модели:
  - Вход сети – трехканальное изображение 224x224 пикселя
  - В качестве функции активации используется «положительная срезка» (Rectified Linear Unit, ReLU)
  - Использование dropout-слоев (обнуление выходов нейронов с вероятностью 0.5)
  - Использование слоев пространственного объединения с перекрытиями (overlapping pooling)
  - Локальная нормализация выходов (Local Response Normalization, LRN) – нормализация выходных значений по размерности, соответствующей глубине выходной карты признаков



AlexNet (3)

- Сложность модели:
  - Сеть содержит 62.3 млн. параметров
  - Прямой проход требует выполнения ~1 миллиарда операций
  - Сверточные слои, на которые приходится 6% всех параметров, производят 95% вычислений

- Особенности обучения:
  - Высокая скорость обучения за счет использования функции активации ReLU
  - Увеличение количества данных (data augmentation) за счет применения операций сдвига и зеркального отражения
  - Обучение на двух видеокартах

OverFeat (1)

- OverFeat – глубокая модель, построенная на базе AlexNet, которая решает одновременно задачи классификации изображений, локализации и детектирования объектов в рамках конкурса ILSVRC
- Локализация предполагает определение положения одного объекта – построение окаймляющего прямоугольника
- Отличия детектирования объектов от локализации:
  - Объектов может быть произвольное количество
  - Изображения содержат объекты небольшого размера
  - При отсутствии объектов предполагается предсказание класса, которому принадлежит фон
  - Разные показатели качества детектирования и локализации

* Sermanet P., Eigen D., Zhang X., Mathieu M., Fergus R., LeCun Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. – 2013. – [https://arxiv.org/pdf/1312.6229.pdf].

OverFeat (2)

- Авторы предлагают две модели, которые очень похожи по архитектуре на AlexNet:
  - «Быстрая» модель (Fast Model)

Layer	1	2	3	4	5	6	7	Output (8)
Stage	conv+max	conv+max	conv	conv	conv+max	full	full	full
K channels	96	256	512	1024	1024	3072	4096	1000
Filter size	11x11	5x5	3x3	3x3	3x3	–	–	–
Convolution stride	4x4	1x1	1x1	1x1	1x1	–	–	–
Pooling size	2x2	2x2	–	–	2x2	–	–	–
Pooling stride	2x2	2x2	–	–	2x2	–	–	–
Zero-Padding size	–	–	1x1x1x1	1x1x1x1	1x1x1x1	–	–	–
Spatial input size	231x231	24x24	12x12	12x12	12x12	6x6	1x1	1x1

AlexNet без пространственного объединения с перекрытиями и локальной нормализации выходов

OverFeat (3)

- Авторы предлагают две модели, которые очень похожи по архитектуре на AlexNet:
  - «Точная» модель (Accurate Model)

Layer	1	2	3	4	5	6	7	8	Output (9)
Stage	conv+ max	conv+ max	conv	conv	conv	conv+max	full	full	full
K channels	96	256	512	512	1024	1024	4096	4096	1000
Filter size	7x7	7x7	3x3	3x3	3x3	3x3	–	–	–
Convolution stride	2x2	1x1	1x1	1x1	1x1	1x1	–	–	–
Pooling size	3x3	2x2	–	–	–	3x3	–	–	–
Pooling stride	3x3	2x2	–	–	–	3x3	–	–	–
Zero-Padding size	–	–	1x1x1x1	1x1x1x1	1x1x1x1	1x1x1x1	–	–	–
Spatial input size	221x221	36x36	15x15	15x15	15x15	15x15	5x5	1x1	1x1

Цветом отмечены отличия от «быстрой» модели

OverFeat (4)

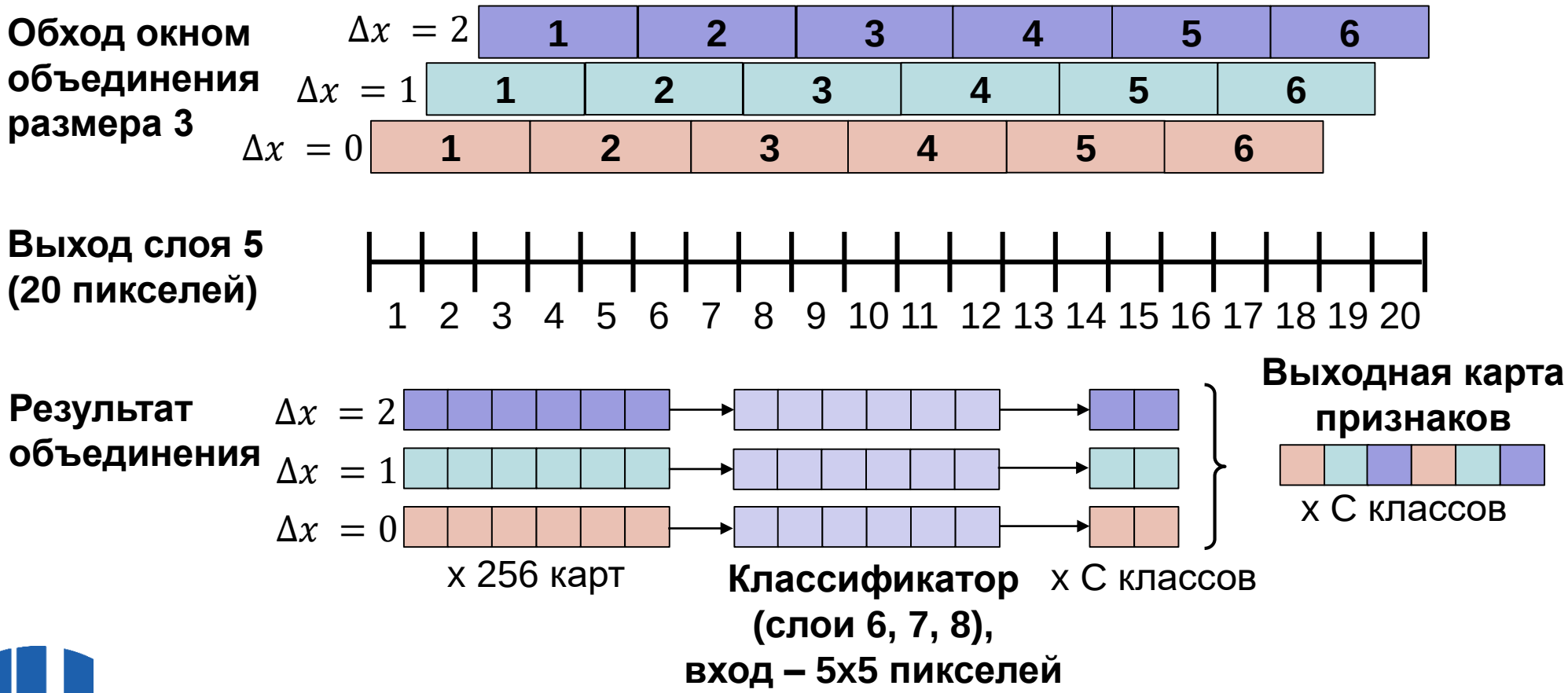
- Отличия от модели AlexNet:
  - Отсутствие пространственного объединения с перекрытиями (замена размера ядра с 3x3 на 2x2)
  - Отсутствие локальной нормализации выходов на первом и третьем слоях
  - Применение многомасштабной классификации изображения (multiscale classification)
    - Объекты на изображении разного размера
    - Идея – классифицировать разные масштабы изображения и принимать интегральное решение
    - Используется 6 разных масштабов входного изображения, для которых строятся карты признаков (выход слоя 5). Полученные карты признаков объединяются и передаются на вход классификатору, который формирует финальное решение о принадлежности изображения классу

OverFeat (5)

- Схема объединения карт признаков, полученных на разных масштабах изображения:
  - Получаем выходные карты признаков с пятого слоя модели
  - Применяем к каждой полученной карте признаков операцию пространственного объединения по максимуму без перекрытий с ядром 3x3. Операция выполняется 9 раз (3x3) для всевозможных сдвигов ядра по горизонтали и вертикали $\Delta x, \Delta y \in \{0,1,2\}$, в результате формируется 9 карт признаков
  - Каждая карта подается на вход классификатору (слои 6, 7, 8). Карты признаков имеют разный размер, а размер входа классификатора фиксирован, поэтому классификатор применяется в манере «скользящего» окна (sliding window)
  - Форма выходных карт приводится к трехмерному тензору (две пространственных размерности и количество классов)

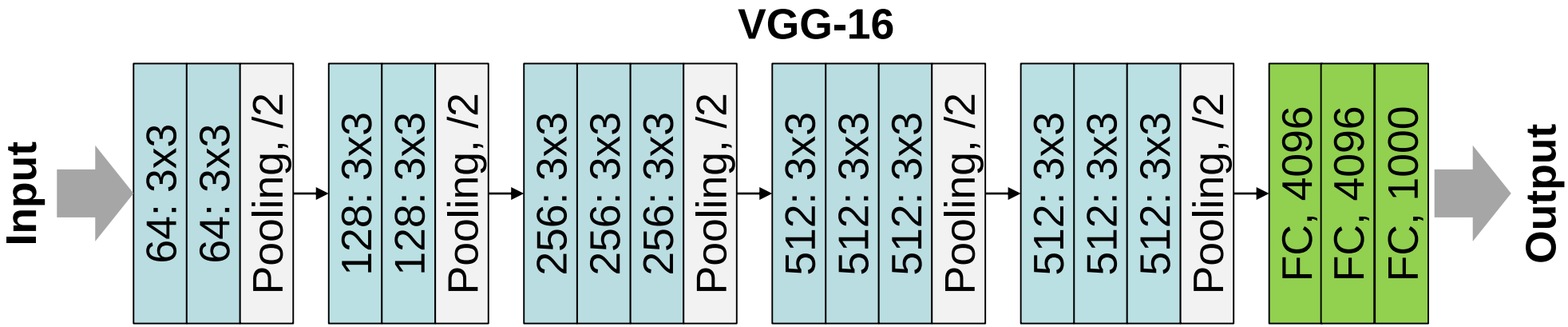
OverFeat (6)

- Схема объединения карт признаков, полученных на разных масштабах изображения:



VGG-16, 19 (1)

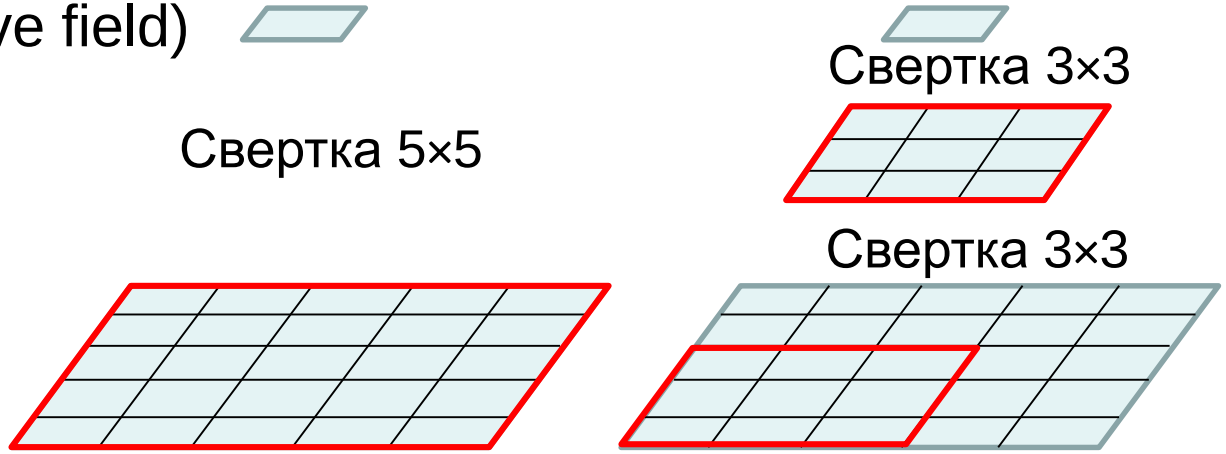
- VGG-* является улучшением модели AlexNet, принципиальное отличие состоит в том, что большие ядра сверточных фильтров (11 и 5) заменены последовательностью сверток размера 3x3



* Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. – 2014. – [https://arxiv.org/pdf/1409.1556.pdf].

VGG-16, 19 (2)

- Свертку с фильтром 5x5 можно заменить двумя последовательными свертками с фильтрами размера 3x3
- При этом формируется сеть с меньшим числом параметров (25 vs. 18), но с тем же размером входа и рецептивного поля (receptive field)



- VGG-19 (16 сверточных слоев) содержит большее количество сверточных слоев по сравнению VGG-16. Количество блоков, содержащих последовательность сверток и операцию пространственного объединения одинаковое

ResNet-50, 101, 152 (1)

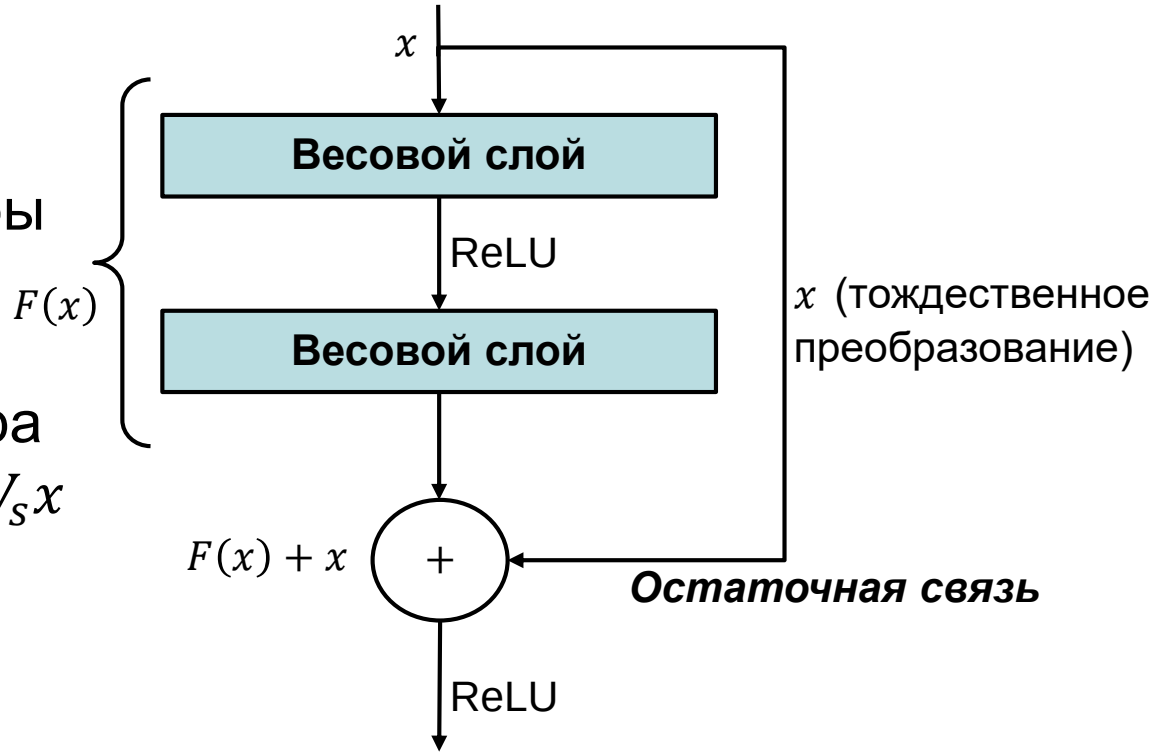
- К началу 2015 года общая тенденция в разработке глубоких моделей состоит в увеличении количества сверточных слоев
- С ростом глубины сети точность насыщается и затем быстро начинает уменьшаться (деградировать)
- ***Проблема деградации глубоких моделей*** не является следствием переобучения модели, добавление дополнительных слоев приводит к еще большему значению тренировочной ошибки из-за затухающих градиентов (vanishing gradients)
- ***Остаточные сети*** (Residual Network, ResNet) решают проблему
- Идея – предположить, что некоторая последовательность слоев сети аппроксимирует не базовое отображение, а остаточное отображение

* He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. – 2015. – [https://arxiv.org/pdf/1512.03385.pdf].

ResNet-50, 101, 152 (2)

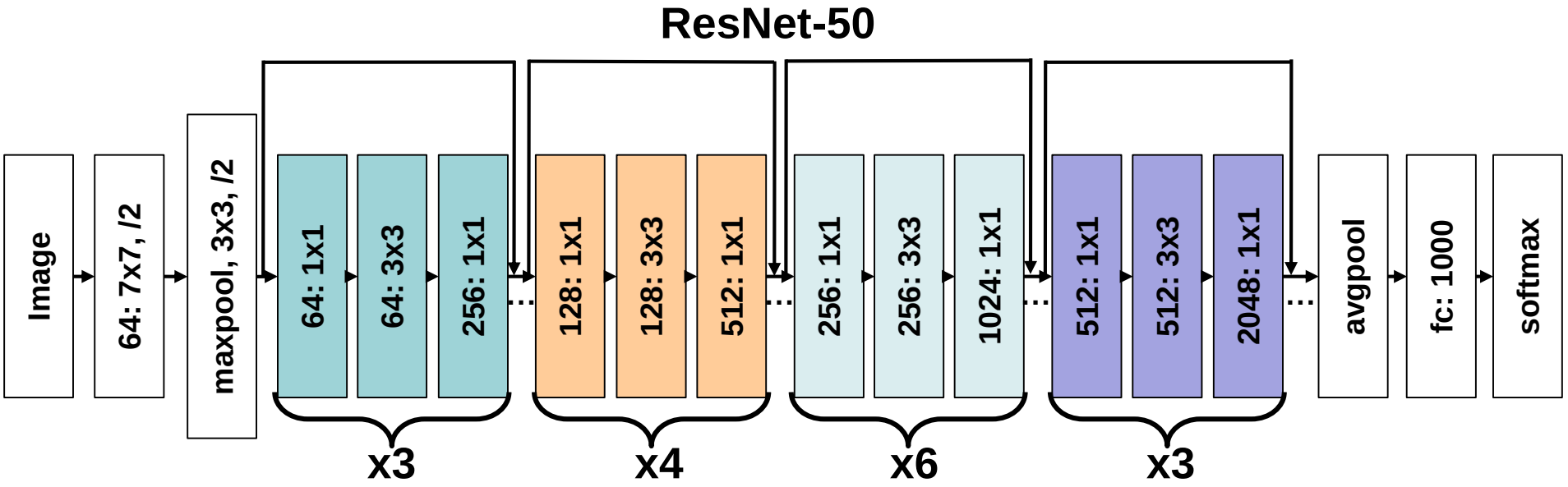
- $H(x)$ – базовое отображение
- $F(x) = H(x) - x$ – остаточное отображение
- Базовое отображение можно представить как поэлементное сложение карт признаков $F(x) + x$

- $F(x)$ и x могут иметь разную размерность, чтобы исправить эту ситуацию достаточно выполнить проекцию входного вектора признаков $y = F(x, W_i) + W_s x$



ResNet-50, 101, 152 (3)

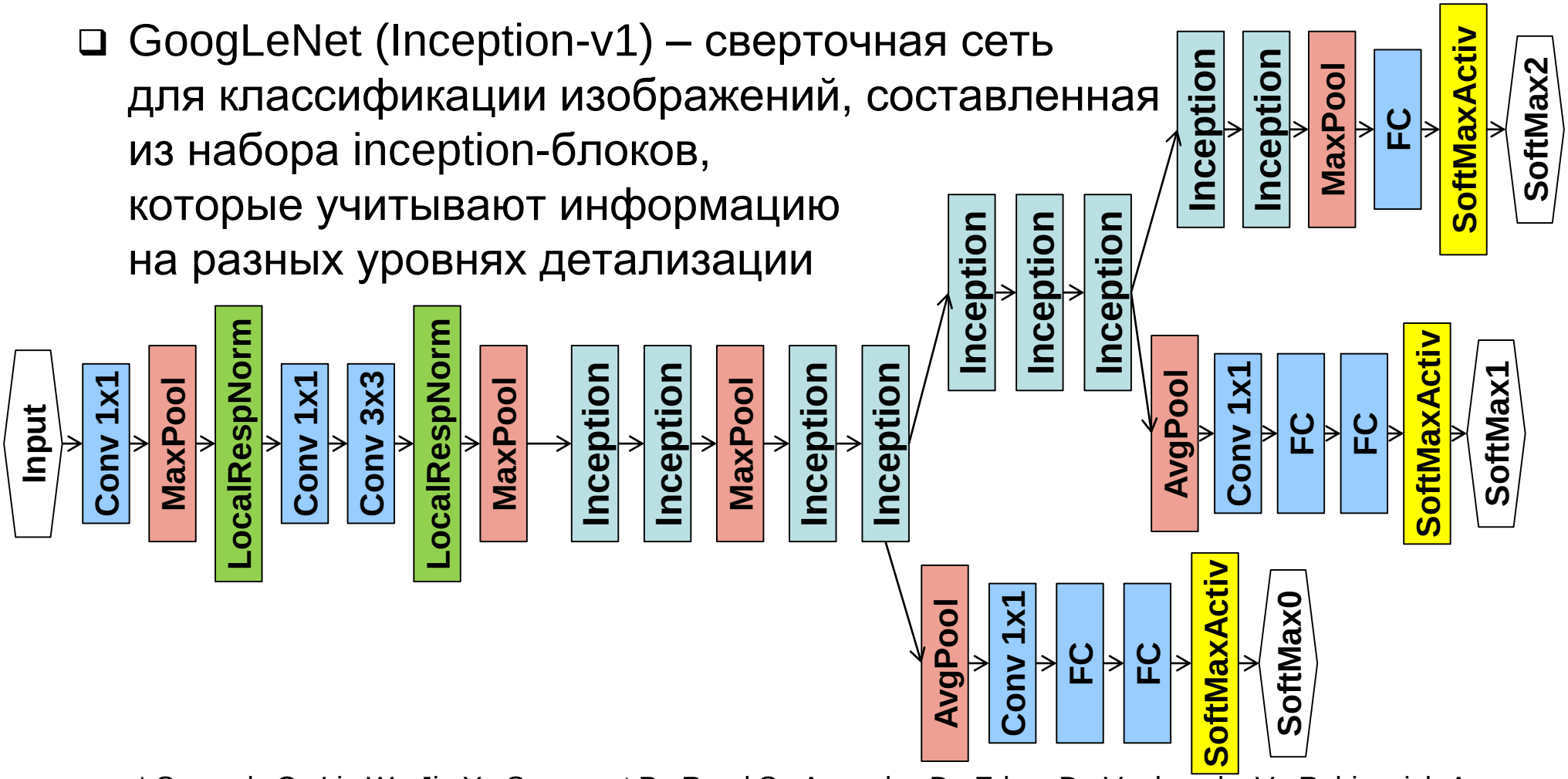
- Модели ResNet-50, 101, 152 построены по принципу наращивания сверточных слоев, проблема деградации моделей решается посредством введения остаточных связей для каждой последовательной тройки сверточных слоев



* He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. – 2015. – [https://arxiv.org/pdf/1512.03385.pdf].

GoogLeNet (1)

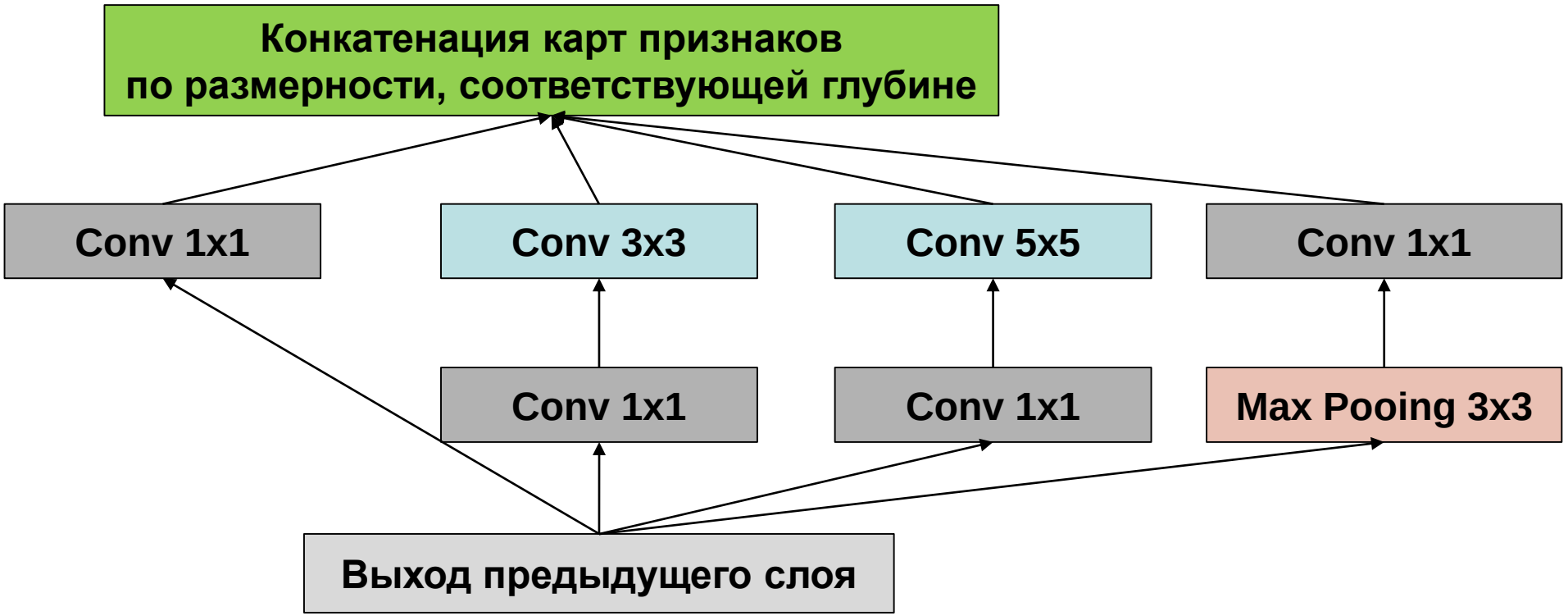
- GoogLeNet (Inception-v1) – сверточная сеть для классификации изображений, составленная из набора inception-блоков, которые учитывают информацию на разных уровнях детализации



* Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going Deeper with Convolutions. – 2014. – [https://arxiv.org/pdf/1409.4842.pdf].

GoogLeNet (2)

- Типовая структура inception-блока:



* Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going Deeper with Convolutions. – 2014. – [https://arxiv.org/pdf/1409.4842.pdf].

GoogLeNet (3)

- GoogLeNet состоит из 9 inception-блоков
- Каждый блок использует свертки разного размера, чтобы получать признаки разного масштаба
- Ядра сверток небольшого размера, поэтому уменьшается количество обучаемых параметров модели. Сеть GoogLeNet содержит примерно в 10 раз меньше параметров, чем AlexNet
- Вспомогательные классификаторы (softmax0 и softmax1) предсказывают класс на основе признаков более низкого уровня. Эти классификаторы позволят «протолкнуть» градиенты к ранним слоям и тем самым уменьшить эффект затухания градиента
- При выводе (inference) ветки, ведущие к вспомогательным выходам, удаляются

Inception-v2

- Inception-v2 – модификация модели GoogLeNet
  - Свертки с ядром 5x5 заменены двумя последовательными свертками с ядрами 3x3
  - В качестве функции активации используется «положительная срезка» (Rectified Linear Unit,ReLU)
  - Перед применением функции активации выполняется ***нормализация по пачке*** обрабатываемых примеров (batch normalization)
    - Для пачки примеров формируется набор карт признаков
    - Для отдельных элементов в карте признаков определяется математическое ожидание и среднеквадратическое отклонение
    - Значение каждого элемента уменьшается на математическое ожидание и делится на среднеквадратическое отклонение

* Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. – 2015. – [https://arxiv.org/pdf/1502.03167.pdf].

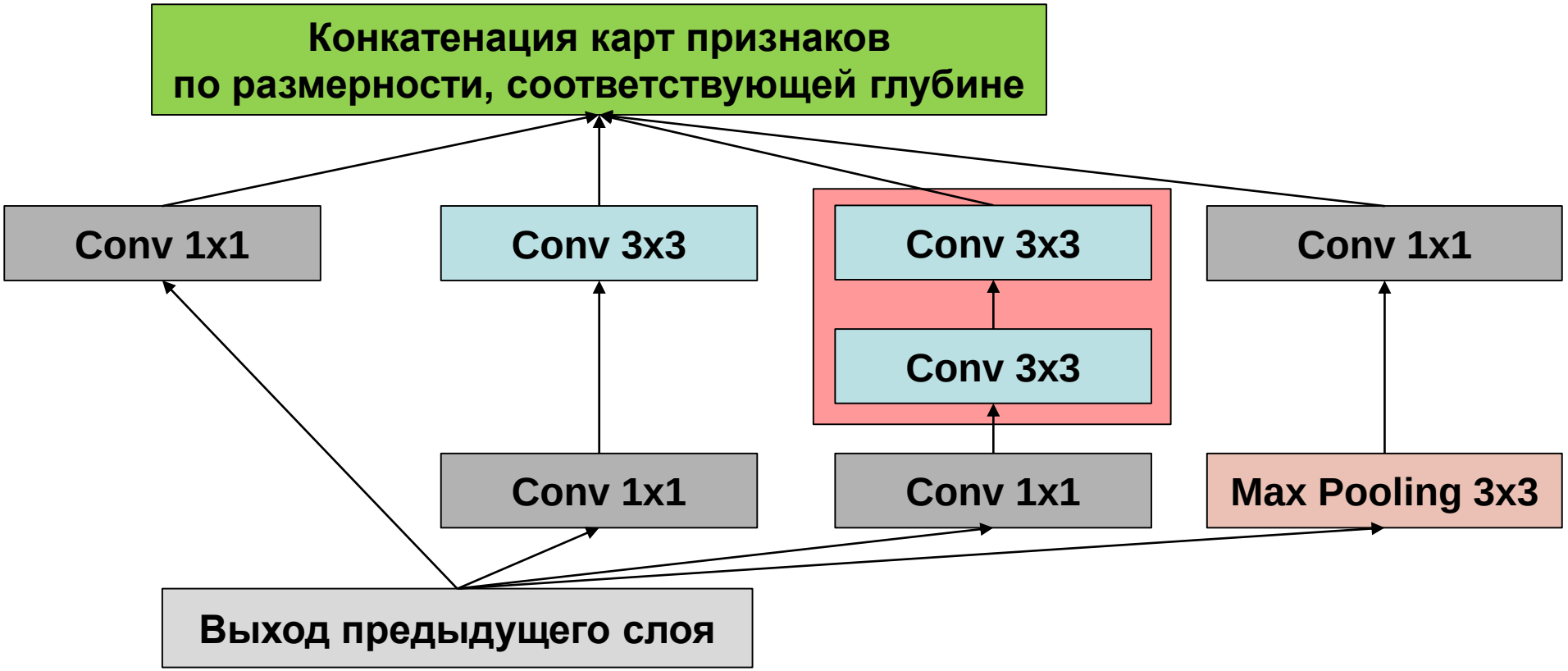
Inception-v3 (1)

- Inception-v3 – модификация модели GoogLeNet
  - Сверточные слои с большими ядрами заменены свертками с ядрами меньшего размера
  - Содержит 3 вида модифицированных inception-блоков
- Принципы построения эффективной модели:
  - Избегать «узких горлышек» в представлении сети, особенно на начальных слоях
  - Пространственную агрегацию следует выполнять по картам признаков более низкой размерности для снижения вычислительной сложности (см. схему inception-блока)
  - Сбалансировать глубину и ширину сети

* Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture for Computer Vision. – 2015. – [https://arxiv.org/pdf/1512.00567.pdf].

Inception-v3 (2)

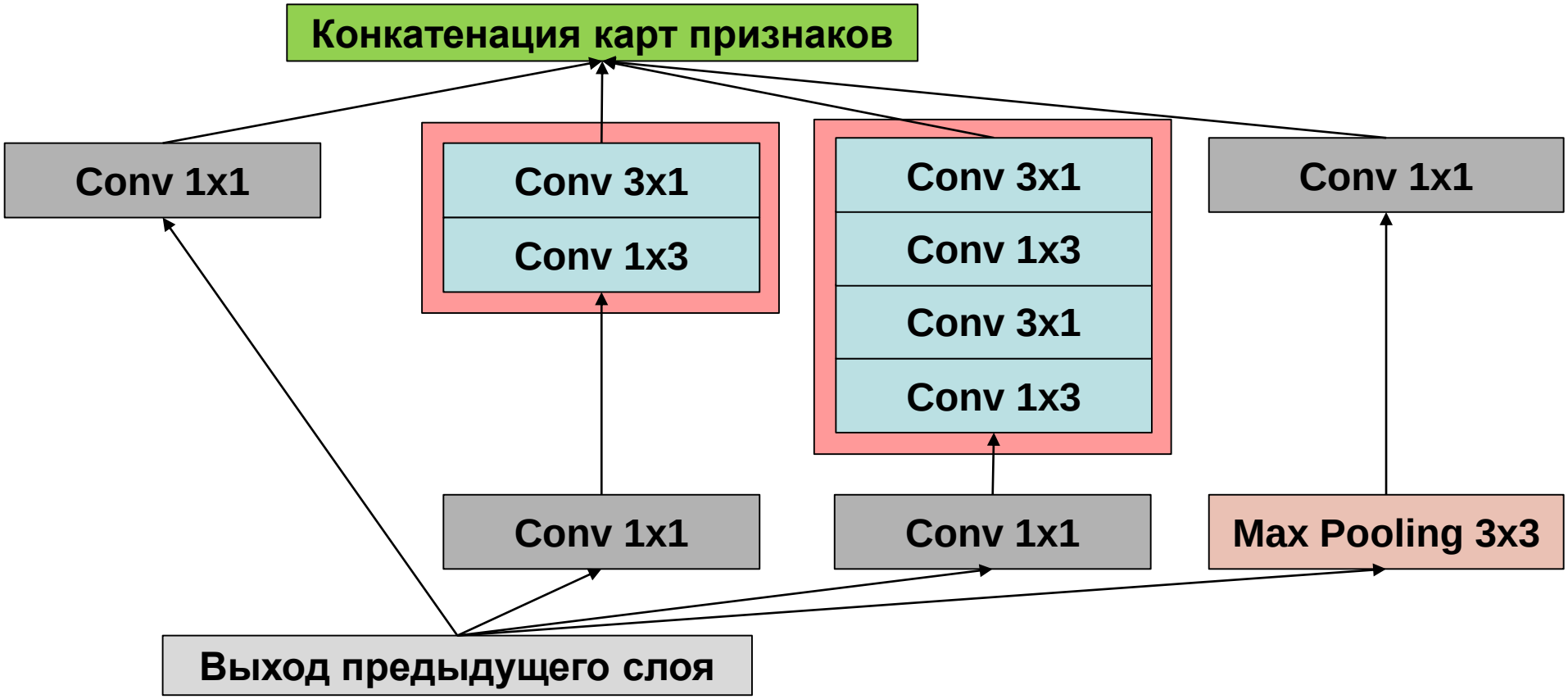
- ***Inception-блок A:*** замена сверток с ядром 5x5 на две свертки с ядрами 3x3



* Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture for Computer Vision. – 2015. – [https://arxiv.org/pdf/1512.00567.pdf].

Inception-v3 (3)

- ***Inception-блок B:*** факторизация сверток NxN на Nx1 и 1xN



* Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture for Computer Vision. – 2015. – [https://arxiv.org/pdf/1512.00567.pdf].

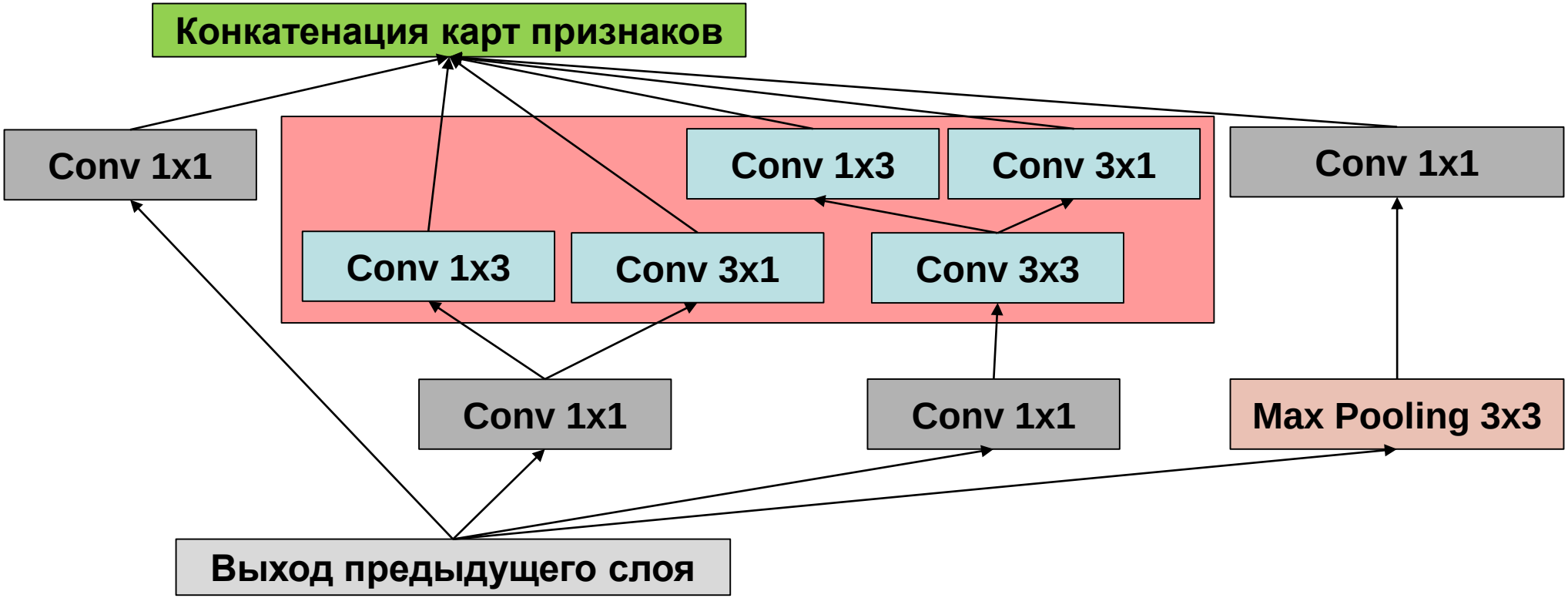
Inception-v3 (4)

- При использовании ***фильтра 3x3*** количество параметров составляет ***3x3=9***
- При использовании ***фильтров 3x1 и 1x3*** количество параметров составляет ***3x1+1x3=6***
- ***Количество параметров уменьшается на 33%***

- ***Количество параметров сокращается для всей сети, вероятность того, что они будут переобучены, меньше, и, сеть может быть глубже!***

Inception-v3 (5)

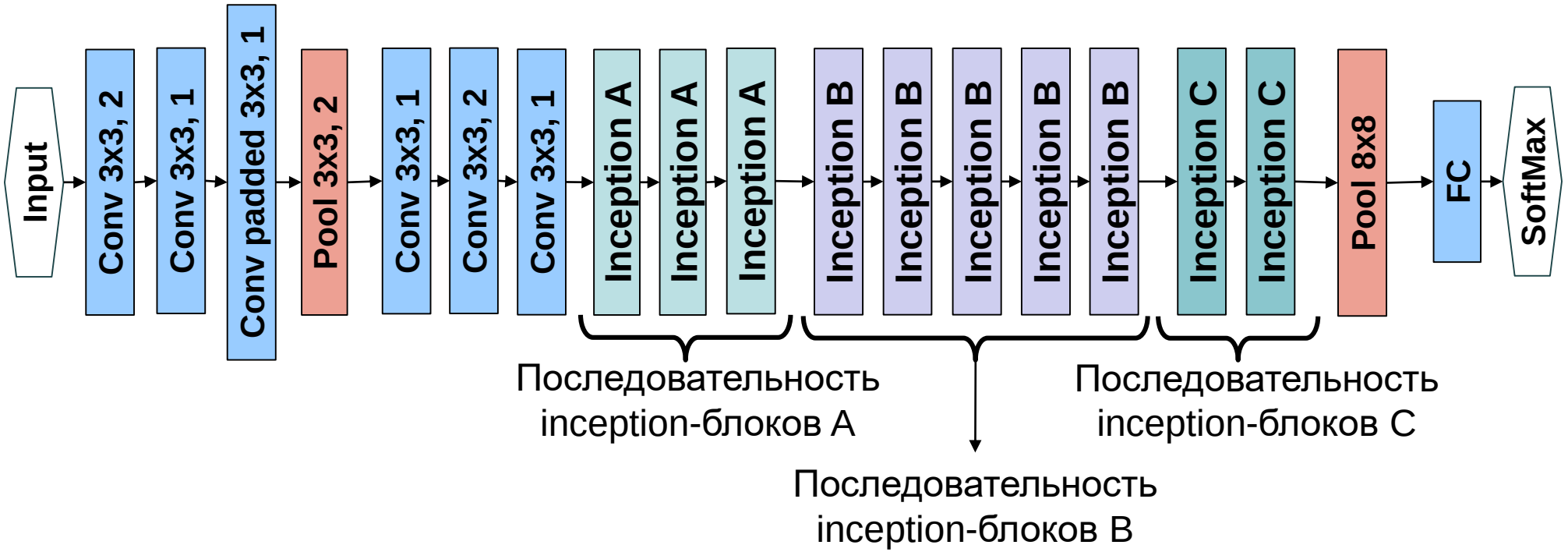
- ***Inception-блок C:*** расширение набора выходов



* Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture for Computer Vision. – 2015. – [https://arxiv.org/pdf/1512.00567.pdf].

Inception-v3 (6)

- Базовая структура модели Inception-v3:



* Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture for Computer Vision. – 2015. – [https://arxiv.org/pdf/1512.00567.pdf], [https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.pdf] (последняя версия статьи).

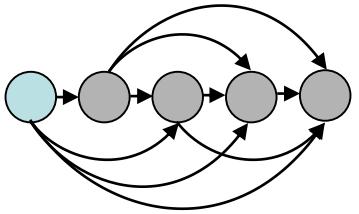
Inception-v3 (7)

- В работе* авторов модели Inception-v3 рассматриваются модификации представленной базовой модели:
  - Изменяется количество inception-блоков разного типа
  - Добавляется нормализация по пачке обрабатываемых изображений
  - Добавляется вспомогательный классификатор
  - Используется механизм регуляризации модели посредством оценки эффекта от прореживания меток в ходе обучения модели* (Model Regularization via Label Smoothing)
  - Реализуется схема эффективного снижения размерности карты признаков* (Efficient Grid Size Reduction)

* Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture for Computer Vision. – 2015. – [https://arxiv.org/pdf/1512.00567.pdf], [https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.pdf] (последняя версия статьи).

DenseNet-121, 169, 201, 264 (1)

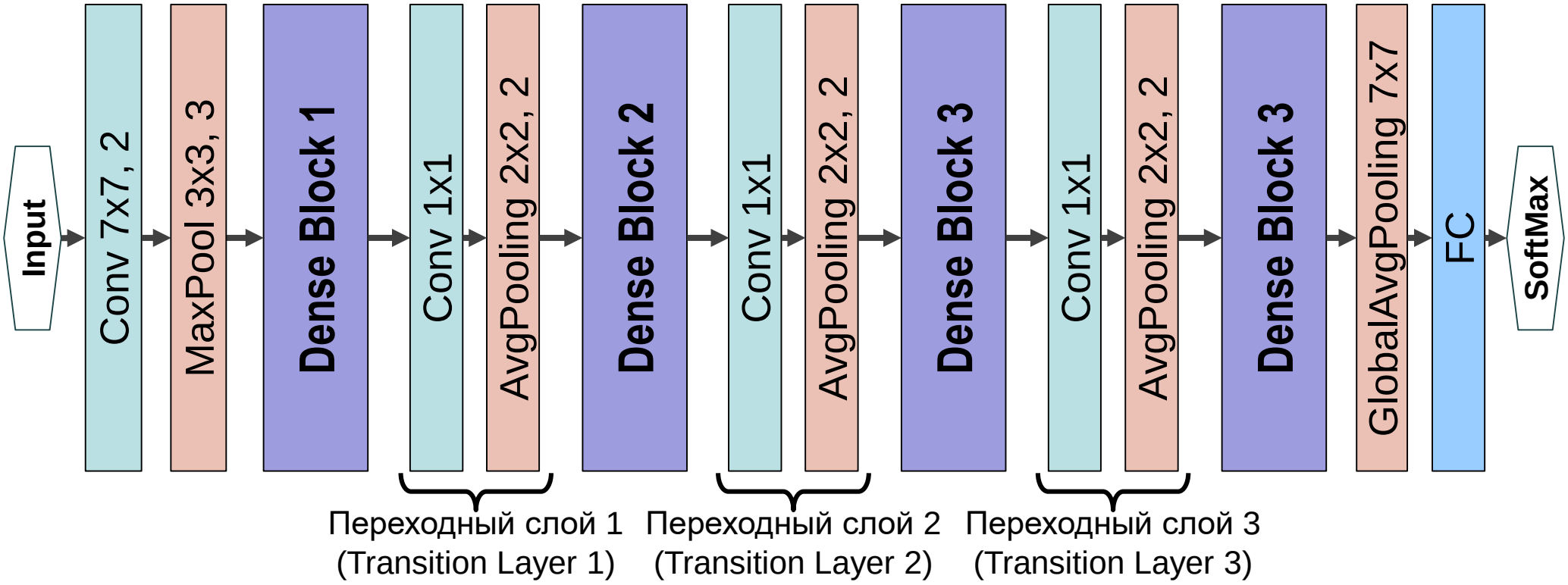
- DenseNet является развитием класса моделей ResNet, направленным на уменьшением количества параметров модели
- В DenseNet каждый слой получает информацию непосредственно от всех предшествующих слоев
- Модель реализуется посредством формирования последовательности ***«плотных» блоков*** (dense block)
  - Каждый блок содержит набор сверточных слоев
  - Вход каждого следующего слоя – конкатенация карт признаков, построенных на предыдущих слоях



* Huang G., Liu Z., Maaten L., Weinberger K.Q. Densely Connected Convolutional Networks. – 2016. – [https://arxiv.org/pdf/1608.06993.pdf].

DenseNet-121, 169, 201, 264 (2)

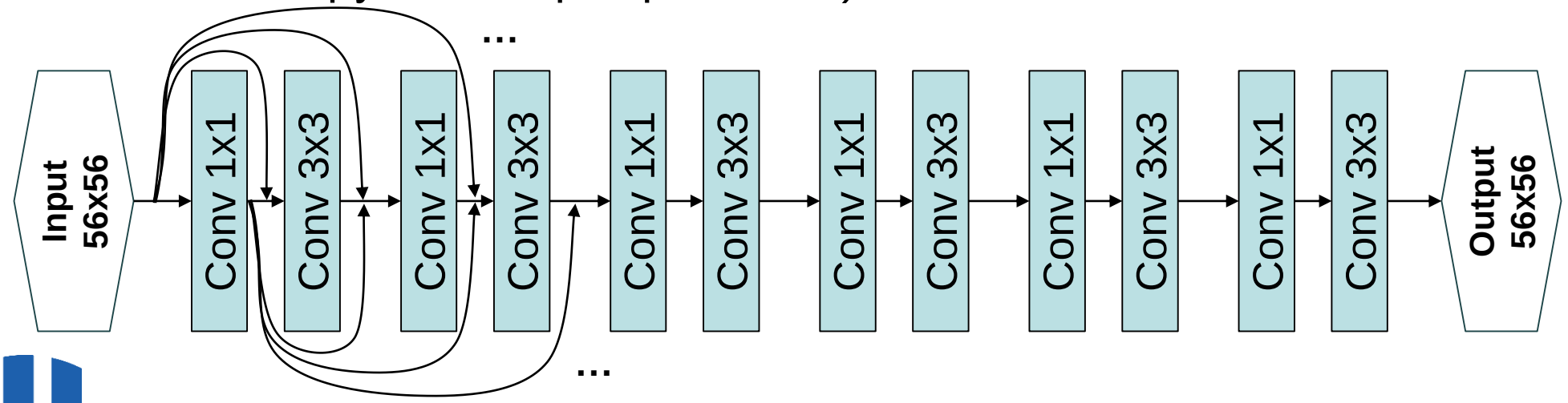
- Общая структура моделей DenseNet:



- Слои между двумя смежными блоками называются ***переходными слоями***, они изменяют размеры карты признаков

DenseNet-121, 169, 201, 264 (3)

- Структура «плотных» блоков сети DenseNet-121:
  - Dense Block 1: 6 x [Conv 1x1, Conv 3x3]
  - Dense Block 2: 12 x [Conv 1x1, Conv 3x3]
  - Dense Block 3: 24 x [Conv 1x1, Conv 3x3]
  - Dense Block 4: 16 x [Conv 1x1, Conv 3x3]
- Схема Dense Block 1 (на схеме дугами показано движение конкатенируемых карт признаков):



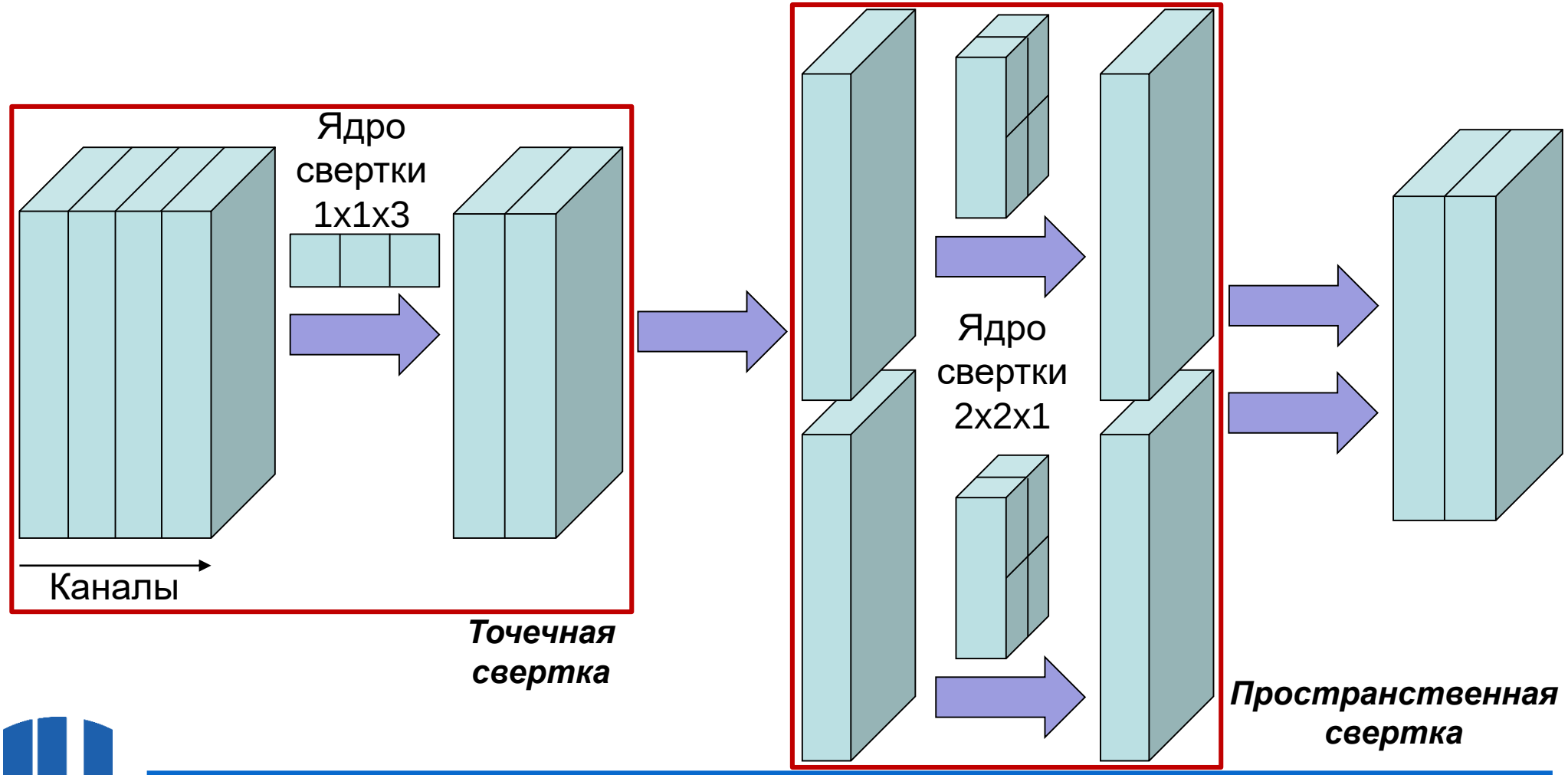
Xception (1)

- Xception (Extreme version of Inception) – модификация модели Inception-v3, в которой используются ***модифицированные отделимые по глубине свертки*** (modified depthwise separable convolution)
- Модифицированные отделимые по глубине свертки состоят из двух преобразований:
  - ***Точечная свертка*** (pointwise convolution) – свертка 1x1 по третьей размерности, соответствующей каналам в карте признаков
  - ***Пространственная свертка по отдельным каналам*** (depthwise convolution) – свертка NxN, применяемая к отдельным каналам карты признаков
- В классическом виде порядок преобразований обратный

* Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].

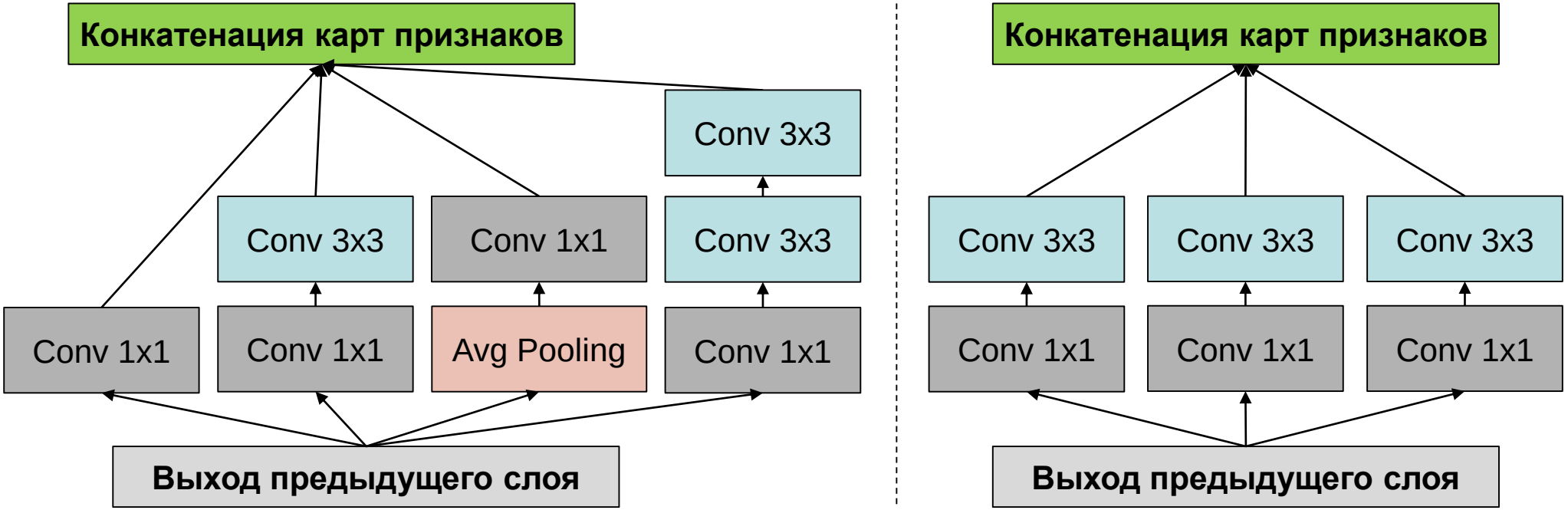
Xception (2)

- Пример модифицированной отделимой по глубине свертки:



Xception (3)

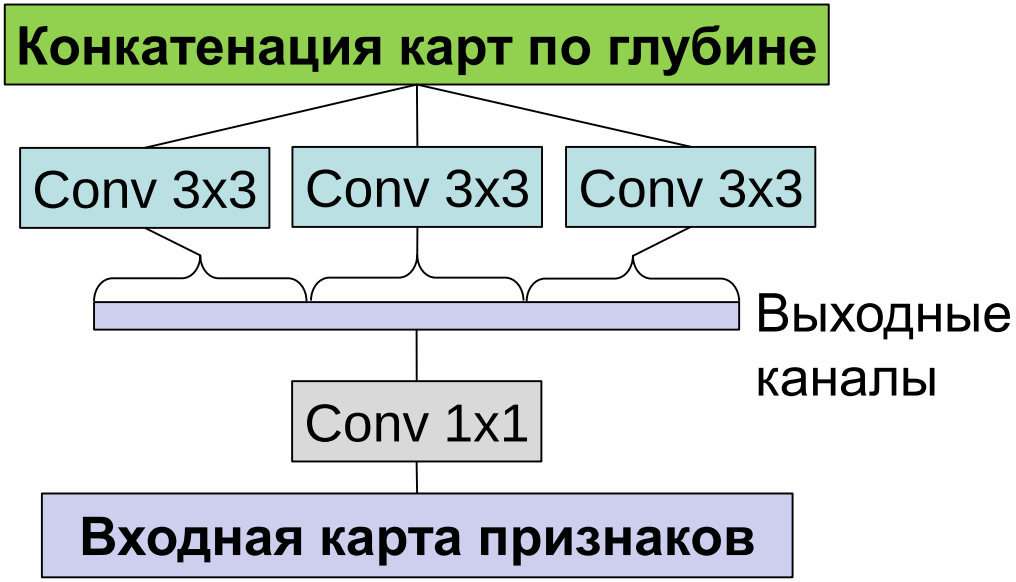
- ***Как использовать?*** Упрощенный inception-блок:
  - Используем только один вид размер ядра свертки 3x3
  - Удалим последовательность преобразований, содержащих пространственное объединение



* Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].

Xception (4)

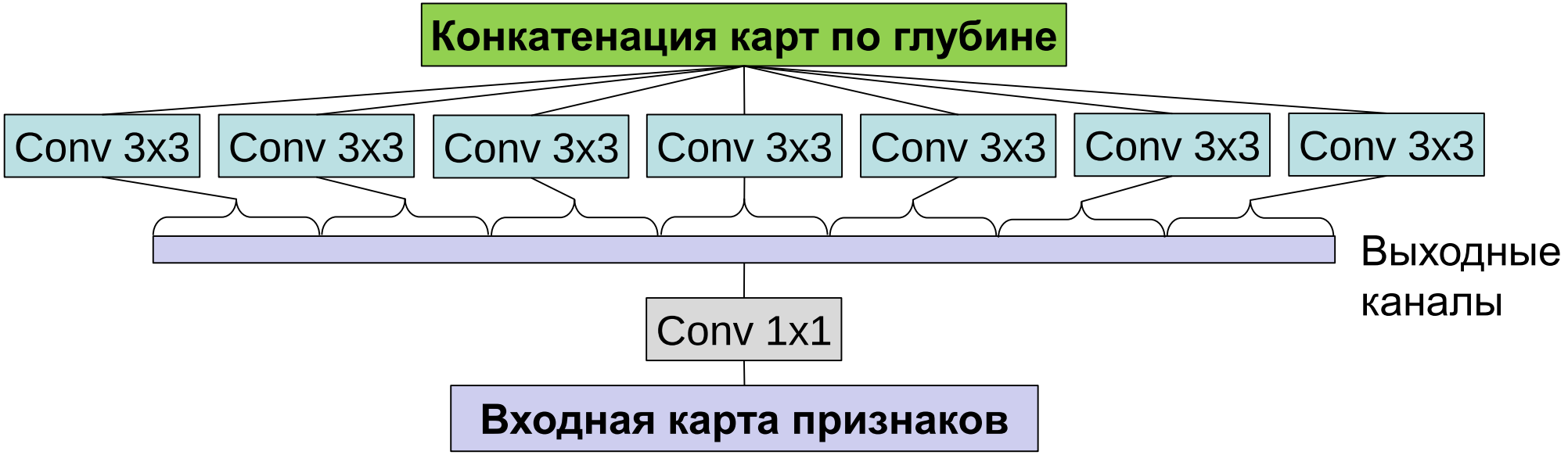
- Более сильная гипотеза – межканальные и пространственные корреляции могут быть отображены полностью раздельно
- Упрощенный модуль можно представить как свертку 1x1, за которой следуют пространственные свертки 3x3, применяемые к непересекающимся сегментам выходных каналов



* Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].

Xception (5)

- «Экстремальная» версия inception-блока основана на приведенной гипотезе
- Сначала используется свертка 1x1 для отображения межканальных корреляций, а затем отдельно отображаются пространственные корреляции каждого выходного канала



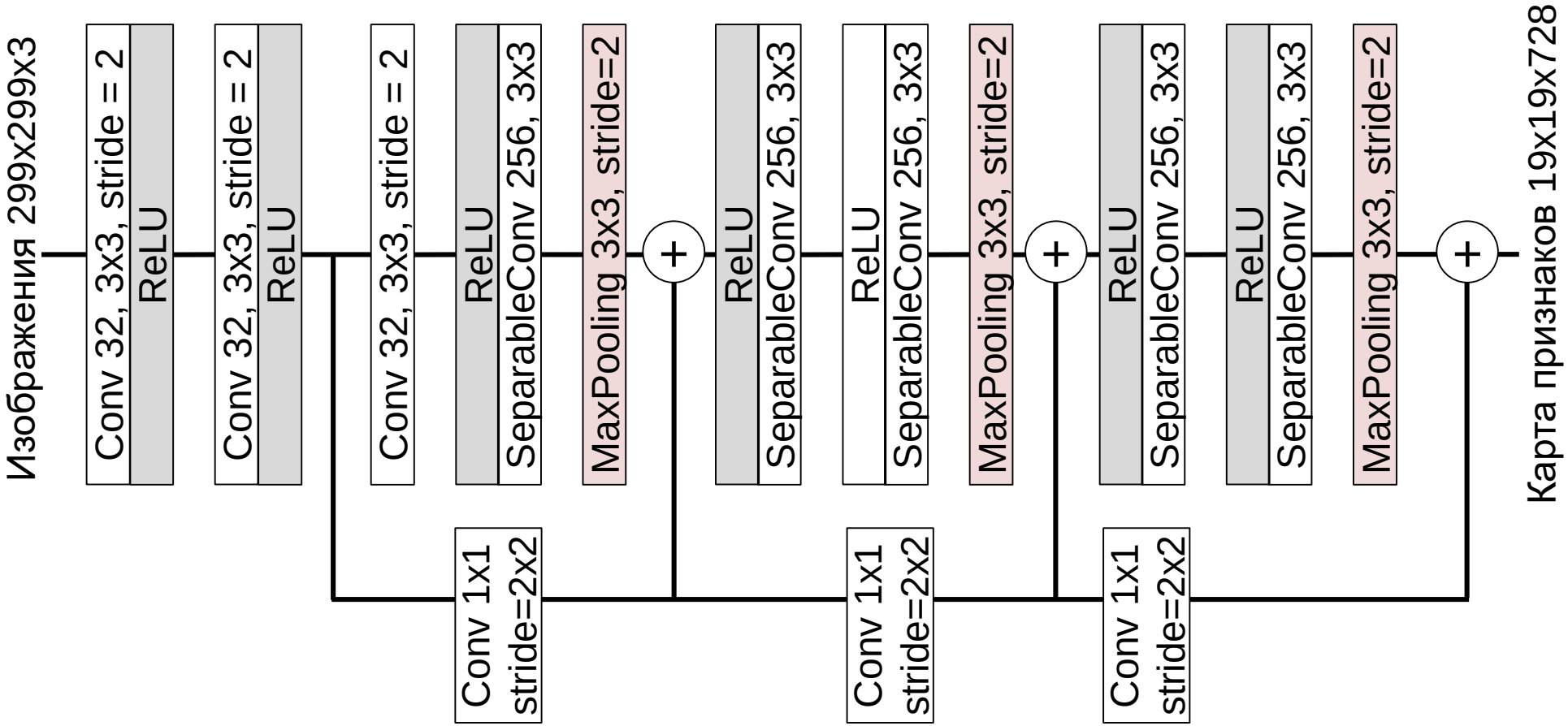
* Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].

Xception (6)

- «Экстремальная» версия inception-блока идентична рассмотренной модифицированной отделимой по глубине свертке
- Архитектура модели Xception построена с использованием «экстремальных» inception-блоков и связей, позволяющих использовать карты признаков с разных уровней детализации

Xception (7.1)

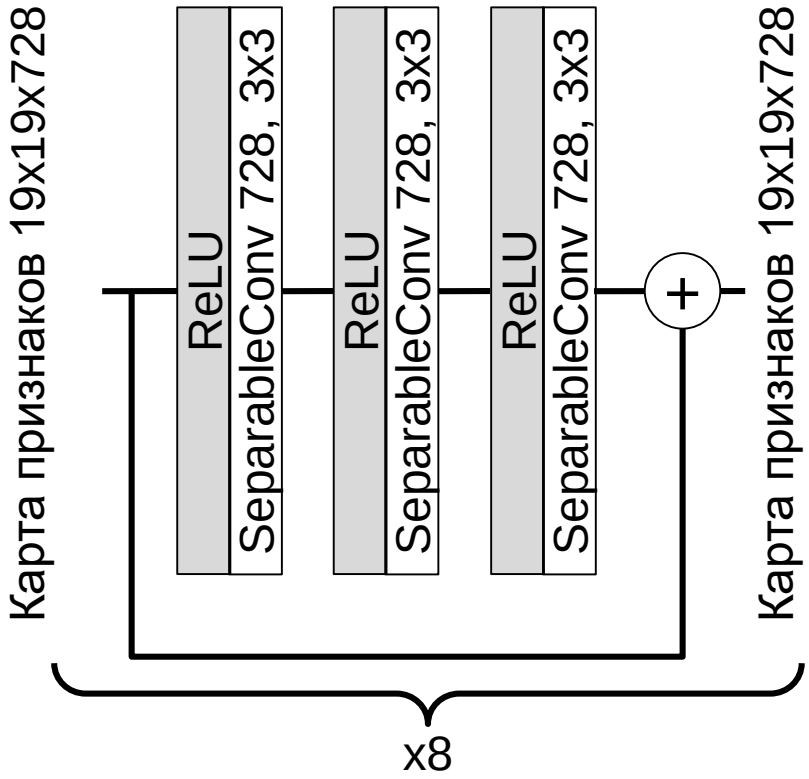
- Архитектура модели Xception (начальная часть, Entry Flow):



* Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].

Xception (7.2)

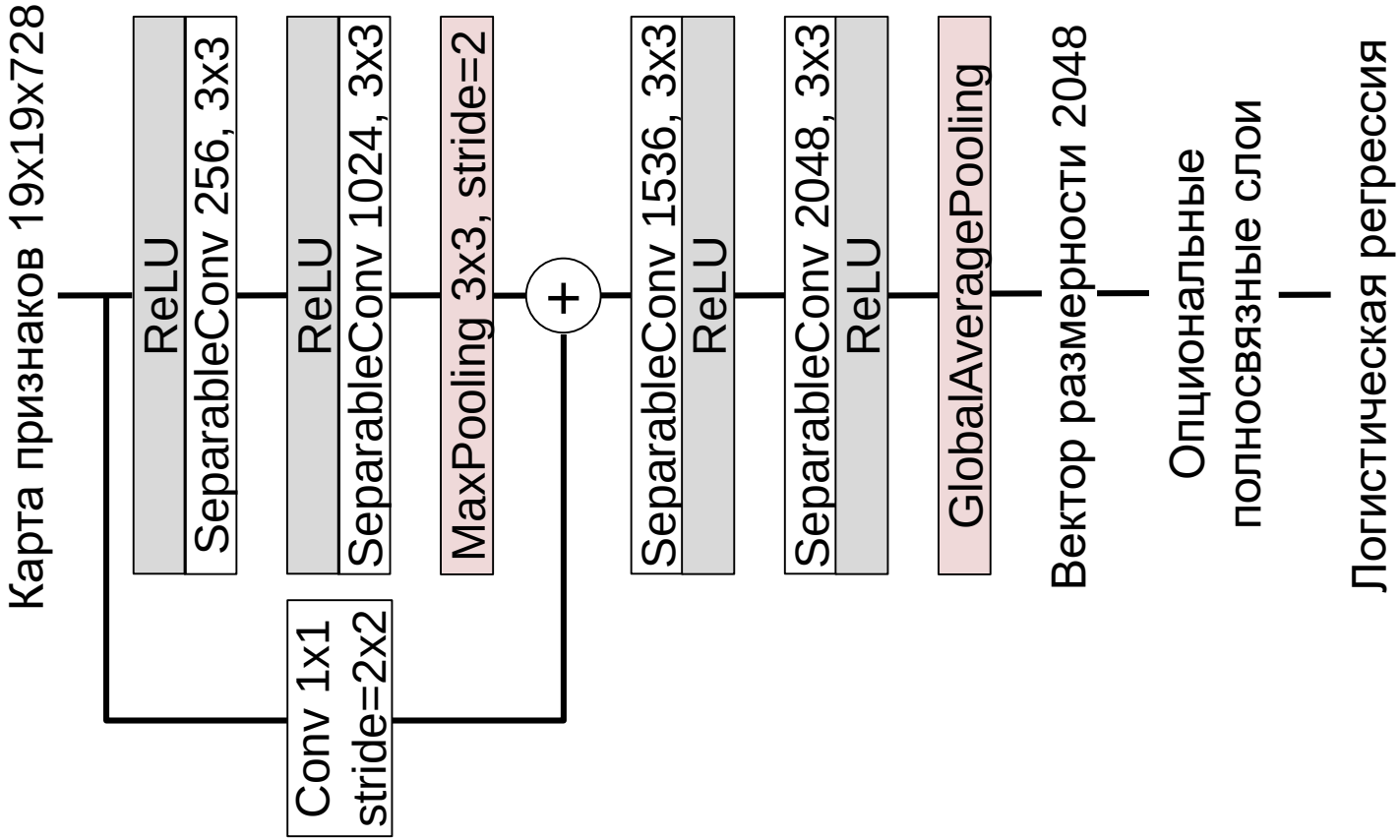
- Архитектура модели Xception (средняя часть, Middle Flow):



* Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].

Xception (7.3)

- Архитектура модели Xception (выходная часть, Exit Flow):



* Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].

MobileNet (1)

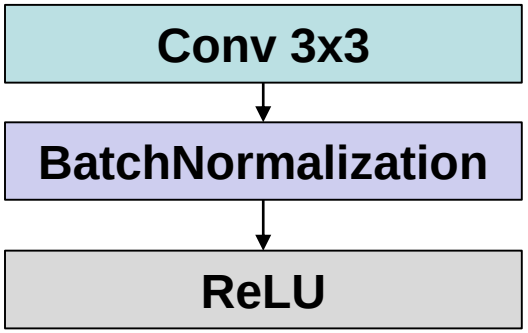
- MobileNets – семейство эффективных глубоких нейросетевых моделей для мобильных и встраиваемых устройств
  - Небольшой размер модели – меньшее число параметров
  - Небольшая вычислительная сложность – меньшее количество операций умножения и сложения
- Модели основаны на отделимых по глубине свертках с целью построения легковесных глубоких нейронных сетей
- Вводится алгоритм подбора модели правильного размера посредством поиска компромисса между сложностью и качеством работы модели

* Howard A.G., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. – 2017. – [https://arxiv.org/pdf/1704.04861.pdf].

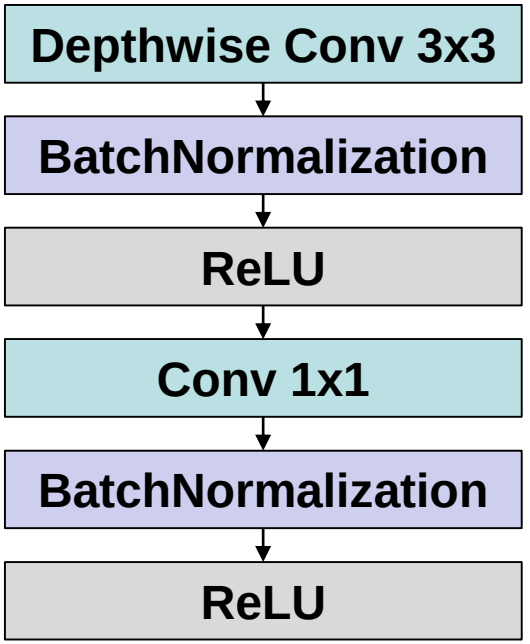
MobileNet (2)

- Структура базового сверточного блока в MobileNet:

Стандартный сверточный блок с нормализацией по пачке



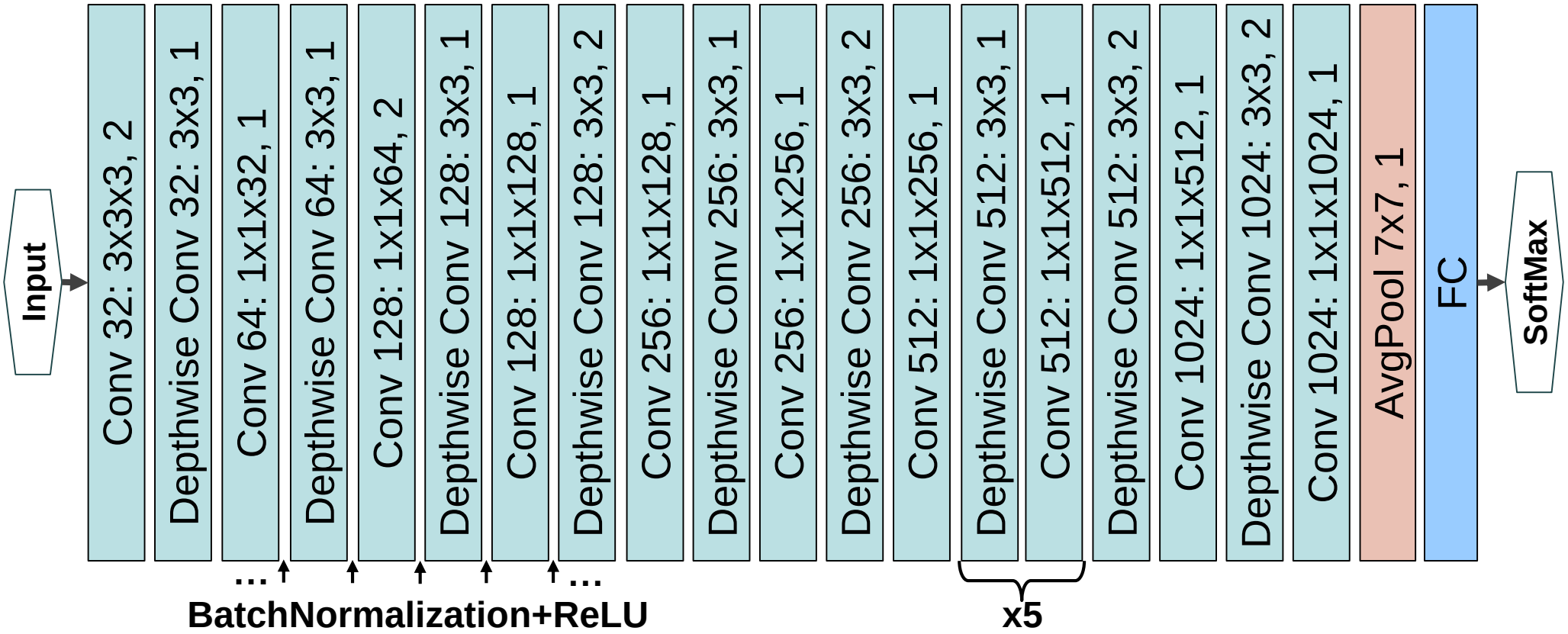
Сверточный блок, построенный на отделимой по глубине свертке



* Howard A.G., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. – 2017. – [https://arxiv.org/pdf/1704.04861.pdf].

MobileNet (3)

- Последовательность слоев модели MobileNet:



* Howard A.G., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. – 2017. – [https://arxiv.org/pdf/1704.04861.pdf].

MobileNet (4)

- Особенности модели MobileNet:
  - Содержит 28 слоев (считая точечные и пространственные свертки как отдельные слои)
  - Отсутствуют операции пространственного объединения после сверточных слоев
  - Для снижения размерности карт признаков используются свертки с шагом 2
  - После каждого сверточного слоя следует нормализация по пачке обрабатываемых изображений и функция активации «положительная срезка» (Rectified Linear Unit, ReLU)

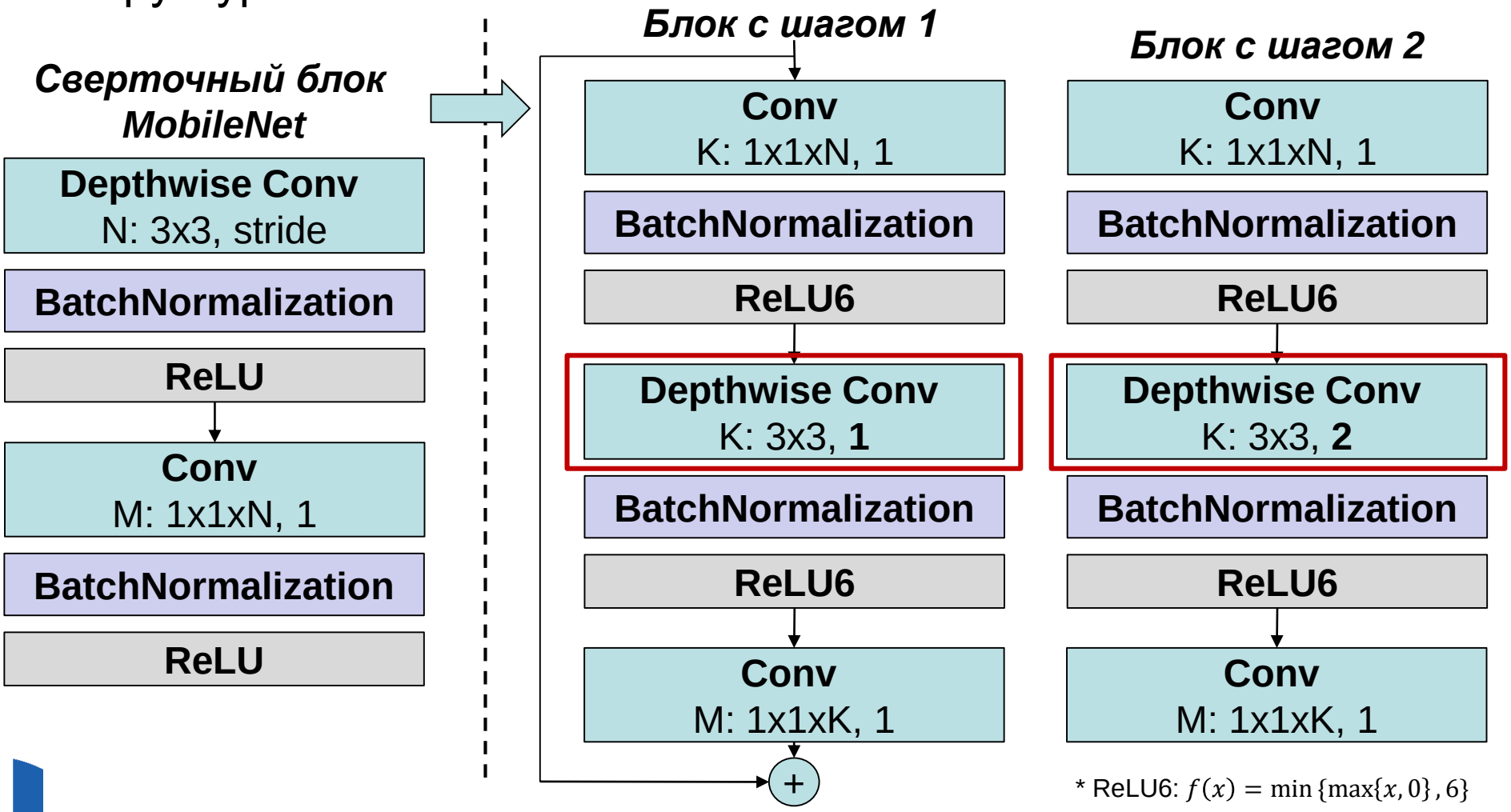
MobileNetV2 (1)

- MobileNetV2 – модификация модели MobileNet, в которой введен блок, имеющий ***инвертированную остаточную структуру*** (inverted residual structure)
- Вводятся два типа строительных блоков:
  - Блок с шагом 1 (Stride=1 block) – инвертированный остаточный блок (inverted residual block or bottleneck block)
  - Блок с шагом 2 (Stride=2 block) – последовательность сверточных слоев, снижающих размер карты признаков
- Каждый блок содержит 3 сверточных слоя:
  - Свертка с ядром 1x1 и функция активации ReLU
  - Пространственная свертка (depthwise convolution)
  - Свертка с ядром 1x1 (!без активации)

* Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. – 2018. – [https://arxiv.org/pdf/1801.04381.pdf].

MobileNetV2 (2)

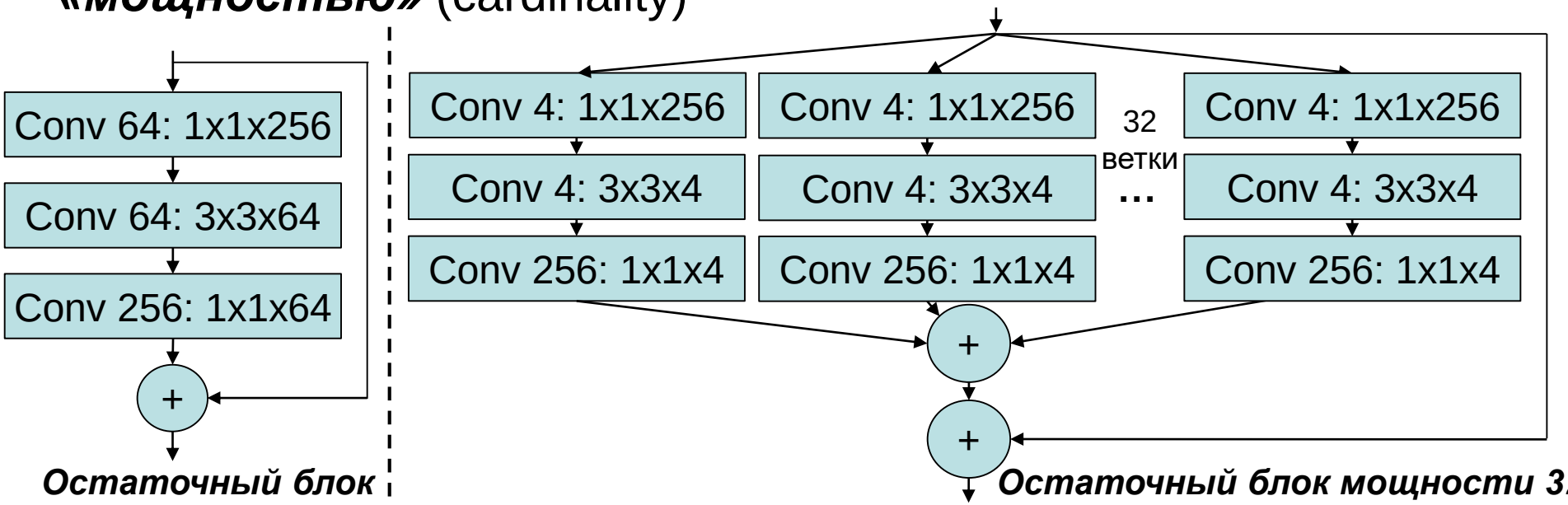
- Структура типовых блоков:



* ReLU6: $f(x) = \min\{\max\{x, 0\}, 6\}$

ResNeXT

- ResNeXT – глубокая сверточная сеть, состоящая из повторяющихся остаточных блоков, которые агрегируют набор преобразований с одинаковой топологией
- Количество веток с одинаковой топологией называется ***«мощностью»*** (cardinality)



* Xie S., Girshick R., Dollar P., Tu Z., He K. Aggregated Residual Transformations for Deep Neural Networks. – 2017. – [https://arxiv.org/pdf/1611.05431v2.pdf].

EfficientNet (1)

- EfficientNets – класс моделей, цель разработки которых сохранить высокое качество решения задачи и повысить эффективность модели (уменьшить количество параметров и снизить вычислительную сложность)
- При правильном конструировании моделей масштабирование по любому размеру сети (глубина, разрешение входного изображения, ширина – количество каналов в картах признаков) приводит к повышению качества решения задачи
- Важно сбалансировать все размеры сети (глубину, разрешение и ширину) во время масштабирования сети для получения высокой точности и эффективности

* Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. – 2019. – [https://arxiv.org/pdf/1905.11946.pdf].

EfficientNet (2)

- Авторы предлагают ***метод составного масштабирования*** модели (compound scaling method)
- Вводится составной коэффициент ϕ для равномерного масштабирования глубины, ширины и разрешения:
  - Глубина (depth): $d = \alpha^{\phi}$
  - Ширина (width): $\omega = \beta^{\phi}$
  - Разрешение (resolution): $r = \gamma^{\phi}$
- Соотношение параметров: $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2,\ \alpha \geq 1,\ \beta \geq 1,\ \gamma \geq 1$

$$FLOPS \cong O(d\omega^2 r^2)$$
$$FLOPS < 2^{\phi}$$

- Коэффициент ϕ – пользовательский параметр, значение которого определяется имеющимися вычислительными ресурсам
- α, β, γ определяют, каким образом ресурсы распределяются между глубиной, шириной и разрешением

EfficientNet (3)

- EfficientNet-B0 – базовая нейронная сеть, построенная с использованием инвертированных остаточных блоков, которые введены в MobileNetV2
- EfficientNet-B1,…,B7 получены в результате поиска оптимального соотношения параметров глубины, ширины и разрешения с использованием предложенного метода масштабирования

СРАВНЕНИЕ КАЧЕСТВА КЛАССИФИКАЦИИ И СЛОЖНОСТИ ГЛУБОКИХ МОДЕЛЕЙ

Тестовый набор данных

- Сравнение результатов качества классификации показано на тестовой выборке набора данных ImageNet
- Приведенные показатели собраны исследователями по результатам конкурса ILSVRC и опубликованы в Интернет [https://paperswithcode.com/sota/image-classification-on-imagenet]

Показатели качества

- Предположим, что N – количество категорий изображений
- Для каждого изображения $I_j, j = \overline{1,S}$ в выборке метод строит вектор достоверностей $p^j = \left(p_1^j, p_2^j, \dots, p_N^j\right)$, где p_i^j – достоверность того, что изображение I_j принадлежит классу i
- ***Точность top-K*** (top-K accuracy) определяется следующим образом:

$$topK = \frac{\sum_{j=1}^{S} 1_{\left\{i_1^j, i_2^j, \dots, i_K^j\right\}}\left(l_j\right)}{S}$$

где $\left\{i_1^j, i_2^j, \dots, i_K^j\right\} \subseteq \{1,2, \dots, N\}$, а $p_{i_1^j}^j, p_{i_2^j}^j, \dots, p_{i_K^j}^j$ – K наибольших достоверностей, l_j – класс, которому принадлежит изображение I_j согласно разметке, $1_{\left\{i_1^j, i_2^j, \dots, i_K^j\right\}}\left(l_j\right)$ – индикаторная функция

Сравнение качества классификации и сложности глубоких моделей (1)

Модель	Год	top-1,%	top-5,%	Количество параметров, млн.
AlexNet	2012	63.3	84.6	60
OverFeat	2013	66.04	86.76	–
VGG-16	2014	74.4	91.9	138
GoogLeNet	2014	69.8	89.9	5
ResNet-101	2015	78.25	93.95	40
Inception-v2	2015	74.8	92.2	11.2
Inception-v3	2015	78.8	94.4	23.8
DenseNet-201	2016	78.54	94.46	20
Xception	2016	79	94.5	22.8
MobileNet-224	2017	70.6	89.5	–
ResNeXT-101 64x4	2017	80.9	95.6	83.6
EfficientNet-B0	2019	76.3	93.2	**5.3**
EfficientNet-B7	2019	**84.4**	**97.1**	66

Рост качества и числа параметров

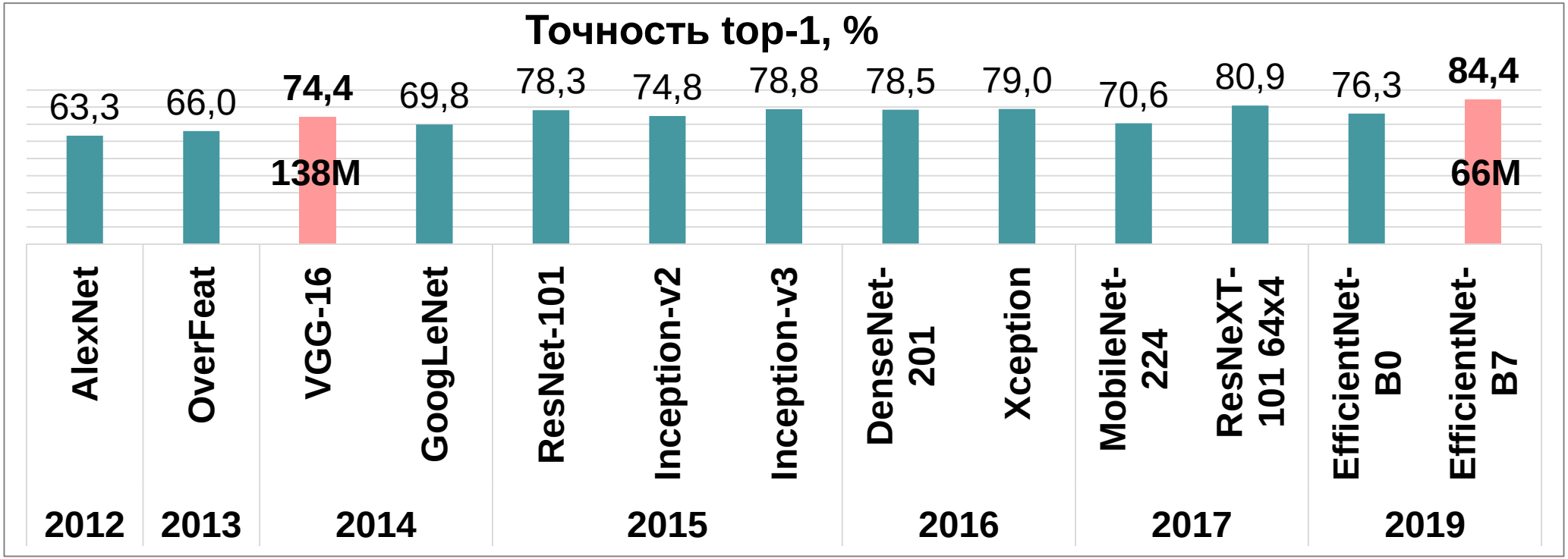
Рост качества и снижение числа параметров

Снижение сложности модели и поиск оптимальной модели

* Image Classification on ImageNet [https://paperswithcode.com/sota/image-classification-on-imagenet].

Сравнение качества классификации и сложности глубоких моделей (2)

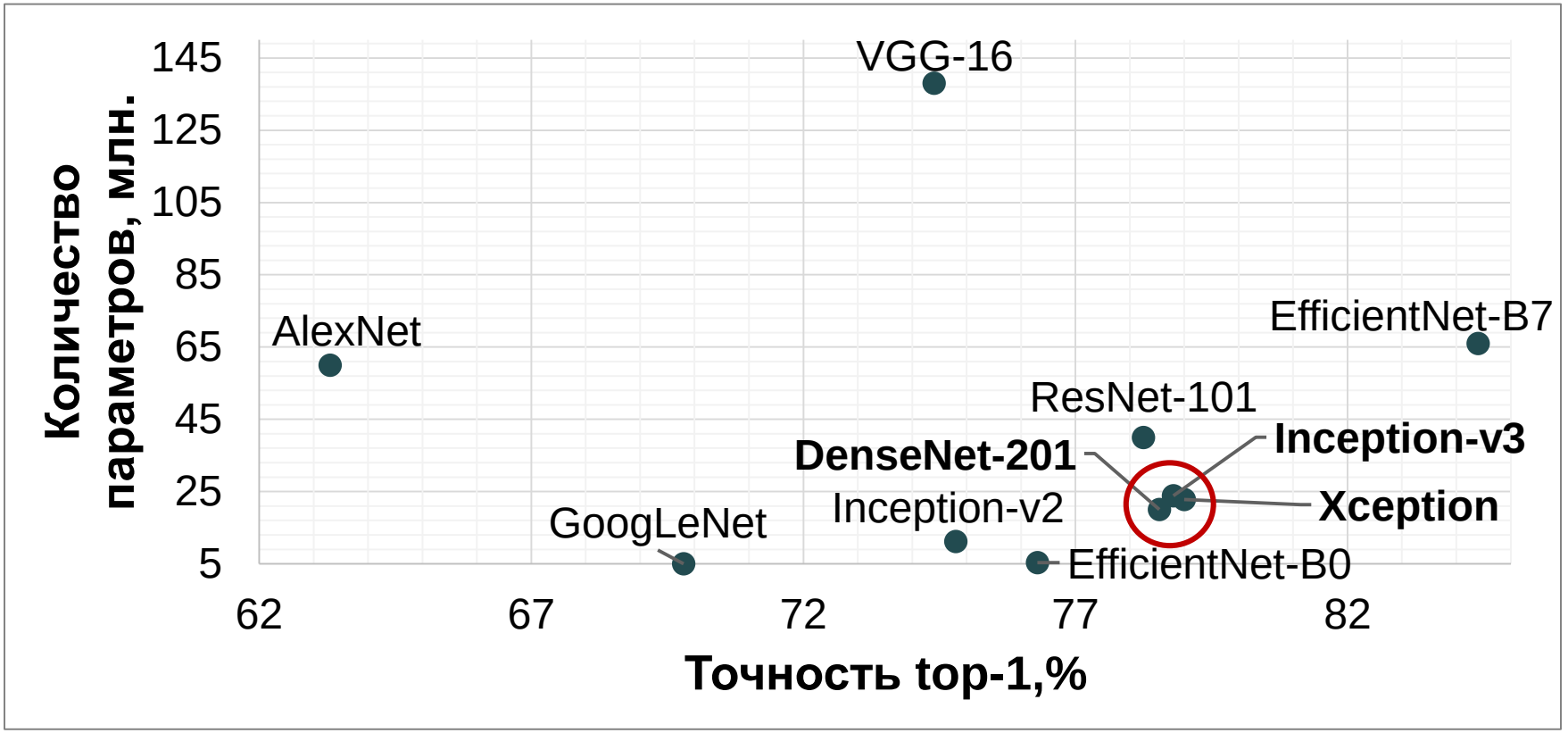
- Изменение точности top-1 на наборе данных ImageNet для избранных моделей:



- ***За 5 лет точность top-1 увеличилась на 10%, а количество параметров уменьшилось в ~2 раза***

Сравнение качества классификации и сложности глубоких моделей (3)

- До 2014 г. цель разработки моделей – повышение качества решения задачи, ***с 2015 г. – повышение эффективности модели и обеспечение роста качества (поиск компромисса)***



Сравнение качества классификации и сложности глубоких моделей (4)

- Примечания:
  - Повышение эффективности модели – снижение вычислительной сложности модели (количества выполняемых операций) и уменьшение размеров (количества параметров) модели
  - Сложность модели напрямую не связана с числом параметров
  - На практике, как правило важна сложность

Заключение

- Множество глубоких моделей для классификации изображений не ограничивается рассмотренными в настоящей лекции, существует множество модификаций базовых архитектур
- В настоящее время большое количество моделей для решения задач из других областей используют описанные архитектуры за счет применения переноса обучения, либо используют базовые строительные блоки рассмотренных моделей (далее в лекциях это будет показано)
- ***Оптимальная модель – компромисс между точностью и сложностью***
  - Точность определяется требованиями, предъявляемыми к решению практической задачи
  - Сложность определяется доступными вычислительными ресурсами и требованиями ко времени выполнения

Основная литература (1)

- Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks // Advances in neural information processing systems. – 2012. – [http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf].
- Sermanet P., Eigen D., Zhang X., Mathieu M., Fergus R., LeCun Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. – 2013. – [https://arxiv.org/pdf/1312.6229.pdf].
- Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. – 2014. – [https://arxiv.org/pdf/1409.1556.pdf].

Основная литература (2)

- Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going Deeper with Convolutions. – 2014. – [https://arxiv.org/pdf/1409.4842.pdf].
- He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. – 2015. – [https://arxiv.org/pdf/1512.03385.pdf].
- Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. – 2015. – [https://arxiv.org/pdf/1502.03167.pdf].
- Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture for Computer Vision. – 2015. – [https://arxiv.org/pdf/1512.00567.pdf], [https://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.pdf].

Основная литература (3)

- Huang G., Liu Z., Maaten L., Weinberger K.Q. Densely Connected Convolutional Networks. – 2016. – [https://arxiv.org/pdf/1608.06993.pdf].
- Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. – 2016. – [https://arxiv.org/pdf/1610.02357.pdf].
- Howard A.G., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. – 2017. – [https://arxiv.org/pdf/1704.04861.pdf].
- Xie S., Girshick R., Dollar P., Tu Z., He K. Aggregated Residual Transformations for Deep Neural Networks. – 2017. – [https://arxiv.org/pdf/1611.05431v2.pdf], [https://ieeexplore.ieee.org/document/8100117].

Основная литература (4)

- Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. – 2018. – [https://arxiv.org/pdf/1801.04381.pdf], [https://ieeexplore.ieee.org/document/8578572].
- Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. – 2019. – [https://arxiv.org/pdf/1905.11946.pdf].

Авторский коллектив

- **Турлапов Вадим Евгеньевич**
  д.т.н., профессор кафедры МОСТ ИИТММ ННГУ
  vadim.turlapov@itmm.unn.ru
- **Васильев Евгений Павлович**
  преподаватель кафедры МОСТ ИИТММ ННГУ
  evgeny.vasiliev@itmm.unn.ru
- **Гетманская Александра Александровна**
  преподаватель кафедры МОСТ ИИТММ ННГУ
  getmanskaya.alexandra@gmail.com
- **Кустикова Валентина Дмитриевна**
  к.т.н., доцент кафедры МОСТ ИИТММ ННГУ
  valentina.kustikova@itmm.unn.ru