

УДК 004.85

ИССЛЕДОВАНИЕ МОДЕЛЕЙ СВЕРТОЧНЫХ И ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧИ УЛУЧШЕНИЯ КАЧЕСТВА ИЗОБРАЖЕНИЙ

© Пахомов Е.В., Солдатова О.П.

*Самарский национальный исследовательский университет
имени академика С.П. Королева, г. Самара, Российская Федерация*

e-mail: evgen.pahomov01@mail.ru

В данной работе проводится детальный анализ современных моделей нейронных сетей для суперсемплирования изображений. Особое внимание уделяется сверточным и генеративно-состязательным нейронным сетям, которые в настоящее время являются одними из наиболее эффективных моделей в этой области.

Для улучшения качества работы модели во время обучения были применены методы аугментации, а также политики изменения скорости обучения. Аугментации вносят разнообразие в обучающие данные, благодаря чему улучшается обобщающая способность сети, а гибкое изменение скорости обучения позволяет более точно настраивать ее параметры.

Наиболее подробно будут рассмотрены сети SRGAN (Super-Resolution Generative Adversarial Network) и ESPCN (Efficient Sub-Pixel Convolutional Neural Network). SRGAN считается одной из лучших сетей на данный момент, показывая высокие метрики оценки качества восстановленных изображений [1]. ESPCN же благодаря простой архитектуре и использованию слоя перемешивания позволяет суперсемплировать изображения в режиме реального времени [2].

Архитектура сети SRGAN состоит из двух основных компонентов – генератора и дискриминатора. Генератор принимает на вход данные и пытается повторить распределение вероятностей набора обучающих данных. Дискриминатор в свою очередь принимает на вход реальные данные и данные, созданные генератором, а затем пытается отличить данные генератора от настоящих. Таким образом, генератор становится все более точным в создании реалистичных данных, а дискриминатор – все более точным в определении их подлинности. Когда генератор сумеет успешно «одурачить» дискриминатор, мы сможем трактовать наш генератор как математическую модель, симулирующую распределение реальных данных.

Функция ошибки, используемая в SRGAN, сложна и состоит из суммы нескольких компонентов. Кроме стандартных функций ошибок для генератора и дискриминатора (соответственно? MSE и бинарная кросс-энтропия), SRGAN также использует функцию ошибок восстановления контента (content loss) и функцию перцептивных ошибок (adversarial loss).

Архитектура сети ESPCN состоит из блоков свертки, блоков функции активации, а также блоков перемешивания. Сверточный блок выполняет извлечение признаков из изображения. Слой перемешивания увеличивает размерность выходного изображения и восстанавливает гладкость текстур. А функции активации добавляют в модель нелинейность. Главной особенностью модели ESPCN является ее вычислительная эффективность, так как субпиксельный слой в конце не увеличивает число обучаемых параметров сети.

При обучении используется оптимизатор Adam, являющийся комбинацией идей Nesterov Momentum и RMSProp [3]. Данный оптимизатор учитывает как первый, так и

второй моменты градиента. Первый момент представляет собой среднее значение градиента, а второй – среднее значение квадрата градиента. Для вычисления оценки первого момента градиента используется экспоненциальное сглаживание, которое позволяет учитывать только последние значения градиента. Это делает Adam более устойчивым к шумам в градиенте. Для вычисления второго момента также используется экспоненциальное сглаживание. Однако здесь учитываются квадраты последних значений градиента. Это позволяет методу учитывать информацию о разбросе градиента. У каждого из моментов есть гиперпараметр, называемый коэффициентом корректировки смещения. Данный коэффициент определяет вес, который следует присвоить первому и второму моменту соответственно. Таким образом, комбинируя методы масштабирования градиентов и адаптивного шага обучения, он позволяет быстрее и точнее находить оптимальные параметры нейронной сети.

Для обучения модели используется датасет DIV2K, содержащий около 1000 изображений в высоком разрешении, а также их аналоги, уменьшенные в 2, 3, 4 раза методом бикубической интерполяции. В ходе исследования 800 изображений было выделено на обучение, 100 – на валидацию и 100 – для тестирования уже обученной модели. Изображения в датасете были собраны из различных ресурсов, таких как альбомы фотографий, интернет-галереи и т. д. Обучение на данном датасете позволили получить модель, способную эффективно и качественно суперсемплировать изображения.

Для проверки эффективности и качества работы моделей использовался датасет Urban100. Этот датасет содержит 100 изображений в высоком разрешении на которых изображен естественный городской ландшафт. На каждом изображении присутствуют различные элементы городской инфраструктуры, такие как здания, дороги, мосты, реки и т. д. Для каждого изображения датасета была подсчитана метрика PSNR, а затем все результаты были усреднены. Средние значения метрики PSNR для сети SRGAN равны 25,66, для ESPCN – 23,12. При этом пиковое потребление оперативной памяти SRGAN равно 52,3 Мб, ESPCN – 22,1, а время затраченное на суперсемплирование одного изображения у SRGAN 2,2 секунды, ESPCN – 1,2 секунды.

В рамках работы были построены графики обучения, которые позволяют визуализировать процесс обучения и проследить за его динамикой. Кроме того, были продемонстрированы примеры суперсемплированных изображений, позволяющие наглядно оценить качество работы моделей на конкретных примерах и выявить ее достоинства и недостатки.

В ходе исследования были произведены измерения потребления памяти и вычислительных ресурсов машины, которые отражены в виде столбчатых диаграмм. Это позволило дать рекомендации по использованию моделей в программном обеспечении.

Библиографический список

1. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. URL: https://openaccess.thecvf.com/content_cvpr_2017/papers/Ledig_PhotoRealistic_Single_Image_CVPR_2017_paper.pdf (дата обращения: 29.05.2023).
2. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. URL: <https://arxiv.org/pdf/1609.05158.pdf> (дата обращения: 29.05.2023).
3. Adam: a method for stochastic optimization. URL: <https://arxiv.org/pdf/1412.6980.pdf> (дата обращения: 30.05.2023).