

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«МОСКОВСКИЙ ИНСТИТУТ ЭЛЕКТРОННОЙ ТЕХНИКИ»

Д Н Е В Н И К П Р А К Т И К И

Обучающийся Ильютченко Павел Сергеевич
(ФИО)

Форма обучения: очная

Кафедра ВМ-1 Группа ВМ-21
(сокращенное название)

Направление подготовки 01.04.04 Прикладная математика

Программа «Цифровая обработка сигналов и изображений»

Вид практики Производственная (Научно-исследовательская работа)

Место практики Кафедра ВМ-1
(организация, подразделение)

Начало «11» февраля 2019 г.

Окончание «9» апреля 2019 г.

Документы сопровождения практики в отдел практики и трудоустройства студентов
(ауд.4354) обучающимся предоставлены:

« _____ » _____ г. _____ (_____)
(подпись сотрудника ОПТС) (ФИО сотрудника ОПТС)

П А М Я Т К А

1. ОСНОВНЫЕ ПОЛОЖЕНИЯ ПО ПРОХОЖДЕНИЮ ПРАКТИКИ

1.1. До начала практики руководитель от кафедры, проводит инструктаж обучающихся и выдает индивидуальные задания по практике.

1.2. По прибытии на место практики обучающийся должен представить руководителю от организации Дневник практики и ознакомить его с содержанием индивидуальных заданий, пройти инструктаж по технике безопасности и противопожарной профилактике, ознакомиться с рабочим местом, правилами технической эксплуатации оборудования и уточнить план прохождения практики.

1.3. Обучающийся во время практики обязан строго соблюдать правила внутреннего распорядка организации. О временном отсутствии на своем месте необходимо ставить в известность руководителя практики от организации.

1.4. Зачет по практике приравнивается к зачетам по теоретическому обучению и учитывается при подведении итогов общей успеваемости обучающихся.

Результаты практики оцениваются путем проведения промежуточной аттестации с выставлением оценок: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

1.5. Обучающиеся, не прошедшие практику какого либо вида по уважительной причине, проходят практику по индивидуальному плану.

Обучающийся, не прошедший практику какого-либо вида по неуважительной причине или не получивший зачета по итогам ее прохождения, признается имеющим академическую задолженность.

2. ПРАВИЛА ВЕДЕНИЯ ДНЕВНИКА ПО ПРАКТИКЕ

Дневник по практике обучающихся имеет единую форму для всех видов практик и является основным документом для текущего и итогового контроля выполнения заданий.

Обучающемуся необходимо:

2.1. Заполнить титульный лист.

2.2. Получить индивидуальное задание на практику.

2.3. Регулярно вести Табель прохождения практики, записывая все, что сделано за весь день по выполнению индивидуального задания.

2.4. По завершению практики составить отчет в соответствии с индивидуальным заданием. Отчет о практике должен содержать сведения о конкретно выполненной работе, а также краткое описание деятельности, выводы и предложения.

2.5. Получить отзыв руководителя практики от организации.

2.6. В установленный срок, обучающийся должен предоставить на кафедру Дневник практики. При отсутствии правильно заполненного Дневника практика не засчитывается.

ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ НА НАУЧНО-ИССЛЕДОВАТЕЛЬСКУЮ РАБОТУ

Тема практики: Фотореалистичное суперразрешение изображения с помощью генеративно-сопоставительной сети

Задание:

1. Теоретическая разработка метода для решения задачи, поставленной для исследования, и его программная реализация.
2. Проведение экспериментов и тестов с целью определения характеристик и уточнения параметров разработанного метода и алгоритма.
3. Отладка и оптимизация разработанного ПО, оформление программного кода.
4. Анализ результатов применения разработанного метода.
5. Подготовка письменного отчета по проведенному исследованию.

Календарный график выполнения работ

Этапы выполнения технического задания	Форма предоставления результатов	Даты отчетов
Задача 1	Устная	5 неделя
Задача 2	Устная	7 неделя
Задачи 3, 4, 5	Отчет	9 неделя

Руководитель практики
от кафедры

(подпись)

(Ф.И.О.)

Студент

(подпись)

(Ф.И.О.)

Требования к содержанию и оформлению отчета:

К окончанию практики студент должен подготовить журнал практики, включающий:

- индивидуальное задание на практику, утвержденное руководителем практики;
- табель-календарь прохождения практики, завизированный руководителем

практики.

- отчет прохождения практики с оценкой руководителя практики.

Отчет по учебной практике включает в себя:

- титульный лист;
- введение с перечислением задач, которые решались в рамках выполнения

индивидуального задания по практике;

- основную часть с описанием результатов выполнения индивидуального задания (по части 1 приводится краткое описание предлагаемого метода решения задачи; по части 2 – описание результатов проведенных экспериментов; по части 3 – обзор программных средств, использованных при проведении исследований заданной направленности, по части 4 – выводы по результатам анализа полученных результатов);

- заключение, содержащее описание возможных направлений дальнейшей работы;
- оглавление;
- список использованных источников.

Примерный объем отчета по научно-исследовательской работе – 12-18 страниц.

Шаблон титульного листа - см. ниже

Т А Б Е Л Ь
п р о х о ж д е н и я п р а к т и к и

Дата	Содержание работы	Подпись руководителя практики от организации
28.02	Теоретическая разработка метода для решения задачи, поставленной для исследования, и его программная реализация	
8.03	Проведение экспериментов и тестов с целью определения характеристик и уточнения параметров разработанного метода и алгоритма	
25.03	Отладка и оптимизация разработанного ПО, оформление программного кода. Анализ результатов применения разработанного метода	
8.04	Подготовка письменного отчета по проведенному исследованию	

МИНОБРНАУКИ РОССИИ
Федеральное государственное автономное образовательное учреждение высшего
образования
«Национальный исследовательский университет
«Московский институт электронной техники»

ОТЧЕТ ПО НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

Направление подготовки — 01.04.04 «Прикладная математика»

Программа — «Цифровая обработка сигналов и изображений»

Выполнил студент Ильютченко Павел Сергеевич
Оценка руководителя практики от кафедры ВМ-1
Козлитин Иван Алексеевич

Группа: ВМ-21

(оценка)

(подпись)

Москва

2019

1. Введение

Очень сложная задача получения изображения с высоким разрешением (HR) из его дубликата с низким разрешением (LR), это называется суперразрешением (SR). SR получил значительное внимание со стороны сообщества исследователей компьютерного зрения и имеет широкий спектр применений.

Неправильно спозиционированная природа недоопределенной проблемы SR особенно выражена для больших коэффициентов масштабирования, для которых детали текстуры в реконструированных SR изображениях обычно отсутствуют. Цель оптимизации SR алгоритмов с учителем обычно сводится к минимизации среднеквадратичной ошибки между восстановленным HR изображением и истинным HR изображением. Это удобно, так как минимизация MSE также максимизирует пиковое отношение сигнал-шум (PSNR), которое является стандартной мерой для оценки и сравнения алгоритмов SR. Однако способность MSE (и PSNR) к улавливанию воспринимаемых различий, как высокодетализированные текстуры, очень ограничена, так как они определены на попиксельной разнице. Это показано на Рисунке 1, где самый высокий PSNR не обязательно отображает лучший результат SR по восприятию. Разница восприятия между SR изображением и оригинальным показывает, что восстановленное изображение не является фотореалистичным, как определено в работе Ferwerda (http://cin.ufpe.br/~in1123/material/vor_hvei03_v20.pdf Фотореалистичное изображение производит тот же визуальный эффект как и реальная сцена).

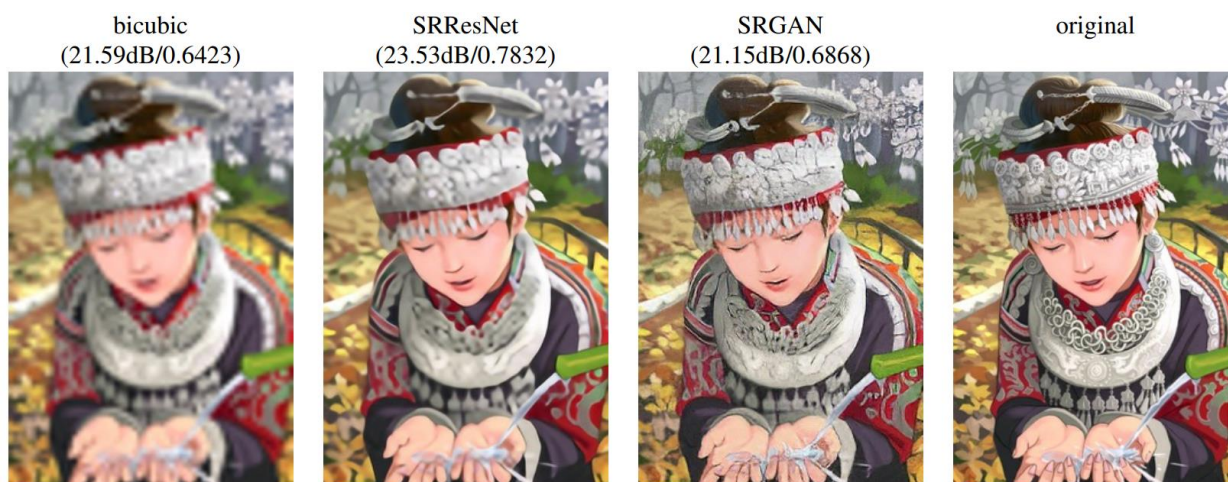


Рисунок 1. Слева направо: бикубическая интерполяция, глубокая остаточная сеть, оптимизированная с МНК, глубокая остаточная генеративно-состязательная сеть,

оптимизированная для потери более чувствительной с человеческому восприятию, оригинальное HR изображение. Соответствующие PSNR и SSIM показаны в скобках. [4x масштабирование]

В этой работе мы предлагаем генеративно состязательную сеть суперразрешения (SRGAN), в которой мы используем глубокую свёрточную сеть с residual связями (ResNet), с пропущенными соединениями и с целью оптимизации отличающейся от MSE. В отличие от предыдущих работ, мы определяем новую потерю восприятия, использующую карты особенностей высокого уровня сети VGG в сочетании с дискриминатором, который поощряет результаты, которые трудно отличить от эталонных изображений HR. Пример фотореалистичного изображения, которое было получено масштабированием в 4 раза показано на Рисунке 2.

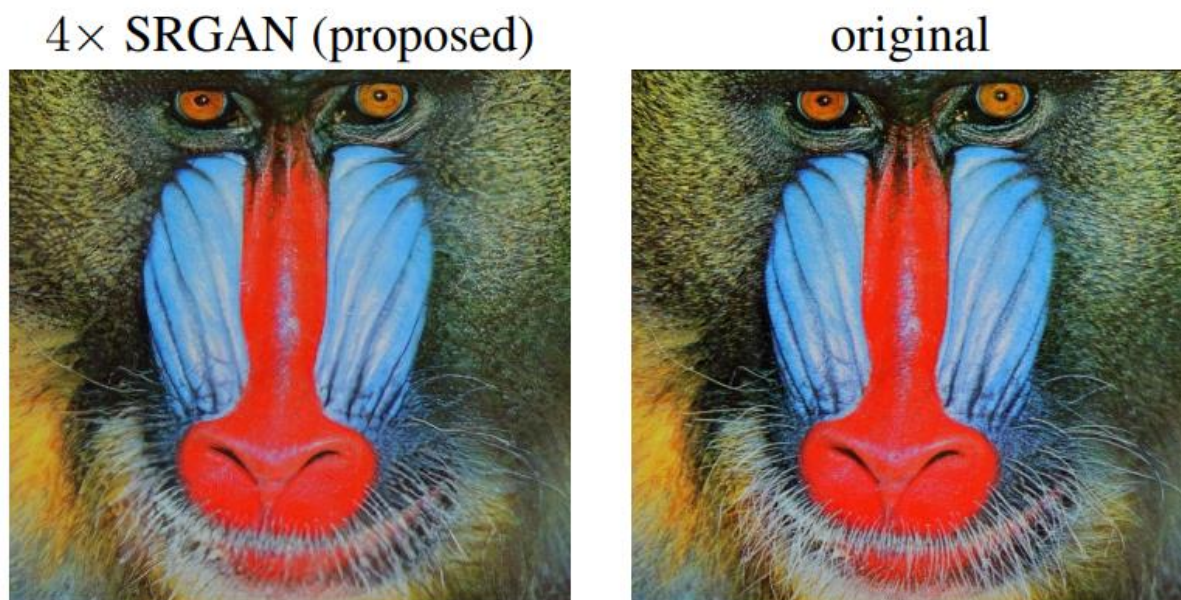


Рисунок 2. Суперразрешение изображения (слева) почти неотличимо от оригинального (справа). [4x масштабирование]

2. Метод

Сконцентрируемся на суперразрешении одного изображения (SISR = Single Image Super Resolution) и не будем затрагивать темы, покрывающие HR изображения по нескольким изображениям. В SISR целью является оценка высокого разрешения, суперразрешимое изображение I^{SR} из изображения низкого разрешения I^{LR} . Здесь I^{LR} дубликат с низким разрешением версии I^{HR} . Изображения высокого разрешения

доступны только во время обучения. В течение обучения, I^{LR} получаются применением фильтра Гаусса перед последующим применением операции понижения дискретизации с коэффициентом r . Для изображения с C цветовыми каналами, мы опишем I^{LR} как вещественный тензор размером $W * H * C$ и I^{HR} , I^{SR} как $rW * rH * C$ соответственно.

Наша конечная цель - это обучить генерирующую функцию G , которая создает по полученному LR входному изображению его соответствующий HR аналог. Для достижения этого, мы обучаем генеративную сеть как прямонаправленную свёрточную нейронную сети G_{θ_G} , зависящую от параметров θ_G . Здесь $\theta_G = \{W_{1:L}, b_{1:L}\}$ обозначает весовые коэффициенты и смещения сети L-го уровня глубокой сети и получаются они путём оптимизации специальной для SR функции потерь l^{SR} . С тренировочными изображениями $I_n^{HR}, n = 1, \dots, N$ с соответствующими $I_n^{LR}, n = 1, \dots, N$, мы решаем задачу оптимизации:

$$\hat{\theta}_G = \arg \min_{\theta_G} \frac{1}{N} \sum_{n=1}^N l^{SR}(G_{\theta_G}(I_n^{LR}), I_n^{HR}) \quad (1)$$

В этой работе мы специально спроектируем функцию потери восприятия l^{SR} как взвешенную комбинацию из нескольких компонент, которые моделируют различные желаемые характеристики восстановленного SR изображения.

2.1. Архитектура состязательной сети

Следуя определению Goodfellow мы также определим сеть дискриминатор D_{θ_D} , которую мы оптимизируем поочередно наряду с G_{θ_G} для решения состязательной мин-макс проблемы:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})} [\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))] \quad (2)$$

Основная идея, заключенная в этой формуле - это позволить тренироваться генеративной модели G с основной целью обмануть дискриминатор D, который тренируется распознавать суперразрешенные изображения от реальных изображений. С этим подходом наш генератор может научиться создавать решения, которые очень похожи на реальные изображения и поэтому их сложно классифицировать дискриминатору D. Это поощряет решения, которые лучше воспринимаются визуально (перцептивно), находящиеся в подпространстве, многообразии естественных изображений. Это контрастирует с SR решениями, прибегающими к минимизации измерений попиксельной ошибки, как MSE.

В центре нашей очень глубокой генеративной сети G, которая отображена на Рисунке 3, в которой B количество residual блоков с идентичным строением. В особенности, мы используем два свёрточных слоя с маленьким ядром 3x3 и с 64-мя картами особенностей, отправляющиеся затем в batch-normalization слои и ParametricReLU, как функцию активации. Мы увеличиваем разрешение входного изображения с помощью двух обученных субпиксельных свёрточных слоёв, предложенных Shi et al. (<https://arxiv.org/pdf/1609.05158.pdf>)

Чтобы определить настоящие HR изображения от сгенерированных SR образцов, мы обучаем сеть дискриминатор. Архитектура показана на Рисунке 3. Мы следуем архитектурным рекомендациям резюмированным Radford et al. (<https://arxiv.org/pdf/1511.06434.pdf>), и используем LeakyReLU активацию ($\alpha = 0.2$), и избегаем максимального пулинга во всей сети. Сеть дискриминатор обучена решать задачу максимизации в Выражении 2. Она содержит 8 свёрточных слоёв с увеличивающимся размером ядер от 3x3 с фактором 2, от 64 до 512 фильтров в ядре, как в сети VGG. В свёртке используются strides для уменьшения разрешения изображения, каждый раз количество особенностей удваивается. В результате 512 карт особенностей проходят через два полносвязных слоя и в конце сигмоидная функция активации для получения вероятностей классификации.

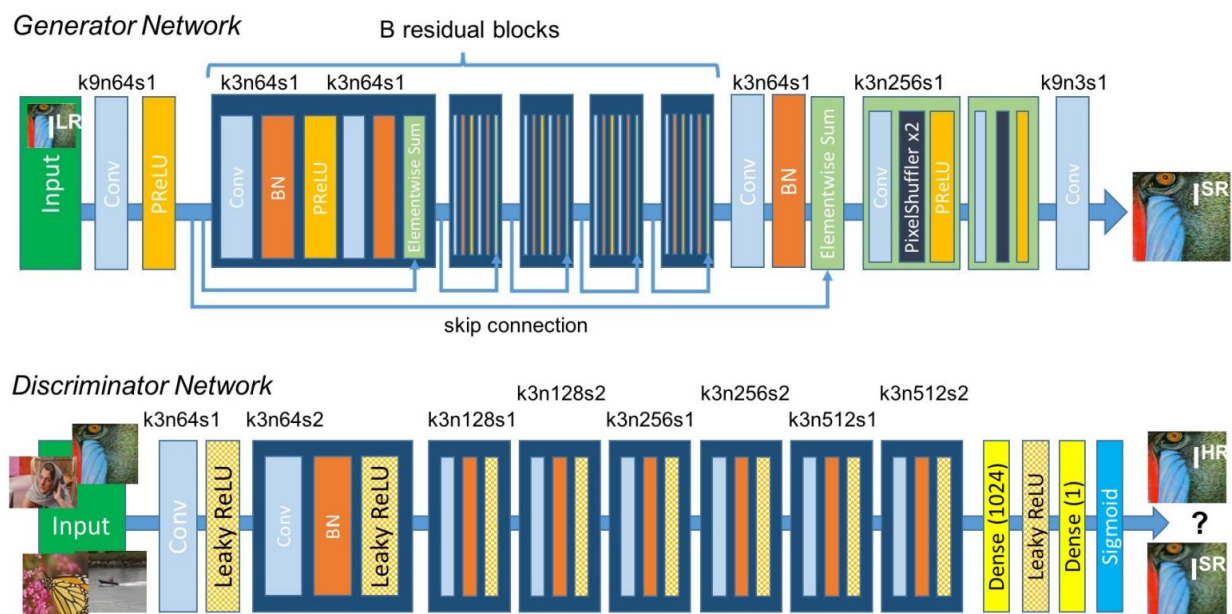


Рисунок 3. Архитектура Генерирующей и Дискриминирующей сети с соответствующими размерами ядер (k), количества карт особенностей (n) и шагами (s), которые показаны для каждого свёрточного слоя.

2.2 Функция потери восприятия

Предназначение нашей функции потери восприятия l^{SR} имеет решающее значение для производительности нашей генерирующей сети. Пока l^{SR} обычно смоделировано на основе MSE, мы улучшаем ее с помощью предложений из Johnson et al. (<https://arxiv.org/pdf/1603.08155.pdf>) и Bruna et al. (<https://arxiv.org/pdf/1511.05666.pdf>) и конструируем функцию потерь, которая оценивает решение с точки зрения воспринимательных характеристик. Мы сформулируем потери восприятия, как взвешенная сумма потери содержания l_X^{SR} и потери состязательной компоненты l_{Gen}^{SR} как:

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}} \quad (3)$$

perceptual loss (for VGG based content losses)

Далее мы опишем возможные реализации для потери содержания и состязательной потери.

2.2.1 Потеря содержания

Попиксельная потеря MSE записывается как:

$$l_{MSE}^{SR} = \frac{1}{r^2 W H} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2 \quad (4)$$

Это наиболее часто используемая задача оптимизации для SR изображения во многих современных подходах. Однако, при достижении особенно высокого PSNR, решениям на основе MSE оптимизации часто не хватает высокочастотных компонент, и это приводит к неудовлетворительному решению с точки зрения восприятия, оно содержит чрезмерно гладкие текстуры (см. Рисунок 1).

За место того чтобы использовать на попиксельные потери мы будем опираться на идеи Gatys et al. (Style-transfer), Bruna et al. и Johnson et al. и использовать функцию потерь, которая ближе к сходству восприятия. Мы определим VGG потерю, основанную на слоях активации ReLU предобученной 19-ти слойной VGG модели описанной Simonyan и Zisserman. Под $\phi_{i,j}$ мы будем полагать карту особенностей, полученную на j – ом слое свёртки (после активации) перед i – ым слоем максимального пулинга в сети VGG19, которые мы считаем данными. Определим VGG потерю как евклидово расстояние между представлениями особенностей восстановленного изображения $G_{\theta_G}(I^{LR})$ и опорного изображения I^{HR} :

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2 \quad (5)$$

Здесь $W_{i,j}$ и $H_{i,j}$ описывают размерности соответствующих карт особенностей в сети VGG.

2.2.2 Состязательная потеря

В добавок к потерям содержания, описанным выше, мы также добавляем генеративную компоненту в наш GAN для потери восприятия. Это побуждает нашу сеть отдавать предпочтение решениям, которые лежат во множестве естественных изображений, чтобы обмануть дискриминирующую сеть. Генеративные потери l_{Gen}^{SR} определены на вероятностях дискриминатора $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ по всем обучающим образцам:

$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (6)$$

Здесь $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ это вероятность, что восстановленное изображение $G_{\theta_G}(I^{LR})$ это естественное HR изображение. Для лучшего градиентного подхода мы минимизируем $-\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$, вместо $\log[1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$.

3. Эксперименты

3.1. Данные и измерения сходства

Мы производили эксперименты на трех широко используемых эталонных наборах данных Set5, Set14 и BSD100, набор тестирования BSD300. Все эксперименты производились с коэффициентом масштабирования 4x между низко- и высоко-разрешенными изображениями. Это соответствует 16x сокращению количества пикселей на изображении. Для честного сравнения, все сообщенные PSNR и SSIM измерения были высчитаны на у-канале центральной области, удалены полосы шириной 4 пикселя с каждой границы изображений, используя daala пакет (<https://github.com/xiph/daala> (commit: 8d03668)). Суперразрешенные изображения для сравнения методов включали алгоритмы ближайшего соседа, бикубический, SRCNN и SelfExSR, были взяты из онлайн материалов, дополняющих Huang et al. (<https://github.com/jbhuang0604/SelfExSR>) и для DRCN от Kim et al.

(<https://cv.snu.ac.kr/research/DRCN/>). Результаты, полученные с помощью SRResNet (для потерь: l_{MSE}^{SR} и $l_{VGG/2.2}^{SR}$) и варианты SRGAN доступны онлайн (<https://twitter.box.com/s/lcue6vlrd01ljkdttdkhmfvk7vtjhetog>). Статические тесты были произведены в виде парных двусторонних тестов Вилкоксона со знаком ранга, и значимость определялась при $p < 0.05$.

Также может заинтересовать независимая разработка GAN-основанных решений на GitHub (<https://github.com/david-gpu/srez>). Однако это только экспериментальные результаты на ограниченном наборе лиц, что является более ограниченной и легкой задачей.

3.2. Детали обучения и параметры

Мы обучали все сети на NVIDIA Tesla M40 GPU, используя случайные образцы из 350 тысяч изображений из ImageNet датасета. Эти изображения отличаются от тестовых изображений. Мы получили LR изображения путем уменьшения разрешения HR изображений (BGR, $C = 3$), используя бикубическое ядро с масштабом $r = 4$. Для каждого минибатча мы вырезаем 16 случайных 96×96 HR подизображений из различных тренировочных изображений. Заметим, что мы можем применять генерирующую модель к изображениям любого размера, так как она полностью свёрточная. Мы отобразили пространство LR выходных изображений к отрезку $[0, 1]$ и для HR изображений к $[-1, 1]$. Потеря MSE были рассчитана на изображениях с яркостью в диапазоне $[-1, 1]$. Карты особенностей VGG были также масштабированы с коэффициентом $1/12.75$, чтобы получить VGG потери сравнимые по масштабу с MSE потерями. Это эквивалентно умножению Выражению 5 с масштабирующим коэффициентом ≈ 0.006 . Для оптимизации используется Adam с $\beta = 0.9$. SRResNet сети были обучены с шагом обучения 10^{-4} и проведено 10^6 итераций. Мы используем обученную, основанную на MSE, SRResNet сеть как первоначальный вариант для генератора, при обучении действительного GAN, чтобы избежать нежелательных локальных минимумов. Все варианты SRGAN были обучены в 10^5 итераций с шагом обучения 10^{-4} и другие в 10^5 итераций с меньшим шагом 10^{-5} . Мы чередуем обновление сети генератора и

дискриминатора, что эквивалентно $k = 1$, как используется в алгоритме Goodfellow et al. Наша сеть генератор имеет 16 идентичных ($B = 16$) residual блоков. В течение времени тестов мы отключаем обновление Batch Normalization, чтобы получить выход, который детерминированно зависит только от входа.

3.3. Тестирование средней оценки мнения (Mean opinion score = MOS)

Мы выполнили MOS тест для количественной оценки способности различных подходов восстанавливать визуальное восприятие изображений. В частности, мы опросили 26 оценщиков присвоить интегральный балл от 1 (плохое качество) до 5 (отличное качество) для суперразрешённых изображений. Оценщики оценили 12 версий каждого изображения на датасетах Set5, Set14 и BSD100: метод ближайшего соседа, букубическая интерполяция, SRCNN, SelfExSR, DRCN, ESPCN, SRResNet-MSE, SRResNet-VGG22* (*не оценено на BSD100), SRGAN-MSR*, SRGAN-VGG22*, SRGAN-VGG54 и оригинальное HR изображение. Каждый оценщик оценил 1128 изображений, которые были представлены в случайном порядке. Оценщики были откалиброваны на NN (оценка 1) и HR (оценка 5) версиях 20 изображений из датасета BSD300. В пробном исследовании мы оценили процедуру калибровки и надежность повторного тестирования 26 оценщиков на подмножестве из 10 изображений из BSD100, добавив изображения из различных методов дважды в большой тестовый набор. Мы нашли хорошую надежность и никаких существенных различий между оценками идентичных изображений. Оценщики очень последовательно оценивали NN-интерполированные тестовые изображения как 1 и оригинальные HR изображения как 5 (см. Рисунок 4).

Экспериментальные результаты проведенных MOS-испытаний описаны в Таблице 1, Таблице 2 и на Рисунке 5.

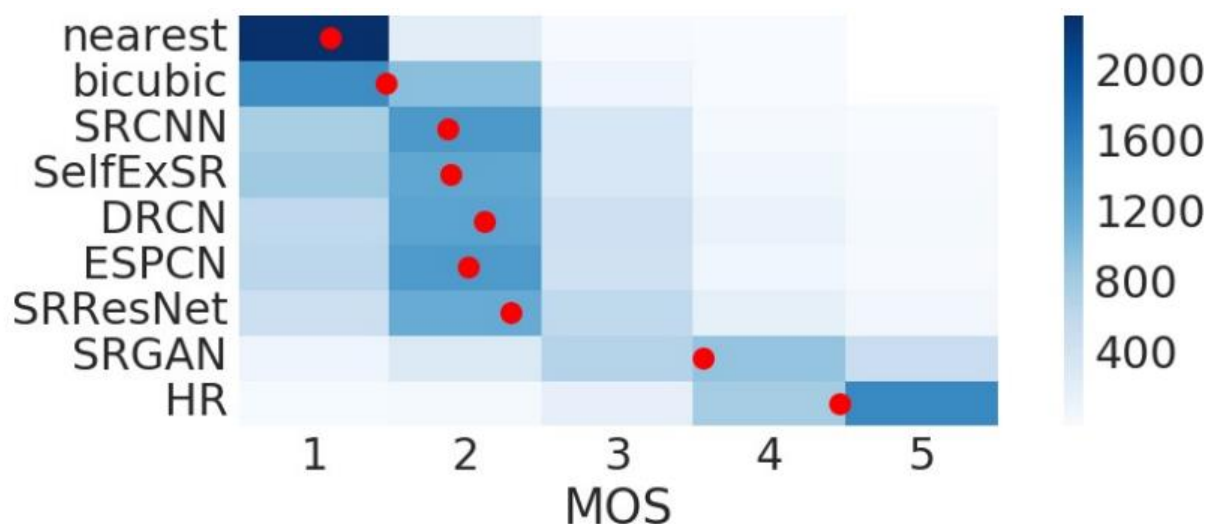


Рисунок 4. Распределение MOS баллов закодированное в цвете на BSD100. Для каждого метода 2600 образцов было оценено (100 изображений \times 26 оценщиков). Среднее значение отображается красным маркером, где ячейки центрированы около значения i . [4x масштабирование]

	SRResNet-		SRGAN-		
	MSE	VGG22	MSE	VGG22	VGG54
Set5					
PSNR	32.05	30.51	30.64	29.84	29.40
SSIM	0.9019	0.8803	0.8701	0.8468	0.8472
MOS	3.37	3.46	3.77	3.78	3.58
Set14					
PSNR	28.49	27.19	26.92	26.44	26.02
SSIM	0.8184	0.7807	0.7611	0.7518	0.7397
MOS	2.98	3.15*	3.43	3.57	3.72*

Таблица 1. Результаты различных функций потерь для SRResNet и состязательных сетей на Set5 и Set14 датасетах. MOS значительно выше ($p < 0.05$) чем с другими потерями в этой категории [4x масштабирование]

Set5	nearest	bicubic	SRCNN	SelfExSR	DRCN	ESPCN	SRResNet	SRGAN	HR
PSNR	26.26	28.43	30.07	30.33	31.52	30.76	32.05	29.40	∞
SSIM	0.7552	0.8211	0.8627	0.872	0.8938	0.8784	0.9019	0.8472	1
MOS	1.28	1.97	2.57	2.65	3.26	2.89	3.37	3.58	4.32
Set14									
PSNR	24.64	25.99	27.18	27.45	28.02	27.66	28.49	26.02	∞
SSIM	0.7100	0.7486	0.7861	0.7972	0.8074	0.8004	0.8184	0.7397	1
MOS	1.20	1.80	2.26	2.34	2.84	2.52	2.98	3.72	4.32
BSD100									
PSNR	25.02	25.94	26.68	26.83	27.21	27.02	27.58	25.16	∞
SSIM	0.6606	0.6935	0.7291	0.7387	0.7493	0.7442	0.7620	0.6688	1
MOS	1.11	1.47	1.87	1.89	2.12	2.01	2.29	3.56	4.46

Таблица 2. Сравнение NN, бикубическая, SRCNN, SelfExSR, DRCN, ESPCN, SRResNet, SRGAN-VGG54 и оригинальное HR изображение на оценочных данных. [4x масштабирование]

3.4. Изучение потери содержания

Мы исследовали влияние различных вариантов потери содержания в потере восприятия для сетей на основе GAN. Конкретно мы исследуем $l^{SR} = l_X^{SR} + 10^{-3}l_{Gen}^{SR}$ для соответствующей потери содержания l_X^{SR} :

- **SRGAN-MSE**: l_{MSE}^{SR} , чтобы исследовать состязательную сеть со стандартной потерей содержания MSE.
- **SRGAN-VGG22**: $l_{VGG/2.2}^{SR}$ с $\phi_{2.2}$, потеря определенная на картах особенностей определяющих низкоуровневые особенности.
- **SRGAN-VGG54**: $l_{VGG/5.4}^{SR}$ с $\phi_{5.4}$, потеря определенная на картах особенностей определяющих высокоуровневые особенности из глубоких слоёв сети с большим потенциалом сфокусированным на содержания из изображений. Дальше эту сеть будем называть **SRGAN**.

Мы также оценивает производительность генераторной сети без состязательной компоненты для двух потерь l_{MSE}^{SR} (**SRResNet-MSE**) и $l_{VGG/2.2}^{SR}$ (**SRResNet-VGG22**). Мы называем SRResNet-MSE как **SRResNet**. Заметим, что при обучении SRResNet-VGG22 мы добавили дополнительную общую вариационную потерю с весом $2 * 10^{-8}$ к $l_{VGG/2.2}^{SR}$. Количественные результаты суммированы в Таблице 1 и визуальные

примеры представлены на Рисунке 5. Даже в сочетании с состязательной потерей MSE обеспечивает решения с наибольшим PSNR, которые воспринимаются довольно гладкими и менее убедительными, чем результаты, достигнутые с компонентами потерь более чувствительными к визуальному восприятию. Это вызвано конкуренцией между потерей содержания на основе MSE и состязательной потерей. Кроме того, мы приписываем этим конкурирующим целям незначительные артефакты реконструкции, которые мы наблюдали в меньшинстве реконструкций на основе SRGAN-MSE. Мы не можем определить существенно лучшую функцию потерь для SRResNet или SRGAN в отношении оценки MOS на Set5. Однако SRGAN-VGG54 значительно превзошел другие SRGAN и SRResNet варианты на Set14 с точки зрения MOS. Мы наблюдали тренд, что используя высокоуровневые карты особенностей VGG $\phi_{5.4}$ дает лучшую детализацию текстуры по сравнению с $\phi_{2.2}$ (см. Рисунок 5).

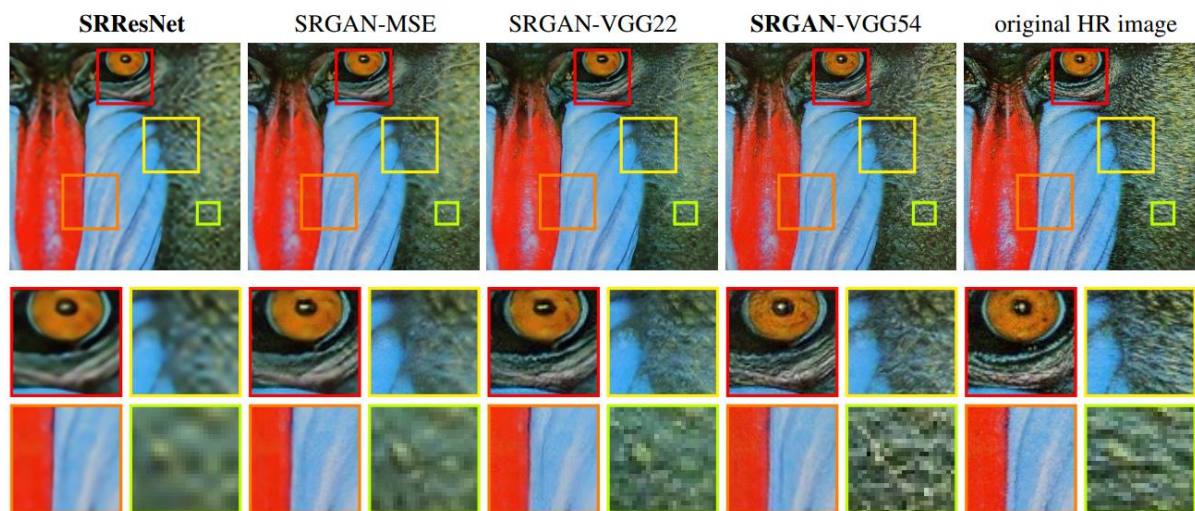


Рисунок 5. *SRResNet*, *SRGAN-MSE*, *SRGAN-VGG22*, *SRGAN-VGG54* результаты восстановления и соответствующее HR изображение [4x масштабирование]

3.5. Результаты финальных сетей

Мы сравниваем производительность **SRResNet** и **SRGAN** с методом ближайшего соседа, бикубической интерполяцией, и четырьмя современными методами. Количественные результаты суммированы в Таблице 2 и подтверждают, что **SRResNet** (с точки зрения PSNR/SSIM) устанавливает новый уровень качества на трех эталонных датасетах.

Далее мы получили рейтинги MOS для **SRGAN** и для всех эталонных методов на **BSD100**. Примеры суперразрешенных изображений с помощью **SRResNet** и **SRGAN** приведены на Рисунке 6. Результаты, показанные в Таблице 2 подтверждают, что **SRGAN** превосходит все эталонные методы с большим отрывом и устанавливает новый уровень качества для фотореалистичных изображений суперразрешения. Все различия MOS (см. Таблицу 2) очень значительны на **BSD100**, за исключением SRCNN vs. SelfExSR. Распределение всех собранных MOS оценок суммированы на Рисунке 4.



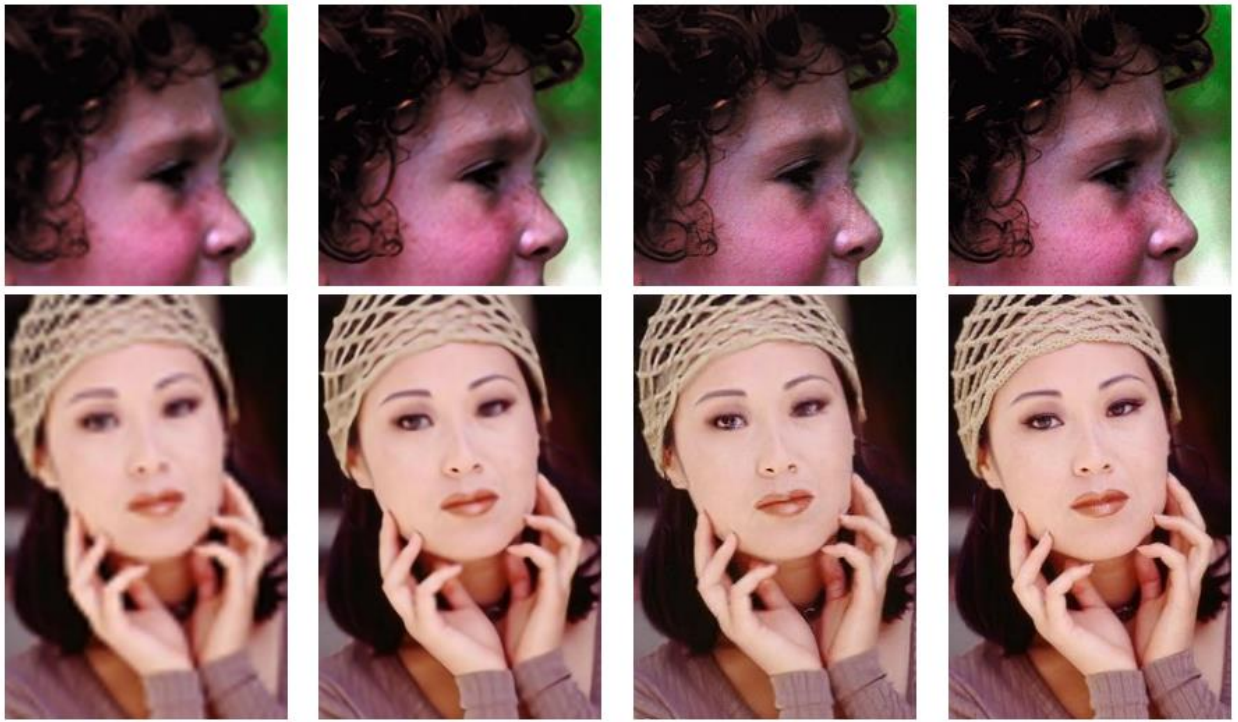


Рисунок 6. Результаты из датасета Set5 используя бикубическую интерполяцию, SRResNet и SRGAN. [4x масштабирование]

LR изображение

HR изображение

SR изображение



Результаты обученной модели

4. Обсуждение авторов и будущая их работа

Мы подтвердили превосходное визуальное восприятие **SRGAN** используя MOS тестирование. Также показали, что стандартные количественные метрики как PSNR и SSIM не в состоянии оценить качество изображения по отношению к зрительной системе человека. Целью этой работы было качество визуального восприятия суперразрешенных изображений, а не вычислительная эффективность. Представленная модель, в отличие от ESPCN в работе Shi et al. (<https://arxiv.org/pdf/1609.05158.pdf>), не оптимизирована для суперразрешения видео в режиме реального времени. Однако предварительные эксперименты с сетевой архитектурой предполагают, что более мелкие сети могут предоставить очень эффективные альтернативы при небольшом уменьшении качества результата. В отличие от Dong et al. (SRCNN), мы сочли полезным более глубокие сетевые архитектуры. Мы предполагаем, что дизайн ResNet оказывает существенное влияние на результаты более глубоких сетей. Мы обнаружили, что даже более глубокие сети ($B > 16$) могут еще улучшить результаты **SRResNet**, однако за счет увеличения времени обучения и тестирования. Далее мы обнаружили, что варианты SRGAN для более глубоких сетей все труднее обучать из-за появления высокочастотных артефактов.

Особое значение при стремлении к фотореалистичным результатам проблемы суперразрешения имеет выбор потери содержания, как показано на Рисунке 5. В этой работе мы нашли, что $l_{VGG/5.4}^{SR}$ дает наиболее убедительный результат для восприятия, атрибутом этого является потенция глубоких слоёв сети для предоставления признаков высокой абстракции вне пространства пикселей. Мы предполагаем, что карты особенностей этих глубоких слоёв фокусируются исключительно на содержании, пока состязательные потери сфокусированы на текстурных деталях, которые являются основным отличием между суперразрешенными изображениями без состязательной потери и фотореалистичными изображениями. Также отметим, что идеальная функции потери зависит от применения. Например, подходы, которые галлюцинируют более мелкие детали, могут быть менее подходящими для медицинского применения или наблюдения. Убедительная в восприятии реконструкция текста или

структурирование сцен является сложной задачей и частью будущих работ. Развитие функций потери содержания, которые описывают пространственное содержание изображения, но более инвариантны к изменениям в пиксельном пространстве, дополнительно улучшат результаты суперразрешения фотореалистичного изображения.

5. Выводы

В этой работе описан новый стандарт для фотореалистичного суперразрешения единственного изображения, под названием **SRGAN**. Этот метод использует все современные подходы свёрточных нейронных сетей, например, residual связи, генеративно-сопоставительные сети для обучения, использование полученных особенностей из слоёв глубоких свёрточных сетей и т.д. Важно заметить, что генерирующая сеть научилась создавать суперразрешенные изображения, не увидев при обучении ни одного HR изображения в таком подходе, таким образом она сама научилась выделять определенные текстуры и цвета для специального восстановления определенных областей. Также подход с функцией потери содержания подчеркивает все ограничения PSNR-сфокусированных подходов и это явно подтверждает MOS тестирование, описанное в данной работе, которое показывает явные преимущества **SRGAN** подхода в создании фотореалистичных изображений.

6. Литература

1. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network [<https://arxiv.org/pdf/1609.04802.pdf>]
2. Keras-GAN [<https://github.com/eriklindernoren/Keras-GAN>]
3. SRGAN, a TensorFlow Implementation [<https://towardsdatascience.com/srgan-a-tensorflow-implementation-49b959267c60>]
4. Is the deconvolution layer the same as a convolutional layer? [<https://arxiv.org/ftp/arxiv/papers/1609/1609.07009.pdf>]
5. Generative Adversarial Nets [<https://arxiv.org/pdf/1406.2661.pdf>]