ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 80 за 2022 г.

ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

В.П. Мещанинов, И.А. Молодецких, Д.С. Ватолин, А.Г. Волобой

Сочетание контрастного обучения и обучения с учителем для обнаружения видео с сверхвысоким разрешением

Рекомендуемая форма библиографической ссылки: Сочетание контрастного обучения и обучения с учителем для обнаружения видео с сверхвысоким разрешением / В.П. Мещанинов [и др.] // Препринты ИПМ им. М.В.Келдыша. 2022. № 80. 13 с. **https://doi.org/10.20948/prepr-2022-80**
https://library.keldysh.ru/preprint.asp?id=2022-80

О р д е н а Л е н и н а
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Р о с с и й с к о й а к а д е м и и н а у к

В.П. Мещанинов, И.А. Молодецких, Д.С. Ватолин, А.Г. Волобой

Сочетание контрастного обучения и обучения с учителем для обнаружения видео со сверхвысоким разрешением

Москва — 2022

Мещанинов В.П., Молодецких И.А., Ватолин Д.С., Волобой А.Г.

Сочетание контрастного обучения и обучения с учителем для обнаружения видео со сверхвысоким разрешением

Обнаружение увеличенного видео является полезным инструментом в мультимедийной криминалистике, однако это сложная задача, требующая применения различных алгоритмов масштабирования и сжатия. Существует множество методов повышения разрешения, включая интерполяцию и методы, основанные на машинном обучении. И все они оставляют уникальные следы. В этой работе предлагается новый метод обнаружения видео со сверхвысоким разрешением, основанный на изучении визуальных представлений с использованием перекрёстной энтропии и контрастной функции потерь. Чтобы объяснить, как метод обнаруживает такие видео, была подробно рассмотрена его структура. В частности, показано, что большинство подходов увеличения обучающей выборки ухудшает результат обучения модели. Благодаря обширным экспериментам с различными наборами данных было продемонстрировано, что предложенный метод эффективно обнаруживает масштабирование даже в сжатых видео и превосходит современные альтернативы. Код и модели доступны по адресу https://github.com/msu-video-group/SRDM.

Ключевые слова: оригинальное разрешение, обнаружение масштабирования, увеличение разрешения, интерполяция, сжатие видео

Viacheslav Pavlovich Meshchaninov, Ivan Andreevich Molodetskikh, Dmitriy Sergeevich Vatolin, Alexey Gennadevich Voloboy

Combining Contrastive and Supervised Learning for Video Super-Resolution Detection

Upscaled video detection is a helpful tool in multimedia forensics, but it’s a challenging task that involves various upscaling and compression algorithms. There are many resolution-enhancement methods, including interpolation and deep-learning based super-resolution, and they leave unique traces. This paper proposes a new upscaled-resolution-detection method based on learning of visual representations using contrastive and cross-entropy losses. To explain how the method detects videos, the major components of our framework are systematically reviewed — in particular, it is shown that most data-augmentation approaches hinder the learning of the method. Through extensive experiments on various datasets, our method has been shown to effectively detects upscaling even in compressed videos and outperforms the state-of-theart alternatives. The code and models are publicly available at https://github.com/msu-video-group/SRDM.

Key words: native resolution, upscaling detection, super-resolution, interpolation, video compression

Исследование выполнено за счет гранта Российского научного фонда № 22-21-00478, https://rscf.ru/project/22-21-00478/

Введение

В последние годы наблюдается большой скачок в качестве методов увеличения разрешения. И нейросетевые модели для этой задачи продолжают улучшаться. Их архитектуры варьируются от трансформеров [17], [22] до рекуррентных [5], [11], [13] и генеративно-состязательных сетей [21], [24], [26], [29], [30]. Эти методы способны генерировать высококачественные видео и изображения, которые человеческий глаз едва может отличить от реальных. Злоумышленники, такие как нечестные продавцы фотоаппаратов и недобросовестные видеооператоры, могут воспользоваться этими возможностями, что приведёт к этическим и юридическим проблемам [25], [33]. Поэтому автоматизированное обнаружение масштабирования видео имеет очень важное значение.

Предыдущие работы [4], [25], [32], [33] были, в основном, сосредоточены на обнаружении классических алгоритмов интерполяции, таких как интерполяция ближайшего соседа, билинейная интерполяция и бикубическая интерполяция [16]. Yang и др. [32] нацелены на методы сверхвысокого разрешения, но их экспериментальная оценка содержит тестовый набор данных всего из 10 видеопоследовательностей и только 3 метода сверхвысокого разрешения, что ограничивает возможность обобщить подход. Кроме того, авторы забыли рассмотреть такой частый случай как сжатые видео.

Как и обнаружение синтетических изображений и видео, обнаружение сверхвысокого разрешения должно быть способно различать сгенерированные данные. Однако на практике методы, разработанные для одной задачи, могут не работать в других ситуациях. Один из относительно новых методов обнаружения синтетических данных [28], например, пытался идентифицировать использование современных методов сверхвысокого разрешения, но он достиг лишь посредственной средней точности 93,6% для исходных изображений и 78,1% для изображений, сжатых в формате JPEG.

В настоящей работе представлен подход к обнаружению как сжатых, так и несжатых видео в увеличенном разрешении. Он включает контрастное обучение [2], [6], [9], [10] и обучение с учителем, что повышает точность результатов преимущественно для сжатых видео, как мы покажем далее. Мы обучили классификатор, создав набор увеличенных и сжатых видео с использованием нескольких моделей сверхвысокого разрешения. Для экспериментальной оценки мы выбрали тестовую часть набора данных REDS [20] и 100 видеопоследовательностей из Vimeo-90K [31], а также 6 моделей сверхвысокого разрешения, способных генерировать как одиночное изображение, так и несколько последовательных кадров видео. Были выбраны модели LGFN [24], RBPN [11], Real-ESRGAN [29], RRN [14], SOF-VSR [27] и Topaz [1]. Мы оценили качество нашего алгоритма на тестах MSU Video Super-Resolution [19] и RealSR [3]. Была получена высокая точность метода (обнаружены 30 из 32 методов увеличения разрешения), и тем самым подтверждена его способность к обобщению.

В итоге можно сказать, что нашим главным вкладом является создание нового метода обнаружения видео с увеличенным разрешением, сочетающий в себе идеи контрастного обучения и обучения с учителем. Также мы представили стратегию и детали конвейера обработки данных.

Предлагаемый метод

В этом разделе мы сначала опишем общую архитектуру предлагаемого метода. Далее будут представлены функция потерь и детали процесса обучения.

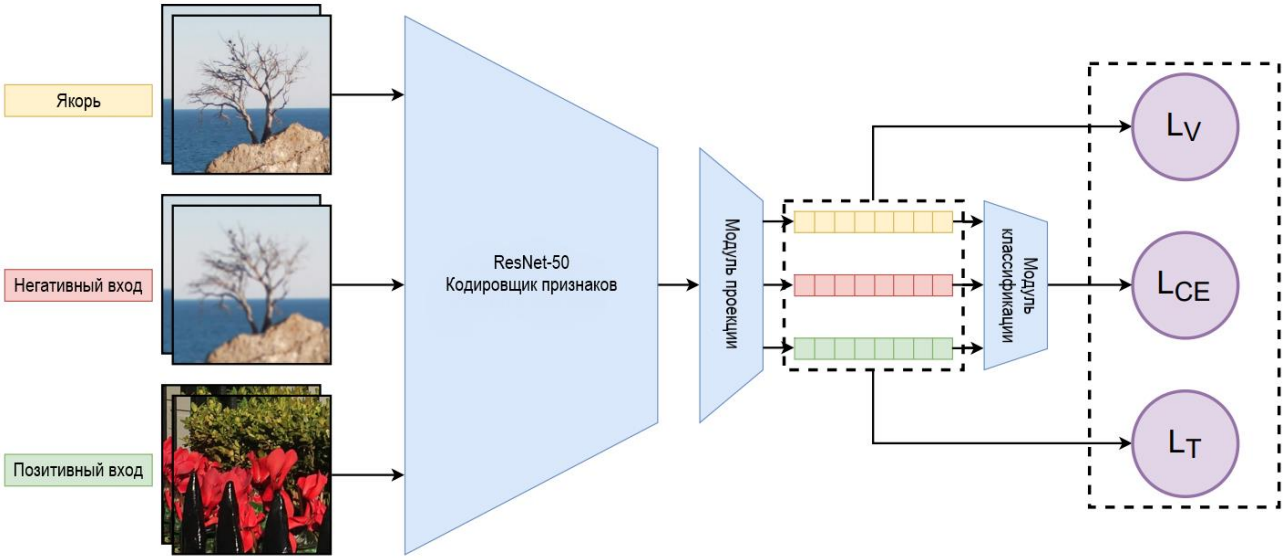

Рис. 1. Архитектура предлагаемого метода

Вдохновлённые недавними алгоритмами контрастного обучения [2], [6], [9], [10], мы предлагаем архитектуру, состоящую из четырёх основных компонентов: конвейер предварительной обработки данных, кодировщик признаков, модуль проекции и модуль классификации. На рис. 1 представлен обзор архитектуры.

- Процесс предобработки данных включает в себя модуль аугментации, который принимает в качестве входных данных два последовательных видеокадра. Затем он случайным образом выбирает квадрат со стороной 224 пикселя и вырезает его из изображения [8]. Как мы покажем далее, комбинация случайного вырезания и сжатия видео имеет решающее значение для достижения хорошего качества. Добавление других модификаций, таких как сжатие JPEG, гауссов шум и размытие по Гауссу, только снижает качество модели.
- Кодировщик признаков на основе нейронной сети $f(x)$ извлекает векторы представления из примеров аугментированных данных. Мы используем ResNet-50 [12] для получения представления $h = f(x)$, где x — пример аугментированных данных, h — выход слоя среднего

пулинга. Так как на вход подаются два кадра, количество каналов во входном слое было увеличено с трёх до шести.

- Модель проекции $g(h)$ — это небольшая нейронная сеть, которая сопоставляет представления с пространством, в котором применяется контрастная функция потерь. Мы используем многослойный перцептрон с тремя скрытыми слоями, чтобы получить проекцию $z = g(h)$. Этот компонент вдохновлён SimCLR [6], который показал, что определение контрастных потерь по z, а не по h, более выгодно.
- Модуль классификации $c(z)$ также представляет собой небольшую нейронную сеть с аналогичной архитектурой. Он сопоставляет проекцию z с вероятностью масштабирования кадров.

Модель оптимизирует сумму трёх функций потерь: перекрёстной энтропии, триплетов и дисперсии. Мы случайным образом извлекаем из видео с реальным разрешением пару последовательных кадров, которые мы считаем якорной выборкой a, после чего мы берём их увеличенную версию, которую считаем негативной выборкой n. Наконец, мы случайно выбираем из другого видео с реальным разрешением пару последовательных кадров и считаем их положительной выборкой p. Получив от кодировщика их представления $f(a)$, $f(n)$ и $f(p)$, проектор отображает их в пространство, в котором применяется контрастная потеря.

В итоге получаем $h_a = g\big(f(a)\big)$, $h_n = g\big(f(n)\big)$ и $h_p = g\big(f(p)\big)$. Пусть $sim(u, v) = \frac{u^t v}{\|u\|\|v\|}$ обозначает скалярное произведение нормализованных векторов u и v, т.е. их косинусное сходство. Тогда потеря триплетов будет следующей:

$$L_T = \frac{1}{N}\sum_{i=1}^{N}(sim\big(h_{a_i}, h_{p_i}\big) - sim\big(h_{a_i}, h_{n_i}\big) + m),$$

где N — это размер батча, а m — расстояние между классами.

Этот критерий имеет тенденцию приближать представление якорного видео по косинусному сходству к представлению случайно выбранного видео, которое имеет совершенно другое содержание. Как правило, это приводит к тому, что представление якорного видео больше отличается от представления масштабированного видео, которое имеет тот же контент, но было обработано методом увеличения разрешения.

Обозначим $H_a = \big[h_{a_1}, \dots, h_{a_N}\big]$, $H_n = \big[h_{n_1}, \dots, h_{n_N}\big]$ и $H_p = \big[h_{p_1}, \dots, h_{p_N}\big]$ как три батча из N векторов размерности d. Также пусть h^j будет вектором, который содержит каждое значение размерности j по всех векторах из $H \in \{H_a, H_n, H_P\}$. Как и в [2], мы определяем регуляризацию дисперсии v как функцию шарнира от стандартного отклонения представлений по всей размерности батча:

$$v(H) = \frac{1}{d}\sum_{j=1}^{d}\max\Big(0, \gamma - S\big(h^j, \varepsilon\big)\Big),$$

где S — регуляризованное стандартное отклонение, определяемое как

$$S(x, \varepsilon) = \sqrt{Var(x) + \varepsilon}.$$

Здесь γ — постоянное целевое значение стандартного отклонения, которое мы зафиксировали в экспериментах равным 1. А ε — малая скалярная величина, предотвращающая численную нестабильность. Затем мы определяем потерю дисперсии следующим образом:

$$L_V = v(H_a) + v(H_n) + v(H_p).$$

Этот критерий способствует тому, чтобы дисперсия в текущем батче равнялась γ по каждому измерению, предотвращая коллапс, при котором все входные данные отображаются в один и тот же вектор.

Мы обозначили $C_a = [ch_{a_1}, \dots, ch_{a_N}]$, $C_n = [ch_{n_1}, \dots, ch_{n_N}]$ и $C_p = [ch_{p_1}, \dots, ch_{p_N}]$ как три набора выхода классификатора до взятия операции максимума, поэтому потеря перекрёстной энтропии равна

$$L_{CE} = \frac{1}{4} CE(C_a, 0) + \frac{1}{4} CE\big(C_p, 0\big) + \frac{1}{2} CE(C_n, 1),$$

где $CE(y, t)$ — перекрёстная энтропия между выходом y и целевым значением t. Мы используем коэффициенты $\frac{1}{4}$, $\frac{1}{4}$, и $\frac{1}{2}$ для якорных, позитивных и негативных выборок соответственно, чтобы устранить несоответствие между видео с оригинальным и с поддельным разрешением.

Итоговая функция потерь представляет собой сумму перекрёстной энтропии, потери триплетов и потери дисперсии:

$$Loss = L_{CE} + L_T + L_V.$$

Для обучения наших сетей мы использовали набор данных REDS [20], состоящий из 240 исходных видеопоследовательностей, каждая из которых содержит 500 кадров. Мы разделили последовательности на 5 блоков по 100 кадров в каждом и взяли по 20 кадров с 10-го по 29-й включительно из каждого блока. Затем мы использовали видеокодек x264 для сжатия исходных видео с равномерно случайным фактором постоянного оценивания (CRF) от 15 до 30 включительно. Эти последовательности составляют набор данных «реального разрешения». Следующим шагом было уменьшение разрешения всех исходных последовательностей с помощью билинейной интерполяции и увеличение их разрешения с использованием шести методов сверхвысокого разрешения: Real-ESRGAN [29], ESRGAN [30], RRN [14], RBPN [11], SOF-VSR [27] и RealSR [15]. Мы сжали увеличенные видео так же, как и оригиналы. Они составляют набор данных «поддельного разрешения».

Таблица 1

Точность распознавания поддельного разрешения

Метод сверхвысокого разрешения	Без сжатия		Со сжатием	
	Точность	**Оценка F1**	**Точность**	**Оценка F1**
Наш метод				
Topaz	0.995	0.959	0.991	0.972
LGFN	0.943	0.934	0.895	0.886
RBPN	0.957	0.941	0.904	0.895
Real-ESRGAN	1.000	0.963	0.967	0.956
RRN	0.977	0.951	0.938	0.945
SOF-VSR-BD	0.981	0.910	0.909	0.930
Сао и др. [4]				
Topaz	0.828	0.864	0.824	0.856
LGFN	0.871	0.888	0.838	0.865
RBPN	0.880	0.893	0.842	0.867
Real-ESRGAN	0.788	0.833	0.927	0.862
RRN	0.914	0.919	0.874	0.884
SOF-VSR-BD	0.900	0.904	0.871	0.884

Экспериментальная оценка

В этом разделе мы сначала исследуем оптимальные настойки для предложенного метода обнаружения сверхвысокого разрешения, а затем представляем экспериментальные результаты, демонстрирующие его эффективность.

Чтобы подготовить данные для извлечения признаков, мы объединяем два последовательных видеокадра. Наш метод включает в себя предварительно обученную на наборе данных ImageNet [7] сеть ResNet-50 в качестве основы, использует трёхслойный перцептрон для проекции представления в 128-мерное скрытое пространство, а также применяет трёхслойный перцептрон для модуля классификации. Мы используем оптимизатор AdamW [18] в течение 300 эпох с коэффициентом скорости обучения 2×10^{-5} и со снижением веса 0,05. Размер батча составляет 64, алгоритм реализует коэффициент скорости обучения с косинусным затуханием с линейным прогревом в 20 эпох, а начальная скорость обучения составляет 5×10^{-6}.

Мы выбрали стандартизированную сбалансированную точность (b-Accuracy) и оценку F1 для тестирования нашей модели. Эти метрики требуют, чтобы детекторы поддерживали крошечный уровень ложноположительных результатов, особенно в практическом случае автоматической проверки

недостоверной информации в социальных сетях. Кроме того, чтобы протестировать наш метод отдельно на каждой модели сверхвысокого разрешения, мы использовали точность.

В наших тестах использовались два набора видеоданных: REDS и Vimeo-90K. REDS содержит реалистичные и динамичные сцены. Vimeo-90K является крупным набором высококачественного видео. Тесты состояли из 30 последовательностей по 500 кадров каждая. Мы взяли тестовую часть набора данных REDS и 100 видео из Vimeo-90K и применили ту же схему генерации данных, что и для набора обучающих данных, включая сжатие. Нашим следующим шагом стало масштабирование видео посредством нескольких методов сверхвысокого разрешения: LGFN [24], RBPN [11], Real-ESRGAN [29], RRN [14], SOF-VSR [27] и Topaz Gigapixel AI [1]. Результаты представлены в таблице 1.

Таблица 2

Оценка на тесте MSU VSR

Часть набора данных	Доля обнаруженных методов сверхвысокого разрешения	
	Наш метод	Cao и др. [4]
Оригинальные видео	1.000	0.375
Board	0.688	0.844
QR	0.688	0.906
Text	0.969	0.688
Tin foil	0.938	0.844
Color lines	0.500	0.844
License-plate numbers	0.906	0.875
Noise	0.875	0.688
Resolution test chart	0.906	0.906

Для оценки возможности обобщения нашего подхода между наборами данных мы использовали набор данных MSU Video Super-Resolution Benchmark [19], который содержит наиболее сложный контент для задачи восстановления: лица, текст, QR-коды, номерные пластины машин, неструктурированные текстуры и мелкие детали. Видеопоследовательности содержат различные типы движений и деградаций. Набор данных включает в себя 10 видео с реальным разрешением и имеет для каждого видео по 32 видео, которые были масштабированы различными методами сверхвысокого разрешения. Считается, что видео имеет поддельное разрешение, если метод обнаруживает не менее 5% всех кадров. Мы протестировали наш метод на всех видео и для каждого посчитали процент обнаруженных методов сверхвысокого разрешения.

Результаты показывают, что метод обнаруживает реальные видео с точностью 100%. В таблице 2 приведены результаты сравнения масштабированных видео.

Таблица 3

Оценка на тесте RealSR

Метод	Точность	
	Наш метод	Сао и др.
Оригинальные видео	0.877	0.734
RealSR	0.993	0.938
Real-ESRGAN	0.952	0.708
ESRGAN	0.996	0.958

Мы также использовали тест RealSR для оценки нашего метода. Мы обучили метод детекции изображений, изменив только входные данные. Вместо двух последовательных кадров модель получает только один кадр. Баланс схемы обучения остался без изменений.

Мы создали увеличенные изображения с помощью 3 современных методов сверхвысокого разрешения: RealSR [3], Real-ESRGAN [29] и ESRGAN [30]. Наша оценка качества работы метода на оригинальных и увеличенных изображениях основана на точности. Результаты приведены в таблице 3.

Таблица 4

Сравнение методов аугментации

Метод аугментации	Без сжатия		Со сжатием	
	b-Accuracy	Оценка F1	b-Accuracy	Оценка F1
Без аугментации	0.924	0.925	0.874	0.883
Размытие	0.923	0.937	0.879	0.885
JPEG	0.914	0.929	0.867	0.878
Гауссов шум	0.918	0.937	0.875	0.885
Вырезание квадрата	0.951	0.952	0.933	0.933

Таблица 4 показывает количественную оценку способности к обобщению обучения с помощью различных методов аугментации. Было обнаружено, что почти все подходы к увеличению данных препятствуют обучению. Мы считаем, что размытие, сжатие JPEG, шум и другие методы изменяют следы сверхвысокого разрешения и создают свои артефакты. Кроме того, мы обнаружили, что вырезание значительно увеличивает качество нашего метода как для сжатых, так и для несжатых видео.

Таблица 5

Оценка на валидационном наборе данных

Метод	Без сжатия		Со сжатием	
	b-Accuracy	Оценка F1	b-Accuracy	Оценка F1
Cao и др. [4]	0.884	0.881	0.867	0.863
MobileNetV2	0.900	0.908	0.894	0.901
MobileNet-CTV	0.929	0.917	0.920	0.916
ResNet	0.930	0.936	0.915	0.912
ResNet-CT	0.935	0.937	0.918	0.920
ResNet-CTV	0.948	0.948	0.925	0.924

Чтобы подтвердить эффективность потери триплетов и потери дисперсии, мы оценили, как они влияют на точность нашей модели. Мы сравнили обычное обучение ResNet-50 с перекрёстной энтропией (ResNet), с добавлением потери триплетов (ResNet-CT), и с добавлением потери дисперсии (ResNet-CTV). Для случая, когда необходима легковесная модель, мы обучили MobileNetV2 [23] и MobileNetV2-CTV. Мы также сравнили наш метод с методом Cao и др. [4], который мы также обучили на тех же данных, что и наши модели. Результаты по точности и оценке F1 представлены в таблице 5.

Таблица 6

Сравнение использования различного количества кадров

Количество кадров	Без сжатия		Со сжатием	
	b-Accuracy	Оценка F1	b-Accuracy	Оценка F1
1	0.925	0.929	0.918	0.920
2	0.948	0.948	0.925	0.924
3	0.951	0.952	0.919	0.915

Мы также оценили, как количество последовательных кадров, которые передаются в качестве входных, влияет на качество нашей модели. В таблице 6 приведены результаты сравнения.

Заключение

В настоящей статье мы предложили новый подход к обнаружению видео со сверхвысоким разрешением, который сочетает в себе контрастное обучение и обучение с учителем. Метод сначала использует основу ResNet для извлечения глубоких признаков, а затем использует легковесные многослойные

перцептроны для проецирования и классификации представлений входных кадров. Наш метод показал хорошие результаты в обширных экспериментах.

Библиографический список

[1] Topaz Gigapixel AI, 2021. https://www.topazlabs.com/gigapixel-ai.

[2] Adrien Bardes, Jean Ponce, and Yann LeCun. Vicreg: Varianceinvariance-covariance regularization for self-supervised learning. arXiv preprint arXiv:2105.04906, 2021.

[3] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 3086-3095, 2019. https://doi.org/10.1109/iccv.2019.00318

[4] Gang Cao, Antao Zhou, Xianglin Huang, Gege Song, Lifang Yang, and Yonggui Zhu. Resampling detection of recompressed images via dualstream convolutional neural network. arXiv preprint arXiv:1901.04637, 2019.

[5] Aman Chadha, John Britto, and M Mani Roja. iseebetter: Spatiotemporal video super-resolution using recurrent generative backprojection networks. Computational Visual Media, 6(3):307-317, 2020. [6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In International conference on machine learning, pages 1597-1607. PMLR, 2020. https://doi.org/10.1007/s41095-020-0175-7

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li FeiFei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248-255. Ieee, 2009. https://doi.org/10.1109/cvpr.2009.5206848

[8] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv: 1708.04552,2017.

[9] Alexey Dosovitskiy, Jost Tobias Springenberg, Martin Riedmiller, and Thomas Brox. Discriminative unsupervised feature learning with convolutional neural networks. Advances in neural information processing systems, 27, 2014. https://doi.org/10.1109/tpami.2015.2496141

[10] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'O6), volume 2, pages 1735-1742. IEEE, 2006. https://doi.org/10.1109/cvpr.2006.100

[11] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Recurrent back-projection network for video super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3897-3906, 2019. https://doi.org/10.1109/cvpr.2019.00402

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770-778, 2016. https://doi.org/10.1109/cvpr.2016.90

[13] Takashi Isobe, Xu Jia, Shuhang Gu, Songjiang Li, Shengjin Wang, and Qi Tian. Video super-resolution with recurrent structure-detail network. In European Conference on Computer Vision, pages 645-660. Springer, 2020.

[14] Takashi Isobe, Fang Zhu, Xu Jia, and Shengjin Wang. Revisiting temporal modeling for video super-resolution. arXiv preprint arXiv:2008.05765, 2020.

[15] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 466-467, 2020. https://doi.org/10.1109/cvprw50498.2020.00241

[16] Robert Keys. Cubic convolution interpolation for digital image processing. IEEE transactions on acoustics, speech, and signal processing, 29(6):1153-1160,1981 https://doi.org/10.1109/tassp.1981.1163711

[17] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 1833-1844, 2021. https://doi.org/10.1109/iccvw54120.2021.00210

[18] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. arXiv preprint arXiv: 1711.05101, 2017.

[19] Eugene Lyapustin, Anastasia Kirillova, Viacheslav Meshchaninov, Evgeney Zimin, Nikolai Karetin, and Dmitriy Vatolin. Towards true detail restoration for super-resolution: A benchmark and a quality metric. arXiv preprint arXiv:2203.08923, 2022.

[20] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 0-0, 2019. https://doi.org/10.1109/cvprw.2019.00251

[21] Jinshan Pan, Haoran Bai, Jiangxin Dong, Jiawei Zhang, and Jinhui Tang. Deep blind video super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 4811-4820, 2021. https://doi.org/10.1109/iccv48922.2021.00477

[22] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran. Image transformer. In International Conference on Machine Learning, pages 4055-4064. PMLR, 2018.

[23] Mark Sandler, Andrew G Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In CVPR, 2018. https://doi.org/10.1109/cvpr.2018.00474

[24] Dewei Su, Hua Wang, Longcun Jin, Xianfang Sun, and Xinyi Peng. Local-global fusion network for video super-resolution. IEEE Access, 8:172443-172456,2020. https://doi.org/10.1109/access.2020.3025780

[25] Yuting Su, Xiao Jin, Chengqian Zhang, and Yawei Chen. Hierarchical image resampling detection based on blind deconvolution. Journal of Visual Communication

and Image Representation, 48:480-490, 2017. https://doi.org/10.1016/j.jvcir.2017.01.009

[26] Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. Tdan: Temporally-deformable alignment network for video super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3360-3369, 2020. https://doi.org/10.1109/cvpr42600.2020.00342

[27] Longguang Wang, Yulan Guo, Li Liu, Zaiping Lin, Xinpu Deng, and Wei An. Deep video super-resolution using hr optical flow estimation. IEEE Transactions on Image Processing, 29:4323-4336, 2020. https://doi.org/10.1109/tip.2020.2967596

[28] Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A Efros. Cnn-generated images are surprisingly easy to spot... for now. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 8695-8704, 2020. https://doi.org/10.1109/cvpr42600.2020.00872

[29] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 1905-1914, 2021. https://doi.org/10.1109/iccvw54120.2021.00217

[30] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In ECCV Workshops (5), 2018. https://doi.org/10.1007/978-3-030-11021-5_5

[31] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. International Journal of Computer Vision, 127(8):1106-1125, 2019. https://doi.org/10.1007/s11263-018-01144-2

[32] Zaixin Yang, Yu Dong, Li Song, Rong Xie, Lin Li, and Yanan Feng. Native resolution detection for 4k-uhd videos. In 2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pages 1-5. IEEE, 2020. https://doi.org/10.1109/bmsb49480.2020.9379531

[33] Qin Zhang, Wei Lu, Tao Huang, Shangjun Luo, Zhaopeng Xu, and Yijun Mao. On the robustness of jpeg post-compression to resampling factor estimation. Signal Processing, 168:107371, 2020. https://doi.org/10.1016/j.sigpro.2019.107371

Оглавление