

УДК 004.89
DOI 10.17513/snt.40113

ПРИМЕНЕНИЕ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ В НЕПАРНОМ ПЕРЕНОСЕ ИЗОБРАЖЕНИЙ

Массеров Д.Д., Массеров Д.А., Лядунов К.А., Перков А.А.

ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет
имени Н.П. Огарёва», Саранск, e-mail: masserovggg@gmail.com

В данной статье рассматриваются несколько методов использования генеративно-сопоставительных сетей для переноса непарных изображений, исследуя различные архитектуры генераторов и дискриминаторов, а также функций потерь и гиперпараметров, которые влияют на качество сгенерированных изображений у моделей без учителя. Целью исследования был анализ современных методов непарного переноса изображений для его дальнейшего совершенствования. Литературный обзор включал в себя в основном базу поиска arXiv preprint arXiv, временной промежуток поиска составил 1985–2024 гг., было проанализировано 37 англоязычных источников, из них 12 указаны в списке литературы. Для проведения исследования были выбраны три модели: циклическая генеративно-сопоставительная сеть, контрастный непарный перенос, фиксированное/обученное самоподобие – из-за их эффективности в области переноса изображений и сходства в архитектуре. В ходе анализа были рассмотрены особенности каждой генеративно-сопоставительной сети для определения преимуществ и недостатков. Авторы установили, что основная разница между методами в том, что они используют разные типы потерь для обучения генеративной сети. Циклическая генеративно-сопоставительная сеть использует циклическую структуру с двумя генераторами и двумя дискриминаторами и потери цикловой согласованности, чтобы обеспечить более точные отображения при большем количестве затрат памяти и времени. Контрастный непарный перенос отказывается от сетей обратного отображения и дискриминаторов, используя контрастное обучение на уровне признаков для более легкой структуры модели. Фиксированное/обученное самоподобие также использует контрастную функцию потерь, но вместо того, чтобы сравнивать признаки в определенном слое, сравнивает пространственно-корреляционные карты для более точных результатов и в то же время быстрого обучения. Были установлены недостатки этих методов, неспособность циклической генеративно-сопоставительной сети значительно менять форму объектов, а недостаток контрастного непарного переноса в том, что модель не может отличать специфические для домена признаки от признаков внешнего вида. Данная статья представляет собой перспективный материал для научных исследований, направленных на улучшение качества сгенерированных изображений и эффективности процесса переноса изображений.

Ключевые слова: генеративно-сопоставительная сеть, циклическая генеративно-сопоставительная сеть, функции потерь, контрастный непарный перенос, фиксированное/обученное самоподобие, цикловая согласованность, уровень признаков, домен, перенос изображений

APPLICATION OF GENERATIVE-ADVERSARIAL NETWORKS IN UNPAIRED IMAGE TRANSFER

Masserov D.D., Masserov D.A., Lyadunov K.A., Perkov A.A.

Ogarev National Research Mordovia State University, Saransk, e-mail: masserovggg@gmail.com

This article discusses several methods for using generative-adversarial networks to transfer unpaired images, exploring different architectures of generators and discriminators, as well as loss functions and hyperparameters that affect the quality of the generated images in teacherless models. The goal of the study was to analyze current methods for unpaired image transfer in order to improve them. The literature review mainly included the arXiv preprint arXiv, the time span of the search was 1985–2024, and it analyzed 37 English-language sources, of which 12 were cited in the reference list. Three models were selected for the study: a cyclic generative-adversarial network, contrastive unpaired transfer, and fixed/learned self-similarity, because of their effectiveness in image transfer and similarity in architecture. The characteristics of each generative-adversarial network were examined to determine advantages and disadvantages. The authors found that the main difference between the methods is that they use different types of losses to train the generative network. The cyclic generative-adversarial network uses a cyclic structure with two generators and two discriminators and cyclic consistency loss to produce more accurate mappings at a higher memory and time cost. The contrastive unpaired transfer network abandons backward mappings and discriminators and uses feature-level contrastive learning for simpler model structure. Fixed/learned self-similarity also uses a contrastive loss function, but instead of comparing features in a given layer, it compares the spatial correlation maps for more accurate results and faster learning time. The shortcomings of these methods have been identified, the inability of the cyclic generative adversarial network to significantly change the shape of the objects, and the contrastive unpaired transfer in that the model cannot distinguish between domain-specific features from appearance features. This paper presents a promising contribution to research efforts aimed at improving the quality of generated images and the efficiency of the image transfer process.

Keywords: generative-adversarial network, cyclic generative-adversarial network, contrast unpaired transfer, fixed/learned self-similarity, loss functions, cyclic consistency, feature level, domain, image transfer

Введение

Генеративно-состязательные сети (GAN) продемонстрировали значительный успех в создании высококачественных изображений в различных областях [1–3], исходя из чего было принято решение использовать именно их. Используя их способность обучаться сложным распределениям и генерировать реалистичные изображения, можно создавать точные переносы «день-ночь».

Актуальность данного исследования многогранна. Во-первых, оно имеет практическое значение для многих отраслей, которые зависят от визуальных данных, например для киностудий, где часто требуется перевести день в ночь, чтобы совместить кадры, снятые в разное время суток. Во-вторых, она может помочь градостроителям в визуализации городских пейзажей при различных условиях освещения, что позволит им принимать более обоснованные решения относительно инфраструктуры и безопасности. Наконец, в сфере безопасности перенос изображения с дневного на ночное может расширить возможности систем видеонаблюдения, позволяя им эффективно работать как днем, так и ночью.

Цель данного исследования – рассмотреть применение современных методов непарного переноса изображений в контексте обучения без учителя. Потенциальные выгоды от успешного выбора реализации этой задачи значительны и охватывают различные отрасли и сферы применения.

Материалы и методы исследования

Литературный обзор включал в себя полнотекстовые оригинальные и обзорные статьи на английском языке через базу поиска arXiv preprint arXiv. Общая методология исследований представлена аналитико-синтетическим, сравнительным, статистическим подходами:

1. Анализ различных архитектур GAN, которые влияют на качество сгенерированных изображений.

2. Сравнение различных архитектур генераторов и дискриминаторов для оценки их эффективности в области переноса изображений (циклическая генеративно-состязательная сеть (CycleGAN), контрастный непарный перенос (CUT), фиксированное/обученное самоподобие (F/LSeSim)).

$$L_{\text{cyc}}(G, F) = E_x [F(G(x)) - x] + E_y [G(F(y)) - y], \quad (1)$$

где $F(G(x))$ – прямая цикловая согласованность;
 $G(F(y))$ – обратная цикловая согласованность.

3. Обоснование возможности улучшения качества сгенерированных изображений путем адаптации GAN для конкретных приложений и дальнейшего исследования.

Результаты исследования и их обсуждение

GAN широко используются в контексте переноса изображений друг в друга благодаря своей способности генерировать похожие изображения. Однако им не хватает контроля над генерируемыми данными.

Сохранение взаимосвязи между входными и выходными изображениями – важнейший аспект переноса изображений. Например, при переносе лошади на зебру должен меняться только внешний вид, а остальные аспекты остаться неизменными. Расчет расстояния или вектора на уровне пикселей не всегда дает удовлетворительные результаты для этой задачи [4–6]. На более абстрактном уровне для сравнения карт признаков или пространственно-корреляционных карт были предложены потери на основе признаков, которые могут эффективно сохранять специфические для данного домена признаки [7].

Цикловая согласованность CycleGAN. Ключевая идея CycleGAN заключается в том, чтобы ввести потерю цикловой согласованности, которая побуждает оба генератора к обучению согласованным отображениям между двумя доменами. На рис. 1 представлен принцип работы цикловой согласованности.

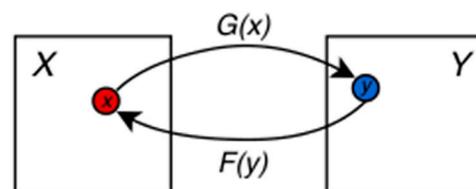


Рис. 1. Схема цикловой согласованности для двух доменов X и Y :
 x – изображение домена X ,
 y – изображение домена Y

Этого можно достичь, пропуская изображения через пары генераторов (G и F), обеспечивая реконструкцию исходного изображения. Математически функция потерь цикловой согласованности может быть определена следующим образом [6]:

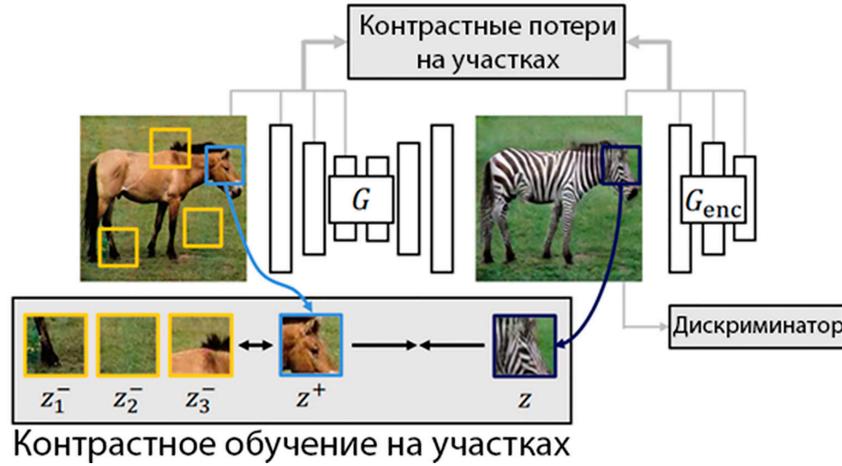


Рис. 2. Контрастное обучение на участках (Patchwise Contrastive Learning) для одностороннего переноса

Потери CycleGAN. Потери для CycleGAN состоят из двух частей: состязательные потери, которые побуждают генераторы производить образцы, неотличимые от реальных образцов дискриминаторов и потерь цикловой согласованности (L_{cyc}). Их можно выразить через следующую формулу:

$$L(G, F, D_X, D_Y) = L_{GANX}(G, D_Y, X, Y) + L_{GANX}(F, D_X, Y, X) + \lambda L_{cyc}(G, F), \quad (2)$$

где $L_{GANX}(G, D_Y, X, Y)$ – функция состязательных потерь домена X ;

$L_{GANX}(F, D_X, Y, X)$ – функция состязательных потерь домена Y ;

λ – относительная важность двух функций, было взято значение 10;

$L_{cyc}(G, F)$ – функция потерь цикловой согласованности.

В модели используются две состязательные сети:

- дискриминатор D_X , отличающий изображения $x, x \in X$ от перенесенных $F(y)$;
- дискриминатор D_Y , отличающий изображения $y, y \in Y$ от перенесенных $G(x)$.

D_X заставляет генератор G переносить изображения из X в неотличимые от домена Y изображения, аналогично D_Y , поощряет F синтезировать близкие к X изображения. Авторы [6] ввели функции потерь цикловой согласованности L_{cyc} (1) и две функции состязательных потерь L_{GANX} и L_{GANX} .

Контрастный непарный перенос (CUT). В области переноса изображения с одного изображения на другое, как показано на рис. 2, задача состоит в том, чтобы преобразовать входное изображение, сохранив его структурное содержание, но изменив его внешний вид в соответствии с целевым доменом. Для этого необходимо разделить содержание, которое должно оставаться неизменным в разных доменах, и внешний вид, который должен быть изменен.

Используя контрастную функцию потерь, такую как InfoNCE [8], данная модель учится связывать соответствующие при-

знаки, одновременно отделяя их от других частей входного изображения или нерелевантного фона. Это побуждает сеть фокусироваться на общих чертах между доменами (например, части и формы объектов), оставаясь инвариантной к различиям (например, текстуры животных).

Генератор и кодировщик вместе создают изображение, по которому можно отследить соответствующие входные данные. Используя многослойное контрастное обучение с использованием патчей и извлекая негативы из входного изображения, метод эффективно сохраняет содержание входных данных.

Потери CUT. В обучении без учителя подход контрастного обучения использовался как на уровне изображений, так и на уровне патчей (участков) [9, 10]. Для рассматриваемой задачи важно понимать, что не только целые изображения должны сохранять сходство содержания, но и соответствующие участки входных и выходных изображений. Эта идея мотивирует использование многослойного обучения на основе патчей.

В CUT выбирают L интересных слов и пропускают карты признаков через небольшую двухслойную MLP-сеть H_p , создавая стек признаков

$$\{\hat{z}_l\}_L = \{H_l(G_{enc}^l(x))\}_L, \quad (3)$$

где $G_{enc}^l(x)$ – выход l -го выбранного слоя; H_l – двухслойная MLP-сеть.

Сама функция потерь PatchNCE приведена ниже:

$$L_{PatchNCE}(G, H, X) = E_{x \sim X} \sum_{l=1}^L \sum_{s=1}^{S_l} l(\hat{z}_l^s, z_l^s, z_l^{S_l^s}), \quad (4)$$

где S_l – количество пространственных расположений в каждом слое;

z_l^s – соответствующий индексу признак,

$z_l^{S_l^s}$ – другие признаки.

Цель обучения CUT. Цель обучения у данной модели двояка: создание реалистичных изображений при сохранении соответствия между признаками на входных и выходных изображениях. Как показано

на рис. 2, задача минимаксной игры предназначена для достижения этого баланса. Кроме того, можно использовать потери PatchNCE для изображений из домена Y , чтобы предотвратить ненужные изменения в генераторе. По сути, эти потери являются обучаемой, специфичной для домена версией потерь идентичности, используемой в предыдущих методах непарного переноса, в том числе в CycleGAN [11].

$$L(G, D, H, X, Y) = L_{GAN}(G, D, X, Y) + \lambda_X L_{PatchNCE}(G, H, X) + \lambda_Y L_{PatchNCE}(G, H, Y) \quad (5)$$

где $L_{GAN}(G, D, X, Y)$ – состязательные потери, $\lambda_X = 1$, если $\lambda_Y = 1$, как было описано в [14],

$L_{PatchNCE}(G, H, X)$ – потери PatchNCE на изображение домена X ,

$L_{PatchNCE}(G, H, Y)$ – потери PatchNCE на изображение домена Y ,

Фиксированное/Обученное Самоподобие (F/LSeSim). Чуанься Чжэн и др. [5] представили новый метод для задач переноса изображений, который фокусируется на явном обучении пространственно-корреляционных карт. Такой перенос изображения сохраняет шаблоны самоподобия в исходном и перенесенном изображениях, независимо от геометрической формы или внешнего вида.

Хотя GAN могут генерировать изображения, соответствующие общему распределению набора данных, они часто не могут сохранить структуру сцены при переносе, если были обучены с только состязательными потерями. Для решения этой проблемы были разработаны различные потери для оценки согласованности содержания, такие как потери при реконструкции изображения на уровне пикселей, потери цикловой согласованности, потери при восприятии на уровне признаков и потери PatchNCE. Однако эти методы все еще страдают от некоторых ограничений. Потери на уровне пикселей не разделяют структуру и внешний вид, в то время как потери на уровне признаков объединяют признаки, характерные для конкретной области. Кроме того, большинство потерь на уровне признаков опираются на фиксированные сети ImageNet, которые могут плохо адаптироваться к произвольным областям.

Несмотря на значительные внешние различия между лошадей и зебр, когда

структуры объектов идентичны (например, одинаковые позы), пространственные шаблоны самоподобия также совпадают, что наглядно представлено на рис. 3.

Оценивая проявления совпадений в самоподобии, можно явно представить структуру в виде нескольких пространственно-корреляционных карт, визуализированных в виде тепловых карт на рис. 3, в и г [5].

Потери F/LSeSim. Фиксированное самоподобие. В предлагаемых авторами [5] фиксированных пространственно-корреляционных потерь, они сравнивают структурное сходство между входным изображением x в определенном домене и его соответствующим переносом \hat{y} в другом домене. Для этого сначала используется сверточная нейронная сеть, такая как VGG16 [5], для извлечения признаков для обоих изображений, в результате чего получаются векторы признаков f_x и $f_{\hat{y}}$. Вместо того чтобы напрямую вычислять расстояние между этими векторами ($f_x - f_{\hat{y}}$), они вводят понятие пространственно-корреляционной карты, математически определяемой как

$$S_{x_i} = (f_{x_i})^T (f_{x_n}), \quad (6)$$

где f_{x_i} – признак точки запроса x_i ;

f_{x_n} – соответствующие признаки в патче точек N_p ;

S_{x_i} – пространственная корреляция между точкой запроса и другими точками в патче.

После этого вся структура изображения представляется с помощью набора пространственно-корреляционных карт $S_x = [S_{x_1}; S_{x_2}; \dots; S_{x_n}]$. Такое представление позволяет проводить сравнения с большей вычислительной эффективностью.

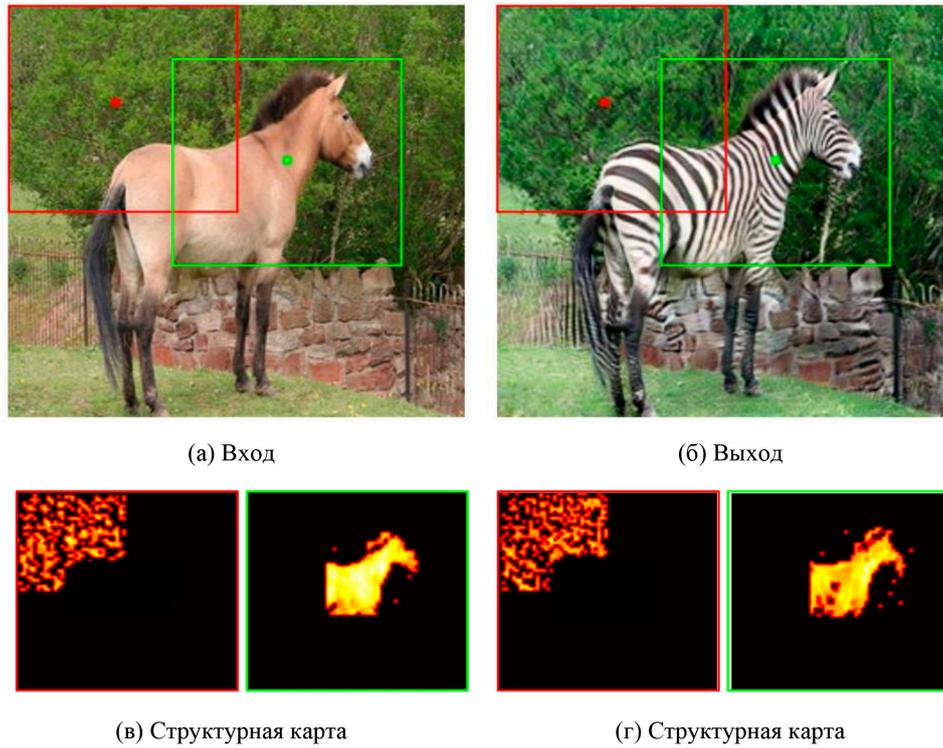


Рис. 3. Обученное пространственно-корреляционное представление кодирует локальную структуру объекта на основе самоподобия [5]

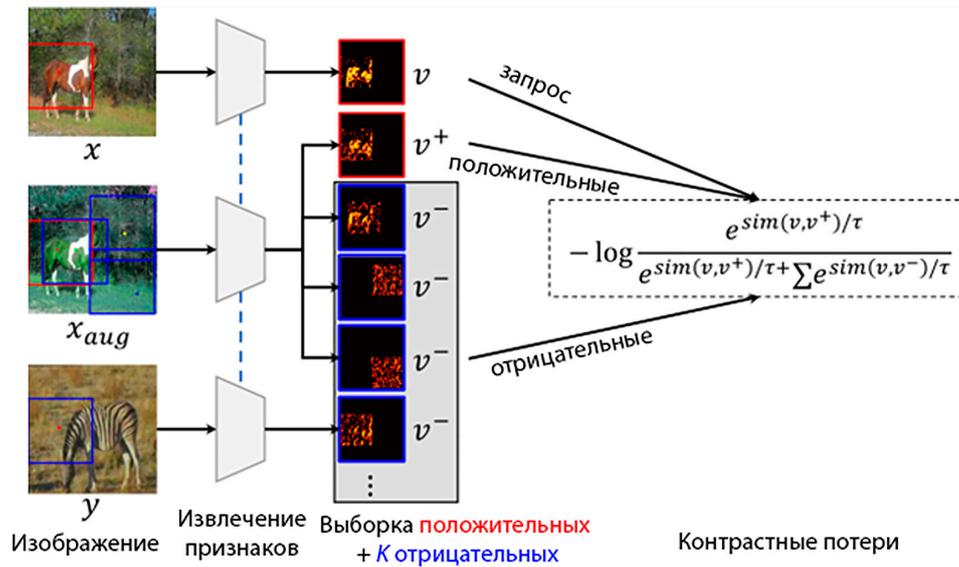


Рис. 4. Контрастное обучение на участках для получения самоподобия [5]

Затем сравниваем карты множественно-структурного сходства между входным x и перенесенным изображением y следующим образом:

$$L_s = d(S_x, S_y), \quad (7)$$

где S_x – набор пространственно-корреляционных карт, представляющий структуру изображения,

S_y – соответствующие S_x пространственно-корреляционные карты в целевом домене.

Для метрики расстояния d существуют два варианта: расстояние $L(S_x - S_y)$ и косинусное расстояние $(1 - \cos(S_x; S_y))$. Первое способствует постоянству пространственного сходства во всех точках участка, а второе – корреляции шаблонов без учета различий в величине S_x и S_y .

Обученное самоподобие. В контексте обучения без учителя авторы [5] предлагают генерировать пары схожих признаков на участках для эффективного обучения. Это достигается путем создания дополненных изображений с помощью структурно-сохраняющих преобразований. Обозначим патч «запроса» как $v = S_{x_i}$. «Положительные» и «отрицательные» образцы патча будут обозначены как $v^+ = S_{x_i}$ и $v^- = S_{K \setminus x_i}$ соответственно.

Запрашиваемый патч позитивно сопоставляется с патчем в той же позиции i в аугментированном изображении x_{aug} . В то же время он отрицательно сопрягается с патчами, отобранными из других позиций в x_{aug} , или патчами из других изображений y .

В LSeSim, как показано на рис. 4, используется контрастная функция потерь, которая поощряет подобие между запросом и положительными образцами и одновременно поощряет несходство с отрицательными образцами.

Для извлечения признаков подаются три изображения, в которых два изображения, x и x_{aug} , с одинаковой структурой, но разным внешним видом, а y – еще одно случайно выбранное изображение. Для каждого запрашиваемого участка в x положительным образцом является соответствующий участок в x_{aug} , а все остальные участки рассматриваются как отрицательные образцы.

Контрастные потери (L) определяются следующим образом:

$$L_c = -\log \frac{e^{\frac{sim(v, v^+)}{\tau}}}{e^{\frac{sim(v, v^+)}{\tau}} + \sum_{k=1}^K e^{\frac{sim(v, v_k^-)}{\tau}}}, \quad (8)$$

$$sim(v, v^+) = \frac{v^T v^+}{v v^+}, \quad (9)$$

$$sim(v, v_k^-) = \frac{v^T v_k^-}{v v_k^-}. \quad (10)$$

где $sim(v, v^{+/-})$ – косинусоидальное подобие между двумя векторами;

K – количество рассматриваемых отрицательных патчей;

k – индекс отрицательных образцов;

τ – температурный параметр, взято значение 0,07 [5].

Подводя итог, для оптимизации сети представления структуры f используется контрастная функция потерь, которая способствует сближению схожих патчей и отнесению несхожих. При этом пространственно-корреляционные потери в (8) используются для сети генератора в процессе обучения.

Цель обучения F/LSeSim. Основной целью является обучение сетей при минимизации следующих потерь:

$$L_D = -E_y [\log D(y)] - E_{\hat{y}} [\log (1 - D(\hat{y}))], \quad (11)$$

$$L_S = L_c, \quad (12)$$

$$L_G = E_y [\log (1 - D(\hat{y}))] + \lambda d(S_x, S_{\hat{y}}). \quad (13)$$

где L_D – состязательные потери дискриминатора;

L_S – контрастные потери сети репрезентации структуры f ;

L_G – потери генеративной сети G ;

λ – гиперпараметр, равный 10 [5].

Сравнительный анализ описанных моделей. Эти методы похожи по архитектуре, но отличаются по критерию потерь. Для репрезентативности сходства и различия сведены в таблицу.

В CycleGAN используется циклическая структура GAN с двумя генераторами и двумя дискриминаторами. CycleGAN также использует потери цикловой согласованности, чтобы входные изображения после прямого и обратного отображения были

как можно ближе к исходным изображениям. Однако из-за двух GAN эта система имеет тяжелую структуру и требует большого объема памяти. В CUT впервые было использовано контрастное обучение для переноса изображений без применения сетей обратного отображения и дополнительных дискриминаторов. Благодаря использованию контрастных потерь структура сети значительно облегчается и упрощается. F/LSeSim также использует контрастные потери, но сравнивает пространственно-корреляционную карту, а не признак в определенном слое. Таким образом, удается избежать зависимости между признаками внешнего вида и признаками, отражающими структуру изображения.

Сравнение CycleGAN, CUT и F/LSeSim на основе теоретических сведений

Метод	CycleGAN	CUT	F/LSeSim
Тип потерь	На уровне пикселей	На уровне признаков	Пространственно-корреляционная карта
Функция потерь	состязательные + цикловая согласованность	состязательные + PatchNCE	состязательные + самоподобие
Набор данных	непарный		
Вклад	Первое применение цикловой согласованности в GAN	Отказ от сетей обратного отображения	Быстрое обучение / точное отображение
Недостатки	Архитектура, требующая наибольших затрат памяти и времени; неспособность значительно менять геометрическую форму объектов; искажения	Модель не способна отличать специфические для домена признаки от признаков внешнего вида; искажения	Искажения
Архитектура	2 G + 2 D	1 G + 1 D	

Заключение

В ходе экспериментальной работы было проведено обширное исследование в области непарного переноса изображений без учителя, были рассмотрены основные преимущества и недостатки, как теоретические, так и практические.

Важно отметить, что данная работа является лишь началом и дальнейшие исследования должны быть направлены на улучшение качества сгенерированных изображений и эффективности процесса переноса изображений, а также на применение полученных результатов в реальной жизни.

Данная статья представляет собой перспективный материал для научных исследований и практических приложений в области обработки изображений, который может быть использован как база для дальнейшего развития в этой области.

Список литературы

- Alqahtani H., Kavakli-Thorne M., Kumar G. Applications of generative adversarial networks (GANs): An updated review // *Archives of Computational Methods in Engineering*. 2021. Vol. 28. P. 525–552. DOI: 10.1007/s11831-019-09388-y.
- Mukherjee A. et al. Generative semantic domain adaptation for perception in autonomous driving // *Journal of big data analytics in transportation*. 2022. Vol. 4. P. 103–117. DOI: 10.1007/s42421-022-00057-4.
- Ren C.X., Ziemann A., Theiler J., Alice M.S. Durieux A.M. Deep snow: synthesizing remote sensing imagery with generative adversarial nets // *Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imagery XXVI*. 2020. DOI: 10.1117/12.2560716.
- Mirza M., Osindero S. Conditional generative adversarial nets // *arXiv preprint arXiv: 1411.1784*. 2014. DOI: 10.48550/arXiv.1411.1784.
- Zheng C., Cham T.J., Cai J. The spatially correlative loss for various image translation tasks // *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021. P. 16402–16412. DOI: 10.1109/CVPR46437.2021.01614.
- Zhu J.Y., Zhang R., Pathak D., Darrell T., Efros A.A., Wang O., Shechtman E. Toward multimodal image-to-image translation // *In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA. 2017. P. 465–476. DOI: 10.48550/arXiv.1711.11586.
- Zhang K. On mode collapse in generative adversarial networks // *Artificial Neural Networks and Machine Learning—ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part II 30*. Springer International Publishing, 2021. P. 563–574. DOI: 10.1007/978-3-030-86340-1_45.
- Oord A., Li Y., Vinyals O. Representation learning with contrastive predictive coding // *arXiv preprint arXiv: 1807.03748*. 2019. DOI: 10.48550/arXiv.1807.03748.
- Bachman P., Hjelm R.D., Buchwalter W. Learning Representations by Maximizing Mutual Information Across Views // *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA. 2019. P. 15535–15545. DOI: 10.48550/arXiv.1906.00910.
- Hénaff O.J., Srinivas A., Fauw J., Razavi A., Doersch C., Eslami S.M. et al. Data-efficient image recognition with contrastive predictive coding // *In Proceedings of the 37th International Conference on Machine Learning (ICML'20)*. 2020. Vol. 119. P. 4182–4192. DOI: 10.48550/arXiv.1905.09272.
- Taigman Y., Polyak A., Wolf L. Unsupervised cross-domain image generation // *arXiv preprint arXiv: 1611.02200*. 2016. DOI: 10.48550/arXiv.1611.02200.
- Park T., Efros A.A., Zhang R., Zhu J.Y. Contrastive learning for unpaired image-to-image translation // *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, Part IX 16*. 2020. P. 319–345. DOI: 10.1007/978-3-030-58545-7_19.