

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АНГАРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ИНСТИТУТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ГЕОФИЗИКИ СО РАН

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ИССЛЕДОВАНИИ СЛОЖНЫХ СТРУКТУР

**МАТЕРИАЛЫ
ЧЕТЫРНАДЦАТОЙ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ
19–24 сентября 2022 г.**

Томск
Издательский Дом Томского государственного университета
2022

НЕПАРНОЕ ПРЕОБРАЗОВАНИЕ ИЗОБРАЖЕНИЯ В ИЗОБРАЖЕНИЕ С ИСПОЛЬЗОВАНИЕМ CYCLEGAN НА ОСНОВЕ ТРАНСФОРМЕРА

Ч. Гу, М.Л. Громов

Томский государственный университет, Томск, Россия
chongyugu@gmail.com

UNPAIRED IMAGE-TO-IMAGE CONVERSION USING CYCLEGAN BASED TRANSFORMER

Ch. Gu, M.L. Gromov

Tomsk State University, Tomsk, Russia
chongyugu@gmail.com

Свёрточные нейронные сети на сегодняшний день являются основой компьютерного зрения благодаря двум основным особенностям: локальной связности и механизму совместного использования параметров. Но сверточные архитектуры активно вытесняются трансформерами [1] в задачах классификации изображений [2], семантической сегментации [3], обнаружении объектов [4] и задачах GAN (Generative adversarial network – *англ.* Генеративно-сопоставительная сеть) [5, 6]. Трансформерные архитектуры обладают широкими возможностями представления свойств объектов и свободны от смещенных оценок свойств. Напротив, свёрточные архитектуры подвержены сильному смещению локальных оценок свойств и пространственной инвариантности из-за совместного использования весовых коэффициентов фильтров.

Предлагаемая нами модель представляет собой модифицированную сеть CycleGAN [7], полностью свободную от сверток, использующую только чистые трансформеры для кодирования. Кодер состоит из двух блоков. Первый блок состоит из модуля множественного внутреннего внимания (multi-head self-attention), а второй блок – из многослойного перцептрона с прямой связью и нелинейностью GELU. Перед обоими блоками встроены слои нормализации. На рис. 1 представлены архитектуры генератора и дискриминатора разработанной нами сети TransCycleGAN.

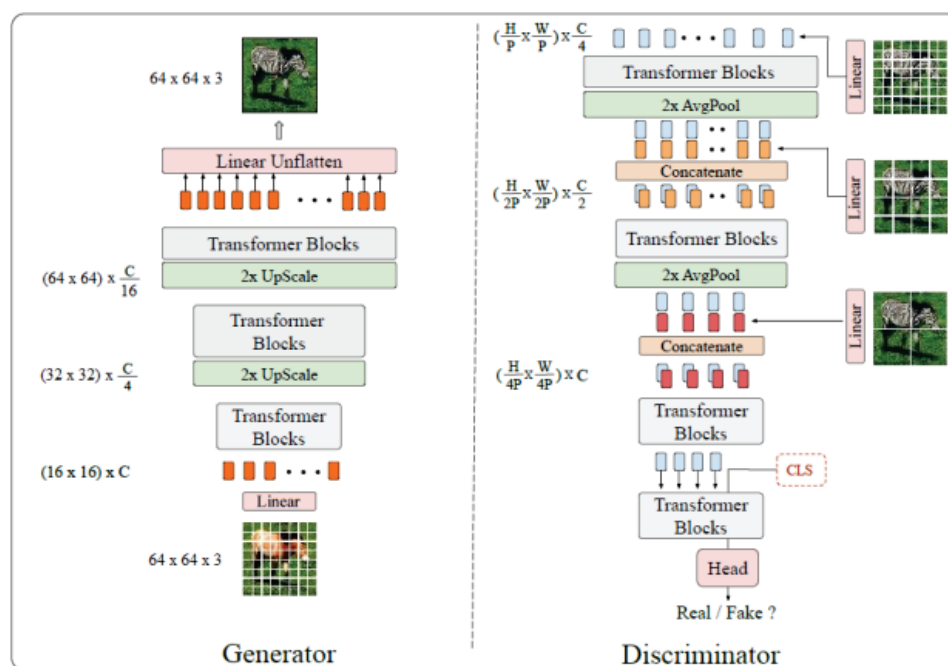


Рис. 1. Архитектуры генератора (слева) и дискриминатора (справа) сети TransCycleGAN

Разработанная сеть была опробована на стандартном наборе данных для преобразования изображения в изображение, размер horse2zebra изменен до 64×64 . Настройки сети были сделаны такими же, как и для TransGAN и использовались потери, введенные в CycleGAN. В процессе обучения применялась аугментация данных: сдвиг с вероятностью 1,0, смена цвета с вероятностью 0,3 и вырезание с вероятностью 1,0. Штраф за градиент не использовался, который обычно используется для обучения GAN, потому что это приводит к серьезным проблемам с обесцвечиванием. Предложенная модель достигла оценки FID [8] 80,54 на horse2zebra 64×64 и 93,05 на zebra2horse 64×64 . Прямое сравнение с другими моделями будет сделано в дальнейшей работе.

Литература

1. *Vaswani A., Shazeer N., Parmar N [et al.]*. Attention is all you need // *Advances in neural information processing systems*. – 2017. – Vol. 30. – P. 5998–6008.
2. *Dosovitskiy A., Beyer L., Kolesnikov A. [et al.]*. An image is worth 16x16 words: Transformers for image recognition at scale [электронный ресурс] // *arXiv.org*, 2021. – URL: <https://arxiv.org/abs/2010.11929> (дата обращения: 30.07.2022).
3. *Bao H., Dong L., Wei F.* BEiT: BERT Pre-Training of Image Transformers [Электронный ресурс] // *arXiv.org*, 2021. – URL <https://arxiv.org/abs/2106.08254> (дата обращения: 30.07.2022).
4. *Caron M., Touvron H., Misra I. [et al.]*. Emerging Properties in Self-Supervised Vision Transformers [Электронный ресурс] // *arXiv.org*, 2021. – URL: <https://arxiv.org/abs/2104.1429> (дата обращения: 30.07.2022).
5. *Jiang Y., Chang S., Wang Z.* TransGAN: Two Pure Transformers Can Make One Strong GAN, and That Can Scale Up [электронный ресурс] // *arXiv.org*, 2021. – URL: <https://arxiv.org/abs/2102.07074> (дата обращения: 30.07.2022).
6. *Lee K., Chang H., Jiang L. [et al.]*. ViTGAN: Training GANs with Vision Transformers [Электронный ресурс] // *arXiv.org*. – URL: <https://arxiv.org/abs/2107.04589> (дата обращения: 30.07.2022).
7. *Zhu J.-Y., Park T., Isola P., Efros A.A.* Unpaired image-to-image translation using cycle-consistent adversarial networks // *Proceedings of the IEEE international conference on computer vision*. – Cambridge: IEEE, 2017. – P. 2223–2232.
8. *Heusel M., Ramsauer H., Unterthiner T., Nessler B., Hochreiter S.* GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium [электронный ресурс] // *arXiv.org*, 2018. – URL: <https://arxiv.org/abs/1706.08500> (дата обращения: 30.07.2022).