

Легковесная нейронная сеть для цветокоррекции изображений с Unpaired режимом обучения

Назаренко А.А. (ИТМО)

Научный руководитель – к.ф.-м.н., доцент, директор ВШЦК, Михайлова Е.Г.

Научный консультант – к.ф.-м.н., доцент ВШЦК, Самарин А.В. (ИТМО)

Введение. Image enhancement — одна из ключевых задач компьютерного зрения, направленная на повышение визуального и общего качества изображений для улучшения восприятия человеком и повышения эффективности последующих задач обработки. Применение предварительного улучшения изображений широко востребовано в таких областях, как медицина, история и общая фотография [1,5,6]. Некачественные изображения возникают из-за условий съёмки, параметров сцены и освещения, что может привести к дисбалансу освещения, чрезмерному или недостаточному контрасту. С развитием глубокого обучения появились эффективные методы автоматической цветокоррекции изображений [2,4], включая тяжёлые модели с высокими вычислительными затратами и лёгкие модели, предназначенные для работы на мобильных устройствах. Область легковесных мало исследована, но достаточно перспективна [7,9]. В рамках данной работы предложена легкая нейронная модель для цветокоррекции изображений, использующая методы обучения без учителя и модель CLIP в качестве дискриминатора для повышения качества обработки.

Основная часть.

Модель нейронной сети должна удовлетворять следующим требованиям:

- 1) высокое качество результатов в сравнении с другими решениями;
- 2) отсутствие артефактов;
- 3) легковесность;
- 4) высокая скорость работы;
- 5) портируемость для применения на мобильном устройстве.

Обучение и оценку будем производить на двух популярных датасетах – MIT Adobe FiveK и FilmSet – с использованием стандартных метрик PSNR и SSIM.

Нейросетевая модель разработана на основе двухступенчатого генератора, основанного на модифицированной архитектуре LFIEM [4]. Первая стадия генератора включает свёрточный кодировщик признаков с тремя слоями и различными активациями (LeakyReLU и ReLU), за которыми следуют два полносвязных слоя. Вторая стадия состоит из применения классических фильтров для таких преобразований, как коррекция баланса белого, контрастности, насыщенности и экспозиции. Вместо нескольких генераторов параметров для различных фильтров, мы использовали лишь один генератор, что снижает сложность модели и повышает её эффективность.

Модель обучается по стандартной схеме генеративно-сопоставительных сетей (GAN) [8]. CLIP, используемый в роли дискриминатора, получает на вход изображения и заранее определённые текстовые подсказки-антонимы, такие как «good photo» или «bad photo». Этот подход позволяет модели более точно оценивать эстетическое качество изображений и генерировать улучшенные параметры фильтров. Преимущество CLIP заключается в возможности оперировать мультимодальными представлениями, что позволяет учесть как визуальные, так и текстовые признаки [3].

Для стабилизации процесса обучения использовалась Consistency Regularization, которая минимизирует разницу между параметрами исходного изображения и его слабых трансформаций. Это позволяет сделать генерацию параметров фильтров более устойчивой к небольшим изменениям данных.

Для более точной оценки качества улучшенных изображений мы предложили методику оценки на непарных данных с использованием моделей Visual Question Answering (VQA). Вопросы включали такие аспекты, как корректность экспозиции, насыщенность и баланс

цветов. Эта методика позволяет объективно оценивать модели даже при отсутствии парных данных, что делает её особенно полезной для задач обучения без учителя.

Таким образом, предложенная модель сочетает лёгкость, высокое качество улучшения изображений и независимость от больших объемов размеченных данных, что делает её перспективным решением для использования на мобильных устройствах и в ограниченных вычислительных условиях.

Выводы. Разработанная легковесная нейросетевая архитектура представляет собой легкое решение для автоматической цветокоррекции изображений, пригодное для мобильных устройств. При этом обеспечивается качество, сопоставимое с современными тяжеловесными подходами. Результаты показали, что предложенная модель демонстрирует качество, сравнимое с современными тяжеловесными моделями, но при этом требует значительно меньше вычислительных ресурсов (всего 47 тысяч обучаемых параметров).

Использование CLIP в роли дискриминатора позволяет добиться более высокой точности улучшения изображений, минимизируя зависимость от парных датасетов. Предложенный метод оценки качества на основе VQA моделей позволяет проводить объективную проверку на небольших объемах данных без привлечения человека для оценки. Дальнейшие исследования могут быть направлены на оптимизацию весов модели и улучшение качества в условиях обучения без учителя.

Список использованных источников:

1. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans / Y.-S. Chen [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2018. — С. 6306–6314.
2. Distort-and-recover: Color enhancement using deep reinforcement learning / J. Park [и др.] // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2018. — С. 5928–5936.
3. Exposure: A white-box photo post-processing framework / Y. Hu [и др.] // ACM Transactions on Graphics (TOG). — 2018. — Т. 37, № 2. — С. 1–17.
4. Tatanov O., Samarin A. LFIEM: Lightweight Filter-based Image Enhancement Model // 2020 25th International Conference on Pattern Recognition (ICPR). — 2021. — С. 873–878.
5. A. Samarin, A. Toropov, A. Dzestelova, A. Nazarenko, E. Kotenko, E. Mikhailova, A. Savelev, and A. Motyko, “Specialized non-local blocks for recognizing tumors on computed tomography snapshots of human lungs,” in 2024 35th Conference of Open Innovations Association (FRUCT), 2024, pp. 659–664.
6. A. M. H. Tiong, J. Li, B. Li, S. Savarese, and S. C. Hoi, “Plug-and-play vqa: Zero-shot vqa by conjoining large pretrained models with zero training,” arXiv preprint arXiv:2210.08773, 2022.
7. Image quality assessment: from error visibility to structural similarity / Z. Wang [и др.] // IEEE transactions on image processing. — 2004. — Т. 13, № 4. — С. 600–612.
8. Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs / V. Bychkovsky [и др.] // The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition. — 2011.
9. C. Yang, M. Jin, X. Jia, Y. Xu, and Y. Chen, “Adaint: Learning adaptive intervals for 3d lookup tables on real-time image enhancement,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 17 522–17 531.