

НАБОР ДАННЫХ ПЕРЕНОСА СТИЛЯ И МОДЕЛИ РЕГРЕССИИ КАЧЕСТВА СТИЛИЗАЦИИ

Абрамов Валентин Андреевич

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: ab.val26@yandex.ru

Научный руководитель — Китов Виктор Владимирович

Перенос стиля — область исследований, в которой для заданного изображения контента (например, фотографии) и изображения стиля (например, картины) ставится задача перерисовать контент в новом стиле. Технология используется в пользовательских и коммерческих сервисах автоматической стилизации изображений (например, *ostagram.me*) и востребована в дизайне, рекламе и социальных сетях. Несмотря на широкий прикладной интерес, для оценки качества стилизации отсутствует общепринятый показатель, согласованный с человеческим восприятием, и поэтому результаты различных работ часто несопоставимы и слабо воспроизводимы.

В данной работе представлен обширный датасет для оценки качества переноса стиля и обучены модели регрессии качества. Датасет включает 50 изображений контента и 50 изображений стиля, каждое изображение стиля представлено в 4 разрешениях (150^2 , 300^2 , 500^2 и 700^2 пикселей). Для всех пар «контент–стиль» сгенерировано 10 000 стилизованных изображений с использованием ArtFlow [1]. Каждое изображение оценено тремя независимыми аннотаторами по шкале от 1 до 10; целевой меткой является среднее этих оценок. Насколько нам известно, представленный датасет является крупнейшим в области оценки качества переноса стиля на основе человеческой разметки.

Чтобы уменьшить влияние различий в цветовой гамме на восприятие стиля, изображения стиля были приведены к цветам контента с помощью согласования среднего и дисперсии в цветовом пространстве CIE LAB. Согласованность разметки подтверждается корреляцией между аннотаторами (коэффициент τ_b Кендалла: 0,36–0,43, $p < 0,01$).

Проведён качественный и количественный анализ факторов, влияющих на воспринимаемое человеком качество стилизации. Основные выводы: (1) наиболее эффективны стили, содержащие сочетание крупных и мелких узоров, согласованное с низко- и высокочастотной структурой контента; (2) размер изображения стиля определяет

средний масштаб узоров, поэтому методы переноса должны устойчиво работать с контентом и стилями разных размеров; (3) разнообразие текстур и цветов в стиле повышает качество, тогда как упрощённые стили чаще приводят к размытию и потере деталей; (4) наличие границ под разными углами в стиле помогает сохранять сложные контуры контента и его узнаваемость.

Предложены две модели регрессии: модель оценки качества по результирующему стилизованному изображению и модель предсказания ожидаемой оценки по входным изображениям контента и стиля без выполнения переноса. Наилучшая архитектура использует предобученный ConvNeXt-Tiny [2] для извлечения признаков и трёхслойный MLP-регрессор.

Для оценки обобщающей способности используется пятикратное разбиение (5-fold), исключаящее пересечение изображений ни по контенту, ни по стилю между обучением и тестом, что снижает риск утечки.

Модель оценки стилизации достигает ошибки $RMSE = 1,43$ на тестовой выборке, что сравнимо с несогласованностью разметки между аннотаторами и позволяет использовать модель для автоматической оценки и сравнения методов стилизации. Модель предсказания оценки по входным изображениям контента и стиля достигает ошибки $RMSE = 1,66$, что позволяет оценивать совместимость изображений контента и стиля без выполнения вычислительно сложного переноса стиля.

Набор данных и модели регрессии доступны публично и могут использоваться для автоматизации оценки методов переноса стиля и подбора лучших стилей для пользовательских фото в сервисах автоматической стилизации изображений.

Литература

1. An J., Huang S., Song Y., Dou D., Liu W., Luo J. ArtFlow: Unbiased Image Style Transfer via Reversible Neural Flows // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 862–871.
2. Liu Z., Mao H., Wu C.-Y., Feichtenhofer C., Darrell T., Xie S. A ConvNet for the 2020s // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. P. 11976–11986.