
Метод мультистилевого рендеринга изображений

A Preprint

Загатин Данилл Ильич
МГУ им. М.В. Ломоносова
Факультет ВМК, кафедра Математических методов прогнозирования
Москва, Россия
Почта тут

Научный руководитель: к.ф.-м.н., доцент Китов Виктор Владимирович
МГУ им. М.В. Ломоносова

Abstract

В работе рассматривается задача произвольной стилизации изображений с использованием сверточных нейронных сетей (CNN). Большинство моделей обучаются для конкретного стиля и успешно передают палитры и текстуры, однако часто не воспроизводят уникальные стилистические особенности, склоняясь к усреднённым мазкам и цветам. На основе методов Гатиса, Джонсона и Гиаси исследуются модификации генератора с кодировщиком стиля и различными способами внедрения эмбедингов: конкатенацией, 1×1-инъекцией, FiLM-блоками и патчевыми методами. Работа выявляет архитектурные ограничения и предлагает улучшения: усиление тренировки кодировщика, раздельное внедрение компонентов стиля и использование альтернативных функций потерь. Разработана мультистилевая лёгкая архитектура с высоким качеством стилизации и гибкостью применения к разнообразным стилям.

Keywords перенос стиля · свёрточные нейронные сети · мультистилевая стилизация · стилиевой эмбединг · FiLM

1 Введение

Перенос художественного стиля (Neural Style Transfer, NST) позволяет преобразовывать изображение, сохраняя его семантическое содержимое и воспроизводя художественные особенности другого изображения. Практическая ценность NST растёт в цифровом искусстве, дизайне, AR/VR и мобильных приложениях, где важны интерактивность, персонализация и низкая стоимость генерации контента.

Классические оптимизационные подходы обеспечивают высокое качество, но требуют длительной итеративной подгонки под каждую новую пару изображений. Быстрые генеративные сети позволяют работать в реальном времени, однако в базовом варианте обучаются под один конкретный стиль, что ограничивает масштабирование. Универсальные методы произвольной стилизации снимают это ограничение, но часто теряют характерные структурные элементы стиля или страдают от артефактов и нестабильного обучения.

В данной работе рассматривается мультистилевая схема рендеринга, где стиливое изображение кодируется в компактный эмбединг и внедряется в генератор различными механизмами: прямой конкатенацией к признаковым картам, 1×1-инъекцией (learnable projection), FiLM-модуляцией, а также через патчевое сопоставление признаков. Такой дизайн объединяет скорость feed-forward архитектур с гибкостью произвольной стилизации.

Наш вклад — структурированное сравнение способов внедрения стиливого эмбединга, практические рекомендации по устойчивому обучению кодировщика стиля и анализ причин деградации качества

(усреднение палитры и мазков, артефакты от локальных сопоставлений). Мы также обсуждаем улучшения: усиление тренировки кодировщика контрастными целями, отдельное внедрение палитры и текстур и альтернативные функции потерь для сохранения локальных деталей.

2 Обзор литературы.

Базовый подход Gatys et al. [2015] формулирует NST как оптимизацию изображения под перцептивные потери на активациях VGG: контент фиксируется через промежуточные признаки, стиль — через матрицы Грама корреляций признаков. Это даёт качественную стилизацию, но непрактично для интерактивных сценариев.

Переход к генераторам реального времени осуществлён в Johnson et al. [2016]: сеть-преобразователь с residual-блоками обучается по перцептивным лоссам и стилизует за один прямой проход. Существенную роль в стабильности генерации сыграла Instance Normalization [Ulyanov et al., 2016]. Однако такие модели обычно требуют отдельного обучения под каждый стиль.

Поддержка произвольных стилей достигается статистическими методами выравнивания признаков: AdaIN выравнивает среднее и дисперсию каналов [Huang and Belongie, 2017], WCT применяет whitening&coloring трансформы [Li et al., 2017]. Работа Ghiasi et al. [2017] использует стиль-кодировщик для предсказания параметров нормализации слоёв генератора, внедряя стиль непосредственно в механизм нормализации. Эти подходы быстры и гибки, но при сильной вариативности стилей могут терять локальные текстуры и характерные мазки.

Локально-структурные (патчевые) методы сопоставляют фрагменты признаков контента и стиля [Chen and Schmidt, 2016] или комбинируют CNN с MRF [Li and Wand, 2016], что помогает передавать текстуры и повторяющиеся мотивы, но повышает требования к памяти и склонно к артефактам. Современные модели внимания усиливают согласование стиля и контента (SANet [Li et al., 2019], AdaAttN [He et al., 2019]), а трансформерные решения (StyTR² [Xia et al., 2022]) улучшают глобальные зависимости ценой усложнения архитектуры и обучения.

Наконец, условные модули (например, FiLM [Perez et al., 2018]) линейно модулируют признаки генератора параметрами, зависящими от условия (стиля), и предоставляют простой общий механизм внедрения стиливой информации; их практическая эффективность чувствительна к качеству стиливого эмбединга и балансу лоссов.

3 Headings: first level

3.1 Headings: second level

3.1.1 Headings: third level

Paragraph

4 Examples of citations, figures, tables, references

4.1 Citations

4.2 Figures

4.3 Tables

4.4 Lists

Список литературы

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576, 2015.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, pages 694–711. Springer, 2016.

- Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022, 2016.
- Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. arXiv preprint arXiv:1703.06868, 2017.
- Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. arXiv preprint arXiv:1705.08086, 2017.
- Golnaz Ghiasi, Honglak Lee, Manjunath Kudlur, Vincent Dumoulin, and Jonathon Shlens. Exploring the structure of a real-time, arbitrary neural artistic stylization network. arXiv preprint arXiv:1705.06830, 2017.
- Tian Qi Chen and Mark Schmidt. Fast patch-based style transfer of arbitrary style. arXiv preprint arXiv:1612.04337, 2016.
- Chuan Li and Michael Wand. Combining markov random fields and convolutional neural networks for image synthesis. arXiv preprint arXiv:1601.04589, 2016.
- Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Style-attentional networks for arbitrary style transfer. arXiv preprint arXiv:1904.08839, 2019.
- Jingwen He, Yezhen Chen, Lin Liu, Yu Li, Xiaodan Jin, Ming-Hsuan Yang, and Alan L. Yuille. Adaptive attention normalization for arbitrary style transfer. arXiv preprint arXiv:1905.01248, 2019.
- Weihao Xia, Yujiu Yang, and Jing-Hao Xue. Stytr²: Image style transfer with transformers. arXiv preprint arXiv:2204.12476, 2022.
- Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In AAAI, pages 3942–3951, 2018.