

Метод мультистилевого рендеринга изображений

Загатин Данилл Ильич

Научный руководитель: к.ф.-м.н., доцент Китов Виктор Владимирович

МГУ им. М. В. Ломоносова

Факультет ВМК, кафедра Математических методов прогнозирования

- Neural Style Transfer (NST): перенос художественного стиля на контент.
- Оптимизационные методы (Gatys) дают высокое качество, но медленные.
- Feed-forward генераторы (Johnson) быстрые, но обычно обучаются на **один стиль**.
- Произвольная стилизация часто **усредняет мазки/палитру** и теряет уникальные элементы.

Цель

Построить лёгкую мультистилевую архитектуру, поддерживающую множество стилей без переобучения.

Вклад

- Сравнение способов внедрения стилового эмбединга в генератор.
- Практические рекомендации по обучению кодировщика стиля.
- Анализ деградаций: усреднение, артефакты локальных сопоставлений.

Постановка задачи

Пусть

$$X_c \subset \mathbb{R}^{H \times W \times 3} \quad (\text{контент}), \quad X_s \subset \mathbb{R}^{H \times W \times 3} \quad (\text{стиль}).$$

Нужно построить отображение:

$$y = f_\theta(x_c, x_s),$$

где y сохраняет семантику x_c и визуальные свойства x_s .

В скрытом пространстве признаков стилевое изображение кодируется эмбедингом:

$$z_s = e_\psi(x_s) \in \mathbb{R}^d.$$

Искомая модель:

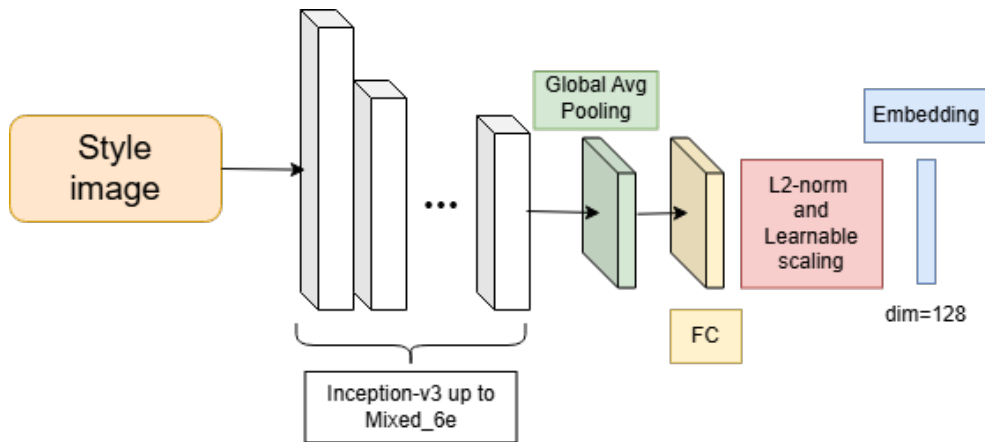
$$f_{\theta}(x_c, x_s) = G_{\theta}(x_c, e_{\psi}(x_s)).$$

- e_{ψ} — **кодировщик стиля**: извлекает компактное представление z_s .
- G_{θ} — **генератор**: преобразует контент с учётом z_s .
- Ключевой вопрос: как **эффективно внедрять** z_s в генератор?

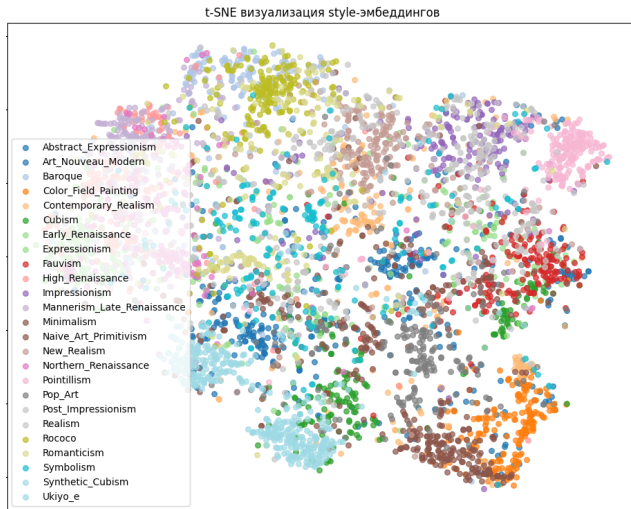
- Основа: модифицированный *Inception-v3*, усечение до *Mixed_6e*.
- После свёрток: Global Average Pooling \rightarrow FC $\rightarrow \mathbb{R}^d$.
- Нормализация и масштабирование:

$$z_s = \alpha \cdot \frac{h}{\|h\|_2}, \quad h = \text{FC}(\text{GAP}(\text{Inception}(x_s))).$$

- Для устойчивости: дообучение на классификации художественных стилей (WikiArt).



Векторные представления стилей



- Симметричная схема: downsampling \rightarrow residual bottleneck \rightarrow upsampling.
- Downsampling: свёртки 9×9 , 3×3 (две со stride=2), InstanceNorm + ReLU.
- Bottleneck: несколько residual-блоков.
- Upsampling: Upsample+Conv $\times 2$, InstanceNorm + ReLU.
- Выход: conv $9 \times 9 \rightarrow$ RGB.

1) Конкатенация

z_s расширяется по $H \times W$ и конкатенируется с картой признаков по каналам (в residual-блоках перед свёртками).

2) 1×1 -инъекция

Стилевой вектор преобразуется в карту признаков и добавляется:

$$F' = F + \text{Conv}_{1 \times 1}(z_s).$$

3) FiLM

$$F' = \gamma(z_s) \cdot F + \beta(z_s),$$

где $\gamma(\cdot), \beta(\cdot)$ — линейные преобразователи эмбединга.

4) Patch-based

Локальное сопоставление патчей признаков (напр., по косинусному сходству):

$$\Phi'(x_c) = \text{PatchMatch}(\Phi(x_c), \Phi(x_s)).$$

Общая функция:

$$\mathcal{L}_{total} = \lambda_c \mathcal{L}_{content} + \lambda_s \mathcal{L}_{style} + \lambda_{TV} \mathcal{L}_{TV}.$$

- **Контент (VGG-19, слой conv4_2 / relu4_2):**

$$\mathcal{L}_{content} = \|\Phi_{4,2}(y) - \Phi_{4,2}(x_c)\|_2^2.$$

- **Стиль (VGG-19, слои conv1_1, conv2_1, conv3_1, conv4_1, conv5_1):**

$$\mathcal{L}_{style} = \sum_{l \in \mathcal{S}} \|G(\Phi_l(y)) - G(\Phi_l(x_s))\|_2^2, \quad \mathcal{S} = \{1_1, 2_1, 3_1, 4_1, 5_1\}.$$

- **Total Variation (TV) — сглаживание и подавление “шумных” артефактов:**

$$\mathcal{L}_{TV}(y) = \sum_{i,j} \left(\|y_{i+1,j} - y_{i,j}\|_2^2 + \|y_{i,j+1} - y_{i,j}\|_2^2 \right).$$

Эксперименты и данные (в процессе)

- Контент: MS COCO (примерно 10 000 изображений).
- Стил: WikiArt (фрагменты; заявлено 50 направлений; 200 стиливых фрагментов).
- Случайные кропы: 128×128 и 256×256 .
- Обучение: Adam, $lr \cdot 10^{-4}$, batch 32, 80 эпох, GPU Tesla P100.
- Две фазы: совместное обучение (θ, ψ) ; затем заморозка G и дообучение e_ψ .



Контентное изображение



Стилевое изображение

Результаты: сравнение по ArtFID (↓)

Стиль	Concat	1×1 inj.	Add	FiLM	Johnson
Импрессионизм	44.6	46.8	41.6	38.6	34.9
Кубизм	48.3	50.4	45.5	42.3	39.7
Постимпрессионизм	42.8	45.1	39.9	36.8	33.5
Экспрессионизм	47.0	49.2	44.1	41.0	37.9
Сюрреализм	45.2	47.3	42.1	39.2	36.1
Абстракция	49.9	52.0	46.8	43.7	40.2
Среднее	46.3	48.5	43.3	40.3	37.0

Вывод: FiLM даёт ArtFID, близкий к Johnson (в среднем ≈ 3 пункта), сохраняя универсальность одной модели для множества стилей.

Примеры работы модели (FiLM)

