



Московский государственный университет имени М.В. Ломоносова

Факультет вычислительной математики и кибернетики

Кафедра интеллектуальных информационных технологий

Лаборатория компьютерной графики и мультимедиа

Гончаренко Дмитрий Александрович

Алгоритм изменения времени суток на изображении

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Научный руководитель:

К.С. Зипа

Москва, 2019

Аннотация

Алгоритм изменения времени суток на изображении относится к классу задач машинного обучения по *переносу изображений*¹. Данная сфера значительно продвинулась благодаря современным вычислительным возможностям, в частности переносу обучения на графические процессоры, GPU. За последние несколько лет появилось немало исследовательских работ на тему переноса изображений, стилей и колоризации. В данной работе рассматриваются современные подходы к переносу изображений на примере изменения времени суток на изображении. Проводится описание нейросетевых моделей и сравнительный анализ серии экспериментов обучения.

Abstract

The algorithm of changing the time of the day on images is a subclass of Machine Learning problems of image translation. This area has advanced significantly due to the modern computing capabilities, in particular the training transfer on GPUs. Over the past few years, many research papers have appeared on the subjects of images translation, styles transferring and colorization. This research reveals modern approaches of image translation on the example of changing the time of day on the image. A description of the neural network models and a comparative quality analysis of a series of training experiments are carried out.

¹ **Перенос изображений** (англ. *image translation*, или *image transferring*) – подвид технологии переноса обучения, позволяющий сохранять и объединять локальные признаки изображений.

Содержание

1	Введение	3
2	Постановка задачи	4
2.1	Цель работы	4
2.2	Решаемые задачи	4
2.3	Формальная постановка задачи	4
3	Обзор существующих методов	5
3.1	Генеративно состязательные сети (GAN)	5
3.2	Автокодировщики (AE)	5
3.3	Вариационные автокодировщики (VAE)	6
3.4	Методы использующие обучение с учителем	7
3.4.1	Перенос изображение и условные генеративные сети (сGAN)	7
3.5	Методы использующие обучение без учителя	8
4	Список литературы	9

1 Введение

Математические описания моделей машинного обучения появились еще в середине 20-го века. А первые попытки их практической реализации начались в конце 50х годов. Спустя полвека задачи машинного обучения все также не теряют своей актуальности. Рост числа работ, открытие исследовательских центров внутри компаний и при университетах показывает, что высокий интерес к данной сфере не ограничивается только наукой. За последние 10 лет появилось множество удобных инструментов для построения самых разнообразных моделей обучения, что оказало немалое влияние на развитие *науки о данных*¹.

Стоит отметить, что не последнюю роль здесь сыграли многократно выросшие вычислительные возможности. Это позволило обучать *нейронные сети*² на персональных компьютерах за доли секунды, для чего ранее могли требоваться дни, а то и недели на специализированных вычислительных устройствах.

Перенос обучения является одной из центральных исследовательских задач современного машинного обучения. Данная область направлена на получение некоторой информации об объекте, сохранение и последующее применение этих знаний к другому объекту, связанному с первым. *Перенос изображений*, в свою очередь, является подклассом переноса обучения применяемым на изображениях. В настоящее время переносу изображений уделяется значительное внимание в исследовательских работах [1], а область применения и методы могут сильно отличаться друг от друга. Так с помощью переноса изображений можно добиться объединения стилей двух изображений [2], колоризации³ черно-белых фотографий [3], объединение локальных признаков объектов и животных [4], увеличение разрешения изображений [6].

В данной исследовательской работе рассматриваются и систематизируются современные подходы переноса изображений, на примере изменения времени суток на изображениях. А также проводится обучение моделей на различных наборах данных и сравнительный анализ скорости и качества полученных результатов.

В качестве инструмента был выбран фреймворк PyTorch⁴, так как он разработан в виде библиотеки для языка Python – основного языка современного машинного обучения [12], и зарекомендовал себя мощным и гибким инструментом для исследования и обучения нейронных сетей.

¹ **Наука о данных** (англ. *data science*) – совокупность процессов и методов направленных на извлечение информации из исследуемых данных.

² **Нейронная сеть** – математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток живого организма. Является мощным современным инструментом машинного обучения. Главная особенность – способность обучаться на предоставленных данных, называемых тренировочными. Помимо исследовательских направлений нейронные сети также активно используются в коммерции, например обработка спама на электронной почте или система рекомендаций в интернет-магазинах.

³ **Колоризация** (англ. *colorization*) – преобразование монохромных изображений в цветные.

⁴ **PyTorch** – библиотека машинного обучения, основанная на Torch. Разработана исследовательской группой в Facebook.

2 Постановка задачи

2.1 Цель работы

В науке о данных одну из ключевых ролей играет непосредственный выбор данных, на которых будет обучаться сеть, и которые будут использоваться для валидации качества модели. Перенос изображений не является исключением. Цель работы –

2.2 Решаемые задачи

Задача исследовательской работы – получить модель, которая будет уметь переносить изображения между *доменами*¹ и способную с приемлемой точностью решить проблему изменения времени суток на изображениях. Проблему изменения времени суток на изображении можно разбить на следующие этапы:

- Исследование существующих методов переноса изображений и анализ решений родственных задач
- Поиск применимых к проблеме тренировочных данных – изображений с различными временем суток
- Классификация тренировочных данных и перенос изображений по доменам
- На основе проведенного исследования и обзора литературы, выбор модели, подходящей для решения поставленной задачи наилучшим образом
- Множественное обучение выбранной модели с различными конфигурациями на наборах данных с помощью библиотеки PyTorch
- Проведение серии экспериментов переноса изображений для получения оценки качества и скорости работы модели

2.3 Формальная постановка задачи

Формально алгоритм изменения времени суток на изображении можно сформулировать следующим образом:

Вход

На вход алгоритму поступают два изображения из различных доменов $x_1 \in X_1$ и $x_2 \in X_2$, где

- X_1 – домен с набором изображений первого типа
- X_2 – домен с набором изображений второго типа

Не ограничивая общности, пусть X_1 соедержит изображения дня, а X_2 соответствует ночи.

Выход

Необходимо построить *кроссдоменные*² отображения $f_{12}: X_1 \rightarrow X_2$ и $f_{21}: X_2 \rightarrow X_1$.

Реализовать функцию, принимающую на вход изображения из разных доменов x_1, x_2 и возвращающую перенесенные изображения x'_2, x'_1 соответственно:

$$x'_2, x'_1 = \text{CrossDomainTranslator}(x_1, x_2), \text{ где } x'_1 \in X_1, x'_2 \in X_2 \quad (1)$$

¹ **Домен** – область, множество, содержащее в себе объекты одного типа.

² **Кроссдоменный перенос** (англ. *cross-domain transfer*) – перенос объектов между доменами

3 Обзор существующих методов

Одними из наиболее популярных алгоритмов, на которых базируются решения многих задач переноса изображений, на сегодняшний день являются генеративно состязательные нейронные сети и вариационные автоэнкодеры.

3.1 Генеративно состязательные сети (GAN)

Генеративно состязательная нейронная сеть (GAN) – комбинация двух нейронных сетей G (генератор) и D (дискриминатор). Генератор подбирает *латентные*¹ параметры для генерации нового объекта. Дискриминатор пытается отличить оригинал от объекта, созданного генератором.

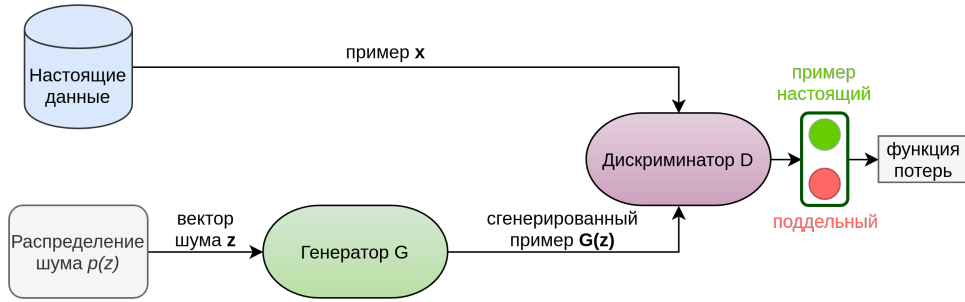


Рис. 1: Схема работы генеративно состязательной сети (GAN)

Формально генератор можно определить как отображение некоторого пространства скрытых параметров \mathcal{Z} , на котором задано априорное распределение $p_z(z)$, в пространство данных \mathcal{X} . Дискриминатор же будет производить отображение \mathcal{X} в отрезок $[0, 1]$ – вероятность того, что пример настоящий. На рис. 1 представлена общая схема работы генеративной сети.

$$\begin{aligned} G: \mathcal{Z} &\rightarrow \mathcal{X} \\ D: \mathcal{X} &\rightarrow [0, 1] \end{aligned} \quad (2)$$

Основной целью генеративно состязательной сети является получение генератором распределения данных p_{gen} , не отличимого дискриминатором от исходного распределения p_{data} . То есть, по сути, заключается в решении задачи оптимизации [11]:

$$\min_G \max_D \mathcal{L}_{GAN}(D, G), \text{ где} \quad (3)$$
$$\mathcal{L}_{GAN}(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

3.2 Автокодировщики (АЕ)

Автоэнкодер (АЕ) – комбинация двух нейронных сетей: энкодера E (кодировщика) и декодера D (декодировщика). Энкодер получает и преобразует данные в сжатый код скрытого пространства. Декодер же старается из этого кода восстановить объект наиболее близко

¹ **Латентный** здесь и далее, то же что и скрытый

к исходному. Математически можно представить автоэнкодер как отображения пространства входных данных \mathcal{X} в латентное пространство \mathcal{Z} и обратно:

$$\begin{aligned} E: \mathcal{X} &\rightarrow \mathcal{Z} \\ D: \mathcal{Z} &\rightarrow \mathcal{X} \\ x &\xrightarrow{E} z \xrightarrow{D} x', \text{ где} \end{aligned} \tag{4}$$

$x \in \mathcal{X}$ — исходное изображение
 $z \in \mathcal{Z}$ — скрытый код
 $x' \in \mathcal{X}$ — восстановленное изображение

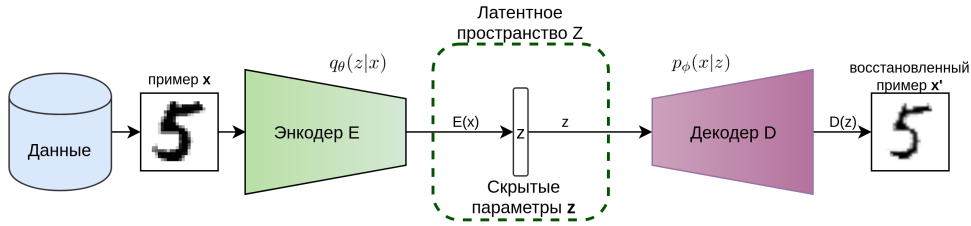


Рис. 2: Схема работы вариационных автоэнкодеров (VAE)

Задача автокодировщика состоит в минимизации разницы между исходным изображением x и восстановленным x' . Для этого вводится функцию потерь \mathcal{L}_{AE} , характеризующая потери при неправильном принятии решений:

$$\mathcal{L}_{AE}(x, x') = \|x - x'\|^2 \tag{5}$$

Основная проблема автоэнкодеров заключается в том, что скрытое пространство, может не быть непрерывным. Из-за чего не получается произвести интерполяцию, и как следствие, их область применения ограничивается. Эту проблему способны решить вариационные приближения.

3.3 Вариационные автокодировщики (VAE)

Вариационный автоэнкодер (VAE) отличается от автоэнкодера непрерывностью латентного пространства и предположениями накладываемыми на распределение данных. Пусть выбор тренировочных данных x находится в некоторой условной зависимости $p(x|z)$ от переменных скрытого пространства z . Кодировщик на вход получает пример x и выдает вектор размерности $2d$ распределения скрытых переменных z , состоящий из двух векторов размерности d : вектора стандартных отклонений σ и вектора средних значений μ . То есть энкодер представляет собой нейронную сеть с параметрами θ и обучается аппроксимировать апостериорное распределение $q_\theta(z|x)$. Декодировщик тоже является нейронной сетью с параметрами ϕ и старается из вектора размерности $2d$ скрытых переменных z получить объект из вариационного распределения над x : $p_\phi(x|z)$. На рисунке 2 представлена верхнеуровневая схема модели вариационных автоэнкодеров.

Задачей вариационных автоэнкодеров является задача оптимизации (6). Необходимо так подобрать параметры θ в $q_\theta(z|x)$, чтобы максимально точно приблизить $p(z|x)$:

$$\arg \min_{\theta} \mathbb{KL}(q_\theta(z|x) \| p(z|x)), \text{ где} \tag{6}$$

$$\mathbb{KL}(q_\theta(z|x) \| p(z|x)) = \mathbb{E}_q[\log q_\theta(z|x)] - \mathbb{E}_q[\log p(x, z)] + \log p(x)$$

\mathbb{KL} – расстояние Кульбака-Лейблера - мера удаленности двух распределений.

Так или иначе, методы реализующие перенос изображений опираются на сети GAN, VAE или их модификации. Рассмотрим следующие алгоритмы с позиции решения предложенной задачи изменения времени суток на изображениях. Будем рассматривать, уже введенные в формальном определении, домены изображений X_1 и X_2 , соответствующие разному времени суток (глава 2.3).

Пусть изображения $x_1, x_2 : x_1 \in X_1, x_2 \in X_2$

3.4 Методы использующие обучение с учителем

Обучение с учителем (англ. *supervised learning*) – способ машинного обучения, в ходе которого входные данные соотносятся с выходными до начала обучения. Обучение происходит на заготовленных парах объектов (x_1, x_2) , называемыми ”стимул-реакция”, которые находятся в доменах в некотором совместном распределении $P_{X_1, X_2}(x_1, x_2)$ ¹. Задача метода – построить функцию внутренней зависимости между примерами, которая затем будет отображать входные изображения желаемым образом.

Успешными примерами применения этого способа для решения задач переноса изображения можно считать эти работы [5, 9].

3.4.1 Перенос изображение и условные генеративные сети (сGAN)

Метод представленный на международной конференции компьютерного зрения (CVPR) в 2017 году исследовательской группой из университета Беркли под именем pix2pix [5] использует для переноса изображений обучение с учителем и условные генеративно состязательные нейронные сети.

Условные GAN являются модификацией над стандартными GAN. От обычных генеративных сетей (3), сGAN отличаются добавочным вектором информации y для генератора $G(z|y) = x'$ и дискриминатора $D(X|y)$. Вектор y может содержать любую уточняющую информацию, например он может содержать метки объектов на изображении, в случае с pix2pix, y – это добавочное входное изображение, видимое генератору и дискриминатору.

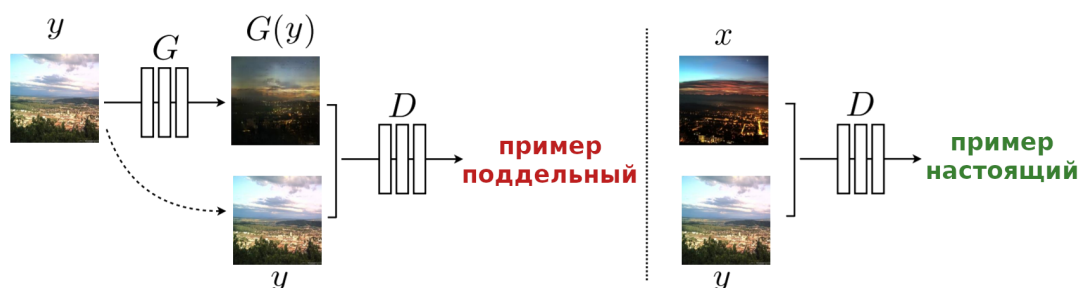


Рис. 3: Схема работы условно генеративно состязательной сети (сGAN)

x – настоящее изображение ночи, соответствующее y

Выше на рис. 3 приведена общая схема работы модели pix2pix. В качестве функции потерь была взята совокупность функций потерь: генеративной сети (3) в случае с сGAN и \mathcal{L}_1 . Сеть решает следующую задачу оптимизации:

$$\arg \min_G \max_D \mathcal{L}_{cGAN}(D, G) + \lambda \mathcal{L}_1(G), \text{ где} \quad (7)$$

¹ **Совместное распределение** – это распределение совместных исходов образованных из нескольких случайных величин.

$$\mathcal{L}_{cGAN}(D, G) = \mathbb{E}_x[\log D(x)] + \mathbb{E}_{y,z}[\log(1 - D(G(z, y)))], \quad \mathcal{L}_1(G) = \sum_{i=1}^n (x - G(z, y))^2$$

λ – параметр, контролирующий вклад \mathcal{L}_1 . При $\lambda = 0$ получаются более четкие изображения, но подверженные артефактам, при больших значениях параметра, будут получаться менее четкие, размытые изображения, но с визуально уменьшенным количеством артефактов [5]. Модель обучалась на датасете¹ [13] парных фотографий (с одинаковых ракурсов) сделанных с уличных вебкамер в разрешении 256x256px в разное время суток. После проведения серии экспериментов стало ясно, что из-за большого числа артефактов и размытости данную модель в своей задаче использовать не получится. Это происходит из-за того, что в датасете [13] используется всего 101 уникальная пара, а на данных, немного отличными от обучающей выборки, точность модели падает. Результаты работы метода на рис. 4.



Рис. 4: Схема работы условно генеративно состязательной сети (сGAN)
 x – настоящее изображение ночи, соответствующее y

Существенным и основным недостатком методов, использующих обучение с учителем, является необходимость хранить объекты доменов по парам. Это сказывается на сложности поиска и сбора тренировочных данных, для применения их на практике. С другой стороны, методы без учителя, лишены этого недостатка, и данные можно получать фактически "из воздуха".

3.5 Методы использующие обучение без учителя

Обучение без учителя (англ. *unsupervised learning*) – совокупность задач машинного обучения, решаемых на неразмеченных данных. В отличие от обучения с учителем, где тренировочные данные находятся в совместном распределении $P_{X_1, X_2}(x_1, x_2)$, в методах без вмешательства учителя требуется найти априорное распределение, выбирая данные (x_1, x_2) из доменов с частным распределением $P_{X_1}(x_1)$ и $P_{X_2}(x_2)$ ².

К наиболее интересным работам, использующим этот метод, можно отнести [4, 7, 8].

¹ Датасет (англ. *dataset*) – то же, что и набор данных

² Частное распределение – это распределение вероятности компонент некоторого множества, без зависимости между компонентами.

4 Список литературы

- [1] **Supervised and Unsupervised Image Translation** – A collection of image to image papers with code, 2019.
- [2] **J.Johnson, A.Alahi, L.Fei-Fei.** *Perceptual Losses for Real-Time Style Transfer and Super-Resolution.* Department of Computer Science, Stanford University, March 2016.
- [3] **R.Zhang, J.-Y.Zhu, P.Isola, X.Geng, A.S.Lin, T.Yu, A.A.Efros.** *Real-Time User-Guided Image Colorization with Learned Deep Priors.* University of California, Berkeley, May 2017.
- [4] **J.-Y.Zhu, T.Park, P.Isola, A.A.Efros.** *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.* Berkeley AI Research (BAIR) laboratory, UC Berkeley, March 2017.
- [5] **P.Isola, J.-Y.Zhu, T.Zhou, A.A.Efros.** *Image-to-image translation with conditional adversarial networks.* Berkeley AI Research (BAIR) laboratory, UC Berkeley, November 2016.
- [6] **Y.Yuan, S.Liu, J.Zhang, Y.Zhang, C.Dong, L.Lin** *Unsupervised Image Super-Resolution using Cycle-in-Cycle Generative Adversarial Networks* Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Graduate School at Shenzhen, Department of Automation, Tsinghua University, September 2018.
- [7] **M.-Y.Liu, T.Breuel, J.Kautz.** *Unsupervised Image-to-Image Translation Networks.* NVIDIA Corporation, March 2017.
- [8] **X.Huang, M.-Y.Liu, S.Belongie, J.Kautz.** *Multimodal Unsupervised Image-to-Image Translation.* Cornell University, NVIDIA Corporation, April 2018.
- [9] **J.-Y.Zhu, R.Zhang, D.Pathak, A.A.Efros.** *Toward Multimodal Image-to-Image Translation.* Berkeley AI Research, UC Berkeley, Adobe Research, November 2017.
- [10] **L.Ma, X.Jia, S.Georgoulis, T.Tuytelaars, L.V.Gool.** *Exemplar Guided Unsupervised Image-to-Image Translation.* Berkeley AI Research, UC Berkeley, Adobe Research, March 2019.
- [11] **С.Николенко, А.Кадурын, Е.Архангельская.** *Глубокое обучение.* СПб.: Питер, 2018.
- [12] **The State of the Octoverse: machine learning** – GitHub: Report, January 2019.
- [13] **Dateset: Paired WebCam Night2Day** – UC Berkley, pix2pix and BycycleGAN projects