УДК 004.9+004.8+004.93+003.24

АВТОМАТИЧНИЙ СИНТЕЗ ТАКТИЛЬНОЇ ГРАФІКИ ЗА ТЕКСТОВОЮ ПІДКАЗКОЮ

Є. А. Джуринський, В. З. Маїк

Українська академія друкарства,
вул. Під Голоском, 19, Львів, 79020, Україна

Проблемним виробничим процесом галузі видавництва інклюзивної продукції залишається друк тактильної графіки, оскільки він супроводжується низкою труднощів: часові та фінансові витрати, відсутність чіткого регламенту, неоднорідність цільової аудиторії читачів, тощо. У статті запропонована інформаційна концепція, яка полягає у тому, щоб використати метод синтезу тактильної графіки за текстовою підказкою у прикладній галузі інклюзивної ілюстрації. Інформаційну модель можна представити як функцію відображення множини тексту у множину тактильної графіки, а завдання моделювання такого відображення є предметом нашого дослідження. Розглянуто та формалізовано поетапний процес моделювання алгоритму вирішення поставленого завдання. Запропонована методика утворюється з наступних етапів: токенізація текстового вмісту (оптимізація представлення), моделювання мови, токенізація вмісту зображення (контекстне представлення), моделювання перетворення послідовностей (тобто seq2seq) текстових токенів у послідовності токенів зображення. Кожен з етапів супроводжується інформацією про результати навчання та оцінювання розроблених моделей. Наведено інформаційну довідку про розроблене програмне забезпечення, що використовувалося під час навчання моделей. Зроблено висновок про успішність та перспективність отриманих результатів дослідження та наведено приклади синтезованих тактильних зображень за текстовою підказкою.

Ключові слова: *інформаційна технологія, штучний інтелект, текстова підказка, модель, критерії оцінювання моделі, методика токенізації, вимоги до ілюстрації, обробка зображень, тактильна графіка, інклюзивна ілюстрація, інклюзивна література, шрифт Брайля.*

Постановка проблеми. Розроблена тактильна графіка у галузі виробництва інклюзивних видань має відповідати визначеним вимогам, що регламентують виконання такої графіки. Тому видавництво потребує від виконавця володіння специфічними навичками створення інклюзивних ілюстрацій, які визначаються як технічними вимогами, так і вимогами до якості інклюзивної ілюстрації. Існуюча проблематика змушує підприємство інклюзивних видань витрачати додаткові виробничі ресурси, збільшуючи собівартість друкованої продукції.

Із розвитком інформаційних технологій, зокрема галузі глибокого машинного навчання, вирішення вищезазначених проблем стало можливим. Чималого розвитку

останнім часом набувають засоби штучного інтелекту [1–3], які дають змогу синтезувати зображення, спираючись на текстову підказку користувача. Підготовка тактильної ілюстрації в галузі інклюзивної літератури передбачає чималу кількість перешкод, що передусім пов'язані із виконанням тактильної ілюстрації [4]. Ці перешкоди можуть бути вирішені за допомогою інформаційних систем синтезу тактильної графіки. Такі системи частково або повністю замінюють ілюстратора тактильної графіки, пошук якого потребує значних часових та фінансових витрат.

Аналіз останніх досліджень та публікацій. Розглянемо наявні методи, які так чи інакше намагаються синтезувати тактильне зображення. Деякі методи прагнуть перетворити фотографію на тактильне представлення, деякі методи не ставлять за мету синтезувати саме тактильне зображення, проте сфери їхнього застосування є дотичними й можуть бути взяті за основу.

Розглянемо перший метод [5–7], що ставить за мету перетворити фотографію на тактильну графіку. Головна ідея методу авторів полягає у застосуванні таких функцій-фільтрів на вихідному зображенні, як операція перетворення кольорового зображення у чорно-біле, операція пошуку країв та метод найближчих k сусідів. Спираючись на отримані авторами результати, можна зазначити, що таке рішення є не зовсім вдалим[1], оскільки сам метод дуже прямолінійний і теоретично працює лише для обмеженої кількості фотографій. Крім того, на нашу думку, зображення, що надається у прикладі цієї роботи, важко назвати вдалим інклюзивним зображенням.

У наступному дослідженні [8] автори поставили за мету передати інформацію про зміст фотографії читачу із вадами зору за допомогою тактильної графіки. З прикладів можна побачити, що запропоноване авторами рішення дійсно добре відображає зміст фотографії, перетворюючи її у тактильну графіку. Крім того, варто відзначити, що правила адаптації інклюзивних ілюстрацій на поданих прикладах не порушуються. Однак, на нашу думку, «вузьким місцем» цього методу є «бібліотека» тактильного зображення, оскільки передбачається, що розширюватися вона буде вручну, що виключає синтез нового зображення (тобто такого, що не належало множині елементів, представлених «бібліотекою»). Також варто відзначити відсутність різноманітності отриманих елементів тактильного зображення: тактильне зображення людей повторюється в усіх сценах, незважаючи на те, що вихідні фотографії є різними. На нашу думку, відсутність різноманітності негативно вплине на зацікавленість читача із вадами зору досліджувати таке зображення. Попри перераховані недоліки, варто зазначити, що такий метод передає зміст вихідної фотографії на високому рівні.

Мета статті – розробити модель синтезу тактильної графіки за текстовою підказкою, яка дозволить пересічному користувачу без навичок у сфері образотворчого мистецтва створювати тактильні ілюстрації, що у відносно легкий спосіб можуть бути вміщені в інклюзивне видання.

Виклад основного матеріалу дослідження. Дослідження складається з декількох етапів, які послідовно моделюють проміжні представлення інформації,

[1] Принаймні з огляду на сучасний стан речей, оскільки робота опублікована у 1997 році.

спрямовані на розв'язання поставленого завдання. Запропоноване рішення являє собою каскадну модель, де перетворення інформації відбувається «згори – вниз». Крім того, оскільки не існує прямого формального відображення текстової інформації у візуальну, машинне навчання передбачає, що моделювання відбувається у невизначений безпосередньо спосіб (за принципом «чорної коробки») із використанням керованого навчання (або навчання з учителем). Як наслідок, чинниками, на які можливо впливати під час моделювання, є вихідна інформація (текст), бажане значення виходу (інклюзивне зображення) та проміжні функціональні відображення, що у той чи інший спосіб перетворюють проміжні представлення інформації.

Визначене завдання першого етапу дослідження – оптимізації текстового вмісту – вирішується методом імовірнісної токенізації, що полягає у стисненні вихідного вмісту в оптимальний з точки зору природної мови спосіб. Як модель імовірнісної токенізації обрано модель BPE [9–10] (або Byte Pair Encoding).

Згідно з [11], необхідно знайти таке загальне розщеплення корпусу Ω із потужністю $\leq N_{\max,t}$, у якого коефіцієнт стискання буде максимальним:

$$V = Arg \max_{\substack{\Omega \in \Omega^* \\ |\Omega| \leq N_{\max,t}}} \mu_C(\Omega),$$

де V – загальне розщеплення корпусу із найбільшим коефіцієнтом стиснення (або словник), $\Omega \in \Omega^*$ – загальне розщеплення корпусу, $N_{max,t}$ – константа, що визначає обмеження у максимальній кількості токенів, з яких складається словник.

Вихідний набір даних, що використовувався під час навчання та подальшого оцінювання отриманої моделі, наведений у таблиці 1, яка містить зведену інформацію про корпуси C_{train} та C_{eval}.

Таблиця 1

Зведена інформація про корпуси

Корпус	Позначення	Кількість речень	Кількість слів	Кількість символів
BrUK [12][2]	C_{train}	37807	132432	3817875
Ukr.fiction.15k [13]	C_{eval}	15000	151595	887809

Модель налаштовувалася наступними параметрами:

- $N_{max,t} = 8192$ – максимальна кількість токенів (зчеплень) у словнику;
- $Seq_{max,t} = 64$ – максимальна кількість токенів (зчеплень) у результаті роботи одного запиту. У разі перебільшення цієї межі речення урізується.

Як критерії оцінювання моделі були обрані метрики $\delta_C(V)$ – коефіцієнт стиснення та $OOVR_C(V)$ (англ. out of vocabulary ratio) – відношення кількості токенів, що не належать словнику, та загальної кількості токенів. Зведені результати оцінювання моделі BPE на корпусі навчання та корпусі оцінювання, за зазначеними вище оцінками, наведені у таблиці 2.

[2] У цьому дослідженні використовуються приклади, перевірені авторами збірки (категорія «good»).

Таблиця 2

Результати оцінювання моделі оптимізації текстового вмісту

Кількість вихідних токенів	Кількість зчеплення	Кількість OOV	$\delta_{C_{eval}}(V)$	$OOVR_{C_{eval}}(V)$
3817875	1188342	13887	3.21	0.00011

У нашій роботі моделювання мови забезпечується керованим навчанням на основі вирішення завдання передбачення замаскованих токенів за допомогою великої моделі мови [14–16]. Процес навчання полягає у надаванні на вхід трансформера [17] послідовності в оптимальному представленні, токени якої із певною імовірністю можуть бути замасковані (тобто замінені на службовий токен). Завдання трансформера в цьому моделюванні – передбачити замаскований токен.

Гіперпараметри моделі, що використовувалися під час процесу навчання моделі мови: розмір пакета (англ. batch size) – 32, коефіцієнт швидкості навчання (англ. learning rate) – 0.0001, коефіцієнт розпаду ваги (англ. weight decay) – 0.001, імовірність маскування токена (англ. mlm probability) – 0.15, алгоритм оптимізації – AdamW [18].

Модель мови конфігурується параметрами, наведеними у таблиці 3.

Таблиця 3

Параметри моделі мови

Параметр	Складова	
	кодер	декодер
Розмір словника	8192	8192
Розмір послідовності	64	64
Кількість шарів	3	3
Розмірність шарів	512	512
Розмірність *FFN*	1024	1024
Кількість голівок уваги *MultiHead*	8	8

Оцінювання моделі здійснюється за допомогою обчислення оцінок моделювання мови, що відображають здатність моделі передбачати токени з огляду на контекст. До таких оцінок належать *перехресна ентропія* (англ. cross-entropy) та *заплутаність* (англ. perplexity), значення яких наводиться у таблиці 4.

Далі йде етап моделювання контекстного представлення зображення тактильної графіки, яке забезпечується токенізацією вихідного зображення за допомогою моделі VQ-VAE [19]. VQ-VAE розглядається як VAE [20], в якому відособлена логарифмічна правдоподібність log$p(x)$ обмежується нижньою межею відособленої ймовірності. З огляду на те, що розподіл $q\left(z=k\middle|x\right)$ є детермінованим, і апріорний розподіл z визначений як рівномірний, дивергенція Кульбака-Лейблера розподілу наближеного розподілу розпізнавання $q_{\varphi}\left(z\middle|x_i\right)$ від істинного апостеріорного розподілу $p_{\theta}\left(z\middle|x_i\right)$ є константною (тобто не підлягає оптимізації), а тому участі у навчанні не бере.

Таблиця 4

Результати оцінювання моделі мови

Корпус		Перехресна ентропія, H	Заплутаність, PPL
назва	позначення		
BrUK	C_{train}	2.114	8.282
Ukr.fiction.15k	C_{eval}	5.709	301.662

Моделювання процесу синтезу зображення методом варіаційного висновку має на увазі навчання кодера, декодера та кодової таблиці, що утворюють модель, відтворювати (або реконструювати) вихідне зображення, зберігаючи можливість відтворення нових зразків.

У цьому досліджені як набір вихідних зображень, відтворенню яких буде вчитися модель, була обрана колекція зображень рослин та тварин, що зберігаються у бібліотеці тактильної графіки APH (American Printing House) [21]. Крім того, навчальний набір даних був розширений власними зразками тактильного зображення у кількості 41-го, збільшуючи загальну кількість зразків до 179-ти. До того ж з наявного набору даних було виокремлено 10 % зображень, що використовуються при оцінюванні моделі, дозволяючи таким чином робити висновки про ефективність моделі.

Під час моделювання процесу синтезу інклюзивного зображення було з'ясовано, що при застосуванні попередньої обробки вихідних зразків зображення ефективність моделі збільшується. Запропонована попередня обробка полягає у квантуванні вихідного значення пікселів зображення (тобто їх кольорів) до одного із значень, що належать попередньо визначеній та обмеженій у розмірі «палітрі». Операція квантування кольору пікселів формально записується таким чином:

$$x = \arg\min_{p \in P} \sqrt{\sum_{i=1}^{n} \left(x_i - p_i\right)^2},$$

де x – вихідне значення кольору пікселя у вигляді вектора розмірності від 1 до 4, P – наперед визначена палітра кольорів, одного із значень кольору якої набуває вислідний піксель, n – розмір вектора кольору пікселя, x – вислідне, або квантоване значення кольору вихідного пікселя.

Модель синтезу зображення, що згадувалася у попередньому розділі, конфігурується наступними параметрами: розмір зображення – 256×256×1, розмір кодової таблиці (латентний простір) – 512, розмір вбудованих векторів латентного простору – 16, кількість прихованих шарів – 5, початкова та кінцева розмірність вершин прихованих шарів – 16.

Наведемо гіперпараметри, що використовувалися під час процесу навчання моделі: розмір пакета – 4, коефіцієнт швидкості навчання – 0.001, коефіцієнт розпаду ваги – 0.001, вартість фіксації (англ. commitment cost) – 0.25, алгоритм оптимізації – AdamW.

Оцінювання моделі, що відображає здатність моделі реконструювати вихідне зображення та синтезувати нове, здійснюється за допомогою обчислення

середньоквадратичної похибки як для простору зображення високої розмірності, так і для латентного простору малої розмірності та дистанції початку Фреше [22] (англ. FID, Frechet inception distance), значення яких наводяться у таблиці 5.

Таблиця 5

Результати оцінювання моделі синтезу зображення

Модель	СКП[3] реконструкції		СКП латентного простору		FID	
	C_{train}	C_{eval}	C_{train}	C_{eval}	C_{train}	C_{eval}
VQ-VAE	0.0134	0.0155	0.0083	0.009	0.6475	0.4663
VQ-VAE + попередня обробка зразків	0.0004	0.0144	0.0007	0.0058	0.1196	0.242

Завершальний етап – моделювання синтезу зображення, обумовленого текстовою підказкою, – забезпечується навчанням вже наперед навчених моделей мови та синтезу зображення здійснювати варіаційний синтез зображення, обумовленого текстовою підказкою, шляхом перетворення послідовностей (тобто seq2seq).

Параметри моделі синтезу зображення, обумовленого текстовою підказкою, що застосовувалися під час навчання, наведені у таблиці 6.

Таблиця 6

Параметри моделі синтезу зображення, обумовленого текстовою підказкою

Параметр	Складова	
	кодер	декодер
Розмір словника	8192	513
Розмір послідовності	64	65
Кількість шарів	3	3
Розмірність шарів	512	512
Розмірність	1024	1024
Кількість голівок уваги	8	8

Гіперпараметри, що використовувалися під час процесу навчання моделі синтезу зображення, обумовленого текстовою підказкою: розмір пакета – 2, коефіцієнт швидкості навчання – 0.0001, коефіцієнт розпаду ваги – 0.001, алгоритм оптимізації – AdamW.

Наприкінці основної частини роботи наведемо відомості про програмне забезпечення, що було розроблене і застосовувалося під час моделювання завдання синтезу тактильної графіки за текстовою підказкою. Програма написана мовою

[3] СКП – середньоквадратична похибка (англ. MSE).

Python3 із використанням бібліотек PyTorch (фреймворку для глибокого навчання), NumPy, Pillow, Transformers, Tokenizers. Було розроблено власний цикл машинного навчання, який застосовувався для кожного окремого етапу дослідження:

```
    # Початок циклу навчання, який триває відведену кількість епох
    for epoch in range(cfg.num_epochs):
        # Ітеративне завантаження пакетів навчального набору даних
        for batch, data in enumerate(dataloader):
            # Обчислення номеру поточного кроку
            step = batch + epoch * len(dataloader)
            # Переведення моделі в режим навчання
            model.train()
            # Обнулення значення градієнту
            optimizer.zero_grad()
            # Надання вихідних даних моделі та отримання результату
            output = cfg.forward(model, data)
            # Обчислення функції втрат
            losses = loss(output, data)
            # Зворотне поширення
            losses[«loss»].backward()
            # Ітерування оптимізатора та планувальника
            optimizer.step()
            scheduler_lr.step()
            # Фіксування статистики
            stats.update(«train», step, epoch, {**losses})
            # Логування стану програми кожні X кроків
            if step > 0 and step % cfg.log_every == 0:
                stats.write_tensorboard(summary_writer, «train»,
flush=True)
            # Валідація моделі кожні X кроків
            if step > 0 and step % cfg.val_every == 0:
                # Переведення моделі в режим оцінювання
                model.eval()
                with torch.no_grad():
                    # Візуалізація прогресу навчання
                    cfg.visualize(model, stats)
            # Збереження моделі кожні X кроків (checkpoint)
            if step > 0 and step % cfg.ckpt_every == 0:
                cfg.save(model)
```

Цикл навчання моделі

Програмний продукт буде використовуватися для наступних досліджень, оскільки він розроблений у такий спосіб, що дозволяє заміняти, розширювати або додавати моделі на кожному з етапів моделювання завдання.

Результати експериментів, що були поставлені на навченій моделі синтезу тактильної графіки за текстовою підказкою, у вигляді синтезованого зображення та за вихідною текстовою підказкою наводяться на рис. 1.

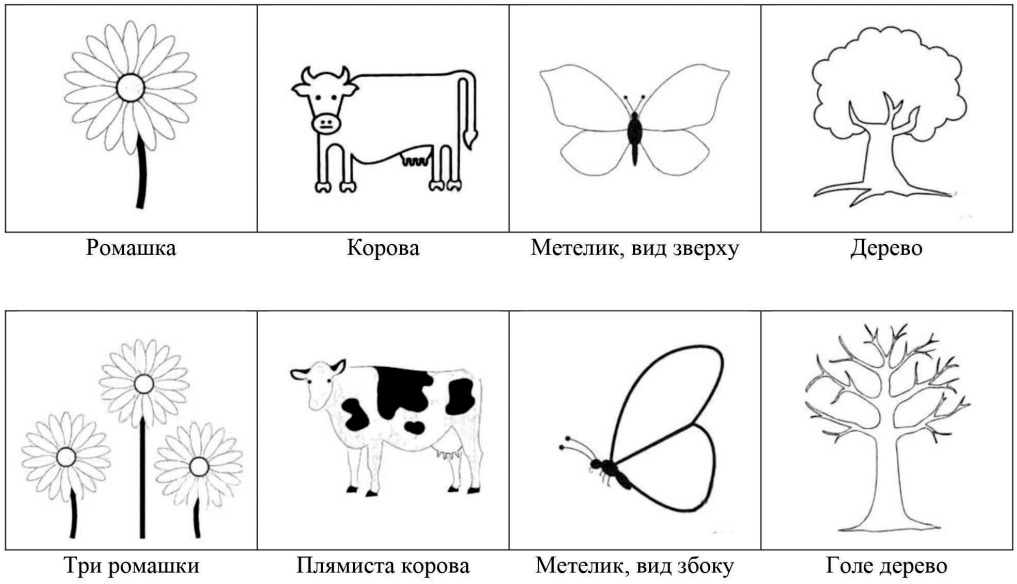

Рис. 1. Приклади синтезованих розробленою моделлю зображень тактильної графіки за текстовою підказкою

Висновки. Розглянуто процес навчання моделі оптимізації текстового вмісту на основі текстового корпусу BrUK, що утворюється із 37807 зразків речень українською мовою. У результаті отриманий словник оптимізованого представлення української мови складається з $|V| = 8192$ токенів, а максимальна можлива кількість токенів, що здатна обробити модель за один запит, становить $Seq_{max,t} = 64$. Крім того, наведено оцінки отриманої моделі, які відображають її ефективність, на основі текстового корпусу Ukr.fiction.15k — колекція текстів українською мовою, що складається з 15000 речень. За результатами оцінювання можна зробити висновок, що модель оптимізації текстового вмісту здатна ефективно виконувати перетворення текстових зразків, що не входили до множини навчального корпусу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Midjourney AI model tool for text-to-image conversion. URL: https://www.midjourney.com/ (access date: 04/05/2023).
2. Stable Diffusion AI model tool for text-to-image conversion. URL: https://stablediffusionweb.com/ (access date: 04/05/2023).
3. DALL·E 2 AI system that can create realistic images and art from a description in natural language. URL: https://openai.com/product/dall-e-2/ (access date: 04/05/2023).
4. Джуринський Є. А., Маїк В. З. Аналіз процесу підготовки ілюстрацій для інклюзивної літератури. Квалілогія книги. 2022. № 1 (41). С. 7−15.

5. Way T., Barner K. Automatic visual to tactile translation - Part I: Human factors, access methods, and image manipulation. *Rehabilitation Engineering, IEEE Transactions on.* 1997. 5. 81−94.
6. Way T., Barner K. Automatic visual to tactile translation. II. Evaluation of the TACTile image creation system. *Rehabilitation Engineering, IEEE Transactions on.* 1997. 5. 95−105.
7. Way T., Barner K. Towards Automatic Generation of Tactile Graphics. Applied Science and Engineering Laboratories. University of Delaware, 1999. 3 p.
8. Pakėnaitė K., Nedelev P., Kamperou E. Communicating Photograph Content Through Tactile Images to People With Visual Impairments. *Front. Comput. Sci.*, 10 January 2022 Sec. Computer Vision. № 3. Doi: https://doi.org/10.3389/fcomp.2021.787735.
9. Zouhar V., Meister C., Luis Gastaldi J., Du L., Vieira T., Sachan M., Cotterell R. A Formal Perspective on Byte-Pair Encoding. ETH Zürich. Johns Hopkins University, 2023. 15 p. Doi: https://doi.org/10.48550/arXiv.2306.16837.
10. Bostrom K., Durrett G. Byte Pair Encoding is Suboptimal for Language Model Pretraining. Department of Computer Science The University of Texas at Austin, 2020. 8 p. Doi: https://doi.org/10.48550/arXiv.2004.03720.
11. Гуляницький Л. Ф., Мулеса О. Ю. Методи комбінаторної оптимізації. Ужгород, 2015. С. 16−26.
12. Браунський корпус української мови. URL: https://github.com/brown-uk/corpus (access date: 01/09/2023).
13. Корпус художньої літератури українською мовою. URL: https://lang.org.ua/static/downloads/corpora/fiction.tokenized.shuffled.txt.bz2 (access date: 01/09/2023).
14. Douglas M. R. Large Language Models. CMSA, Harvard University. 2023. 47 p. Doi: https://doi.org/10.48550/arXiv.2307.05782.
15. Naveed H., Ullah Khan A., Qiu S., Saqib M., Anwar S. A Comprehensive Overview of Large Language Models. University of Engineering and Technology (UET), 2023. 46 p. Doi: https://doi.org/10.48550/arXiv.2307.06435.
16. Hadi Usman M., Al-Tashi. Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects. 2023. 44 p.
17. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., N. Gomez A., Kaiser L., Polosukhin I. Attention Is All You Need. Google Brain. Google Research. 2017. 15 p. Doi: https://doi.org/10.48550/arXiv.1706.03762.
18. Loshchilov I., Hutter F. Decoupled weight decay regularization. University of Freiburg. 2019. 19 p. Doi: https://doi.org/10.48550/arXiv.1711.05101.
19. Aaron van den Oord, Vinyals O., Kavukcuoglu K. Neural Discrete Representation Learning. DeepMind. 2017. 11 p. Doi: https://doi.org/10.48550/arXiv.1711.00937.
20. Kingma D. P., Welling M. Auto-Encoding Variational Bayes. Machine Learning Group Universiteit van Amsterdam. 2013. 14 p. Doi: https://doi.org/10.48550/arXiv.1312.6114.
21. The American Printing House Tactile Library. URL: https://imagelibrary.aph.org/portals/aphb/ (access date: 12/09/2023).
22. Yu Yu, Zhang Weibin, Deng Yun. Frechet Inception Distance (FID) for Evaluating GANs, 2021.

REFERENCES

1. Midjourney AI model tool for text-to-image conversion. Retrieved from https://www.midjourney.com/ (access date: 04/05/2023) (in English).
2. Stable Diffusion AI model tool for text-to-image conversion. Retrieved from https://stablediffusionweb.com/ (access date: 04/05/2023) (in English).
3. DALL·E 2 AI system that can create realistic images and art from a description in natural language. Retrieved from https://openai.com/product/dall-e-2/ (access date: 04/05/2023) (in English).
4. Dzhurynskyi, Ye. A., & Maik, V. Z. (2022). Analiz protsesu pidhotovky iliustratsii dlia inkliuzyvnoi literatury: Kvalilohiia knyhy, 1 (41), 7−15 (in Ukrainian).
5. Way, T., & Barner, K. (1997). Automatic visual to tactile translation - Part I: Human factors, access methods, and image manipulation: Rehabilitation Engineering, IEEE Transactions on, 5, 81−94 (in English).
6. Way, T., & Barner, K. (1997). Automatic visual to tactile translation. II. Evaluation of the TACTile image creation system: Rehabilitation Engineering, IEEE Transactions on, 5, 95−105 (in English).
7. Way, T., & Barner, K. (1999). Towards Automatic Generation of Tactile Graphics. Applied Science and Engineering Laboratories. University of Delaware (in English).
8. Pakėnaitė, K., Nedelev, P., & Kamperou, E. Communicating Photograph Content Through Tactile Images to People With Visual Impairments: Front. Comput. Sci., 10 January 2022 Sec. Computer Vision, 3. Doi: https://doi.org/10.3389/fcomp.2021.787735 (in English).
9. Zouhar, V., Meister, C., Gastaldi, Luis J., Du, L., Vieira, T., Sachan, M., & Cotterell, R. (2023). A Formal Perspective on Byte-Pair Encoding. ETH Zürich. Johns Hopkins University. Doi: https://doi.org/10.48550/arXiv.2306.16837 (in English).
10. Bostrom, K., & Durrett, G. (2020). Byte Pair Encoding is Suboptimal for Language Model Pretraining. Department of Computer Science The University of Texas at Austin. Doi: https://doi.org/10.48550/arXiv.2004.03720 (in English).
11. Hulianytskyi, L. F., & Mulesa, O. Yu. (2015). Metody kombinatornoi optymizatsii. Uzhhorod, 16−26 (in Ukrainian).
12. Braunskyi korpus ukrainskoi movy. Retrieved from https://github.com/brown-uk/corpus (access date: 01/09/2023) (in Ukrainian).
13. Korpus khudozhnoi literatury ukrainskoiu movoiu. Retrieved from https://lang.org.ua/static/downloads/corpora/fiction.tokenized.shuffled.txt.bz2 (access date: 01/09/2023) (in Ukrainian).
14. Douglas, M. R. (2023). Large Language Models. CMSA, Harvard University. Doi: https://doi.org/10.48550/arXiv.2307.05782 (in English).
15. Naveed, H., Ullah Khan, A., Qiu, S., Saqib, M., & Anwar, S. (2023). A Comprehensive Overview of Large Language Models. University of Engineering and Technology (UET). Doi: https://doi.org/10.48550/arXiv.2307.06435 (in English).
16. Hadi Usman, M., & Al-Tashi, Q. (2023). Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects (in English).
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., N. Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. Google Brain. Google Research. Doi: https://doi.org/10.48550/arXiv.1706.03762 (in English).

18. Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. University of Freiburg. Doi: https://doi.org/10.48550/arXiv.1711.05101 (in English).
19. Aaron van den Oord, Vinyals O., Kavukcuoglu K. Neural Discrete Representation Learning. DeepMind. 2017. Doi: https://doi.org/10.48550/arXiv.1711.00937 (in English).
20. Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. Machine Learning Group Universiteit van Amsterdam. Doi: https://doi.org/10.48550/arXiv.1312.6114 (in English).
21. The American Printing House Tactile Library. Retrieved from https://imagelibrary.aph.org/portals/aphb/ (access date: 12/09/2023) (in English).
22. Yu, Yu, Zhang, Weibin, & Deng, Yun. (2021). Frechet Inception Distance (FID) for Evaluating GANs (in English).

doi: 10.32403/0554-4866-2023-2-86-28-39

AUTOMATIC SYNTHESIS OF TACTILE GRAPHICS CONDITIONED BY TEXT PROMPT

Y. A. Dzhurynskyi, V. Z. Mayik

Ukrainian Academy of Printing,
19, Pid Holoskom St., Lviv, 79020, Ukraine
vol_maik@meta.ua

The problem of the development of tactile graphics in the field of inclusive literature publishing lies in the peculiarities of the execution of convex-tactile illustrations. Such a development process requires the performer to possess the basic skills of a fine art specialist and knowledge of the specifics of the technical performance of a tactile image, which are determined by a considerable number of requirements. In addition, the design of the illustration is complicated by additional factors, as they affect the final result of the developed tactile illustration. Such factors may include: the age of the target audience, the genre of the publication, the textual content that complements the illustration, etc. With the development of information technologies, in particular, the field of deep machine learning, solving the above problems has become possible. Recently, artificial intelligence tools [1, 2, 3], which allow synthesizing images based on the user's text prompt, have gained significant development. The proposed information concept is to use the method of synthesis of tactile graphics with a text prompt in the applied field of inclusive illustration. In this way, the information model can be represented as a function of mapping a set of text into a set of tactile graphics, and the emerging task of modelling such a mapping is the subject of research in this paper. The work considers and formalizes the step-by-step process of modelling the algorithm for solving the given problem. The proposed technique consists of the following stages: tokenization of text content (optimization of representation), language modelling, tokenization of image content (contextual representation), modelling of sequence conversion (i.e., seq2seq)

of text tokens into a sequence of image tokens. Each of the stages is accompanied by information about the results of training and evaluation of the developed models. At the end of the main part of the study, an informative note is given about the developed software that was used during model training. It is also noted that the developed software product will be used in subsequent studies related to the topic of this work. To sum up, a conclusion is made about the success and prospects of the obtained research results and examples of synthesized tactile images based on a text prompt are presented.

***Keywords**: information technology, artificial intelligence, text prompt, model, model evaluation criteria, tokenization technique, illustration requirements, image processing, tactile graphics, inclusive illustration, inclusive literature, Braille.*

Стаття надійшла до редакції 30.08.2023.
Received 30.08.2023.