ISSN 2181-9696

Doi Journal 10.26739/2181-9696

ТЕХНИКА ФАНЛАРИ

7 ЖИЛД, 1 СОН

ТЕХНИЧЕСКИЕ НАУКИ

ТОМ 7, НОМЕР 1

TECHNICAL SCIENCES

VOLUME 7, ISSUE 1

ТОШКЕНТ-2024

ТЕХНИКА ФАНЛАРИ

ТЕХНИЧЕСКИЕ НАУКИ | TECHNICAL SCIENCES

№1 (2024) DOI http://dx.doi.org/10.26739/2181-9696-2024-1

Контакт редакций журналов. www.tadqiqot.uz
ООО Tadqiqot город Ташкент,
улица Амира Темура пр.1, дом-2.
Web: http://www.tadqiqot.uz/; Email: info@tadqiqot.uz
Тел: (+998-94) 404-0000

Editorial staff of the journals of www.tadqiqot.uz
Tadqiqot LLC the city of Tashkent,
Amir Temur Street pr.1, House 2.
Web: http://www.tadqiqot.uz/; Email: info@tadqiqot.uz
Phone: (+998-94) 404-0000

МУНДАРИЖА | СОДЕРЖАНИЕ | CONTENT

ISSN: 2181-9696
www.tadqiqot.uz
JOURNAL OF TECHNICAL SCIENCE

Chulliyev Shokhrukh Ibadullayevich
(Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, chulliyevshohruh@gmail.com)
Ishniyazov Odil Olimovich
(Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, oishniyazov@gmail.com)

TEXT-TO-VIDEO SYNTHESIS: BRIDGING LANGUAGE AND VISUALS THROUGH ARTIFICIAL INTELLIGENCE

http://dx.doi.org/10.5281/zenodo.12103974

ABSTRACT

Text-to-video technology is an AI-driven process that converts text into video content. It employs natural language processing and multimedia tools to select visuals, generate audio, and assemble videos automatically. Despite its potential for content creation and accessibility, challenges persist in accurately interpreting complex text and maintaining creative quality. Anticipated advancements include enhanced personalization and real-time generation, fostering applications in education, marketing, and beyond.

Keywords: Text to video,artificial intelligence (AI),natural language processing (NLP),multimedia processing,content creation,automation,video generation,text-to-speech (TTS), computer vision, machine learning, content personalization, accessibility, communication, interactive video, real-time generation, ethical AI, automated content creation, algorithmic video production.

Chulliyev Shohruh Ibadullayevich
(Muhammad Al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti, chulliyevshohruh@gmail.com)
Ishniyazov Odil Olimovich
(Muhammad Al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti, oishniyazov@gmail.com)

MATNDAN VIDEOGA SINTEZ: SUN’IY INTELLEKT ORQALI TIL VA VIZUAL TASVIRLARNI KO‘PAYTIRISH

ANNOTATSIYA

Matnni videoga aylantirish texnologiyasi - bu matnni video kontentiga aylantiruvchi sun'iy intellektga asoslangan jarayon. U vizual tasvirlarni tanlash, audio yaratish va videolarni avtomatik ravishda yig'ish uchun tabiiy tilni qayta ishlash va multimedia vositalaridan foydalanadi. Kontent yaratish va foydalanish imkoniyatiga qaramay, murakkab matnni to'g'ri talqin qilish va ijodiy sifatni saqlab qolishda qiyinchiliklar saqlanib qolmoqda. Kutilayotgan yutuqlarga takomillashtirilgan shaxsiylashtirish va real vaqt rejimida ishlab chiqarish, ta'lim, marketing va boshqa sohalarda ilovalarni rivojlantirish kiradi.

Kalit so‘zlar: matnni videoga o’girish, sun'iy intellekt (AI),tabiiy tilni qayta ishlash (NLP),multimediali ishlov berish,kontent yaratish,avtomatlashtirish,video yaratish,matndan nutqqa o’girish (TTS), kompyuterni ko'rish, mashinani o'qitish, , real vaqtda interaktiv video yaratish, avtomatlashtirilgan kontent yaratish,

---

Чуллиев Шохрух Ибадуллаевич
(Ташкентский университет информационных технологий имени Мухаммада аль-Хорезми, chulliyevshohruh@gmail.com)
Ишниязов Одил Олимович
(Ташкентский университет информационных технологий имени Мухаммада аль-Хорезми, oishniyazov@gmail.com)

СИНТЕЗ ТЕКСТА В ВИДЕО: ОБЪЕДИНЕНИЕ ЯЗЫКА И ВИЗУАЛЬНЫХ ЭФФЕКТОВ С ПОМОЩЬЮ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

АБСТРАКТ

Технология преобразования текста в видео - это управляемый искусственным интеллектом процесс, который преобразует текст в видеоконтент. Она использует обработку естественного языка и мультимедийные инструменты для автоматического выбора визуальных элементов, генерации аудио и сборки видео. Несмотря на потенциал для создания контента и доступности, сохраняются проблемы с точной интерпретацией сложного текста и поддержанием креативного качества. Ожидаемые достижения включают улучшенную персонализацию и генерацию в режиме реального времени, что способствует применению в образовании, маркетинге и за его пределами.

Ключевые слова: преобразование текста в видео,искусственный интеллект (AI), обработка естественного языка (NLP), обработка мультимедиа, создание контента, автоматизация, генерация видео, преобразование текста в речь (TTS), компьютерное зрение, машинное обучение, персонализация контента, доступность, коммуникация, интерактивное видео, генерация в реальном времени, этичный искусственный интеллект, автоматизированное создание контента, алгоритмическое производство видео.

Text-to-video technology involves converting written text or scripts into video content using artificial intelligence (AI) and various multimedia tools. This technology utilizes natural language processing (NLP), machine learning, and computer vision algorithms to generate video content automatically from textual inputs. Here's an overview of how text-to-video technology typically works:

1. Text Processing: The process begins with the input of text, which can be in the form of a script, description, or narration. This text is parsed and analyzed using natural language processing algorithms to understand its context, structure, and intended meaning.

2. Content Selection:Based on the analyzed text, the system selects relevant images, videos, animations, or graphics that match the content of the text. It might also identify scenes, backgrounds, or visual elements suitable for the video.

3. Audio Generation:Text-to-video platforms often generate a voiceover or speech synthesis from the text using text-to-speech (TTS) technology. This creates a narration or dialogue for the video.

4. Video Compilation:The system assembles the chosen visual elements, images, videos, animations, and graphics to create a coherent video sequence that aligns with the narrative or script. Transitions, effects, and other visual enhancements may also be added to enhance the video's quality.

5. Rendering and Output: Once the video sequence is created, the system renders the final output, generating a video file that can be saved, shared, or further edited if necessary.

Text-to-video technology finds applications in various fields:

- Content Creation: It's used by content creators, marketers, and social media managers to quickly generate engaging videos for different platforms without needing extensive video production skills.

- E-Learning and Training: Text-to-video tools can convert educational content or training materials into video lectures or tutorials, enhancing the learning experience for students or employees.

- Automated Video Production:Enterprises may use this technology to automate video production workflows for generating personalized videos, product demos, or promotional content at scale.

- Accessibility: Text-to-video can aid in making content accessible by providing video versions of text-based information for individuals with disabilities.

However, as of my last update in January 2022, while text-to-video technology has made significant advancements, there might still be limitations in accurately interpreting and converting complex or nuanced textual content into high-quality videos. Additionally, the generated videos might lack the creativity and human touch that skilled video editors or content creators bring. The field continues to evolve, so there might be further advancements beyond this point.

Certainly! Text-to-video technology has undergone continual development, and since my last update in January 2022, there might have been advancements and trends in this field:

- AI Advancements:AI models, particularly in natural language processing and computer vision, have likely become more sophisticated. Enhanced language understanding and improved image/video recognition capabilities can lead to better text-to-video conversion accuracy and quality.
- Personalization and Customization:Text-to-video platforms might offer more customization options, allowing users to tailor videos with specific styles, branding, or templates to align with their preferences or brand identity.
- Real-Time Generation: Advancements might enable real-time or near-real-time generation of videos from text inputs. This could be particularly useful for live events, social media updates, or rapid content creation needs.
- Improved Voice Synthesis: Text-to-speech (TTS) technology may have improved, providing more natural-sounding and expressive voices, enhancing the overall quality of voiceovers in generated videos.
- Interactive Elements: Future developments might integrate interactive elements within generated videos, allowing viewers to engage with the content, make choices, or explore additional information within the video itself.
- Enhanced Video Editing Capabilities: While the process is largely automated, text-to-video platforms might offer more options for manual editing or fine-tuning of generated videos to meet specific requirements or creative preferences.
- AI Ethics and Bias Mitigation:As AI technology evolves, there's a growing focus on ensuring ethical use and mitigating biases in content generation. Efforts might include improved algorithms to reduce biases and maintain ethical standards in video creation.
- Integration with Other Technologies: Text-to-video tools could integrate with augmented reality (AR), virtual reality (VR), or mixed reality (MR) technologies, enabling the creation of more immersive and interactive video experiences.
- Broader Applications:Beyond marketing and e-learning, text-to-video technology might expand into various industries like healthcare, journalism, entertainment, and customer service, offering innovative ways to communicate information and engage audiences.

It's important to note that the pace of technological advancements can vary, and while these trends suggest potential directions, the actual developments in text-to-video technology may have evolved further or taken different paths after my last update.

Creating a full-fledged text-to-video program involves a complex combination of natural language processing, computer vision, and multimedia manipulation skills. Below is a simplified

Python-based example using the moviepy library to create a basic text-to-video program that generates a video from text:

Firstly, make sure you have moviepy installed. You can install it via pip:

```
pip install moviepy
```

Here's an example code snippet that takes a text input and generates a video with that text

```
from moviepy.editor import *
def text_to_video(text, output_filename='output_video.mp4'):
    # Generate a clip with the text
    txt_clip = TextClip(text, fontsize=70, color='white')
     # Set duration (5 seconds in this example)
    duration = 5  # Change the duration as needed
    # Set the duration of the clip
    txt_clip = txt_clip.set_duration(duration)
    # Set the size of the clip (width, height)
    txt_clip = txt_clip.set_pos('center')
      # Create a video file from the text clip
   video = CompositeVideoClip([txt_clip])
   # Set the resolution of the video (width, height)
   video = video.set_fullscreen()
   # Write the video file
   video.write_videofile(output_filename, codec='libx264', fps=24)
# Example usage:
text_input = "This is an example text for the video."
text_to_video(text_input)
```

This code creates a basic video using the moviepy library with the specified text input and saves it as 'output_video.mp4' in the current directory.

For a more advanced text-to-video program that involves more sophisticated processing (such as integrating images, background music, more complex text parsing, etc.), it would require deeper integration of various libraries and APIs for natural language processing, image processing, and video editing.

Creating a comprehensive text-to-video application would involve substantial development effort, including handling audio, selecting appropriate visuals, implementing a user interface, and more. This example provides a starting point for a simple text-to-video conversion using Python.

List of references:

1. "Text-to-Video: Generating Descriptive Video Content from Text"Authors: Haoran Wang, Yueh-Hua Wu, Ming-Ting Sun.Published in: 2019 IEEE International Conference on Multimedia and Expo (ICME)

2. "Text-to-Video Synthesis via Adversarial Cross-Modal Retrieval"Authors: Yitong Li, Chenliang Xu.Published in: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV)

3. "Generative Adversarial Text-to-Image Synthesis",Authors: Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee,Published in: 2016 Proceedings of The 33rd International Conference on Machine Learning

4. "Learning to Generate Images from Text Descriptions",Authors: Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee,Published in: 2016 arXiv preprint arXiv:1511.02793

5. "A Review on Text to Video Synthesis Techniques",Authors: Renuka Devi K., S. Kalaivani,Published in: 2020 International Conference on Communication and Electronics Systems (ICCES)

6. "Deep Text-to-Video Synthesis with Natural Language",Authors: Hsuan-I Ho, Wei-Chen Chiu, Yu-Sheng Chen, Chu-Song Chen,Published in: 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)

ISSN 2181-9696

Doi Journal 10.26739/2181-9696

ТЕХНИКА ФАНЛАРИ

7 ЖИЛД, 1 СОН

ТЕХНИЧЕСКИЕ НАУКИ

ТОМ 7, НОМЕР 1

TECHNICAL SCIENCES

VOLUME 7, ISSUE 1

Контакт редакций журналов. www.tadqiqot.uz
ООО Tadqiqot город Ташкент,
улица Амира Темура пр.1, дом-2.
Web: http://www.tadqiqot.uz/; Email: info@tadqiqot.uz
Тел: (+998-94) 404-0000

Editorial staff of the journals of www.tadqiqot.uz
Tadqiqot LLC The city of Tashkent,
Amir Temur Street pr.1, House 2.
Web: http://www.tadqiqot.uz/; Email: info@tadqiqot.uz
Phone: (+998-94) 404-0000