



Extract and explain contents clearly

Length limit reached. DeepSeek can only read the first 86%.



The document titled "**A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks**" provides an extensive overview of the applications of Transformer models across various domains in deep learning. Here's a clear and concise extraction and explanation of its contents:

1. Introduction

- **Transformers** are deep neural networks that use **self-attention mechanisms** to understand contextual relationships in sequential data.
- Unlike traditional models like **RNNs** and **LSTMs**, Transformers excel at handling long-range dependencies and enable parallel processing.
- Transformers have gained significant attention due to their success in **Natural Language Processing (NLP)** and their potential in other fields like **computer vision**, **audio processing**, **healthcare**, and **IoT**.
- The paper aims to fill the gap in existing surveys by providing a comprehensive analysis of Transformer models from 2017 to 2022, focusing on five key application domains: **NLP**, **Computer Vision**, **Multi-Modality**, **Audio & Speech Processing**, and **Signal Processing**.

2. Preliminaries

- **Transformer Architecture:** Introduced in 2017 by Vaswani et al., the Transformer model uses **self-attention** and **multi-head attention** to process sequential data.
 - **Attention Mechanism:** Focuses on important parts of the input sequence, allowing the model to capture long-range dependencies.
 - **Multi-head Attention:** Processes different parts of the input sequence in parallel, improving the model's ability to capture diverse features.
- The Transformer consists of **encoder** and **decoder** modules, which are stacked multiple times depending on the task.

3. Research Methodology

- The survey collects and analyzes recent research on Transformers from 2017 to 2022.
 - The authors identified the top five application domains and systematically searched for relevant papers using keywords like "Transformer," "attention," and "deep learning."
 - Over 600 Transformer models were reviewed, and a taxonomy was proposed to classify them based on their tasks and application domains.
-

4. Related Work

- Several existing surveys focus on specific aspects of Transformers, such as architecture, efficiency, or performance in specific fields like NLP or computer vision.
 - This survey differentiates itself by providing a comprehensive analysis of Transformer applications across multiple domains, including **multi-modality**, **audio & speech**, and **signal processing**.
-

5. Transformer Applications

- The survey identifies the major fields where Transformers are applied:
 - **NLP**: 40% of Transformer models.
 - **Computer Vision**: 31% (divided into **natural image processing** and **medical image processing**).
 - **Multi-Modality**: 15%.
 - **Audio & Speech Processing**: 11%.
 - **Signal Processing**: 4%.
 - The analysis highlights the proportion of attention each field has received and identifies areas for further research.
-

6. Application-Based Classification Taxonomy

- The authors propose a taxonomy to classify Transformer models based on their tasks within each application domain. The taxonomy is divided into five main areas:

6.1 Natural Language Processing (NLP)

- **Tasks**: Language translation, text classification, question answering, text summarization, text generation, natural language reasoning, and automated symbolic reasoning.
- **Key Models**:
 - **Transformer (2017)**: The original model for language translation.
 - **BERT**: Bidirectional Encoder Representations for tasks like question answering.

- **GPT**: Generative Pre-trained transformer for text generation.

- **T5**: Text-to-Text Transfer Transformer for multiple NLP tasks.

6.2 Computer Vision

- **Tasks**: Image classification, object detection, image segmentation, and image generation.

- **Key Models**:

- **ViT (Vision Transformer)**: Applies Transformers to image classification.

- **DETR**: Detection Transformer for object detection.

- **Swin Transformer**: Efficient Transformer for image segmentation.

6.3 Multi-Modality

- **Tasks**: Visual question answering, image-text generation, and multi-modal classification.

- **Key Models**:

- **CLIP**: Contrastive Language-Image Pre-training for image-text understanding.

- **DALL-E**: Generates images from text descriptions.

6.4 Audio & Speech Processing

- **Tasks**: Speech recognition, speech separation, and speech classification.

- **Key Models**:

- **Wav2Vec 2.0**: Self-supervised model for speech recognition.

- **Conformer**: Combines CNNs and Transformers for speech processing.

6.5 Signal Processing

- **Tasks**: Wireless signal processing and medical signal processing.

- **Key Models**:

- **SigT**: Transformer for wireless signal detection.

- **Three-tower Transformer**: For EEG signal classification in epilepsy detection.

7. Future Prospects and Challenges

- **Wireless Networks & Cloud Computing**: Transformers can improve tasks like interference detection, 5G/6G networks, and cloud workload forecasting.

- **Medical Image & Signal Processing**: Challenges include limited 3D image datasets and computational complexity in high-resolution imaging.

- **Reinforcement Learning**: Transformers can enhance sequential decision-making in RL tasks.

- **Other Prospects**: Transformers have potential in **protein modeling**, **bug fixing in software**, and **generative art**.

8. Conclusion

- Transformers have revolutionized deep learning by outperforming traditional models in handling sequential data.
- The survey provides a comprehensive taxonomy of Transformer applications across five key domains and highlights future research directions.
- Transformers are expected to continue expanding into new fields, shaping the future of AI and deep learning.

Key Takeaways

- **Transformers** are versatile models that have significantly impacted various domains, especially **NLP** and **computer vision**.
- The survey provides a structured taxonomy to understand Transformer applications and identifies areas for future research.
- Transformers are expected to play a crucial role in emerging fields like **wireless networks**, **healthcare**, and **reinforcement learning**.

This document serves as a valuable resource for researchers and practitioners looking to explore the potential of Transformers in diverse applications.

explain abstract clearly



The **abstract** of the paper "**A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks**" provides a concise summary of the key points and objectives of the study. Here's a clear explanation of its contents:

1. What is a Transformer?

- A **Transformer** is a type of **deep neural network** that uses a **self-attention mechanism** to understand the relationships between elements in sequential data (e.g., words in a sentence or pixels in an image).
- Unlike traditional models like **Recurrent Neural Networks (RNNs)** or **Long Short-Term Memory (LSTM)**, Transformers excel at capturing **long-range dependencies** and can process data in **parallel**, making them faster and more efficient.

2. Why are Transformers Important?

- Transformers have shown **remarkable success** in **Natural Language Processing (NLP)** tasks, such as language translation and text generation.
 - Beyond NLP, Transformers have been applied to a wide range of fields, including **computer vision**, **audio and speech processing**, **healthcare**, and the **Internet of Things (IoT)**.
 - Their ability to handle complex, sequential data has made them a **groundbreaking technology** in artificial intelligence.
-

3. What is the Goal of This Survey?

- While there are many surveys on Transformers, most focus on specific areas (e.g., NLP or computer vision) or compare model architectures and performance.
 - This survey aims to **fill the gap** by providing a **comprehensive overview** of Transformer applications across **multiple domains**.
 - The authors analyze Transformer models proposed between **2017 and 2022**, focusing on the **top five application areas**:
 1. **Natural Language Processing (NLP)**
 2. **Computer Vision**
 3. **Multi-Modality** (combining text, images, audio, etc.)
 4. **Audio and Speech Processing**
 5. **Signal Processing**
-

4. What Does the Survey Do?

- The survey **identifies** the most influential Transformer models in each domain and **classifies** them based on the tasks they perform.
 - A **taxonomy** is proposed to organize these models, making it easier to understand their contributions and applications.
 - The goal is to **shed light** on the potential of Transformers and inspire researchers to explore new applications and improve existing models.
-

5. Why is This Survey Valuable?

- It serves as a **reference source** for researchers who want to understand how Transformers are used in different fields.
 - By highlighting the **strengths** and **limitations** of Transformer models, the survey encourages further innovation and refinement of this technology.
 - It also discusses **future possibilities**, helping researchers identify new areas where Transformers can be applied.
-

Key Takeaways from the Abstract

- **Transformers** are powerful models that have transformed deep learning, especially in **NLP** and **computer vision**.
 - This survey provides a **comprehensive analysis** of Transformer applications across **five key domains**.
 - It aims to inspire researchers to explore the **full potential** of Transformers and develop new models for emerging challenges.
-

In summary, the abstract sets the stage for the survey by explaining what Transformers are, why they are important, and how the paper will provide a broad and detailed overview of their applications in deep learning.

explain introduction section clearly



The **Introduction** section of the paper "**A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks**" provides the background, motivation, and objectives of the study. Here's a clear and detailed explanation of its contents:

1. The Rise of Deep Neural Networks (DNNs)

- **Deep Neural Networks (DNNs)** have become the **dominant infrastructure** for solving machine learning tasks in artificial intelligence.
 - Different types of DNNs are used for specific tasks:
 - **Convolutional Neural Networks (CNNs)** for image processing.
 - **Recurrent Neural Networks (RNNs)** for sequential data, especially in **Natural Language Processing (NLP)**.
-

2. Limitations of Traditional Models

- **RNNs** and their variants (e.g., **LSTMs**) have limitations:
 - **Short-term dependencies:** They struggle to capture long-range relationships in data.
 - **Sequential processing:** They process data one step at a time, which is slow and hinders parallel computation.
 - **Vanishing/exploding gradients:** These issues make training difficult, especially for long sequences.
- While **LSTMs** improved memory and gradient issues, they still face challenges in **sequential**

processing and **context understanding**.

3. The Emergence of Transformers

- **Transformers** were introduced in 2017 to address the limitations of traditional sequence-to-sequence models.
 - Key features of Transformers:
 - **Self-attention mechanism:** Allows the model to focus on important parts of the input sequence, capturing long-range dependencies.
 - **Parallel processing:** Unlike RNNs, Transformers can process entire sequences simultaneously, making them faster and more efficient.
 - **Encoder-decoder architecture:** The model consists of stacked encoder and decoder layers, which process input and generate output, respectively.
-

4. Why Transformers are Revolutionary

- Transformers have achieved **state-of-the-art performance** in NLP tasks like **language translation**, **text summarization**, and **question answering**.
 - Their success has inspired researchers to apply Transformers to other domains, including:
 - **Computer Vision:** Transformers are used for tasks like image classification and object detection.
 - **Audio and Speech Processing:** Transformers improve speech recognition and audio analysis.
 - **Healthcare:** Transformers are applied to medical image analysis and signal processing.
 - **IoT:** Transformers are used for tasks like sensor data analysis and predictive maintenance.
-

5. The Need for This Survey

- While there are many surveys on Transformers, most focus on specific areas (e.g., NLP or computer vision) or compare model architectures and performance.
 - There is a **lack of a comprehensive survey** that covers the **major application domains** of Transformers and analyzes their impact on various tasks.
 - This survey aims to **fill this gap** by providing a **broad and systematic analysis** of Transformer models from 2017 to 2022.
-

6. Objectives of the Survey

- **Identify the top five application domains** for Transformers:
 1. **Natural Language Processing (NLP)**
 2. **Computer Vision**
 3. **Multi-Modality** (combining text, images, audio, etc.)

4. Audio and Speech Processing

5. Signal Processing

- **Analyze the impact** of influential Transformer models in these domains.
 - **Propose a taxonomy** to classify Transformer models based on their tasks and application areas.
 - **Highlight future possibilities** and challenges for Transformers, encouraging further research and innovation.
-

7. Structure of the Paper

The paper is organized as follows:

1. **Introduction:** Provides background and motivation for the survey.
 2. **Preliminaries:** Explains key concepts like the Transformer architecture and attention mechanisms.
 3. **Research Methodology:** Describes how the survey was conducted.
 4. **Related Work:** Compares this survey with existing surveys on Transformers.
 5. **Transformer Applications:** Analyzes Transformer models in the top five application domains.
 6. **Taxonomy:** Proposes a classification framework for Transformer models.
 7. **Future Prospects:** Discusses potential research directions and challenges.
 8. **Conclusion:** Summarizes the key findings and contributions of the survey.
-

Key Takeaways from the Introduction

- **Transformers** are a **breakthrough** in deep learning, overcoming the limitations of traditional models like RNNs and LSTMs.
 - They have been successfully applied to **NLP, computer vision, audio processing, healthcare, and IoT.**
 - This survey provides a **comprehensive analysis** of Transformer applications across multiple domains, aiming to inspire further research and innovation.
-

In summary, the **Introduction** sets the stage for the survey by explaining the importance of Transformers, their advantages over traditional models, and the need for a comprehensive analysis of their applications. It also outlines the structure and objectives of the paper.

explain preliminaries section clearly



The **Preliminaries** section of the paper "**A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks**" provides the foundational concepts necessary to understand the rest of the

survey. Here's a clear and detailed explanation of its contents:

1. Transformer Architecture

- The **Transformer** model was introduced in 2017 by Vaswani et al. for **machine translation** tasks.
 - It consists of two main components:
 - **Encoder**: Processes the input sequence and extracts features.
 - **Decoder**: Generates the output sequence based on the encoder's output.
 - The key innovation of the Transformer is the **self-attention mechanism**, which allows the model to focus on different parts of the input sequence to capture long-range dependencies.
-

2. Attention Mechanism

- The **attention mechanism** was first introduced in the 1990s and has since become a crucial component in deep learning.
- In the context of Transformers, attention allows the model to **focus on important parts** of the input sequence, improving its ability to understand context.
- The **scaled dot-product attention** is the specific type of attention used in Transformers. It involves three matrices:
 - **Query (Q)**: Represents the current word or element being processed.
 - **Key (K)**: Represents the words or elements in the sequence that the model attends to.
 - **Value (V)**: Contains the actual information from the sequence.
- The attention score is calculated using the formula:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

where d_k is the dimension of the key vectors.

3. Multi-head Attention

- **Multi-head attention** is an extension of the attention mechanism that allows the model to focus on **different parts** of the input sequence simultaneously.
 - Instead of computing a single attention score, the model computes multiple attention scores (or "heads") in parallel.
 - Each head learns to focus on different aspects of the input, and the results are concatenated and linearly transformed to produce the final output.
 - This approach improves the model's ability to capture diverse features and relationships in the data.
-

4. Encoder Module

- The **encoder** module of the Transformer consists of several layers, each containing two main components:
 - **Multi-head attention layer:** Computes attention scores for the input sequence.
 - **Feed-forward layer:** Processes the output of the attention layer.
 - The encoder also includes **residual connections** and **layer normalization** to stabilize training and improve performance.
-

5. Decoder Module

- The **decoder** module is similar to the encoder but includes an additional **masked multi-head attention layer**.
 - The masked attention layer ensures that the model only attends to **previous elements** in the sequence when generating the output, preventing it from "cheating" by looking at future elements.
 - The decoder also uses the encoder's output to generate the final output sequence.
-

6. Key Advantages of Transformers

- **Parallel processing:** Unlike RNNs, Transformers can process entire sequences simultaneously, making them faster and more efficient.
 - **Long-range dependencies:** The self-attention mechanism allows Transformers to capture relationships between distant elements in the sequence.
 - **Scalability:** Transformers can be scaled up to handle large datasets and complex tasks, making them suitable for a wide range of applications.
-

Key Takeaways from the Preliminaries

- The **Transformer** model is built on the **self-attention mechanism**, which allows it to focus on important parts of the input sequence.
 - **Multi-head attention** enables the model to capture diverse features and relationships in the data.
 - The **encoder** and **decoder** modules work together to process input sequences and generate output sequences.
 - Transformers are **faster, more efficient**, and **better at capturing long-range dependencies** than traditional models like RNNs and LSTMs.
-

In summary, the **Preliminaries** section provides the foundational knowledge needed to understand how Transformers work and why they are so effective. It explains the key components of the Transformer architecture, including the attention mechanism, multi-head attention, and the roles of the encoder and decoder modules.



The **Research Methodology** section of the paper "**A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks**" explains how the authors conducted their comprehensive survey. Here's a clear and detailed explanation of its contents:

1. Objective of the Survey

- The goal of the survey is to provide a **comprehensive analysis** of Transformer models and their applications across various domains.
- The authors aim to identify the **top five application domains** for Transformers and analyze the impact of influential models in these areas.

2. Data Collection Process

- The authors employed a **two-fold strategy** to gather relevant research papers:

1. Searching Established Databases:

- They used several established search engines and digital libraries (e.g., AAAI, ACM, CVPR, IEEE, etc.) to find papers related to Transformers.
- Keywords used for the search included: "Transformer," "attention," "self-attention," "deep learning," "artificial intelligence," and specific domain-related terms (e.g., "NLP," "computer vision").

2. Evaluating and Filtering Papers:

- The authors reviewed the collected papers and eliminated those that were **irrelevant** or did not focus on Transformer-based models.
- They selected papers that proposed **novel Transformer models** or significant improvements to existing models.

3. Identification of Application Domains

- Based on their search and expert discussions, the authors identified the **top five application domains** for Transformers:

1. Natural Language Processing (NLP)

2. Computer Vision

3. Multi-Modality (combining text, images, audio, etc.)

4. Audio and Speech Processing

□ Signal Processing

3. Signal Processing

- These domains were chosen because they represent the areas where Transformers have had the most significant impact.
-

4. Systematic Search for Transformer Models

- For each application domain, the authors performed a **systematic search** to find Transformer-based models.
 - They used specific keywords related to each domain (e.g., "image processing" for computer vision, "speech recognition" for audio processing) to ensure a focused search.
 - This process yielded a large number of papers, which were then reviewed to identify **relevant Transformer models**.
-

5. Selection Criteria for Models

- The authors selected Transformer models based on the following criteria:
 1. **Novelty:** Models that introduced new approaches or significantly improved existing methods.
 2. **Impact:** Models that had a high citation rate and were widely accepted by the research community.
 3. **Performance:** Models that achieved state-of-the-art results in their respective tasks.
 4. **Real-world Applications:** Models that were applied to practical, real-world problems.
 5. **Theoretical and Practical Significance:** Models that generated significant interest in the AI community.
-

6. Taxonomy Development

- After selecting the models, the authors developed a **taxonomy** to classify them based on their **application domains** and **tasks**.
 - The taxonomy organizes Transformer models into **five main categories** (NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing) and further divides them based on specific tasks (e.g., language translation, image classification, speech recognition).
-

7. Analysis of Selected Models

- The authors conducted an **in-depth analysis** of the selected Transformer models, focusing on:
 - **Architecture:** How the model is structured (e.g., encoder-decoder, multi-head attention).
 - **Pre-training:** Whether the model was pre-trained on large datasets.
 - **Datasets:** The datasets used for training, fine-tuning, and testing.
 - **Performance:** The model's performance on specific tasks and benchmarks.

- This analysis helps to highlight the **strengths** and **limitations** of each model and provides insights into their practical applications.
-

8. Future Prospects and Challenges

- The authors also discussed **future directions** and **challenges** for Transformer models in each application domain.
 - This includes potential areas for improvement, such as **scalability**, **efficiency**, and **generalization** to new tasks.
-

Key Takeaways from the Research Methodology

- The survey was conducted using a **systematic and comprehensive approach** to identify and analyze Transformer models across five key application domains.
 - The authors used **specific criteria** to select models that were novel, impactful, and relevant to real-world applications.
 - A **taxonomy** was developed to classify Transformer models based on their tasks and application areas.
 - The analysis provides valuable insights into the **strengths**, **limitations**, and **future potential** of Transformer models.
-

In summary, the **Research Methodology** section outlines the systematic process used by the authors to gather, evaluate, and analyze Transformer models. It explains how the survey was conducted, the criteria for selecting models, and the development of a taxonomy to organize the findings. This section provides a clear framework for understanding the rest of the survey.

explain Related work section clearly



The **Related Work** section of the paper "**A Comprehensive Survey on Applications of Transformers for Deep Learning Tasks**" provides a comparative analysis of existing surveys on Transformers and highlights the unique contributions of this survey. Here's a clear and detailed explanation of its contents:

1. Purpose of the Related Work Section

- The goal of this section is to **contextualize** the current survey within the broader landscape of existing literature on Transformers.
- The authors compare their survey with **17 significant surveys** on Transformers, focusing on their **scope**, **focus**, and **contributions**.

- This comparison helps to **identify gaps** in the existing literature and **justify the need** for this comprehensive survey.
-

2. Overview of Existing Surveys

- The authors reviewed over **50 survey papers** on Transformers and selected **17** for in-depth analysis.
 - These surveys primarily focus on:
 - **Architecture and Efficiency:** Analyzing the design and computational efficiency of Transformer models.
 - **Specific Fields:** Focusing on applications in **NLP** or **computer vision**.
 - **Performance Comparisons:** Comparing the performance of different Transformer models on specific tasks.
-

3. Key Findings from Existing Surveys

- **Architecture and Efficiency:**
 - Some surveys, like **Lin et al. (2022)**, provide a taxonomy of Transformer models based on architectural modifications, pre-training methods, and applications.
 - Others, like **Tay et al. (2023)**, compare the computational power and memory efficiency of different Transformer models.
 - **NLP:**
 - Surveys like **Brasoveanu & Andonie (2020)** focus on visualizing Transformer architectures for NLP tasks.
 - **Wang et al. (2020a)** explore the use of Transformers for text summarization.
 - **Computer Vision:**
 - **Khan et al. (2022)** provide a comprehensive overview of Transformer models in computer vision, classifying them based on popular recognition tasks.
 - **Han et al. (2023)** analyze the advantages and disadvantages of Vision Transformers (ViTs) and their performance on different datasets.
 - **Multi-Modality:**
 - **Ruan & Jin (2022)** categorize Transformer models for vision-language tasks and compare their performance.
 - **Shin et al. (2022)** explore Transformers in multi-modal visual-linguistic tasks.
-

4. Gaps in Existing Surveys

- **Limited Scope:** Most existing surveys focus on **specific fields** (e.g., NLP or computer vision) or **specific aspects** (e.g., architecture or performance).

• **Lack of Comprehensive Surveys:** There is a lack of surveys that provide a broad and systematic

- **Lack of Comprehensive Coverage.** There is a lack of surveys that provide a **broad and systematic analysis** of Transformer applications across **multiple domains**.
 - **Neglected Areas:** Some application areas, such as **signal processing** and **cloud computing**, are not adequately covered in existing surveys.
-

5. Unique Contributions of This Survey

- **Comprehensive Coverage:** This survey covers **five major application domains** (NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing), providing a **holistic view** of Transformer applications.
 - **Task-Based Taxonomy:** The authors propose a **taxonomy** to classify Transformer models based on their **tasks** within each application domain.
 - **In-Depth Analysis:** The survey provides a detailed analysis of **significant Transformer models**, including their **architecture, pre-training methods, datasets, and performance**.
 - **Future Prospects:** The survey discusses **future directions** and **challenges** for Transformer models, encouraging further research and innovation.
-

6. Comparison with Existing Surveys

- The authors provide a **comparative table** (Table 2) that summarizes the similarities and differences between this survey and the 17 selected surveys.
 - Key comparisons include:
 - **Fields of Application:** This survey covers a broader range of domains compared to existing surveys.
 - **Focus:** While existing surveys often focus on architecture or specific tasks, this survey provides a **task-based analysis** across multiple domains.
 - **Unique Aspects:** This survey includes **signal processing** and **cloud computing**, which are not covered in most existing surveys.
-

Key Takeaways from the Related Work Section

- Existing surveys on Transformers are **focused and specialized**, often limited to specific fields or aspects.
 - This survey **fills the gap** by providing a **comprehensive and systematic analysis** of Transformer applications across **multiple domains**.
 - The proposed **taxonomy** and **in-depth analysis** of significant models make this survey a **valuable resource** for researchers and practitioners.
-

In summary, the **Related Work** section highlights the **limitations** of existing surveys and **justifies the need for this comprehensive survey**. It provides a **comparative analysis** of existing literature and

needed for this comprehensive survey. It provides a **comparative analysis** of existing literature and outlines the **unique contributions** of this survey, setting the stage for the detailed analysis that follows.

Approach Fields of

Application Similarities Differences

Q Fournier

et al.

(Fournier

et al., 2021)

Performance

/Architecture

- A classification of the trans~~for~~mers is suggested, and this classification is based on attention mechanism modification or architecture modification
- This paper surveyed the different alternatives of the standard transformers that are more efficient in terms of time and memory complexities, and these alternatives are categorized by either modifying the attention mechanism or the network architecture. Their classification is based on the change in architecture or change in attention mechanism, while our classification is driven by application areas.

T. Lin et

al. (Lin

et al., 2022)

Performance

/Architecture

- Proposed taxonomy of X~~for~~mers covering several fields
- This existing survey compared X-formers from architectural modification, pre-training, and a very small range of application perspectives, while our survey deeply focuses on popular tasks under each field of application.
- The wireless/medical signal processing and cloud computing tasks application were missing in this exciting survey, while our survey c



The table you provided compares the **approach**, **fields of application**, **similarities**, and **differences** between the current survey and two existing surveys: **Fournier et al. (2021)** and **Lin et al. (2022)**. Here's a clear explanation of the comparison:

1. Comparison with Fournier et al. (2021)

Approach

- **Fournier et al. (2021)**: Focuses on **performance and architecture** of Transformers.
- **Current Survey**: Focuses on **application areas** and **tasks** of Transformers.

Fields of Application

- **Fournier et al. (2021)**: Primarily discusses **efficient alternatives** to standard Transformers.
- **Current Survey**: Covers **five major application domains** (NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing).

Similarities

- Both surveys propose a **classification** of Transformer models.
- Both discuss **modifications** to Transformer architectures.

Differences

- **Fournier et al. (2021)**:
 - Focuses on **efficiency** (time and memory complexity) of Transformer models.
 - Classifies models based on **architectural modifications** (e.g., changes to the attention mechanism or network architecture).
- **Current Survey**:
 - Focuses on **application areas** and **tasks** performed by Transformer models.
 - Classifies models based on their **application domains** (e.g., NLP, Computer Vision) rather than architectural changes.
 - Provides a broader and more comprehensive analysis of Transformer applications.

2. Comparison with Lin et al. (2022)

Approach

- **Lin et al. (2022)**: Focuses on **performance and architecture** of Transformers.
- **Current Survey**: Focuses on **application areas** and **tasks** of Transformers.

Fields of Application

- **Lin et al. (2022)**: Covers **several fields** but with a focus on architectural modifications and pre-training.
- **Current Survey**: Covers **five major application domains** (NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing).

Similarities

- Both surveys propose a **taxonomy** of Transformer models.
- Both discuss **architectural modifications** and **pre-training** methods.

Differences

- **Lin et al. (2022):**
 - Focuses on **architectural modifications**, **pre-training**, and a **small range of applications**.
 - Does not cover **wireless/medical signal processing** or **cloud computing** tasks.
- **Current Survey:**
 - Focuses on **popular tasks** within each application domain.
 - Includes **wireless/medical signal processing** and **cloud computing**, which are missing in Lin et al. (2022).
 - Provides a deeper and more comprehensive analysis of Transformer applications across multiple domains.

Key Takeaways from the Comparison

- **Fournier et al. (2021)** and **Lin et al. (2022)** focus on **architectural modifications** and **efficiency** of Transformer models.
- The **current survey** differentiates itself by focusing on **application areas** and **tasks**, providing a **broader and more comprehensive analysis** of Transformer models.
- The current survey includes **wireless/medical signal processing** and **cloud computing**, which are not covered in the existing surveys.

Summary Table

Approach	Fields of Application	Similarities	Differences
Fournier et al. (2021)	Performance / Architecture	<ul style="list-style-type: none">• Proposes a classification of Transformers based on architectural modifications.	<ul style="list-style-type: none">• Focuses on efficiency (time and memory complexity).
		<ul style="list-style-type: none">• Discusses modifications to attention mechanism and network architecture.	<ul style="list-style-type: none">• Classifies models based on architectural changes, not application areas.
Lin et al. (2022)	Performance /	<ul style="list-style-type: none">• Proposes a taxonomy of Transformers covering several	<ul style="list-style-type: none">• Focuses on architectural modifications and pre-training

		<ul style="list-style-type: none"> • Discusses architectural modifications and pre-training methods. 	<ul style="list-style-type: none"> • Does not cover wireless/medical signal processing or cloud computing tasks.
Current Survey	Application Areas / Tasks	<ul style="list-style-type: none"> • Proposes a taxonomy of Transformers based on application domains and tasks. 	<ul style="list-style-type: none"> • Focuses on application areas and tasks, not just architecture or efficiency.
		<ul style="list-style-type: none"> • Discusses architectural modifications and pre-training methods. 	<ul style="list-style-type: none"> • Includes wireless/medical signal processing and cloud computing tasks.

In summary, the **current survey** stands out by focusing on **application areas** and **tasks**, providing a **comprehensive analysis** of Transformer models across multiple domains, including areas like **signal processing** and **cloud computing** that are not covered in existing surveys.

Y. Tay et

al. (Tay

et al., 2023)

Performance

/Architectural

- Proposed a taxonomy considering the primary use case of transformer models in language and vision domains .

- This existing survey compared the computational power and memory efficiency of transformer models, whereas our survey focuses on deep learning tasks and applications.

- This exciting survey focused on language and vision domain only, while we cover other top five fields of transformer applications: NLP, computer vision, multi-modality, audio/speech, and signal processing.

A. M. P.

Brasoveanu

et al.

(Brasoveanu

& Andonie,

2020)

Natural

language

Processing NLP

- Explain transformer architecture and explain its features.
- Our survey describes the transformer model and the significant models' working processing for a range of tasks. However, this existing paper focused on visualization techniques used to explain the most recent transformer architectures and explored two large tool classes to explain the inner workings of Transformers.
- we covered five fields of transformer applications: NLP, computer vision, multi-modality, audio/speech, and signal processing and this exciting survey focused on the models for NLP only



The table you provided compares the **approach**, **fields of application**, **similarities**, and **differences** between the current survey and two existing surveys: **Tay et al. (2023)** and **Brasoveanu & Andonie (2020)**. Here's a clear explanation of the comparison:

1. Comparison with Tay et al. (2023)

Approach

- **Tay et al. (2023)**: Focuses on **performance and architecture** of Transformers.
- **Current Survey**: Focuses on **application areas** and **tasks** of Transformers.

Fields of Application

- **Tay et al. (2023)**: Primarily discusses **language and vision domains**.
- **Current Survey**: Covers **five major application domains** (NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing).

Similarities

- Both surveys propose a **taxonomy** of Transformer models.
- Both discuss the **primary use cases** of Transformer models.

Differences

- **Tay et al. (2023)**:
 - Focuses on **computational power** and **memory efficiency** of Transformer models.
 - Limited to **language and vision domains**.
- **Current Survey**:
 - Focuses on **deep learning tasks** and **applications** across multiple domains.

- Covers **NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing**, providing a broader scope.
-

2. Comparison with Brasoveanu & Andonie (2020)

Approach

- **Brasoveanu & Andonie (2020)**: Focuses on **visualizing Transformer architectures** for NLP tasks.
- **Current Survey**: Focuses on **application areas** and **tasks** of Transformers.

Fields of Application

- **Brasoveanu & Andonie (2020)**: Primarily discusses **NLP**.
- **Current Survey**: Covers **five major application domains** (NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing).

Similarities

- Both surveys explain the **Transformer architecture** and its features.
- Both discuss the **working process** of Transformer models.

Differences

- **Brasoveanu & Andonie (2020)**:
 - Focuses on **visualization techniques** to explain Transformer architectures.
 - Limited to **NLP**.
 - **Current Survey**:
 - Describes the **working process** of significant Transformer models across a **range of tasks**.
 - Covers **five fields of Transformer applications**, providing a more comprehensive analysis.
-

Key Takeaways from the Comparison

- **Tay et al. (2023)** and **Brasoveanu & Andonie (2020)** focus on **specific aspects** of Transformers (e.g., efficiency, visualization) and are limited to **specific domains** (e.g., language, vision, NLP).
 - The **current survey** differentiates itself by focusing on **application areas** and **tasks**, providing a **broader and more comprehensive analysis** of Transformer models across **multiple domains**.
-

Summary Table

Approach	Fields of Application	Similarities	Differences
----------	-----------------------	--------------	-------------

Tay et al. (2023)	Performance / Architecture	<ul style="list-style-type: none"> Proposes a taxonomy of Transformers based on primary use cases. 	<ul style="list-style-type: none"> Focuses on computational power and memory efficiency.
		<ul style="list-style-type: none"> Discusses the primary use cases of Transformer models. 	<ul style="list-style-type: none"> Limited to language and vision domains.
Brasoveanu & Andonie (2020)	Natural Language Processing (NLP)	<ul style="list-style-type: none"> Explains Transformer architecture and its features. 	<ul style="list-style-type: none"> Focuses on visualization techniques for NLP tasks.
		<ul style="list-style-type: none"> Discusses the working process of Transformer models. 	<ul style="list-style-type: none"> Limited to NLP.
Current Survey	Application Areas / Tasks	<ul style="list-style-type: none"> Proposes a taxonomy of Transformers based on application domains and tasks. 	<ul style="list-style-type: none"> Focuses on deep learning tasks and applications across multiple domains.
		<ul style="list-style-type: none"> Explains Transformer architecture and its features. 	<ul style="list-style-type: none"> Covers NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing.

In summary, the **current survey** stands out by focusing on **application areas** and **tasks**, providing a **comprehensive analysis** of Transformer models across multiple domains, including areas like **signal processing** and **cloud computing** that are not covered in existing surveys.

W Guan et al. (Wang et al., 2020a) Natural language Processing□NLP

- Survey an application area of transformers, which is text summarization, which is one of the application areas covered in our survey
- The authors propose a transformer-based summarizer that

solves the issues of standard transformers that cannot take a long text as an input. They survey different use cases of applying transformers to different text summarization tasks and they only cover text summarization. no proposed transformers have been built in our survey.

R Kumar

(Kaliyar,

2020)

Natural

language

Processing□NLP

- Discussion of different NLP downstream tasks that BERT performs. BERT is covered in our survey as well as the different NLP tasks
- Survey different techniques on using BERT as a word□embedder against traditional word-embedding techniques.

Their survey is only focused as using transformers as a tool

for embedding text

F Acheampong et al.

(Acheampong et al.,

2021)

Natural

language

Processing□NLP

- Survey different transformer architectures that accomplish the emotion detection task. We do the same, the application of different transformers to the same type of task
- Survey the application of transformer architecture to a single application area but in too much detail, which is emotion detection from text-based data, a form of sentiment analysis but the goal is to extract fine-grained emotion from the data. The task of sentiment analysis is covered in our survey, but we didn't cover especially the task of detecting emotions on different levels and not just as a binary classification task as usually done in sentiment analysis



The table you provided compares the **approach**, **fields of application**, **similarities**, and **differences** between the current survey and three existing surveys: **Wang et al. (2020a)**, **Kaliyar (2020)**, and **Acheampong et al. (2021)**. Here's a clear explanation of the comparison:

1. Comparison with Wang et al. (2020a)

Approach

- **Wang et al. (2020a)**: Focuses on **text summarization** using Transformers.
- **Current Survey**: Focuses on **application areas** and **tasks** of Transformers.

Fields of Application

- **Wang et al. (2020a)**: Primarily discusses **text summarization** in NLP.
- **Current Survey**: Covers **five major application domains** (NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing).

Similarities

- Both surveys cover **text summarization** as an application area of Transformers.

Differences

- **Wang et al. (2020a)**:
 - Proposes a **transformer-based summarizer** that addresses the issue of handling long texts.
 - Focuses **only on text summarization** and does not cover other NLP tasks or application domains.
- **Current Survey**:
 - Covers **text summarization** as one of many tasks within the NLP domain.
 - Provides a **broader analysis** of Transformer applications across multiple domains.

2. Comparison with Kaliyar (2020)

Approach

- **Kaliyar (2020)**: Focuses on using **BERT** as a word-embedding tool in NLP.
- **Current Survey**: Focuses on **application areas** and **tasks** of Transformers.

Fields of Application

- **Kaliyar (2020)**: Primarily discusses **word embedding** techniques using BERT.

- **Current Survey:** Covers **five major application domains** (NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing).

Similarities

- Both surveys discuss **BERT** and its applications in NLP.

Differences

- **Kaliyar (2020):**
 - Focuses on **word embedding** techniques and compares BERT with traditional methods.
 - Limited to **text embedding** and does not cover other NLP tasks or application domains.
 - **Current Survey:**
 - Covers **BERT** as one of many Transformer models used in NLP.
 - Provides a **broader analysis** of Transformer applications across multiple domains.
-

3. Comparison with Acheampong et al. (2021)

Approach

- **Acheampong et al. (2021):** Focuses on **emotion detection** using Transformers.
- **Current Survey:** Focuses on **application areas** and **tasks** of Transformers.

Fields of Application

- **Acheampong et al. (2021):** Primarily discusses **emotion detection** in NLP.
- **Current Survey:** Covers **five major application domains** (NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing).

Similarities

- Both surveys discuss the application of Transformers to **emotion detection** and **sentiment analysis**.

Differences

- **Acheampong et al. (2021):**
 - Focuses on **fine-grained emotion detection** from text-based data.
 - Provides a **detailed analysis** of emotion detection but does not cover other NLP tasks or application domains.
 - **Current Survey:**
 - Covers **sentiment analysis** as one of many tasks within the NLP domain.
 - Provides a **broader analysis** of Transformer applications across multiple domains.
-

Key Takeaways from the Comparison

- Wang et al. (2020a), Kaliyar (2020), and Acheampong et al. (2021) focus on **specific tasks** within NLP (e.g., text summarization, word embedding, emotion detection).
- The **current survey** differentiates itself by focusing on **application areas** and **tasks**, providing a **broader and more comprehensive analysis** of Transformer models across **multiple domains**.

Summary Table

Approach	Fields of Application	Similarities	Differences
Wang et al. (2020a)	Natural Language Processing (NLP)	<ul style="list-style-type: none">• Covers text summarization as an application area of Transformers.	<ul style="list-style-type: none">• Focuses only on text summarization and proposes a transformer-based summarizer. <ul style="list-style-type: none">• Does not cover other NLP tasks or application domains.
Kaliyar (2020)	Natural Language Processing (NLP)	<ul style="list-style-type: none">• Discusses BERT and its applications in NLP.	<ul style="list-style-type: none">• Focuses on word embedding techniques using BERT. <ul style="list-style-type: none">• Does not cover other NLP tasks or application domains.
Acheampong et al. (2021)	Natural Language Processing (NLP)	<ul style="list-style-type: none">• Discusses emotion detection using Transformers.	<ul style="list-style-type: none">• Focuses on fine-grained emotion detection from text-based data. <ul style="list-style-type: none">• Does not cover other NLP tasks or application domains.
Current Survey	Application Areas / Tasks	<ul style="list-style-type: none">• Covers text summarization, word embedding, and emotion detection as part of NLP.	<ul style="list-style-type: none">• Focuses on deep learning tasks and applications across multiple domains. <ul style="list-style-type: none">• Covers NLP, Computer Vision, Multi-Modality, Audio & Speech, and Signal Processing.

In summary, the **current survey** stands out by focusing on **application areas** and **tasks**, providing a **comprehensive analysis** of Transformer models across multiple domains, including areas like **signal processing** and **cloud computing** that are not covered in existing surveys.

R Gruet□zemacher et
al. (Gruet□zemacher
& Paradice,
2022)

Natural

language

Processing□NLP

- Survey the progress of trans□formers in the text-mining ap□plication area. We do cover in our survey the progress of transformers on a wide variety of tasks
- Tackle the different transformers on how they can be used as text miners for organizations that have huge amounts of un□structured data against traditional NLP text-mining techniques

J. Selva et
al. (Selva
et al., 2023)

Computer

Vision

- This paper is an overview of transformers developed for modeling images and video data
- This survey focuses solely on image and video data. Models are compared based on their performance in video classification, it does not cover any other applications. The paper pro□poses a taxonomy of various transformer models based on their recurrence properties, memory capacities, and architectural de□sign

K. S.

Kalyan et
al. (Subra□manyam
et al.,
2021a)

Natural

language

Processing□Medical

- This paper provides an overview of the developed transformer-based BPLMs for a wide range of NLP tasks, including Natural language inference, Entity extraction, Relation extraction, Semantic textual similarity, Text classification, Question answering, and Text summarization
 - This survey addresses only transformer-based biomedical pre-trained language models, which restricts its scope to the specific field of biomedical natural language processing. The taxonomy does not distinguish models based on the type of application they are used for, but rather based on the dataset of pre-training, the embedding type, and other criteria such as the targeted language
- K. Han et al. (Han et al., 2023)
- Computer Vision
- Categorized vision transformer models based on different tasks
 - This existing paper analyzed transformer models' advantages and disadvantages, and efficient transformer methods for the backbone network, while our survey categorizes transformer models based on tasks and summarize downstream tasks and commonly used dataset.
 - While our survey paper classified computer vision tasks into two segments: natural image processing & medical image processing and then focused on popular computer vision like visual question answering, classification, segmentation, question answering, and so on, then this existing paper focused on high/mid-level vision, low-level vision, and video processing computer vision tasks.
 - This survey focused on computer vision tasks only, while we covered other four fields of applications-NLP, Multi-modal, Audio/speech, and signal processing besides computer visio



Y. Xu et

al. (Xu

et al., 2022)

Computer

Vision

- The survey covers the fields

of computer vision and multi-modal in a similar fashion to
our survey

- This survey focuses primarily on recent advancements in computer vision by comparing the performance of different transformer models. Specifically, this study discusses four areas of research: advances in the design of the ViT models for image classification, high-level vision tasks (such as object detection and semantic segmentation), low-level vision tasks (such as super-resolution, and image generation), and multimodal learning (such as visual question answering (VQA), image captioning)

J Li et al.

(Li et al.,

2023)

Computer

Vision

- Comparative analysis of transformer models is presented in this paper for several tasks involved in medical vision.

Several criteria are considered

when comparing papers, including the type of dataset, the type of input data, and the architecture of the model

- This paper describes in detail several transformer models that have been developed for medical images; however, it does not provide information regarding medical signals

F Shamshad

et al.

(Shamshad

et al., 2023)

Computer

Vision-medical

- A review of a number of transformer models with a focus on some tasks related to medical images and different image modalities, and a description of the datasets used for these tasks

- This paper compares deep learning models starting with CNNs and moving up to vision transformers. In this paper, medical image modalities and several medical computer vision tasks are discussed to compare papers through the specification of datasets used and also provide an overview of models' performance. In this paper, the comparison is based solely on medical images; medical signals are not considered

Salman

Khan et

al. (Khan

et al., 2022)

Computer

Vision

- A overview of existing transformer computer vision models

and classified the models based

on popular tasks

- While this existing survey paper compared the popular techniques in terms of architectural design and experimental value,

while our survey worked based on popular tasks and applications.

- In the computer vision section, we put a special focus on Medical image tasks besides natural image processing.

- This survey focused on computer vision tasks only, while we covered other four fields of applications, namely NLP, Multi-modal, Audio/speech, and signal processing besides computer

vision

L. Ruan et

al. (Ruan &

Jin, 2022)

Multi-modal(NLP-CV)

- Categorize transformer vision-language models based on tasks and summarize downstream

tasks and commonly used video

dataset

- This existing survey focused on multi-modal(NLP-CV) tasks only, while we covered other four fields of applications-NLP,

Computer vision, Audio/speech, and signal processing besides

multi-modal

Continued on next page

Table 2 – continued from previous page

Approach Fields of

Application Similarities Differences

A Shin et

al. (Shin

et al., 2022)

Multi-modal

(Perfor□mance

/Architectur□e)

- They survey transformers for multi-modal tasks, which we do also include in our different application tasks
- Cover only one application area in detail, which is multimodal visual-linguistic tasks



Time flies when we're having a great chat! We've exceeded the length limit. Please start a new chat!

Y. Xu et

al. (Xu

et al., 2022)

Computer

Vision

- The survey covers the fields of computer vision and multi-modal in a similar fashion to our survey
- This survey focuses primarily on recent advancements in com□puter vision by comparing the performance of different trans□former models. Specifically, this study discusses four areas of research: advances in the design of the ViT models for image classification, high-level vision tasks (such as object detection and semantic segmentation), low-level vision tasks (such as super-resolution, and image generation), and multimodal learn□ing (such as visual question answering (VQA), image caption□ing)

J Li et al.

(Li et al.,

2023)

Computer

Vision

- Comparative analysis of trans□former models is presented in this paper for several tasks

involved in medical vision.

Several criteria are considered

when comparing papers, including the type of dataset, the type of input data, and the architecture of the model

- This paper describes in detail several transformer models that have been developed for medical images; however, it does not provide information regarding medical signals

F Shamshad

et al.

(Shamshad

et al., 2023)

Computer

Vision-Medical

- A review of a number of transformer models with a focus on some tasks related to medical images and different image modalities, and a description of the datasets used for these tasks

- This paper compares deep learning models starting with CNNs and moving up to vision transformers. In this paper, medical image modalities and several medical computer vision tasks are discussed to compare papers through the specification of datasets used and also provide an overview of models' performance. In this paper, the comparison is based solely on medical images; medical signals are not considered

Salman

Khan et

al. (Khan

et al., 2022)

Computer

Vision

- A overview of existing transformer computer vision models and classified the models based on popular tasks

- While this existing survey paper compared the popular techniques in terms of architectural design and experimental value, while our survey worked based on popular tasks and applications.

- In the computer vision section, we put a special focus on Medical image tasks besides natural image processing.

- This survey focused on computer vision tasks only, while we covered other four fields of applications, namely NLP, Multi-modal, Audio/speech, and signal

processing besides computer

vision

L. Ruan et

al. (Ruan &

Jin, 2022)

Multi-modal(NLP-CV)

- Categorize transformer vision-language models based on tasks and summarize downstream tasks and commonly used video dataset

- This existing survey focused on multi-modal(NLP-CV) tasks only, while we covered other four fields of applications-NLP, Computer vision, Audio/speech, and signal processing besides multi-modal

Continued on next page

Table 2 – continued from previous page

Approach Fields of

Application Similarities Differences

A Shin et

al. (Shin

et al., 2022)

Multi-modal

(Performance

/Architecture)

- They survey transformers for multi-modal tasks, which we do also include in our different application tasks
- Cover only one application area in detail, which is multimodal visual-linguistic tasks



Time flies when we're having a great chat! We've exceeded the length limit. Please start a new chat!



New chat

Message DeepSeek

AI-generated, for reference only