

Identified 3 related articles per member and provided a summary of articles with gap analysis.

1. Syeda Anshrah Gillani (1337–2021): Focus on Generative Models for Text to Image

Related Articles:

"Generative Adversarial Text-to-Image Synthesis" by Scott Reed et al.

1. **Summary:** This paper presents a generative adversarial network (GAN) for generating images from text descriptions. It proposes an architecture that learns both a mapping from the text space to image space and the relationship between text and visual features.
2. **Gap Analysis:** The paper mainly focuses on GAN-based architectures but could benefit from exploring transformer-based approaches, which have seen improvements in generative tasks.

"AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks" by Tao Xu et al.

1. **Summary:** AttnGAN utilizes attention mechanisms to improve the fidelity of generated images by allowing the network to focus on specific parts of the text description during the generation process.
2. **Gap Analysis:** While AttnGAN improves the visual quality of images, its dependence on detailed annotations may limit the model's scalability. Exploring approaches with less reliance on annotated data could be beneficial.

"T2F: Text to Face Generation Using Deep Neural Networks" by Sadegh M. Z. et al.

1. **Summary:** This paper focuses on generating realistic facial images from text descriptions using a novel architecture. It combines image generation with text understanding in a human face context.
2. **Gap Analysis:** The paper focuses on a narrow domain of face generation, which restricts its generalizability to more complex objects and scenes. Expanding this method to more diverse image domains is a potential gap.

Gap Analysis Summary:

The literature suggests that GANs have been widely used for text-to-image synthesis. However, existing models may struggle with large, complex images and often rely on detailed annotations. A potential gap lies in combining transformer models with GANs to better handle diverse, less-structured datasets with minimal reliance on annotations.

2. Umeema Mujeeb (2396–2021): Focus on Model Architecture and Evaluation Techniques

Related Articles:

"Deep Attentional Multimodal Fusion for Text-to-Image Generation" by Lu et al.

1. **Summary:** This paper presents a method to fuse multiple modalities (text, image features) for better synthesis of images from text. The model leverages deep attention to handle various input features efficiently.
2. **Gap Analysis:** While the multimodal fusion approach works well for detailed descriptions, it can be computationally expensive and may struggle with longer, more complex descriptions.

"StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks" by Han Zhang et al.

1. **Summary:** StackGAN generates high-resolution images from text descriptions by stacking two GANs, one generating rough sketches and the other refining them into photorealistic images.
2. **Gap Analysis:** The challenge with StackGAN lies in the fact that the coarse-to-fine architecture may fail to capture intricate details in more abstract or artistic image generation tasks. Investigating more granular architectural refinements could help.

"Unsupervised Text-to-Image Generation with Generative Adversarial Networks" by Xintao Duan et al.

1. **Summary:** This paper explores unsupervised learning techniques in text-to-image generation, where the model learns from unpaired datasets (images without associated text).
2. **Gap Analysis:** While unsupervised learning enables greater flexibility in training data, it may lead to lower image quality and less coherent outputs. Further exploration of semi-supervised learning techniques might close this gap.

Gap Analysis Summary:

The focus on attention mechanisms and unsupervised learning in the papers could be promising, but current architectures are computationally expensive, and image quality may be inconsistent. There's potential for improvement in handling less-structured data and refining text-image coherence, especially in artistic or abstract domains.

3. Maheen Ali (1589–2021): Focus on Evaluation Metrics and Real-world Applications

Related Articles:

"Text2Image: Generating Images from Captions" by Alireza Shabanian et al.

- 1.
1. **Summary:** This paper presents a text-to-image generation system that evaluates output images based on how well they align with the textual description. It focuses on automatic evaluation metrics to measure image-text similarity.
2. **Gap Analysis:** While automatic evaluation metrics are valuable, they often fail to capture subjective elements like image quality and creativity. More human-centered evaluation techniques could enhance this approach.

"Visual Semantic Role Labeling for Text-to-Image Generation" by Yujia Li et al.

1. **Summary:** The paper proposes using semantic role labeling (SRL) to enhance the text-to-image generation process, allowing the model to better understand the underlying semantics in a description.
2. **Gap Analysis:** The reliance on SRL can limit flexibility and adaptability, particularly when generating images from vague or creative descriptions. Expanding this technique to handle more abstract text could be a valuable direction.

"Revisiting Text-to-Image Synthesis via a Convolutional Attention GAN" by Bo Zhao et al.

1. **Summary:** This paper enhances text-to-image generation through convolutional attention mechanisms, improving the model's ability to focus on relevant parts of the description and generate images with fine details.
2. **Gap Analysis:** The convolutional attention approach may not scale well to extremely complex images with large scene diversity. Exploring methods that allow the model to handle varying levels of detail and abstraction could be a fruitful avenue.

Gap Analysis Summary:

Evaluating text-to-image generation systems remains a challenge, as current methods are often insufficient to capture subjective quality. There is an opportunity to integrate more human-based evaluation metrics, as well as to refine models for better handling abstract or creative text descriptions, improving real-world applications.

Conclusion:

In summary, the gap analysis across all three members' research reveals several key areas for improvement in the text-to-image field, such as:

- **Annotation and Dataset Complexity:** Reducing reliance on annotated data and exploring semi-supervised methods.
- **Text-Image Coherence:** Enhancing the connection between the textual input and visual features, especially for complex or abstract text.
- **Evaluation Metrics:** Developing more comprehensive evaluation methods that go beyond quantitative metrics and incorporate subjective, human-centered assessment.

THE END