

LAND OF THE CURIOS



CLIP FOR IMAGE STYLE TRANSFER: EXPLORE TEXT-IMAGE CORRELATIONS

CVPR SEMINAR PRESENTATION

Presented by Nadine Cyizere Bisanukuli

February 22, 2024

Examiners:

- » Prof. Zhi-Song Liu, Lappeenranta-Lahti University of Technology (LUT)
- » Doctor Jun Xiao, Hong Kong Polytechnic

OUTLINE

1 Introduction and Background

Introduction

Background and Challenges

Architecture

Objectives

2 Related Work

Image-Driven Style Transfer

Text-Driven Style Transfer

Attention-Driven Style Transfer

3 Proposed Methods

CLIP for text-image correlation

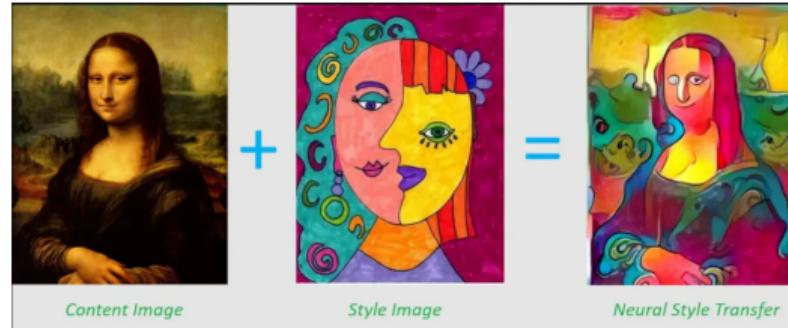
Clip for style transfer

Loss Function

4 Next Steps

INTRODUCTION

Style transfer: the process of applying artistic styles to images.



Applications: Film production, Fashion design, Virtual reality, social media



Film Production



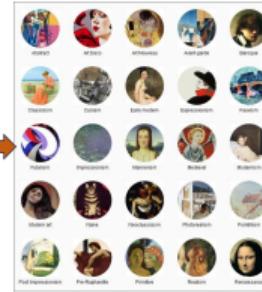
Fashion Design

BACKGROUND AND CHALLENGES

Traditional methods rely on reference images,
limiting creativity and applicability



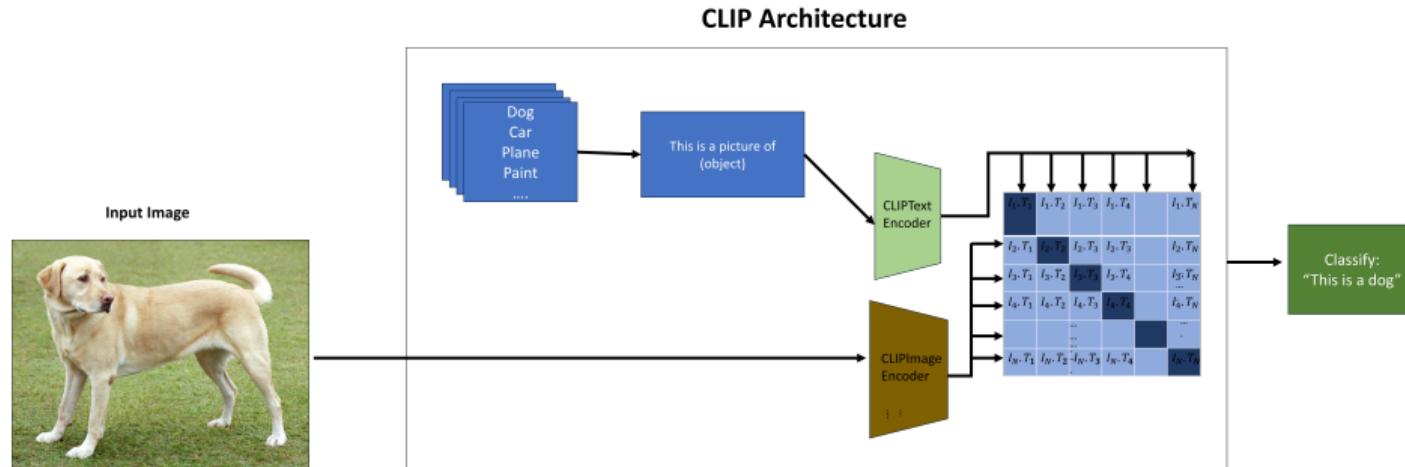
Domain specificity hinders generalization to arbitrary styles.



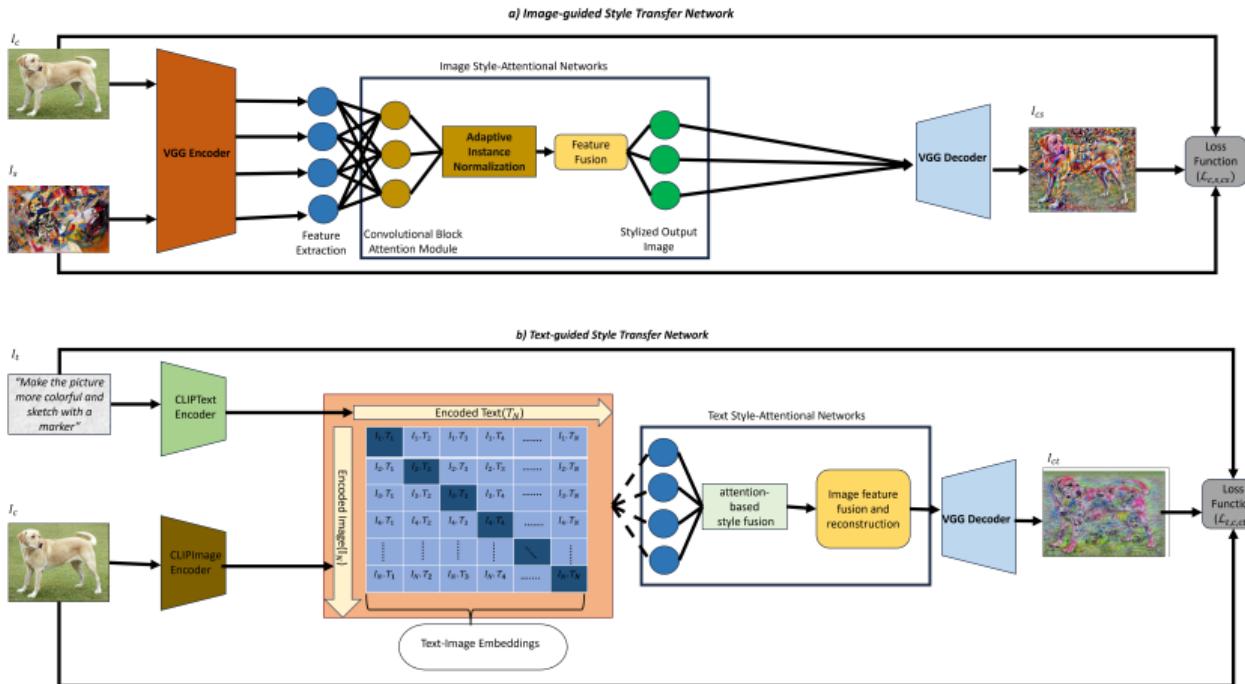
- » Domain specificity hinders generalization to arbitrary styles.
- » Computational efficiency.
- » CLIP presents new possibilities for manipulating styles using text descriptions

CLIP ARCHITECTURE

This thesis will dive deep into the intersection of Contrastive Language-Image Pre-training (CLIP) [3] of the original target image and image style transfer. This will leverage the text-image correlations.



STYLE TRANSFER ARCHITECTURE



Objectives:

1. Explore the effectiveness of CLIP for image style transfer.
2. Leverage text-image correlations to enhance image style transfer.
3. Conduct extensive experiments and analysis on text-driven style transfer.

Scope:

- » Focus on using the CLIP model for image style transfer, excluding other models or methods.
- » Evaluate the proposed method using available datasets.
- » Computational efficiency is not the primary focus.
- » Limited exploration of CLIP for tasks beyond image style transfer.

OUTLINE

1 Introduction and Background

Introduction

Background and Challenges

Architecture

Objectives

2 Related Work

Image-Driven Style Transfer

Text-Driven Style Transfer

Attention-Driven Style Transfer

3 Proposed Methods

CLIP for text-image correlation

Clip for style transfer

Loss Function

4 Next Steps

RELATED WORK

Style transfer in Computer Vision has seen significant advancements. This was done through the Exploration of various techniques and methodologies.

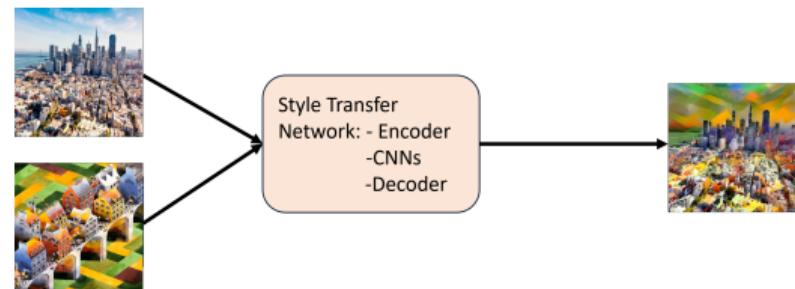
1. Image-Driven Style Transfer:

Advantages:

- » Utilizes of CNN.
- » Easy transfer of styles across multiple domains which makes creation and alteration easy.

Disadvantages:

- » Fixed styles in feed-forward methods.
- » computational efficiency and quality.
- » Limit generalization.

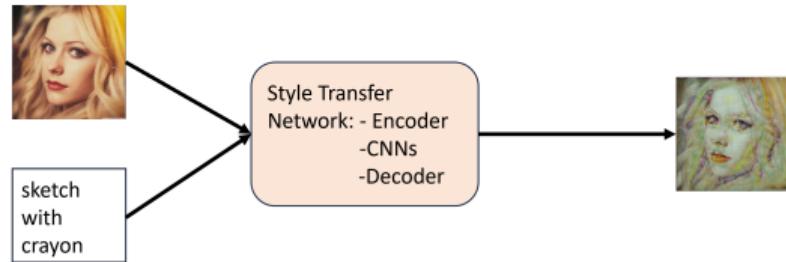


2. Text-Driven Style Transfer: Advantages:

- » Textual descriptions for image transformation.
- » CLIP for text-image correlations.
- » Flexibility for arbitrary styles.

Disadvantages:

- » Limitations in manipulating images beyond trained domains.
- » Content preservation for complex styles.
- » Computational complexity



CONT'D RELATED WORK

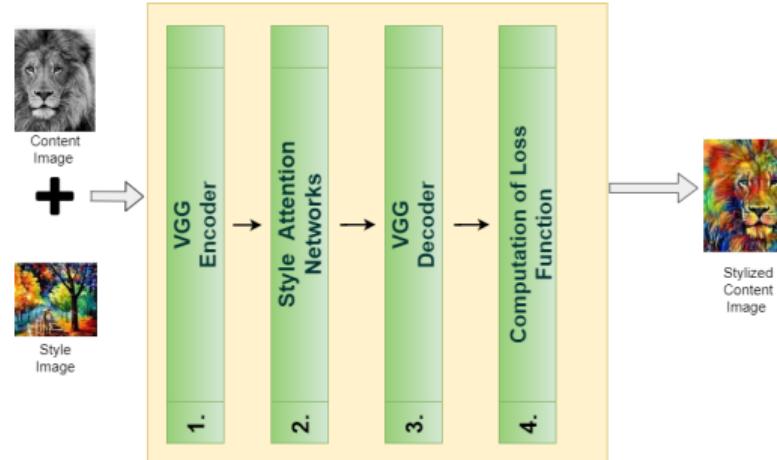
3. Attention-Driven Style Transfer:

Advantages:

- » Attention mechanisms to capture key styles.
- » Quality and diversification of output.
- » Adaptability and scalability of features.

Disadvantages:

- » Optimizing attention mechanisms



OUTLINE

1 Introduction and Background

Introduction

Background and Challenges

Architecture

Objectives

2 Related Work

Image-Driven Style Transfer

Text-Driven Style Transfer

Attention-Driven Style Transfer

3 Proposed Methods

CLIP for text-image correlation

Clip for style transfer

Loss Function

4 Next Steps

PROPOSED METHOD

1. CLIP for text-image correlation

2. Clip for style transfer

- 0.1 Choose content and Style Image
- 0.2 Preprocess Image(Encoders, and other steps)
- 0.3 Random Image generation(starting point for the style transfer)
- 0.4 Apply the CLIP [2] model(to extract content features) to the random Image while constantly evaluating Loss function.

CONT'D PROPOSED METHOD

3. VGG Loss

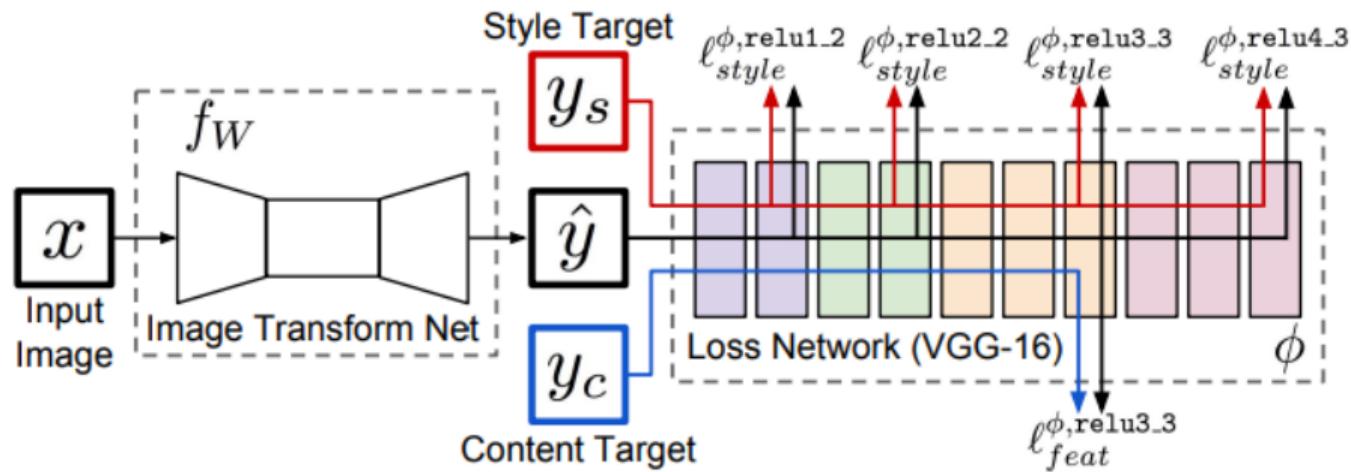


Figure: VGG Architecture [1]

OUTLINE

1 Introduction and Background

Introduction

Background and Challenges

Architecture

Objectives

2 Related Work

Image-Driven Style Transfer

Text-Driven Style Transfer

Attention-Driven Style Transfer

3 Proposed Methods

CLIP for text-image correlation

Clip for style transfer

Loss Function

4 Next Steps

NEXT STEPS

1. Experiments on attention for text-driven style transfer
2. Full evaluation, including VGG loss and CLIP text-image correlation

CONCLUSION

THANK YOU!

References

- [1] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. "Perceptual Losses for Real-Time Style Transfer and Super-Resolution". In: *Computer Vision – ECCV 2016*. Ed. by Bastian Leibe et al. Cham: Springer International Publishing, 2016, pp. 694–711.
- [2] Gihyun Kwon and Jong Chul Ye. "CLIPstyler: Image Style Transfer with a Single Text Condition". In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022, pp. 18041–18050. DOI: [10.1109/CVPR52688.2022.01753](https://doi.org/10.1109/CVPR52688.2022.01753).
- [3] Or Patashnik et al. "StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery". In: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021, pp. 2065–2074. DOI: [10.1109/ICCV48922.2021.00209](https://doi.org/10.1109/ICCV48922.2021.00209).



LUT
University