Presentation by:
Hari Haran A, 513121104304,
Dept. of CSE,
Thanthai Periyar Government Institute of Technology, Vellore-02.

IMAGE DEEP LEARNING

ANNA UNIVERSITY **NAAN MUDHALVAN – GENERATIVE AI** **06 – 08 -2024**

Abstract

Image animation consists of generating a video sequence so that an object in a source image is animated according to the motion of a driving video. Our framework addresses this problem without using any human bodies), our method can be applied to any object of this class. To achieve this, we decouple appearance and motion information using a self-supervised formulation. To support complex motions, we use a representation consisting of a set of learned keypoints along with their local affine transformations. A generator network models occlusions arising during target motions and comannotation or prior information about the specific object to animate. Once trained on a set of videos depicting objects of the same category (e.g. faces, happiness the appearance extracted from the source image and the motion derived from the driving video. Our framework scores best on diverse benchmarks and on a variety of object categories

Overview

- Introduction
- Problem
- Literary Review
- Theoretical
- Objectives
- Hypothesis
- Methodology
- Implementation
- Result
- Conclusion
- Thank You

Introduction

Definition:
Image animation refers to the creation of animated sequences from static images, using the motion patterns observed in a separate driving video.
Importance and Applications:
Applications include video synthesis, special effects in film, augmented reality, virtual reality, and more.
Framework Introduction:
Our framework aims to create realistic animations without the need for annotations or prior object information, using a self-supervised learning approach.

Problems

PROBLEM 1

Traditional methods require annotations or prior knowledge about the object, which is labor-intensive and limits generalization.

PROBLEM 2

Handling various types of motions and ensuring the animated object looks natural.

PROBLEM 3

Need for a method that can generalize across different object categories.

Literary Review

01

- **Existing Methods:**
- Review of traditional image animation techniques that rely on annotated data and prior object-specific information.

0 2

Introduces self-supervised learning for decoupling appearance and motion.

0 3

Limitations:

- Dependence on manual annotations.
- Limited ability to generalize across different types of objects and motions.

0 4

Uses a keypoint-based representation and local affine transformations to model complex motions.

Theoretical Framework

Self-Supervised Learning:
Learning motion and appearance features without labeled data.
Utilizes the intrinsic properties of the data for training.

Keypoint-Based Representation:
Keypoints are used to capture important features of the object.
Self-supervised learning is used to identify these keypoints.

Local Affine Transformations:
Models the motion around each keypoint using local transformations.
Allows for a broader range of motion types compared to using keypoints alone.

Objectives

Generalized Framework:

- Develop a framework that can animate any object within a specific category using a driving video.

High-Quality Animation:

- Ensure the framework performs well across various benchmarks and object categories, maintaining the quality of the animation.

Hypothesis

Decoupling Motion and Appearance:
Hypothesize that self-supervised learning can effectively separate motion and appearance features.

Handling Complex Motions:
Keypoints and local affine transformations can model a wide range of motion types.

Generator Network Efficiency:
A generator network can successfully merge appearance and motion information to create realistic animations.

Methodology

Data Collection:

Use a large collection of videos with objects from the same category.

Reconstruction:

The model is trained to reconstruct videos by combining a single frame with a learned latent motion representation.

Frame Pairs:

Observe pairs of frames (source and driving) to encode motion.

Keypoint Displacements and Local Affine Transformations:

Encode motion as a combination of these two elements.

Implementation

Motion Estimation Module:

Dense Motion Field Prediction: Predicts a dense motion field.

Reference Frame: Assume an abstract reference frame and estimate transformations.

Keypoints and Sparse Trajectories: Use keypoints to approximate transformations.

Local Affine Transformations: Model motion in the neighborhood of each keypoint.

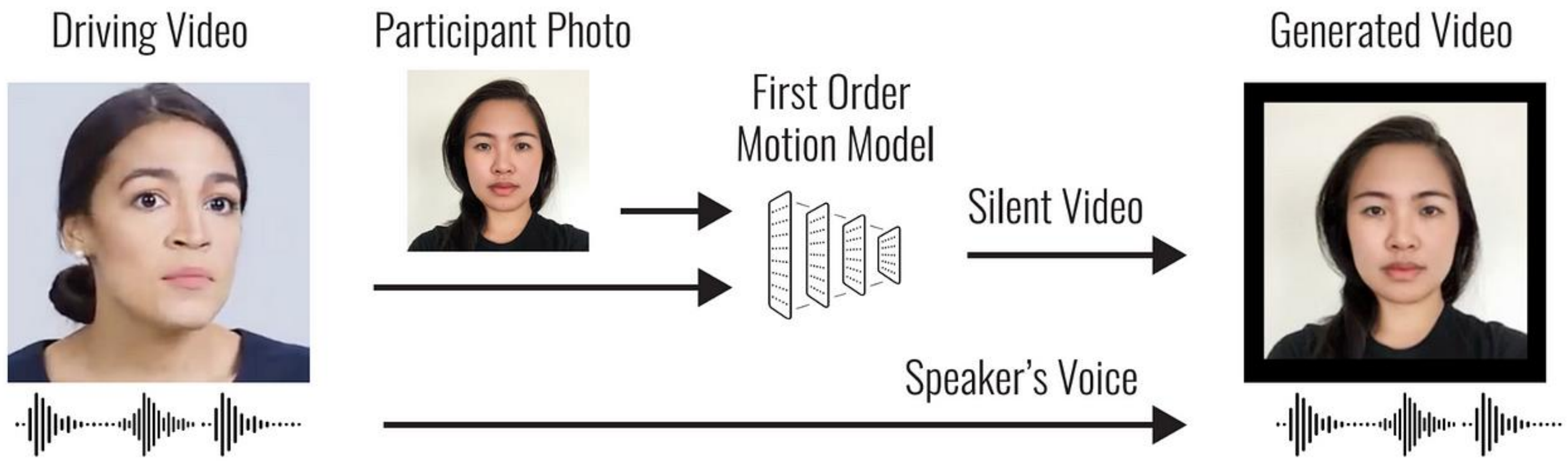
Image Generation Module:

Dense Motion Field Combination: Combines local approximations to create a dense motion field.

Occlusion Mask: Indicates which parts of the driving image can be reconstructed by warping the source image and which need inpainting.

Generator Network: Warps the source image according to the dense motion field and inpatients occluded parts to create the final animated image.

Result



Performance Evaluation:
Present quantitative results on diverse benchmarks.
Metrics used for evaluation (e.g., SSIM, LPIPS).

Comparison with Existing Methods:
Show how your framework outperforms others in terms of quality and generalization.

Conclusion

Summary:
Recap the framework and its key benefits, emphasizing the lack of need for annotations and the ability to generalize across various object categories.

Key Findings:
Highlight major results and observations from your experiments.

Future Work:
Discuss potential improvements, such as enhancing the realism of animations and extending the framework to more object categories and motion types.

Thank You

Presentation by: Hari Haran A

Anna University | Generative AI | 2024